On a probabilistic approach to synthesize control policies from example datasets ⋆,⋆⋆

Davide Gagliardi[1], Giovanni Russo [a]

[a] *Department of Information and Electrical Engineering and Applied Mathematics, University of Salerno, Italy (e-mail: giovarusso@unisa.it)*

Abstract

This paper is concerned with the design of control policies from example datasets. The case considered is when just a black box description of the system to be controlled is available and the system is affected by actuation constraints. These constraints are not necessarily fulfilled by the (possibly, noisy) example data and the system under control is not necessarily the same as the one from which these data are collected. In this context, we introduce a number of theoretical results to compute a control policy from example datasets that: (i) makes the behavior of the closed-loop system similar to the one illustrated in the data; (ii) guarantees compliance with the constraints. We recast the control problem as a finite-horizon optimal control problem and give an explicit expression for its optimal solution. Moreover, we turn our findings into an algorithmic procedure. The procedure gives a systematic tool to compute the policy. The effectiveness of our approach is illustrated via a numerical example, where we use real data collected from test drives to synthesize a control policy for the merging of a car on a highway.

1 Introduction

Model-based control is a key paradigm to design control systems: the design of platoons, fault tolerant and biochemical systems are just few of the frontier applications where this approach has been successfully used. Unfortunately, detailed mathematical models in the form of e.g. differential/difference equations, are not always available and, when available, can be hard to identify. Hence, a paradigm that is becoming increasingly popular is to synthesize control policies directly from data, see e.g. (De Persis and Tesi, 2020; Van Waarde et al., 2020; Hou and Wang, 2013) and references therein. As noted in these papers, this approach aims at designing policies while bypassing the need to devise/identify a mathematical model and can be useful in applications where first-principle models cannot be obtained and/or system identification is too computationally expensive.

An appealing framework to design controllers from data is that of using example datasets. This *learning from demonstrations* approach involves learning actions by observing an expert (Hanawal et al., 2019; Wabersich and Zeilinger, 2018; Zhu et al., 2020). In this context, we consider the key challenge of designing control policies from noisy example datasets for systems affected by actuation constraints. Motivated by this, we present [2] a set of technical results to synthesize policies directly from examples that might not satisfy the actuation constraints. Situations where our results are of interest naturally arise in the context of e.g. autonomous urban driving, where one is often interested in designing policies from driving examples, while ensuring that the control (e.g. speed, acceleration) signal fulfills certain properties with some acceptable/design probability level (Vitus and Tomlin, 2013). We now survey some related works on data-driven control and learning from examples.

Data-driven control. As noted in e.g. (De Persis and Tesi, 2020; Van Waarde et al., 2020), work on data-driven control can be traced back to (Ziegler and Nichols, 1942) and their results on the tuning of PID controllers. Recently, driven by the explosion in the *amount* of available data, the problem of finding control policies from datasets has gained increasing attention, see e.g. (Hou and Xu, 2009). For example, (Van Waarde et al., 2020)

⋆ **This is an authors' version of the work that is published in Automatica, Vol. 137, 2022. Changes were made to this version by the publisher prior to publication. The final version of record is available at https://doi.org/10.1016/j.automatica.2021.110121**

⋆⋆This publication has emanated from research conducted with the financial support of Science Foundation Ireland under Grant number 16/ RC/3872

[1] Work done in part while at the School of Electrical and Electronic Engineering, University College Dublin, Ireland.

[2] An early version of the results presented at the 21st IFAC World Congress (Gagliardi and Russo, 2020).

studies data-driven control in LTI systems when the data available are not persistently exciting and where, therefore, it is impossible to perform subspace identification. Recent works also include (Tanaskovic et al., 2017), where a direct data-driven design approach is introduced for discrete-time stabilizable systems with Lipschitz nonlinearities and e.g. (Markovsky and Rapisarda, 2007; Gonçalves da Silva et al., 2019; Baggio et al., 2019) that consider the problem of designing controllers for systems that have an underlying linear dynamics. We also recall here (De Persis and Tesi, 2020) that derives a parametrization of linear feedback systems using data-dependent linear matrix inequalities. Besides works on e.g. data-based tuning of PID controllers (Keel and Bhattacharyya, 2008) other remarkable results have been obtained by taking inspiration from the rich literature on Model Predictive Control (MPC). These include (Rosolia and Borrelli, 2018), where an MPC learning algorithm is introduced for iterative tasks when the system dynamics is partially known, (Salvador et al., 2018) where a data-based predictive control algorithm is presented for unknown time-invariant systems and (Coulson et al., 2019a,b) that, by taking a behavioral systems perspective, introduce a data-enabled predictive control algorithm for data generated by LTI systems.

Learning from example datasets. At their roots, learning from demonstration techniques largely rely on inverse optimal control (Bryson, 1996). Nowadays, these techniques, which essentially aim at finding/improving a policy from an initial demonstration dataset (Bertsekas, 2021), are recognized as a convenient framework to learn parametrized policies from *success stories* (Argall et al., 2009) and potential applications include planning (Englert et al., 2017) and medical prescriptions learning (Xu and Paschalidis, 2019). There is then no surprise that, over the years, a number of techniques have been developed to tackle the problem of learning parametrized control policies from demonstrations, mainly in the context of Markov Decision Processes. Results include (Hanawal et al., 2019), where parametrized policies consistent with data for a Markov decision process are learned via regularized logistic regression and regret bounds are given, (Ratliff et al., 2009), which leverages a linear programming approach, (Ratliff et al., 2006) which relies on a maximum margin approach, (Ziebart et al., 2008) that makes use of the maximum entropy principle, (Ramachandran and Amir, 2007) that formalizes the problem via Bayesian statistics and (Abbeel and Ng, 2004), where rewards are learned by expressing them as a combination of known features. We also recall the recent (Edwards et al., 2019), where an approach based on characterizing causal effects of latent actions is proposed to tackle imitation learning problems. A Bayesian approach (Peterka, 1981) to dynamical systems is also at the basis of works such as (Kárný, 1996; Kárný and Guy, 2006; Herzallah, 2015; Pegueroles and Russo, 2019; Guy et al., 2018; Kárný and Kroupa, 2012) see also (Kappen et al., 2012; Todorov, 2009, 2007) and references therein, which formalize (via the so-called Kullback-Leibler divergence, see Section 2.1 for the definition) the control problem as the problem of minimizing a cost that captures the discrepancy between an ideal probability density function and the actual probability density function of the system under control. Further, we recall (Guan et al., 2014), which also consider these Kullback-Leibler control problems without constraints. In such a work, by leveraging an average cost formulation, the probability mass function for the state transitions is found. Finally, we also recall (Russo, 2021; Garrabe and Russo, 2022), where policies are obtained from the minimization of similar costs by leveraging multiple, specialized, data sources.

Contributions of this paper

We introduce a framework to design policies from example datasets for systems having actuation constraints and for which just a probabilistic description is known. With our results, we aim at synthesizing control policies that make the behavior of the closed loop system *similar* (in a sense defined in Section 3.1) to the one seen in the example data, while still fulfilling the system-specific actuation constraints. To the best of our knowledge, the formulation presented here is novel and these are the first results that, simultaneously: (i) do not require assumptions on the linearity of the underlying stochastic dynamics that is generating the data; (ii) allow for the example data to be noisy and collected from a system that is not necessarily the same as the one that is under control; (iii) do not need the actuation constraints to be necessarily fulfilled in the example data; (iv) do not need an a-priori parametrization of the policy. Our key technical contributions are summarized as follows:

- we recast the control problem as the problem of *reshaping* certain probability density functions that can be obtained from data. This leads to formulate a constrained infinite dimensional, finite-horizon, optimal control problem. The decision variables of the problem are probability densities and the constraints are linear functionals in these densities;
- we introduce a number of theoretical results to tackle the control problem. Specifically, we prove that the finite-horizon optimal control problem can be split into infinite-dimensional sub-problems, with each of these sub-problems having a probability density (i.e. the control policy) as decision variable. After showing that each of the sub-problems is convex, we solve them explicitly. By leveraging this, we show that the control problem can be solved iteratively via a backward recursion. This leads to an explicit expression of the optimal policy;
- the results are turned into an algorithmic procedure to compute the optimal control policy;
- we illustrate the effectiveness of the results via a numerical example that involves the use of real data.

The paper is organized as follows. After introducing some mathematical preliminaries in Section 2, we formalize the control problem (Problem 1, in Section 3). This problem is solved with the technical results presented in Section 4 (some of the proofs are in the appendix). Specifically, in this section we first introduce, and discuss, an auxiliary problem (i.e. Problem 2). With Lemma 1 we then give an explicit expression for the optimal solution of Problem 2 and this result is used iteratively to prove Theorem 1, which gives the optimal solution to Problem 1. We conclude Section 4 by highlighting a connection between our approach and maximum entropy. The results are turned into an algorithmic procedure (Section 5) and a numerical example is presented in Section 6. Concluding remarks are given in Section 7.

2 Mathematical Preliminaries

Sets, as well as operators, are denoted by *calligraphic* characters, while vector quantities are denoted in **bold**. Random (row) vectors (i.e. a multidimensional random variables) are denoted by upper-case bold letters and their realization is denoted by lower-case bold letters. For example, $\mathbf{Z}$ denotes a multi-dimensional random variable and its realization is denoted by $\mathbf{z}$. The *probability density function* (or simply *pdf* in what follows) of a continuous $\mathbf{Z}$ is denoted by $f(\mathbf{z})$. The support of $f(\mathbf{z})$ is denoted by $\mathcal{S}(f)$ and, analogously, the expectation of a function $\mathbf{h}(\cdot)$ of $\mathbf{Z}$ is indicated with $\mathbb{E}_f[\mathbf{h}(\mathbf{Z})]$ and defined as $\mathbb{E}_f[\mathbf{h}(\mathbf{Z})] := \int_{\mathcal{S}(f)} \mathbf{h}(\mathbf{z}) f(\mathbf{z}) d\mathbf{z}$. For notational convenience, whenever it is clear from the context, we omit the domain of integration in the integral. We also remark here that: (i) whenever we apply the averaging operator to a given function, we use an upper-case letter for the function argument as this is a random vector; (ii) to stress the linearity of certain functionals or operators with respect to a specific argument, we include that argument in square brackets. The *joint* pdf of two (non independent) random vectors, say $\mathbf{Z}$ and $\mathbf{Y}$, is denoted by $f(\mathbf{z}, \mathbf{y})$. The *conditional* probability density function (or *cpdf* in what follows) of $\mathbf{Z}$ with respect to $\mathbf{Y}$ is denoted by $f(\mathbf{z}|\mathbf{y})$ and, whenever the context is clear, we use the shorthand notation $\tilde{f}_{\mathbf{Z}}$. Finally, given $\mathcal{Z} \subseteq \mathbb{R}^{n_z}$, its *indicator function* is denoted by $\mathbb{1}_{\mathcal{Z}}(\mathbf{z})$. That is, $\mathbb{1}_{\mathcal{Z}}(\mathbf{z}) = 1$, $\forall \mathbf{z} \in \mathcal{Z}$ and 0 otherwise. We also make use of the internal product between tensors, which is denoted by $\langle \cdot, \cdot \rangle$, while $\mathcal{A} \setminus \mathcal{B}$ is the set difference between $\mathcal{A}$ and $\mathcal{B}$. We indicate (ordered) countable sets as $\{w_k\}_{\mathcal{K}} := \{w_k\}_{k=k_1}^{k_n}$, where w_k is the generic element belonging to the set and k_1, k_n are the indices of the first and last element respectively.

2.1 The Kullback-Leibler divergence

The control problem is stated (see Section 3.1) in terms of the Kullback-Leibler (KL, (Kullback and Leibler, 1951)) divergence, formalized with the following:

Definition 1 *Consider two pdfs, $\phi(\mathbf{z})$ and $g(\mathbf{z})$, with the former being absolutely continuous with respect to the latter. Then, the KL-divergence of $\phi(\mathbf{z})$ with respect to $g(\mathbf{z})$, is* $\mathcal{D}_{KL}(\phi(\mathbf{z})||g(\mathbf{z})) := \int_{\mathcal{S}(\phi)} \phi(\mathbf{z}) \ln\left(\frac{\phi(\mathbf{z})}{g(\mathbf{z})}\right) d\mathbf{z}$.

Intuitively, $\mathcal{D}_{\mathrm{KL}}(\phi(\mathbf{z})||g(\mathbf{z}))$ is a measure of the proximity of the pair of pdfs, $\phi(\mathbf{z})$ and $g(\mathbf{z})$. We now give a property of the KL-divergence, also known as *chain rule*, used in the proof of Theorem 1.

Property 1 *Let $\mathbf{Z}$ and $\mathbf{Y}$ be two random vectors and let $\phi(\mathbf{y}, \mathbf{z})$ and $g(\mathbf{y}, \mathbf{z})$ be two joint pdfs. Then, the following identity holds:* $\mathcal{D}_{KL}(\phi(\mathbf{y}, \mathbf{z})||\, g(\mathbf{y}, \mathbf{z})) = \mathcal{D}_{KL}(\phi(\mathbf{y})||\, g(\mathbf{y})) + \mathbb{E}_{\phi}[\mathcal{D}_{KL}(\phi(\mathbf{z}|\mathbf{Y})||\, g(\mathbf{z}|\mathbf{Y}))]$.

Proof:the chain rule can be found in e.g. (Cover and Thomas, 2006) and its proof follows from the definition of $\mathcal{D}_{KL}$, the conditioning and independence rules for pdfs. We give a self-contained proof in the appendix. □

3 Formulation of the Control Problem

Let: (i) $\mathcal{K} := \{k\}_{k=1}^{n}$, $\mathcal{K}_0 := \mathcal{K} \cup \{0\}$ and $\mathcal{T} := \{t_k : k \in \mathcal{K}_0\}$; (ii) $\mathbf{x}_k \in \mathbb{R}^{n_x}$ and $\mathbf{u}_k \in \mathbb{R}^{n_u}$ be, respectively, the system state and input at time $t_k \in \mathcal{T}$; (ii) $\mathbf{\Delta}_k := (\mathbf{x}_k, \mathbf{u}_k)$ be the dataset collected from the system at time $t_k \in \mathcal{T}$ and $\mathbf{\Delta}^k$ the dataset collected from $t_0 \in \mathcal{T}$ up to time $t_k \in \mathcal{T}$ ($t_k > t_0$). See also (Peterka, 1981), where it is shown that the system behavior can be described via the joint pdf of the observed dataset, say $f(\mathbf{\Delta}^n)$. Moreover, by making the standard assumption that Markov's property holds (Kárný, 1996), the chain rule for pdfs leads to the following factorization for $f(\mathbf{\Delta}^n)$:

$$f(\mathbf{\Delta}^n) = \prod_{k \in \mathcal{K}} f(\mathbf{x}_k|\mathbf{u}_k, \mathbf{x}_{k-1})\, f(\mathbf{u}_k|\mathbf{x}_{k-1})\, f(\mathbf{x}_0)\,. \quad (1)$$

Throughout this work we refer to (1) as the *probabilistic description of the closed loop system*, or we simply say that (1) is our *closed loop* system.

Remark 1 *The cpdf $f(\mathbf{x}_k|\mathbf{u}_k, \mathbf{x}_{k-1})$ describes the system behavior at time t_k, given the previous state and the input at time t_k. The input is generated by the cpdf $f(\mathbf{u}_k|\mathbf{x}_{k-1})$. This is a* randomized control policy *returning the input given the previous state. Note that the initial conditions are embedded in the probabilistic system description through the prior $f(\mathbf{x}_0)$.*

In the rest of the paper we use the following *shorthand* notations: $\tilde{f}^k_{\mathbf{X}} := f(\mathbf{x}_k|\mathbf{u}_k, \mathbf{x}_{k-1})$, $\tilde{f}^k_{\mathbf{U}} := f(\mathbf{u}_k|\mathbf{x}_{k-1})$, $\tilde{f}^k := \tilde{f}^k_{\mathbf{X}} \tilde{f}^k_{\mathbf{U}}$, $f_0 := f(\mathbf{x}_0)$ and $f^n := f(\mathbf{\Delta}^n)$. Hence, (1) can be compactly written as

$$f^n = \left(\prod_{k \in \mathcal{K}} \tilde{f}^k_{\mathbf{X}} \tilde{f}^k_{\mathbf{U}} \right) f_0. \quad (2)$$

3.1 The control problem

Our goal is to synthesize, from an example dataset, say $\mathbf{\Delta}_e^n$, the control pdf/policy, i.e. $\tilde{f}_{\mathbf{U}}^k$, that: (i) makes the closed loop system *similar* to the behavior illustrated in the dataset; (ii) satisfies the system actuation constraints even if these are not fulfilled by the examples. We specify the behavior illustrated in the examples through the reference pdf $g\left(\mathbf{\Delta}_e^n\right)$ extracted from the example dataset. By means of Markov's property and the chain rule for pdfs we have $g\left(\mathbf{\Delta}_e^n\right) := \prod_{k\in\mathcal{K}} g\left(\mathbf{x}_k|\mathbf{u}_k,\mathbf{x}_{k-1}\right) g\left(\mathbf{u}_k|\mathbf{x}_{k-1}\right) g\left(\mathbf{x}_0\right)$. Again, by setting $\tilde{g}_{\mathbf{X}}^k := g\left(\mathbf{x}_k|\mathbf{u}_k,\mathbf{x}_{k-1}\right)$, $\tilde{g}_{\mathbf{U}}^k := g\left(\mathbf{u}_k|\mathbf{x}_{k-1}\right)$, $\tilde{g}^k := \tilde{g}_{\mathbf{X}}^k \tilde{g}_{\mathbf{U}}^k$, $g_0 := g\left(\mathbf{x}_0\right)$ and $g^n := g\left(\mathbf{\Delta}_e^n\right)$ we get:

$$g^n = \left(\prod_{k\in\mathcal{K}} \tilde{g}_{\mathbf{X}}^k \tilde{g}_{\mathbf{U}}^k\right) g_0. \tag{3}$$

The pdf $\tilde{g}_{\mathbf{X}}^k$ is not necessarily the same as the pdf $\tilde{f}_{\mathbf{X}}^k$. This allows to consider situations of practical interest where e.g. the system used to collect the example dataset is not necessarily the same as the system under control [3].

Remark 2 *Within this paper, a dataset is a sequence of data. In certain applications one might have access to a collection of datasets, which can be leveraged to compute the above pdfs. In e.g. autonomous driving applications one might run multiple test drives and, in this case, the pdfs can be computed from the collection of the speed profiles (i.e. the collection of datasets) obtained from each test drive. This observation is exploited in Section 6.*

The control problem, formalized next, can then be recast as the problem of designing $\tilde{f}_{\mathbf{U}}^k$ so that f^n (i.e. the probabilistic description of the closed loop system) approximates g^n (i.e. the joint pdf extracted from the example dataset). This is formally stated as follows:

Problem 1 *Let, $\forall k \in \mathcal{K}$:*

(i) n_e^k *and* n_l^k *be positive integers;*
(ii) $h_{\mathbf{u},j}^k : \mathcal{U}_k \subseteq \mathbb{R}^{n_u} \mapsto \mathbb{R}$, $j = 1,\dots,n_e^k + n_l^k$, *be measurable mappings from* $\mathcal{U}_k \subseteq \mathbb{R}^{n_u}$ *into* $\mathbb{R}$;
(iii) $H_{\mathbf{u},j}^k \in \mathbb{R}$, $j = 1,\dots,n_e^k + n_l^k$, *be constants;*
(iv) $h_{\mathbf{u},0}^k(\mathbf{u}) := \mathbb{1}_{\mathcal{U}_k}(\mathbf{u})$ *and* $H_{\mathbf{u},0}^k := 1$.

Find $\left\{\left(\tilde{f}_{\mathbf{U}}^k\right)^*\right\}_{\mathcal{K}} := \left\{f^*\left(\mathbf{u}_k|\mathbf{x}_{k-1}\right)\right\}_{\mathcal{K}}$ *such that:*

$$\begin{aligned}\left\{\left(\tilde{f}_{\mathbf{U}}^k\right)^*\right\}_{\mathcal{K}} \in \arg \min_{\left\{\tilde{f}_{\mathbf{U}}^k\right\}_{\mathcal{K}}} \quad & \mathcal{D}_{KL}\left(f^n||g^n\right) \\ \text{s.t.:} \quad & c_{\mathbf{u},j}^k\left[\tilde{f}_{\mathbf{U}}^k\right] = 0,\ \forall j \in \mathcal{E}_0^k, k \in \mathcal{K}, \\ & c_{\mathbf{u},j}^k\left[\tilde{f}_{\mathbf{U}}^k\right] \leq 0,\ \forall j \in \mathcal{I}^k, k \in \mathcal{K}\end{aligned} \tag{4}$$

where

$$c_{\mathbf{u},j}^k\left[\tilde{f}_{\mathbf{U}}^k\right] := \mathbb{E}_{\tilde{f}_{\mathbf{U}}^k}\left[h_{\mathbf{u},j}^k\left(\mathbf{U_k}\right)\right] - H_{\mathbf{u},j}^k, \tag{5}$$

and where $\mathcal{E}_0^k := \mathcal{E}^k \cup \{0\}$ *(with* $\mathcal{E}^k := \{j\}_1^{n_e^k}$*),* $\mathcal{I}^k := \{j\}_{n_e^k+1}^{n_e^k+n_l^k}$ *are the equality and inequality constraints index sets at time* t_k *respectively.*

Following Definition 1, the cost in (4) is well defined if f^n is absolutely continuous with respect to g^n. That is, as for other control results based on the minimization of the KL-divergence, the cost is well defined if the former pdf is 0 whenever the latter pdf is 0. The constraints in Problem 1 are specific to the system under control and we do not make any assumption on the fact that the behavior observed in the example dataset fulfills these constraints. We make the following remarks:

Remark 3 *Since the pdf* $\tilde{g}_{\mathbf{X}}^k$ *can be different from* $\tilde{f}_{\mathbf{X}}^k$*, we have that* $\tilde{g}_{\mathbf{U}}^k$ *is not, in general, the optimal solution to Problem 1. Further, since* $\tilde{g}_{\mathbf{U}}^k$ *might not satisfy the constraints in Problem 1, this pdf might not be feasible for the problem. The explicit expression for the optimal solution to Problem 1 is instead given in Theorem 1 and, as we shall see, it depends on* $\tilde{g}_{\mathbf{U}}^k$.

Remark 4 *The control problem is cast as the problem of designing* $\tilde{f}_{\mathbf{U}}^k$ *so that* $\mathcal{D}_{KL}\left(f^n||g^n\right)$ *is minimized, subject to the constraints. In this sense, the policy is designed so that* f^n approximates g^n *while still fulfilling the constrains. With Theorem 1 we give an explicit expression for the optimal solution* $\left\{\left(\tilde{f}_{\mathbf{U}}^k\right)^*\right\}_{\mathcal{K}}$*. The control input, say* $\mathbf{u_k}$*, applied to the system at time-step* k *is obtained by sampling from the pdf* $\left(\tilde{f}_{\mathbf{U}}^k\right)^*$.

As noted in (Pegueroles and Russo, 2019), in the special case when in Problem 1 there are no constraints and the pdfs f^n, g^n are normal distributions with zero mean, then the control policy solving the problem has the same update rules as the Linear Quadratic Regulator. The introduction of the constraints formalized in Problem 1 can be useful in situations of practical interest where the actuation capabilities of the system are different from the actuation capabilities of the system used to collect the example data and/or in situations where the example data are collected without necessarily knowing the

[3] As noted in (Peterka, 1981) the probabilistic descriptions are the *most general description of a system from the viewpoint of an outer observer*. These pdfs can be computed from the available data as e.g. empirical distributions

specific constraints of the system under control.Also, by embedding actuation constraints into the problem formulation and by solving the resulting problem, one can *export* the policy (by e.g. using it to generate example datasets) that has been synthesized on a given system to other systems having different actuation capabilities.

As we shall see, the solution to Problem 1 depends on the conditional pdf $\tilde{f}^k_{\mathbf{X}}$, i.e. the cpdf describing the behavior of the system under control, which therefore needs to be properly estimated. Hence, as for other data-driven control approaches that rely solely on the available data, these need to be sufficiently *informative*. We refer to e.g. (Van Waarde et al., 2020; De Persis and Tesi, 2020; van Waarde et al., 2020; Colin et al., 2020) for recent results on data informativity and to (van Waarde, 2021; Karlin and Studden, 1966) for results on optimal experimental design and persistency of excitation.

Remark 5 *In Problem 1, the constraints are formalized as expectations and can be equivalently written as* $\int_{\mathcal{S}(\tilde{f}^k_{\mathbf{U}})} \tilde{f}^k_{\mathbf{U}}\, h^k_{\mathbf{u},j}(\mathbf{u})\, d\mathbf{u} = H^k_{\mathbf{u},j}$. *The equality and inequality constraints, and their number, can change over time. In Problem 1,* n^k_e *and* n^k_l *denote, respectively, the number of equality and inequality constraints at time-step* k *(see the definitions of the sets* $\mathcal{E}^k$ *and* $\mathcal{I}^k$ *in the problem statement).Finally, the first equality constraint is a normalization constraint on the solution of the problem.*

Remark 6 *The constraints in* (4) *can be used to guarantee properties on e.g. the moments of the cdpfs* $\left(\tilde{f}^k_{\mathbf{U}}\right)^*$, *i.e. on the optimal solution of the problem (see also Remark 10 and Remark 11). Additionally, we note that the inequality constraints in (4) can be used to capture* bound constraints *of the form* $\mathbb{P}(\mathbf{U_k} \in \bar{\mathcal{U}}_k) \geq 1-\varepsilon$, *where* $\bar{\mathcal{U}}_k \subset \mathcal{U}_k$ *and* $\varepsilon \geq 0$. *Indeed, note that:* $\mathbb{P}(\mathbf{U_k} \in \bar{\mathcal{U}}_k) = \int_{\mathcal{S}(\tilde{f}^k_{\mathbf{U}})} \mathbb{1}_{\bar{\mathcal{U}}_k}(\mathbf{u}_k)\tilde{f}^k_{\mathbf{U}} d\mathbf{u}_k = \mathbb{E}_{\tilde{f}^k_{\mathbf{U}}}\left[\mathbb{1}_{\bar{\mathcal{U}}_k}(\mathbf{U}_k)\right]$. *Hence, the constraint* $\mathbb{P}(\mathbf{U_k} \in \bar{\mathcal{U}}_k) \geq 1-\varepsilon$ *can be written as a constraint of the form of these in (4). Typically, in works on* safe learning *for stochastic systems, see e.g. (McKinnon and Schoellig, 2019; Nakka et al., 2021),* ε *is a small constant and hence these types of constraints model the fact that the probability that the control variable is outside some (e.g. desired) set* $\bar{\mathcal{U}}_k$ *is smaller than some* acceptable ε. *When* $\varepsilon = 0$, *the constraint amounts at imposing that* $\mathbb{P}(\mathbf{U_k} \in \bar{\mathcal{U}}_k) = 1$, *thus implying that the pdf* $\tilde{f}^k_{\mathbf{U}}$ *is zero outside the set* $\bar{\mathcal{U}}_k$.

Remark 7 *In relation to Remark 6, we also note how the constraint* $\mathbb{P}(\mathbf{U_k} \in \bar{\mathcal{U}}_k) = \mathbb{E}_{\tilde{f}^k_{\mathbf{U}}}\left[\mathbb{1}_{\bar{\mathcal{U}}_k}(\mathbf{U}_k)\right] \geq 1-\varepsilon$ *is linear, and hence convex, in the decision variable (i.e.* $\tilde{f}^k_{\mathbf{U}}$*) even if the indicator function is not convex, see e.g. (Vitus and Tomlin, 2013). This implies that, in Problem 1, these constraints can be handled without resorting to bound approximations. These approximations are typically used in the literature to handle the intrinsic non-convexity of the constraint arising in problems where the decision variable is* $\mathbf{U}_k$, *see e.g. (Vitus and Tomlin, 2013; Moser et al., 2017). Note indeed that, while* $\mathbb{E}_{\tilde{f}^k_{\mathbf{U}}}\left[\mathbb{1}_{\bar{\mathcal{U}}_k}(\mathbf{U}_k)\right]$ *is convex in* $\tilde{f}^k_{\mathbf{U}}$, *this is not convex in* $\mathbf{U}_k$.

4 Technical Results

We now introduce our main results. The key result behind the algorithm of Section 5 is Theorem 1. The proof of this theorem, given in this section, makes use of two technical results (i.e. Lemma 1 and Lemma 2). With our first result, i.e. Lemma 1, we tackle an *auxiliary* problem that is iteratively solved within the proof of Theorem 1. This auxiliary problem is formalized next.

Problem 2 *Let:*

(i) n_e, n_l *be two positive integers;*
(ii) $\mathbf{Z}$ *be a random vector, with* $\mathbf{z} \in \mathcal{Z} \subseteq \mathbb{R}^{n_z}$;
(iii) $g(\mathbf{z})$ *and* $f(\mathbf{z})$ *be two pdfs having support* $\mathcal{Z}$;
(iv) $\alpha : \mathcal{Z} \mapsto \mathbb{R}$ *be a mapping from* $\mathcal{Z}$ *into* $\mathbb{R}$, *which is integrable under the measure given by* $f(\mathbf{z})$;
(v) $h_j : \mathcal{Z} \mapsto \mathbb{R}$, $j = 1, \ldots, n_e + n_l$, *be measurable mappings from* $\mathcal{Z}$ *into* $\mathbb{R}$;
(vi) $H_j \in \mathbb{R}$, $j = 1, \ldots, n_e + n_l$, *be a constant;*
(vii) $h_0(\mathbf{z}) := \mathbb{1}_{\mathcal{Z}}(\mathbf{z})$ *and* $H_0 := 1$.

Find the pdf $f^*(\mathbf{z})$ *such that:*

$$f^*(\mathbf{z}) \in \arg\min_{f(\mathbf{z})} \mathcal{L}\left(f(\mathbf{z})\right) \quad \text{s.t.: } c_j\left[f(\mathbf{z})\right] = 0,\ \forall j \in \mathcal{E}_0, \quad c_j\left[f(\mathbf{z})\right] \leq 0,\ \forall j \in \mathcal{I} \tag{6}$$

where:

$$\mathcal{L}\left(f(\mathbf{z})\right) := \mathcal{D}_{KL}\left(f(\mathbf{z})||g(\mathbf{z})\right) + \int f(\mathbf{z})\,\alpha(\mathbf{z})\,d\mathbf{z}, \tag{7}$$

with

$$c_j\left[f(\mathbf{z})\right] := \int f(\mathbf{z})\ h_j(\mathbf{z})\ d\mathbf{z} - H_j, \tag{8}$$

and $\mathcal{E}_0 := \mathcal{E} \cup \{0\}$, $\mathcal{E} := \{j\}_1^{n_e}$, $\mathcal{I} := \{j\}_{n_e+1}^{n_e+n_l}$.

In Lemma 1 we give the optimal solution of Problem 2. Such a result is used in Theorem 1 to find the solution to Problem 1. Consider the special case where, in Problem 1: (i) the function $\alpha(\cdot)$ is constant in $\mathcal{Z}$; (ii) the support $\mathcal{Z}$ is compact and $g(\mathbf{z})$ is uniform. Then, the cost in Problem 2 becomes $\mathcal{L}(f(\mathbf{z})) = \int f(\mathbf{z}) \log f(\mathbf{z}) d\mathbf{z}$. Hence, in this case, Problem 2 becomes a constrained entropy maximization problem. See also Remark 14. We consider feasible sets of constraints that satisfy the following constraint qualification condition (CQC):

Definition 2 *Consider the set of linear constraints*

$$\begin{cases} c_i\left[f(\mathbf{z})\right] = 0,\ i = 1, \dots, n_e \\ c_j\left[f(\mathbf{z})\right] \leq 0,\ j = 1, \dots, n_l. \end{cases}$$

The Slater's CQC (or simply Slater's condition in what follows) is said to hold for such a set if there exists a pdf, say $\bar{f}(\mathbf{z})$, such that $c_i[\bar{f}(\mathbf{z})] = 0$, $\forall i \in \{1, \dots, n_e\}$ and $c_j[\bar{f}(\mathbf{z})] < 0$, $\forall j \in \{1, \dots, n_l\}$.

Remark 8 *Slater's condition is also known in the literature on infinite-dimensional optimization problems as a* regularity *condition on the constraint set (Rockafeller, 1976; Ben-Tal et al., 1988). For a convex infinite-dimensional optimization problem, the fulfillment of such a condition guarantees strong duality (Ben-Tal et al., 1988).*

We make the following:

Assumption 1 *The constraints sets in (4) and in (6) are feasible and satisfy Slater's condition.*

Remark 9 *Assumption 1 is widely used in the literature on e.g. infinite-dimensional optimization (Ben-Tal et al., 1988; Fan, 1968), divergences optimization and cross-entropy problems (Singh and Vishnoi, 2014; Bot et al., 2005). If the problems involve discrete distributions, then checking feasibility and the Slater's condition implies solving (finite dimensional) systems of equalities and inequalities, see e.g. (Duffin et al., 1956; Fan, 1975; Hiebert, 1980; Censor and Elfving, 1982). For problems with continuous variables, an approach to check the condition consists in building an* initial *pdf for which equality constraints are satisfied and inequality constraints are satisfied strictly. This approach has been used in Section 6 to verify the fulfillment of Assumption 1 for our numerical example.*

We are now ready to introduce the next result, which gives a solution to Problem 2.

Lemma 1 *Consider Problem 2. Then:*

(R1) *the problem has a unique solution and this is given by the pdf*

$$f^*(\mathbf{z}) = g(\mathbf{z})\ \frac{e^{-\left\{\alpha(\mathbf{z}) + \sum_{j \in \mathcal{I}_a(f^*(\mathbf{z})) \setminus \{0\}} \lambda_j^* h_j(\mathbf{z})\right\}}}{e^{1+\lambda_0^*}}. \tag{9}$$

In (9), λ_j^* is the Lagrange multiplier associated to the constraint $c_j\left[f(\mathbf{z})\right]$ and $\mathcal{I}_a\left(f^*(\mathbf{z})\right)$ is the active index set defined as

$$\mathcal{I}_a\left(f\right) := \mathcal{E}_0 \cup \{j \in \mathcal{I} : c_j\left[f\right] = 0\}. \tag{10}$$

In (9) the Lagrange multipliers

$$\boldsymbol{\lambda}^* := [\lambda_0^*, \lambda_1^*, \dots, \lambda_{n_e+n_l}^*]^T,$$

can be computed by solving the optimization problem

$$\begin{aligned} \boldsymbol{\lambda}^* \in \arg\ \max_{\boldsymbol{\lambda}}\ & \mathcal{L}^D\left(\boldsymbol{\lambda}\right) \\ \text{s.t.:}\ & \lambda_j \text{ free},\ \forall j \in \mathcal{E}_0, \\ & \lambda_j \geq 0,\ \forall j \in \mathcal{I} \end{aligned} \tag{11}$$

where

$$\mathcal{L}^D\left(\boldsymbol{\lambda}\right) := -\langle \boldsymbol{\lambda}, \mathbf{H} \rangle - \int g\left(\mathbf{z}\right)\ e^{-\{1+\alpha(\mathbf{z})+\langle \boldsymbol{\lambda}, \mathbf{h}(\mathbf{z}) \rangle\}} d\mathbf{z}. \tag{12}$$

(R2) *Moreover, the corresponding minimum is $\mathcal{L}\left(f^*(\mathbf{z})\right) = -\left(1 + \sum_{j \in \mathcal{I}_a(f^*(\mathbf{z}))} \lambda_j^*\, H_j\right)$.*

Proof: See the appendix. □

Before introducing the next technical result, we make the following remarks on Lemma 1.

Remark 10 *The equality constraints in* (8) *can be used to impose parametric prescriptions on the solution. For example, one could impose that $f^*(\mathbf{z})$ has the central moment of order i equal to some $m_{\mathbf{Z}}^i$. This is equivalent to impose that the solution satisfies $\mathbb{E}_f\left[\mathbf{Z}^i\right] = m_{\mathbf{Z}}^i$, which in turn can be formalized as $c_i\left[f(\mathbf{z})\right] := \int f\left(\mathbf{z}\right)\ \mathbf{z}^i\ d\mathbf{z} - m_{\mathbf{Z}}^i$. These types of equality constraints, also arise within the literature on the approximation of spectral density functions with respect to the KL-divergence under moment constraints, where linear systems are typically considered. See e.g. (Georgiou and Lindquist, 2003; Pavon and Ferrante, 2006; Zhu and Baggio, 2019).*

Remark 11 *The inequality constraints in* (8) *can also be used to assign properties to the solution: with these constraints, one could express bounds on the expected value of any function of $\mathbf{Z}$, say $h(\mathbf{Z})$. For instance, the rectangular bound $\underline{m}_{\mathbf{Z}}^2 \leq \mathbb{E}_f\left[\mathbf{Z}^2\right] \leq \overline{m}_{\mathbf{Z}}^2$ can be formalized with the pair of inequality constraints:*

$$\begin{cases} c_a\left[f(\mathbf{z})\right] := \int f\left(\mathbf{z}\right)\ \mathbf{z}^2\ d\mathbf{z} - \ \overline{m}_{\mathbf{Z}}^2, \\ c_b\left[f(\mathbf{z})\right] := -\int f\left(\mathbf{z}\right)\ \mathbf{z}^2\ d\mathbf{z} + \ \underline{m}_{\mathbf{Z}}^2. \end{cases}$$

Remark 12 *The Lagrange multiplier λ_0 in* (9) *can be expressed as a function of all the other Lagrange multipliers. This can be done by imposing the normalization constraint, i.e. $c_0\left[f^*(\mathbf{z})\right] = 0$, on the pdf solving the problem, i.e.* (9)*. This yields:*

$$1 + \lambda_0 = \ln\left(\int g\left(\mathbf{z}\right)\ e^{-\left\{\alpha(\mathbf{z}) + \sum_{j \in \mathcal{I}_a(f^*(\mathbf{z})) \setminus \{0\}} \lambda_j h_j(\mathbf{z})\right\}} d\mathbf{z}\right),$$

which can also be obtained by imposing the stationarity condition of the Lagrange dual function with respect to λ_0. Also, the expression for λ_0 above could be directly embedded in the dual cost function (12), yielding $\mathcal{L}^D(\boldsymbol{\lambda}) = -\sum_{j\in\mathcal{E}\cup\mathcal{I}}\lambda_j H_j - \ln\left(\int g(\mathbf{z})\, e^{-\left\{\alpha(\mathbf{z})+\sum_{j\in\mathcal{E}\cup\mathcal{I}}\lambda_j h_j(\mathbf{z})\right\}} d\mathbf{z}\right)$*, and thus reducing by one the dimension of the search space of the dual problem.*

We now introduce the following technical result that is also used in the proof of Theorem 1.

Lemma 2 *Let f^n and g^n be the joint pdfs defined in* (2) *and* (3)*, respectively. Then:*

$$\mathcal{D}_{KL}(f^n||g^n) = \mathcal{D}_{KL}\left(f^{n-1}||g^{n-1}\right) + \mathbb{E}_{f^{n-1}}\left[\mathcal{D}_{KL}\left(\tilde{f}^n||\tilde{g}^n\right)\right],$$

where $\tilde{f}^n$ and $\tilde{g}^n$ are the conditional pdfs defined as in Section 3, i.e. $\tilde{f}^n := \tilde{f}^n_{\mathbf{X}}\tilde{f}^n_{\mathbf{U}}$ and $\tilde{g}^n := \tilde{g}^n_{\mathbf{X}}\tilde{g}^n_{\mathbf{U}}$.

Proof: The result is obtained from Property 1 (see the appendix for a proof of this property) by setting $\mathbf{Y} := [\mathbf{X}_0, \mathbf{U}_1, \mathbf{X}_1, \ldots, \mathbf{U}_{n-1}, \mathbf{X}_{n-1}]$ and $\mathbf{Z} := [\mathbf{U}_n, \mathbf{X}_n]$. □

The main result behind the algorithm of Section 5, the proof of which makes use of the above technical results, is presented next.

Theorem 1 *Consider Problem 1. Then:*

(R1) *The control policy at time instant t_k, $\left(\tilde{f}^k_{\mathbf{U}}\right)^*$, composing $\left\{\left(\tilde{f}^k_{\mathbf{U}}\right)^*\right\}_{\mathcal{K}}$ solving the problem is given by*

$$\left(\tilde{f}^k_{\mathbf{U}}\right)^* = \tilde{g}^k_{\mathbf{U}} \frac{e^{-\{\hat{\omega}(\mathbf{u}_k,\mathbf{x}_{k-1})+\sum_{j\in\mathcal{I}_a{}^k\setminus\{0\}}\left(\lambda^k_{\mathbf{u},j}\right)^* h^k_{\mathbf{u},j}(\mathbf{u}_k)\}}}{e^{1+\left(\lambda^k_{\mathbf{u},0}\right)^*}}, \tag{13}$$

where:

- *$\hat{\omega}(\cdot,\cdot)$ is generated via the backward recursion*

$$\hat{\omega}(\mathbf{u}_k, \mathbf{x}_{k-1}) = \hat{\alpha}(\mathbf{u}_k, \mathbf{x}_{k-1}) + \hat{\beta}(\mathbf{u}_k, \mathbf{x}_{k-1}), \tag{14}$$

with

$$\begin{aligned}\hat{\alpha}(\mathbf{u}_k, \mathbf{x}_{k-1}) &:= \mathcal{D}_{KL}\left(\tilde{f}^k_{\mathbf{X}}||\tilde{g}^k_{\mathbf{X}}\right),\\ \hat{\beta}(\mathbf{u}_k, \mathbf{x}_{k-1}) &:= -\mathbb{E}_{\tilde{f}^k_{\mathbf{X}}}\left[\ln\hat{\gamma}(\mathbf{X}_k)\right],\end{aligned} \tag{15}$$

and terminal conditions $\hat{\beta}(\mathbf{u}_n, \mathbf{x}_{n-1}) = 0$, $\hat{\alpha}(\mathbf{u}_n, \mathbf{x}_{n-1}) = \mathcal{D}_{KL}\left(\tilde{f}^n_{\mathbf{X}}||\tilde{g}^n_{\mathbf{X}}\right)$;

- *$\hat{\gamma}(\cdot)$ in* (15) *is defined as*

$$\ln\hat{\gamma}(\mathbf{x}_{k-1}) := \sum_{j\in\mathcal{I}_a{}^k}\ln\left(\hat{\gamma}^k_{\mathbf{u},j}(\mathbf{x}_{k-1})\right), \tag{16}$$

with

$$\hat{\gamma}^k_{\mathbf{u},0}(\mathbf{x}_{k-1}) = \exp\left\{\left(\lambda^k_{\mathbf{u},0}\right)^* + 1\right\},\ \ \hat{\gamma}^{n+1}_{\boldsymbol{u},0}(\mathbf{x}_n) = 1, \tag{17}$$

and

$$\hat{\gamma}^k_{\mathbf{u},j}(\mathbf{x}_{k-1}) := \exp\left\{\left(\lambda^k_{\mathbf{u},j}\right)^* H^k_{\mathbf{u},j}\right\},\ \hat{\gamma}^{n+1}_{\mathbf{u},j}(\mathbf{x}_n) = 1, \tag{18}$$

$\forall j \in \mathcal{E}^k \cup \mathcal{I}^k$;

- *$\left(\lambda^k_{\mathbf{u},j}\right)^*$ in (13) is the Lagrange multiplier associated to the constraint $c^k_{\mathbf{u},j}$, while $\mathcal{I}_a{}^k$ is the active index set associated to $\left(\tilde{f}^k_{\mathbf{U}}\right)^*$. In particular, the vector of Lagrange multipliers $\left(\boldsymbol{\lambda}^k_{\mathbf{u}}\right)^* := \left[\left(\lambda^k_{\mathbf{u},0}\right)^*, \left(\lambda^k_{\mathbf{u},1}\right)^*, \ldots, \left(\lambda^k_{\mathbf{u},n^k_e+n^k_l}\right)^*\right]^T$ can be computed solving:*

$$\begin{aligned}\left(\boldsymbol{\lambda}^k_{\mathbf{u}}\right)^* \in \arg\max_{\boldsymbol{\lambda}^k_{\mathbf{u}}}\ & \mathcal{L}^D\left(\boldsymbol{\lambda}^k_{\mathbf{u}}\right)\\ \text{s.t.:}\ & \lambda^k_{\mathbf{u},j}\ \text{free},\ \forall j\in\mathcal{E}^k\\ & \lambda^k_{\mathbf{u},j}\geq 0,\ \forall j\in\mathcal{I}^k,\end{aligned} \tag{19}$$

where

$$\begin{aligned}\mathcal{L}^D\left(\boldsymbol{\lambda}^k_{\mathbf{u}}\right) = &-\sum_{j\in\mathcal{E}^k\cup\mathcal{I}^k}\lambda^k_{\mathbf{u},j}H^k_{\mathbf{u},j} -\\ &\ln\Big(\int\tilde{g}^k_{\mathbf{U}}\, e^{-\hat{\omega}(\mathbf{u}_k,\mathbf{x}_{k-1})}\\ &\quad e^{-\left\{\sum_{j\in\mathcal{E}^k\cup\mathcal{I}^k}\lambda^k_{\mathbf{u},j}\,h^k_{\mathbf{u},j}(\mathbf{u}_k)\right\}} d\mathbf{u}_k\Big),\end{aligned}$$

with $\left(\lambda^k_{\mathbf{u},0}\right)^*$ given by

$$\begin{aligned}\left(\lambda^k_{\mathbf{u},0}\right)^* = \ln\Bigg(&\int\tilde{g}^k_{\mathbf{U}}e^{-\hat{\omega}(\mathbf{u}_k,\mathbf{x}_{k-1})}\\ &e^{-\left\{\sum_{j\in\mathcal{I}_a{}^k\setminus\{0\}}\left(\lambda^k_{\mathbf{u},j}\right)^* h^k_{\mathbf{u},j}(\mathbf{u}_k)\right\}} d\mathbf{u}_k\Bigg) - 1.\end{aligned} \tag{20}$$

(R2) *Moreover, the corresponding minimum at time t_k is given by:*

$$B^*_k := -\mathbb{E}_{p^{k-1}_{\mathbf{X}}}\left[\ln\hat{\gamma}(\mathbf{X}_{k-1})\right]. \tag{21}$$

where $p^k_{\mathbf{X}}$ denotes the pdf of the state at time t_k (i.e. $p^k_{\mathbf{X}} := f(\mathbf{x}_k)$).

Before giving the proof we make the following remark:

Remark 13 *The policy solving Problem 1, i.e.* $\left(\tilde{f}^k_{\mathbf{U}}\right)^*$, *directly depends on* $\tilde{g}^k_{\mathbf{U}}$. *This is the policy extracted from the examples and a natural choice is to estimate it from the available data (see also the example in Section 6). In principle, one could use a* $\tilde{g}^k_{\mathbf{U}}$ *that, while not being extracted from the examples, embeds design preferences that might be known a-priori. If full knowledge of the system (and of its constraints) is available, one can build a* synthetic *joint pdf* g^n *that embeds the desired design properties. While our results can be used in this* ideal *situation, we remark here that building such synthetic pdf is not always possible in situations of practical interest where the example data are collected from a system that is not the same and/or does not have the same constraints as the system under control (or the constraints of the system under control are not fully known). Finally, another choice, for pdfs having compact supports, is to set* $\tilde{g}^k_{\mathbf{U}}$ *equal to the uniform distribution.*

Proof: we prove the result by induction and the proof is organized in steps. First, in **Step 1**, we leverage Lemma 2 to show that Problem 1 can be split into sub-problems, where the optimization sub-problem for the last iteration, i.e. $k = n$, can be solved independently on the others. We then (**Step 2**) make use of Lemma 1 to find an explicit solution for the sub-problem at $k = n$. Once this is done, we update the cost of Problem 1 with the minimum found by solving the sub-problem at $k = n$ and show, in **Step 3**, that the original problem can be again broken down into sub-problems. This time, the sub-problem at iteration $k = n-1$ can be solved independently on the others. We then solve this sub-problem and note, in **Step 4**, how for all the remaining time-steps the structure of the optimization remains the same. From this, the desired conclusions are drawn.

Before proceeding with the proof note that, for notational convenience, we use the shorthand notation $\left\{\mathbf{C}^k_{\mathbf{u}}\right\}$ to denote the set of constraints of Problem 1 at iteration k. We also denote by $\left\{\mathbf{C}^k_{\mathbf{u}}\right\}_{\mathcal{K}}$ the set of constraints over $\mathcal{K}$ and by $\left\{\mathbf{C}^k_{\mathbf{u}}\right\}_{k=1}^{n-1}$ the constraints from t_1 up to t_{n-1}.

Step 1. Note that, following Lemma 2, Problem 1 can be re-written as

$$\begin{aligned} &\min_{\left\{\tilde{f}^k_{\mathbf{U}}\right\}_{\mathcal{K}}} \ \mathcal{D}_{\mathrm{KL}}\left(f^n||g^n\right) = \\ &\text{s.t.:} \ \left\{\mathbf{C}^k_{\mathbf{u}}\right\}_{\mathcal{K}} \\ &= \min_{\left\{\tilde{f}^k_{\mathbf{U}}\right\}_{k=1}^{n-1}} \left\{\mathcal{D}_{\mathrm{KL}}\left(f^{n-1}||g^{n-1}\right) + B^*_n\right\} \\ &\quad \text{s.t.:} \ \left\{\mathbf{C}^k_{\mathbf{u}}\right\}_{k=1}^{n-1} \end{aligned} \tag{22}$$

where:

$$B^*_n := \min_{\tilde{f}^n_{\mathbf{U}}} B_n \quad \text{s.t.:} \ \mathbf{C}^n_{\mathbf{u}} \tag{23a}$$

and

$$B_n := \mathbb{E}_{f^{n-1}}\left[\mathcal{D}_{\mathrm{KL}}\left(\tilde{f}^n||\tilde{g}^n\right)\right]. \tag{23b}$$

That is, Problem 1 can be approached by solving first the optimization of the last iteration of the horizon $\mathcal{K}$ (the term B_n in (22)) and then by taking into account the result from this optimization problem in the optimization up to iteration $n-1$.

Step 2. We first observe that, for B_n defined in (23a):

$$B_n = \mathbb{E}_{f^{n-1}}\left[\mathcal{D}_{\mathrm{KL}}\left(\tilde{f}^n||\tilde{g}^n\right)\right] = \mathbb{E}_{p^{n-1}_{\mathbf{X}}}\left[\mathcal{D}_{\mathrm{KL}}\left(\tilde{f}^n||\tilde{g}^n\right)\right].$$

The above expression was obtained by noticing that, by definition of $\tilde{f}^n$ and $\tilde{g}^n$ (see soon before (2) and (3)), $\mathcal{D}_{\mathrm{KL}}\left(\tilde{f}^n||\tilde{g}^n\right)$ is a function of the previous state and, to stress this in the notation, we let $\hat{A}\left(\cdot\right) := \mathcal{D}_{\mathrm{KL}}\left(\tilde{f}^n||\tilde{g}^n\right)$. Hence, B_n can be written as

$$B_n = \mathbb{E}_{p^{n-1}_{\mathbf{X}}}\left[\mathcal{D}_{\mathrm{KL}}\left(\tilde{f}^n||\tilde{g}^n\right)\right] = \mathbb{E}_{p^{n-1}_{\mathbf{X}}}\left[\hat{A}\left(\mathbf{X}_{n-1}\right)\right], \tag{24}$$

and the sub-problem in (23a) becomes:

$$B^*_n = \min_{\tilde{f}^n_{\mathbf{U}}} \mathbb{E}_{p^{n-1}_{\mathbf{X}}}\left[\hat{A}\left(\mathbf{X}_{n-1}\right)\right] \quad \text{s.t.:} \ \mathbf{C}^n_{\mathbf{u}} \tag{25}$$

Now, note that

$$\begin{aligned} &\min_{\tilde{f}^n_{\mathbf{U}}} \ \mathbb{E}_{p^{n-1}_{\mathbf{X}}}\left[\hat{A}\left(\mathbf{X}_{n-1}\right)\right] = \mathbb{E}_{p^{n-1}_{\mathbf{X}}}\left[A^*_n\right], \\ &\text{s.t.:} \ \mathbf{C}^n_{\mathbf{u}} \end{aligned} \tag{26}$$

where

$$A^*_n := \min_{\tilde{f}^n_{\mathbf{u}}} \hat{A}(\mathbf{x}_{n-1}) \quad \text{s.t.:} \ \mathbf{C}^n_{\mathbf{u}} \tag{27}$$

Also, the equality in (26) was obtained by using the fact that the expectation operator is linear and the fact that the decision variable (i.e. $\tilde{f}^n_{\mathbf{U}}$) is independent on the pdf over which the expectation is performed (i.e. $p^{n-1}_{\mathbf{X}}$).

Following (26), we can obtain B^*_n by solving (27) and then by averaging A^*_n over $p^{n-1}_{\mathbf{X}}$. We now focus on solving problem (27). From (24), we get:

$$\hat{A}\left(\mathbf{x}_{n-1}\right) = \int \tilde{f}^n_{\mathbf{U}}\left[\ln\left(\frac{\tilde{f}^n_{\mathbf{U}}}{\tilde{g}^n_{\mathbf{U}}}\right) + \hat{\alpha}\left(\mathbf{u}_n, \mathbf{x}_{n-1}\right)\right] d\mathbf{u}_n, \tag{28a}$$

$$\hat{\alpha}\left(\mathbf{u}_n, \mathbf{x}_{n-1}\right) := \mathcal{D}_{\mathrm{KL}}\left(\tilde{f}^n_{\mathbf{X}}||\tilde{g}^n_{\mathbf{X}}\right). \tag{28b}$$

In turn, (28a) can be compactly written as:

$$\hat{A}(\mathbf{x}_{n-1}) = \mathcal{D}_{\mathrm{KL}}\left(\tilde{f}^n_{\mathbf{U}}||\tilde{g}^n_{\mathbf{U}}\right) + \int \tilde{f}^n_{\mathbf{U}}\, \hat{\alpha}\left(\mathbf{u}_n, \mathbf{x}_{n-1}\right) d\mathbf{u}_n,$$

where we used the definition of KL-divergence.

Hence, Lemma 1 can be used to solve the optimization problem in (27). Indeed by applying Lemma 1 with: $\mathbf{Z} = \mathbf{U}_n$, $f(\mathbf{z}) = \tilde{f}^n_{\mathbf{U}}$, $g(\mathbf{z}) = \tilde{g}^n_{\mathbf{U}}$, $\alpha(\cdot) = \hat{\alpha}(\cdot, \mathbf{x}_{n-1})$, $h_j(\mathbf{z}) = h^n_{\mathbf{u},j}(\mathbf{u}_n)$, $H_j = H^n_{\mathbf{u},j}$, $c_j[\cdot] = c^n_{\mathbf{u},j}[\cdot]$, $\lambda_j = \lambda^n_{\mathbf{u},j}$, $\mathcal{E} = \mathcal{E}^n$, $\mathcal{I} = \mathcal{I}^n$, we get the following solution to (27):

$$\left(\tilde{f}^n_{\mathbf{U}}\right)^* = \tilde{g}^n_{\mathbf{U}} \frac{e^{-\{\hat{\alpha}(\mathbf{u}_n, \mathbf{x}_{n-1}) + \sum_{j \in {\mathcal{I}_a}^n \setminus \{0\}} \left(\lambda^n_{\mathbf{u},j}\right)^* h^n_{\mathbf{u},j}(\mathbf{u}_n)\}}}{e^{1+\left(\lambda^n_{\mathbf{u},0}\right)^*}}, \tag{29}$$

where ${\mathcal{I}_a}^n$ is the active set index associated to $\left(\tilde{f}^n_{\mathbf{U}}\right)^*$. In the above pdf, $\left(\lambda^n_{\mathbf{u},j}\right)^*, j \in \mathcal{E}^n_0 \cup \mathcal{I}^n$ are the Lagrange multipliers at the last iteration $k = n$. Now, following Lemma 1 and Remark 12, the Lagrange multipliers $(\boldsymbol{\lambda}^n_{\mathbf{u}})^* = \left[\left(\lambda^n_{\mathbf{u},0}\right)^*, \left(\lambda^n_{\mathbf{u},1}\right)^*, \ldots, \left(\lambda^n_{\mathbf{u},n^n_e+n^n_l}\right)^*\right]^T$ are computed by solving

$$\begin{aligned}(\boldsymbol{\lambda}^n_{\mathbf{u}})^* \in \arg\max\ & \mathcal{L}^D\left(\boldsymbol{\lambda}^n_{\mathbf{u}}\right) \\ \text{s.t.:}\ & \lambda^n_{\mathbf{u},j} \text{ free}, \ \forall j \in \mathcal{E}^n, \\ & \lambda^n_{\mathbf{u},j} \geq 0, \ \forall j \in \mathcal{I}^n\end{aligned}$$

choosing $\left[\left(\lambda^n_{\mathbf{u},1}\right)^*, \ldots, \left(\lambda^n_{\mathbf{u},n^n_e+n^n_l}\right)^*\right]^T$ so that $\left(\tilde{f}^n_{\mathbf{U}}\right)^*$ is feasible. In the above expression $\mathcal{L}^D\left(\boldsymbol{\lambda}^n_{\mathbf{u}}\right)$ is defined as:

$$\begin{aligned}\mathcal{L}^D\left(\boldsymbol{\lambda}^n_{\mathbf{u}}\right) = -\sum_{j \in \mathcal{E}^n \cup \mathcal{I}^n} \lambda^n_{\mathbf{u},j} H^n_{\mathbf{u},j} - \ln\Bigg(\int \tilde{g}^n_{\mathbf{U}}\, e^{-\hat{\alpha}(\mathbf{u}_n, \mathbf{x}_{n-1})} \\ \cdot\, e^{-\left\{\sum_{j \in \mathcal{E}^n \cup \mathcal{I}^n} \lambda^n_{\mathbf{u},j} h^n_{\mathbf{u},j}(\mathbf{z})\right\}} d\mathbf{u}_n\Bigg),\end{aligned}$$

and $\left(\lambda^n_{\mathbf{u},0}\right)^*$ can be obtained from all the other Lagrange multipliers by normalizing (29), i.e.:

$$\begin{aligned}\left(\lambda^n_{\mathbf{u},0}\right)^* + 1 &= \ln\Big(\int \tilde{g}^n_{\mathbf{U}} e^{-\hat{\alpha}(\mathbf{u}_n, \mathbf{x}_{n-1})} \\ &\qquad e^{-\left\{\sum_{j \in {\mathcal{I}_a}^n \setminus \{0\}} \left(\lambda^n_{\mathbf{u},j}\right)^* h^n_{\mathbf{u},j}(\mathbf{u}_n)\right\}} d\mathbf{u}_n\Big) \\ &= \ln\left(\hat{\gamma}^n_{\mathbf{u},0}\left(\mathbf{x}_{n-1}\right)\right).\end{aligned}$$

Moreover, from Lemma 1, the minimum of the problem in (27) is given by:

$$\hat{A}^*_n = -\left(1 + \sum_{j \in {\mathcal{I}_a}^n} \left(\lambda^n_{\mathbf{u},j}\right)^* H^n_{\mathbf{u},j}\right),$$

which, using the definitions in (17) and (18), can be equivalently written as

$$\hat{A}^*_n = -\left[\sum_{j \in {\mathcal{I}_a}^n} \ln\left(\hat{\gamma}^n_{\mathbf{u},j}\left(\mathbf{x}_{n-1}\right)\right)\right] = -\ln\hat{\gamma}\left(\mathbf{x}_{n-1}\right).$$

Thus, we get:

$$B^*_n = -\mathbb{E}_{p^{n-1}_{\mathbf{X}}}\left[\ln\hat{\gamma}\left(\mathbf{X}_{n-1}\right)\right]. \tag{30}$$

Step 3. Note now that the B^*_n in (30) only depends on $\mathbf{X}_{n-1}$ and therefore the original problem (22) can be split, following Lemma 2, as

$$\begin{aligned}&\min_{\{\tilde{f}^k_{\mathbf{U}}\}_{k=1}^{n-1}} \left\{\mathcal{D}_{\mathrm{KL}}\left(f^{n-1}||g^{n-1}\right) + B^*_n\right\} \\ &\text{s.t.:}\quad \left\{\mathbf{C}^k_{\mathbf{u}}\right\}_{k=1}^{n-1} \\ &= \min_{\{\tilde{f}^k_{\mathbf{U}}\}_{k=1}^{n-2}} \left\{\mathcal{D}_{\mathrm{KL}}\left(f^{n-2}||g^{n-2}\right) + B^*_{n-1}\right\} \\ &\qquad \text{s.t.:}\quad \left\{\mathbf{C}^k_{\mathbf{u}}\right\}_{k=1}^{n-2}\end{aligned} \tag{31}$$

where:

$$B^*_{n-1} := \min_{\tilde{f}^{n-1}_{\mathbf{U}}} B_{n-1} \quad \text{s.t.: } \mathbf{C}^{n-1}_{\mathbf{u}} \tag{32a}$$

and

$$B_{n-1} := \mathbb{E}_{f^{n-2}}\left[\mathcal{D}_{\mathrm{KL}}\left(\tilde{f}^{n-1}||\tilde{g}^{n-1}\right)\right] + B^*_n. \tag{32b}$$

We approach the above problem in the same way we used to solve the problem in (25). We do this by finding a function, $\hat{A}\left(\mathbf{x}_{n-2}\right)$, such that $B_{n-1} = \mathbb{E}_{p^{n-2}_{\mathbf{X}}}\left[\hat{A}\left(\mathbf{X}_{n-2}\right)\right]$. Once this is done, we then get

$$A^*_{n-1} := \min_{\tilde{f}^{n-1}_{\mathbf{U}}} \hat{A}(\mathbf{x}_{n-2}) \quad \text{s.t.: } \mathbf{C}^{n-1}_{\mathbf{u}} \tag{33}$$

and obtain B^*_{n-1} as $B^*_{n-1} := \mathbb{E}_{p^{n-2}_{\mathbf{X}}}\left[A^*_{n-1}\right]$. To this end we first note that the following identities

$$\mathbb{E}_{p^{n-1}_{\mathbf{X}}}\left[\varphi\left(\mathbf{X}_{n-1}\right)\right] = \mathbb{E}_{p^{n-2}_{\mathbf{X}}}\left[\mathbb{E}_{\tilde{f}^{n-1}}\left[\varphi\left(\mathbf{X}_{n-1}\right)\right]\right], \tag{34a}$$

$$\mathbb{E}_{\tilde{f}^{n-1}_{\mathbf{X}}}\left[\varphi\left(\mathbf{X}_{n-1}\right)\right] = \mathbb{E}_{\tilde{f}^{n-2}_{\mathbf{X}}}\left[\mathbb{E}_{\tilde{f}^{n-1}}\left[\varphi\left(\mathbf{X}_{n-1}\right)\right]\right], \tag{34b}$$

hold for any function φ of $\mathbf{X}_{n-1}$. Therefore, by means of

(30) and (34a) we obtain, from (32b):

$$
\begin{aligned}
B_{n-1} &= \mathbb{E}_{f^{n-2}}\left[\mathcal{D}_{\mathrm{KL}}\left(\tilde{f}^{n-1}||\tilde{g}^{n-1}\right)\right] + B_n^* \\
&= \mathbb{E}_{p_{\mathbf{x}}^{n-2}}\left[\mathcal{D}_{\mathrm{KL}}\left(\tilde{f}^{n-1}||\tilde{g}^{n-1}\right)\right] + B_n^* \\
&= \mathbb{E}_{p_{\mathbf{x}}^{n-2}}\left[\mathcal{D}_{\mathrm{KL}}\left(\tilde{f}^{n-1}||\tilde{g}^{n-1}\right)\right] - \mathbb{E}_{p_{\mathbf{x}}^{n-2}}\left[\mathbb{E}_{\tilde{f}^{n-1}}\left[\ln\hat{\gamma}\left(\mathbf{X}_{n-1}\right)\right]\right] \\
&= \mathbb{E}_{p_{\mathbf{x}}^{n-2}}\left[\underbrace{\mathcal{D}_{\mathrm{KL}}\left(\tilde{f}^{n-1}||\tilde{g}^{n-1}\right) + \mathbb{E}_{\tilde{f}^{n-1}}\left[-\ln\hat{\gamma}\left(\mathbf{X}_{n-1}\right)\right]}_{=:\hat{A}(\mathbf{X}_{n-2})}\right],
\end{aligned}
$$

and the term $\hat{A}\left(\mathbf{x}_{n-2}\right)$ can hence be recognized. Now, following the same reasoning we used to obtain $\hat{A}(\mathbf{x}_{n-1})$, we explicitly write $\hat{A}(\mathbf{x}_{n-2})$ in compact form as

$$
\begin{aligned}
&\hat{A}\left(\mathbf{x}_{n-2}\right) = \\
&\mathcal{D}_{\mathrm{KL}}\left(\tilde{f}^{n-1}||\tilde{g}^{n-1}\right) + \mathbb{E}_{\tilde{f}_{\mathbf{U}}^{n-1}}\left[\mathbb{E}_{\tilde{f}_{\mathbf{X}}^{n-1}}\left[-\ln\hat{\gamma}\left(\mathbf{X}_{n-1}\right)\right]\right] = \\
&\int \tilde{f}_{\mathbf{U}}^{n-1}\left\{\ln\left(\frac{\tilde{f}_{\mathbf{U}}^{n-1}}{\tilde{g}_{\mathbf{U}}^{n-1}}\right) + \hat{\omega}\left(\mathbf{u}_{n-1},\, \mathbf{x}_{n-2}\right)\right\} d\mathbf{u}_{n-1}\,,
\end{aligned}
\tag{35}
$$

where $\hat{\omega}\left(\mathbf{u}_{n-1},\, \mathbf{x}_{n-2}\right) = \hat{\alpha}\left(\mathbf{u}_{n-1},\, \mathbf{x}_{n-2}\right) + \hat{\beta}\left(\mathbf{u}_{n-1},\, \mathbf{x}_{n-2}\right)$ and

$$
\begin{aligned}
\hat{\alpha}\left(\mathbf{u}_{n-1},\, \mathbf{x}_{n-2}\right) &:= \mathcal{D}_{\mathrm{KL}}\left(\tilde{f}_{\mathbf{X}}^{n-1}||\tilde{g}_{\mathbf{X}}^{n-1}\right), \\
\hat{\beta}\left(\mathbf{u}_{n-1},\, \mathbf{x}_{n-2}\right) &:= -\mathbb{E}_{\tilde{f}_{\mathbf{X}}^{n-1}}\left[\ln\hat{\gamma}\left(\mathbf{X}_{n-1}\right)\right].
\end{aligned}
$$

The last expression we found for $\hat{A}(\mathbf{x}_{n-2})$ in (35) enables us to use Lemma 1 in order to solve the optimization problem in (33). This time, by applying Lemma 1 with $\mathbf{Z} = \mathbf{U}_{n-1}$, $f(\mathbf{z}) = \tilde{f}_{\mathbf{U}}^{n-1}$, $g(\mathbf{z}) = \tilde{g}_{\mathbf{U}}^{n-1}$, $\alpha(\cdot) = \hat{\omega}(\cdot, \mathbf{x}_{n-2})$, $h_j(\mathbf{z}) = h_{\mathbf{u},j}^{n-1}(\mathbf{u}_{n-1})$, $H_j = H_{\mathbf{u},j}^{n-1}$, $c_j[\cdot] = c_{\mathbf{u},j}^{n-1}[\cdot]$, $\lambda_j = \lambda_{\mathbf{u},j}^{n-1}$, $\mathcal{E} = \mathcal{E}^{n-1}$, $\mathcal{I} = \mathcal{I}^{n-1}$, we get the following solution to (33):

$$
\left(\tilde{f}_{\mathbf{U}}^{n-1}\right)^* = \tilde{g}_{\mathbf{U}}^{n-1}\frac{e^{-\{\hat{\omega}(\mathbf{u}_{n-1},\mathbf{x}_{n-2}) + \sum_{j\in\mathcal{I}_a{}^{n-1}\setminus\{0\}}\left(\lambda_{\mathbf{u},j}^{n-1}\right)^* h_{\mathbf{u},j}^{n-1}(\mathbf{u}_{n-1})\}}}{e^{1+\left(\lambda_{\mathbf{u},0}^{n-1}\right)^*}},
$$

where $\mathcal{I}_a{}^{n-1}$ is the active set index associated to $\left(\tilde{f}_{\boldsymbol{U}}^{n-1}\right)^*$ and the Lagrange multipliers can again be obtained from Lemma 1. Namely:

- $\left(\boldsymbol{\lambda}_{\mathbf{u}}^{n-1}\right)^* = \left[\left(\lambda_{\mathbf{u},0}^{n-1}\right)^*, \left(\lambda_{\mathbf{u},1}^{n-1}\right)^*, \ldots, \left(\lambda_{\mathbf{u},n_e^n+n_l^n}^{n-1}\right)^*\right]^T$ are the solution to

$$
\begin{aligned}
\left(\boldsymbol{\lambda}_{\mathbf{u}}^{n-1}\right)^* \in \arg\,\max\; & \mathcal{L}^D\left(\boldsymbol{\lambda}_{\mathbf{u}}^{n-1}\right) \\
\text{s.t.:}\;\; & \lambda_{\mathbf{u},j}^{n-1} \text{ free}, \quad \forall j \in \mathcal{E}^{n-1}, \\
& \lambda_{\mathbf{u},j}^{n-1} \geq 0, \quad \forall j \in \mathcal{I}^{n-1}
\end{aligned}
$$

where

$$
\begin{aligned}
\mathcal{L}^D\left(\boldsymbol{\lambda}_{\mathbf{u}}^{n-1}\right) = &-\textstyle\sum_{j\in\mathcal{E}^{n-1}\cup\mathcal{I}^{n-1}} \lambda_{\mathbf{u},j}^{n-1} H_{\mathbf{u},j}^{n-1} \\
&- \ln\Big(\textstyle\int \tilde{g}_{\mathbf{U}}^{n-1}\; e^{-\hat{\omega}(\mathbf{u}_{n-1},\,\mathbf{x}_{n-2}) + } \\
&e^{-\left\{\sum_{j\in\mathcal{E}^{n-1}\cup\mathcal{I}^{n-1}} \lambda_{\mathbf{u},j}^{n-1} h_{\mathbf{u},j}^{n-1}(\mathbf{u}_{n-1})\right\}} d\mathbf{u}_{n-1}\Big);
\end{aligned}
$$

- $\left(\lambda_{\mathbf{u},0}^{n-1}\right)^*$ is given by

$$
\begin{aligned}
&\left(\lambda_{\mathbf{u},0}^{n-1}\right)^* + 1 = \ln\Big\{\textstyle\int \tilde{g}_{\mathbf{U}}^{n-1} e^{-\{\hat{\omega}(\mathbf{u}_{n-1},\,\mathbf{x}_{n-2})\}} \\
&e^{-\left\{\sum_{j\in\mathcal{I}_a{}^{n-1}\setminus\{0\}}\left(\lambda_{\mathbf{u},j}^{n-1}\right)^* h_{\mathbf{u},j}^{n-1}(\mathbf{u}_{n-1})\right\}} d\mathbf{u}_{n-1}\Big\} \\
&= \ln\left\{\hat{\gamma}_{\mathbf{u},0}^{n-1}\left(\mathbf{x}_{n-2}\right)\right\}.
\end{aligned}
$$

Moreover, from Lemma 1 we also obtain:

$$
\begin{aligned}
B_{n-1}^* &= -\mathbb{E}_{\tilde{f}_{\mathbf{X}}^{n-2}}\left[\textstyle\sum_{j\in\mathcal{I}_a{}^{n-1}} \ln\hat{\gamma}_{\mathbf{u},j}^{n-1}\left(\mathbf{X}_{n-2}\right)\right] \\
&= -\mathbb{E}_{\tilde{f}_{\mathbf{X}}^{n-2}}\left[\ln\hat{\gamma}\left(\mathbf{X}_{n-2}\right)\right].
\end{aligned}
$$

Step 4. The proof can then be concluded by observing that, using B_{n-1}^* in (31), the optimization can again be split in sub-problems, with the *last* sub-problem (i.e. the sub-problem corresponding to $k = n-2$) being independent from the others and having the same structure as the problem we solved at $k = n-1$. Hence, the solution at the generic iteration k, i.e. $\left(\tilde{f}_{\mathbf{U}}^{k}\right)^*$, will have the same structure as $\left(\tilde{f}_{\mathbf{U}}^{n-1}\right)^*$, with the functions $\hat{\alpha}(\cdot,\cdot)$, $\hat{\beta}(\cdot,\cdot)$, $\hat{\omega}(\cdot)$ given by (14) - (15) and the Lagrange multipliers given by (19) - (20). This leads to the expression in (13) with the minimum given in (21). Finally, for the last iteration ($k = n$) we note from (28b) that $\hat{\beta}\left(\mathbf{u}_n,\, \mathbf{x}_{n-1}\right) = 0$ which, by means of (16), implies that $\hat{\gamma}_{\mathbf{u},j}^{n+1}\left(\mathbf{x}_n\right) = 1$, $\forall j$. This gives the terminal conditions in (17) and (18). The proof is then completed. □

Finally, we close the section with the following remark:

Remark 14 *From the proof of Theorem 1, we get that, at time-step k, $\left(\tilde{f}_{\mathbf{U}}^{k}\right)^*$ is obtained as*

$$
\begin{aligned}
\left(\tilde{f}_{\mathbf{U}}^{k}\right)^* \in \arg\min_{\tilde{f}_{\mathbf{U}}^{k}}\; & \mathcal{D}_{KL}\left(\tilde{f}_{\mathbf{U}}^{k}||\tilde{g}_{\mathbf{U}}^{k}\right) + \mathbb{E}_{\tilde{f}_{\mathbf{U}}^{k}}\left[\hat{\omega}\left(\mathbf{U}_k, \mathbf{X}_{k-1}\right)\right] \\
\text{s.t.:}\;\; & c_{\mathbf{u},j}^{k}\left[\tilde{f}_{\mathbf{U}}^{k}\right] = 0, \quad \forall j \in \mathcal{E}_0^k, \\
& c_{\mathbf{u},j}^{k}\left[\tilde{f}_{\mathbf{U}}^{k}\right] \leq 0, \quad \forall j \in \mathcal{I}^k
\end{aligned}
$$

where $\hat{\omega}\left(\cdot,\, \cdot\right)$ is defined as in Theorem 1. If for the above problem: (i) $\hat{\omega}\left(\cdot,\, \cdot\right)$ is constant in the first argument; (ii) $\tilde{g}_{\mathbf{U}}^{k}$ is a uniform distribution on a compact support. Then,

its minimizer can also be found via the constrained maximum entropy problem:

$$\left(\tilde{f}^k_{\mathbf{U}}\right)^* \in \arg\max_{\tilde{f}^k_{\mathbf{U}}} -\mathbb{E}_{\tilde{f}^k_{\mathbf{U}}}\left[\ln \tilde{f}^k_{\mathbf{U}}\right]$$
$$\text{s.t.: } c^k_{\mathbf{u},j}\left[\tilde{f}^k_{\mathbf{U}}\right] = 0, \quad \forall j \in \mathcal{E}^k_0,$$
$$c^k_{\mathbf{u},j}\left[\tilde{f}^k_{\mathbf{U}}\right] \leq 0, \quad \forall j \in \mathcal{I}^k$$

5 An algorithm from Theorem 1

Theorem 1 gives an explicit expression for synthesizing the control policy and we leverage this to turn the result into an algorithmic procedure. The algorithm introduced here takes as input $g\left(\mathbf{\Delta}^n_e\right)$, $\tilde{f}^k_{\mathbf{X}} := f(\boldsymbol{x}_k|\boldsymbol{u}_k, \boldsymbol{x}_{k-1})$ and the constraints of Problem 1 (if any). The pdfs, as further illustrated in the next section, can be obtained from the data. Given this input, the algorithm outputs $\left\{\left(\tilde{f}^k_{\mathbf{U}}\right)^*\right\}_{k\in\mathcal{K}}$ solving Problem 1. Then, at each k, the control input, $\mathbf{u_k}$, applied to the system is obtained by sampling from $\left(\tilde{f}^k_{\mathbf{U}}\right)^*$. The key steps of the proposed algorithm are summarized as pseudo-code in Algorithm 1.

Algorithm 1 Pseudo-code

Inputs:
$g\left(\mathbf{\Delta}^n_e\right)$, $\tilde{f}^k_{\mathbf{X}}$ and constraints of Problem 1
Output:
$\left\{\left(\tilde{f}^k_{\mathbf{U}}\right)^*\right\}_{k\in\mathcal{K}}$ solving Problem 1
Initialize
$\hat{\gamma}^{n+1}_{\mathbf{u},j}\left(\mathbf{x}_n\right) = 1, \forall j;$
$\hat{\gamma}\left(\mathbf{x}_n\right) \leftarrow \exp\left[\sum_j \ln\left(\hat{\gamma}^{n+1}_{\mathbf{u},j}\left(\mathbf{x}_n\right)\right)\right];$
for $k = n$ to 1 **do**
$\hat{\alpha}\left(\mathbf{u}_k, \mathbf{x}_{k-1}\right) \leftarrow \int \tilde{f}^k_{\mathbf{X}} \ln \frac{\tilde{f}^k_{\mathbf{X}}}{\tilde{g}^k_{\mathbf{X}}}\, d\mathbf{x}_k;$
$\hat{\beta}\left(\mathbf{u}_k, \mathbf{x}_{k-1}\right) \leftarrow -\int \tilde{f}^k_{\mathbf{X}} \ln\left(\hat{\gamma}\left(\mathbf{x}_k\right)\right) d\mathbf{x}_k;$
$\hat{\omega}\left(\mathbf{u}_k, \mathbf{x}_{k-1}\right) \leftarrow \hat{\alpha}\left(\mathbf{u}_k, \mathbf{x}_{k-1}\right) + \hat{\beta}\left(\mathbf{u}_k, \mathbf{x}_{k-1}\right);$
Compute $\left[\left(\lambda^k_{\mathbf{u},1}\right)^*, \ldots, \left(\lambda^k_{\mathbf{u},n^k_e+n^k_i}\right)^*\right]^T$ by solving (19) and $\left(\lambda^k_{\mathbf{u},0}\right)^*$ using (20);
Compute the control policy:

$$\left(\tilde{f}^k_{\mathbf{U}}\right)^* \leftarrow \tilde{g}^k_{\mathbf{U}} \frac{e^{-\{\hat{\omega}(\mathbf{u}_k, \mathbf{x}_{k-1}) + \sum_{j\in\mathcal{I}_a{}^k\setminus\{0\}} \left(\lambda^k_{\mathbf{u},j}\right)^* h^k_{\mathbf{u},j}(\mathbf{u}_k)\}}}{e^{1+\left(\lambda^k_{\mathbf{u},0}\right)^*}};$$

Prepare variables for the next iteration, $k-1$:
$\hat{\gamma}^k_{\mathbf{u},0}\left(\mathbf{x}_{k-1}\right) \leftarrow \exp\left\{\left(\lambda^k_{\mathbf{u},0}\right)^* + 1\right\}$
$\hat{\gamma}^k_{\mathbf{u},j}\left(\mathbf{x}_{k-1}\right) \leftarrow \exp\left\{\left(\lambda^k_{\mathbf{u},j}\right)^* H^k_{\mathbf{u},j}\right\} \quad j \in \mathcal{E}^k \cup \mathcal{I}^k$
$\hat{\gamma}\left(\mathbf{x}_{k-1}\right) \leftarrow \exp\left[\sum_{j\in\mathcal{I}_a{}^k} \ln\left(\hat{\gamma}^k_{\mathbf{u},j}\left(\mathbf{x}_{k-1}\right)\right)\right]$
end for

Algorithm 1 is used to compute the control policy for the example of the next section.

6 Numerical example

We now illustrate the effectiveness of the results via a numerical example where Algorithm 1 is used to synthesize, from data measured during test drives, a policy for the merging of a car on a highway. Specifically, we leverage our results to synthesize a policy that makes the behavior of the controlled car similar to the behavior seen in the examples (collected through test drives) while satisfying some desired constraints on the control variable. We first describe the scenario considered and the experimental set-up. Then, we describe the data we collected and the process we used to compute the pdfs for the algorithm. Finally, we discuss the results.

Scenario description and experimental set-up. The scenario we considered is schematically illustrated in Figure 1, where a car is merging onto a highway. The

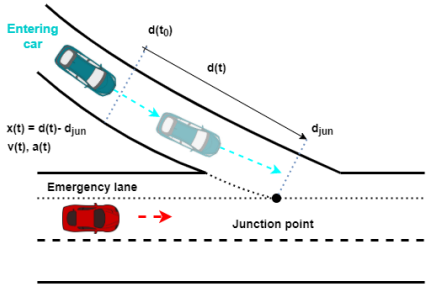

Fig. 1. The scenario considered in Section 6: the light-blue vehicle is trying to merge on a highway.

stretch of road where our experiments took place is outside University College Dublin (UCD) and the highway is *Stillorgan Road* in Dublin 4 (see Figure 2). Data were collected from a Toyota Prius using an OBD2 connection with a smartphone running the Android apps *Hybrid Assistant* and its reporting tool *Hybrid Reporter*[4]. We collected the car GPS position and its longitudinal speed using the hardware-in-the-loop architecture of (Griggs et al., 2019). The apps on the smartphone provided *raw data* in a `.txt` file with each line reporting the quantities measured at a given time instant (the sampling period was of approximately 0.25s). Data were imported in Matlab and the car position was localized by cross-referencing these data with the road information from *OpenStreetMap* (OpenStreetMap contributors, 2017).

Collecting the data. We performed 100 test drives. In each of the tests, data were collected within a 300 meters observation window starting 200 m before the junction (i.e. after the UCD entrance). The raw data were processed to obtain the speed, acceleration and jerk profiles

[4] See http://hybridassistant.blogspot.com/

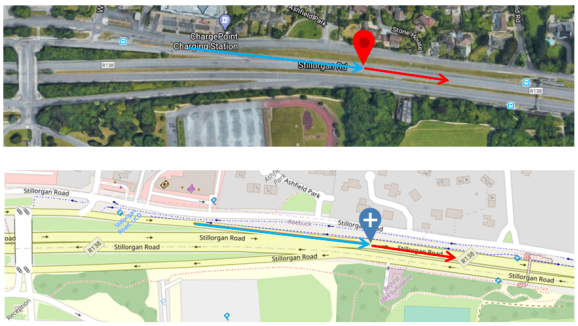

Fig. 2. Area for the experiments: map view (from Google maps) and *OpenStreetMap* representation.

of the car as a function of the distance traveled within the observation window (see left panels in Figure 3). Following the notation introduced in Section 3 (see also Remark 2) the sequence of data collected from each test drive is a dataset. For notational convenience, we simply term the collection of these 100 datasets as *complete* dataset in what follows. The vertical line in each panel highlights the physical location of the junction of Figure 2. From the complete dataset we extracted a subset of profiles that would serve as examples. In particular, we selected the profiles with the lowest root mean square (RMS) value for the jerk, which is typically associated to a comfortable driving style, see e.g. (Bae et al., 2019). We used as example driving profiles, those having a RMS value for the jerk of at most 0.16. This gave the subset of 20 driving profiles (i.e. a collection of 20 datasets) shown in the right panels of Figure 3. We simply term the collection of these 20 datasets as example dataset.

Computing the pdfs. We let $\mathbf{x}_k$ be the position of the car at the k-th time-step (i.e. $\mathbf{x}_k = d_k$) and $\mathbf{u}_k$ be its longitudinal speed (i.e. $\mathbf{u}_k = v_k$). Following (Deng et al., 2019), we computed the joint probability density functions as empirical distributions to obtain $f(\mathbf{x}_{k-1}, \mathbf{u}_k)$ and $g(\mathbf{x}_{k-1}, \mathbf{u}_k)$ from the complete and the example dataset respectively (see Figure 4). Following the same process, we also obtained $f(\mathbf{x}_k, \mathbf{x}_{k-1}, \mathbf{u}_k)$ and $g(\mathbf{x}_k, \mathbf{x}_{k-1}, \mathbf{u}_k)$ and conditioned these joint pdfs to get $f(\mathbf{x}_k|\mathbf{x}_{k-1}, \mathbf{u}_k) = f(\mathbf{x}_k, \mathbf{x}_{k-1}, \mathbf{u}_k)/f(\mathbf{x}_{k-1}, \mathbf{u}_k)$ and $g(\mathbf{x}_k|\mathbf{x}_{k-1}, \mathbf{u}_k) = g(\mathbf{x}_k, \mathbf{x}_{k-1}, \mathbf{u}_k)/g(\mathbf{x}_{k-1}, \mathbf{u}_k)$. We then assumed $\tilde{f}^k_{\mathbf{X}}$ and $\tilde{g}^k_{\mathbf{X}}$, two inputs to Algorithm 1, to be normal distributions and estimated their parameters via least squares. This yielded $\tilde{f}^k_{\mathbf{X}} \sim \mathcal{N}\left(a_c \mathbf{x}_{k-1} + b_c \mathbf{u}_k, \sigma_c^2\right)$, $\tilde{g}^k_{\mathbf{X}} \sim \mathcal{N}\left(a_e \mathbf{x}_{k-1} + b_e \mathbf{u}_k, \sigma_e^2\right)$ with $a_c = 0.9820$, $b_c = 0.2591\,s$, $\sigma_c^2 = 2.6118\,m^2$ and $a_e = 0.9811$, $b_e = 0.2723\,s$, $\sigma_e^2 = 1.7622\,m^2$. Finally, $\tilde{g}^k_{\mathbf{U}}$ was obtained from the empirical pdfs as $\tilde{g}^k_{\mathbf{U}} = g\left(\mathbf{x}_{k-1}, \mathbf{u}_k\right)/\left(\int g\left(\mathbf{x}_{k-1}, \mathbf{u}_k\right) d\mathbf{u}_k\right)$.

Remark 15 *The Gaussian assumption for $\tilde{f}^k_{\mathbf{X}}$ and $\tilde{g}^k_{\mathbf{X}}$ is inspired by (Nguyen et al., 2017; Moser et al., 2015). These works considered the problem of making short term predictions of vehicles' trajectories. See also references therein and our concluding remarks in Section 7.*

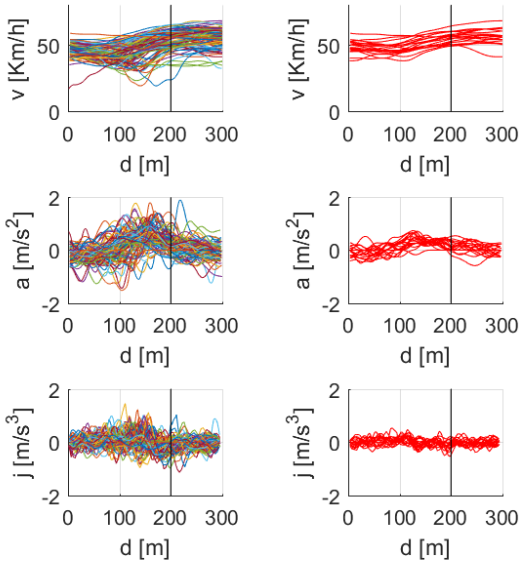

Fig. 3. Left panels: driving profiles from the complete dataset of 100 trips. Right panels: the subset of 20 profiles used as examples. Acceleration and jerk computed from the data.

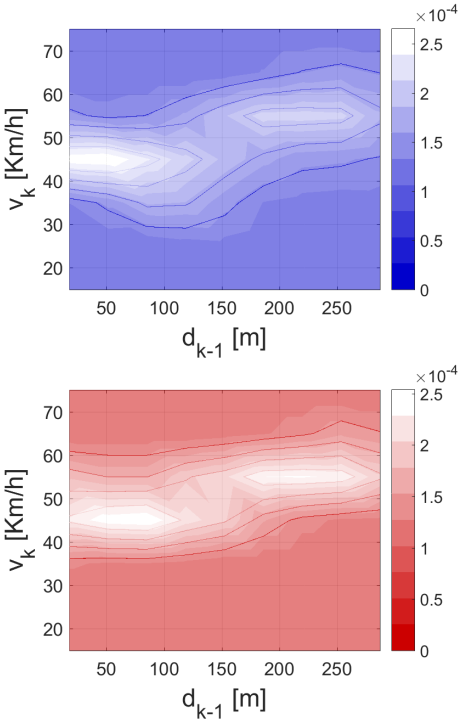

Fig. 4. *Heat-maps* for $f(\mathbf{x}_{k-1}, \mathbf{u}_k)$ (top panel) and $g(\mathbf{x}_{k-1}, \mathbf{u}_k)$ (bottom panel).

Synthesis of the control policy. Given this set-up, we used Algorithm 1 to solve Problem 1 and hence to synthesize from the examples a control policy allowing the car to merge on the highway. When synthesizing the control policy, we imposed, at each k, the following set

of constraints:

$$c^k_{\mathbf{u},j}\left[\tilde{f}^k_{\mathbf{U}}\right] = 0, \quad j \in \{0,1,2\}, \tag{36a}$$

where:

$$c^k_{\mathbf{u},0}\left[\tilde{f}^k_{\mathbf{U}}\right] = \mathbb{E}_{\tilde{f}^k_{\mathbf{U}}}\left[\mathbb{1}_{\mathcal{U}_k}(\mathbf{U}_k)\right] - H^k_{\mathbf{u},0}, \tag{36b}$$

$$c^k_{\mathbf{u},1}\left[\tilde{f}^k_{\mathbf{U}}\right] = \mathbb{E}_{\tilde{f}^k_{\mathbf{U}}}\left[U_k\right] - H^k_{\mathbf{u},1}, \tag{36c}$$

$$c^k_{\mathbf{u},2}\left[\tilde{f}^k_{\mathbf{U}}\right] = \mathbb{E}_{\tilde{f}^k_{\mathbf{U}}}\left[U^2_k\right] - H^k_{\mathbf{u},2}, \tag{36d}$$

with:

$$H^k_{\mathbf{u},0} = 1, \tag{36e}$$

$$H^k_{\mathbf{u},1} = \mathbb{E}_{\tilde{g}^k_{\mathbf{U}}}\left[U_k\right], \tag{36f}$$

$$H^k_{\mathbf{u},2} = 4\left(\mathbb{E}_{\tilde{g}^k_{\mathbf{U}}}\left[U^2_k\right] - \mathbb{E}^2_{\tilde{g}^k_{\mathbf{U}}}\left[U_k\right]\right) + \mathbb{E}^2_{\tilde{g}^k_{\mathbf{U}}}\left[U_k\right]. \tag{36g}$$

At each k, the fulfillment of the first constraint, corresponding to $j = 0$, guarantees that the solution to the problem is a pdf (this is the normalization constraint). Instead, the fulfillment of the other two constraints in (36) guarantees that, at each k: (i) the expected value of the control variable of the closed loop system is the same as the one seen in the example dataset (constraint corresponding to $j = 1$); and (ii) the variance of the control variable of the closed loop system is 4 times the variance of the control variable seen in the example dataset (constraint corresponding to $j = 2$). Making the variance of the control variable of the closed loop system larger than the control variable from the example dataset corresponds in accounting for a reduced liability on the example dataset, allowing the closed loop system to depart from the behavior seen in the examples. Moreover, we note that the constraints in (36) satisfy Assumption 1. Indeed any pdf with the first two moments satisfying the constraints fulfills the assumption. We used Algorithm 1 to compute the solution to Problem 1 and approximated these pdfs via the Maximum Entropy Principle. This resulted into the truncated Gaussians of Figure 5. In such a figure, it is clearly shown how these pdfs have higher variance than the corresponding $\tilde{g}^k_{\mathbf{U}}$ extracted from the examples. In the figure, for clarity, $\tilde{g}^k_{\mathbf{U}}$ is also represented as a truncated Gaussian that has the same mean and variance as the pdf extracted from the data. At each time-step, the control input, $\mathbf{u}_k$, applied to the car was obtained by sampling from the pdfs of Figure 5 (in green). In particular, by sampling the $\mathbf{u}_k$'s as the mean value of the random variable generated by these pdfs, we obtained the speed profile for the controlled car illustrated in Figure 6.

7 Conclusions

We considered the problem of synthesizing control policies from noisy example datasets for systems affected by

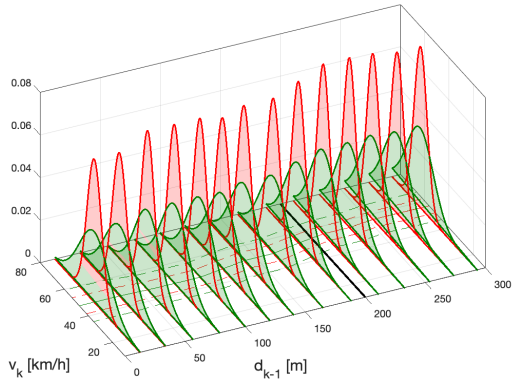

Fig. 5. The pdfs $\tilde{g}^k_{\mathbf{U}}$ (in red) together with the pdfs obtained from Algorithm 1 (green). For the sake of clarity in the figure, the pdfs are not shown for each iteration (the policies shown here are representative for all the other time-steps). The continuous line on the speed/distance plane denotes the expectation of the pdfs/policies, while the dashed lines represent the confidence interval corresponding to the standard deviation from the examples (red) and from Algorithm 1 (green). These are shown for each time-step. Colors online.

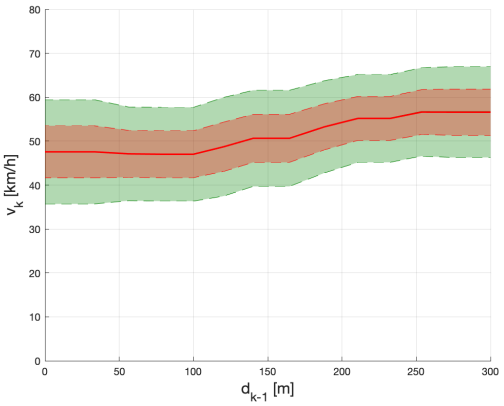

Fig. 6. In green: speed profile from $\left(\tilde{f}^k_{\mathbf{U}}\right)^*$. The bold line is the average speed profile and the shaded areas represent the confidence intervals corresponding to the standard deviation. For comparison, the corresponding speed profile from the example dataset is also shown in red. As expected, the bold red line overlaps with the bold green line. Colors online.

actuation constraints. To tackle this problem, we introduced a number of technical results to explicitly compute the policy directly from certain pdfs obtained from the data and in compliance with the constraints. The optimal policy obtained with our results allows to approximate the behavior seen in the examples, while simultaneously fulfilling the system-specific actuation constraints. The results were also turned into an algorithmic procedure and their effectiveness was illustrated via a use-case. The use-case involved the synthesis, from measured data collected during test drives, of a control policy allowing a car to merge on a highway. We are currently

investigating whether our methodology can be extended to consider other divergences (Basseville, 2013) rather than the KL-divergence in the cost of Problem 1. Our future work will be aimed at extending the results presented in this paper by considering: (i) the introduction of chance constraints on the state variable; (ii) cost functionals that do not only aim at tracking the behavior in the examples but also minimize additional costs; (iii) the use of concepts from data informativity and optimal experimental design to obtain sufficiently *informative* data for the framework developed here. Finally, we will explore the possibility of devising *automated* fitting procedures to extract suitable pdfs from the data in order to enable an end-to-end pipeline for our results.

Acknowledgments

The authors wish to acknowledge Dr. Herzallah (Aston University) for her comments on an earlier version of this manuscript during her visit at UCD, Dr. Guy and Prof. Kárný (both at the Institute of Information Theory and Automation at the Czech Academy of Sciences) for the insightful discussions on the results. GR would also like to thank Prof. Bullo at UCSB for reading an earlier version of this work. Five anonymous referees made several helpful comments and suggestions, which led to improvements over the originally submitted manuscript.

A Appendix: proofs of the technical results

A.1 Proof of Property 1

To prove this result we start from the definition of KL-divergence. In particular:

$$\mathcal{D}_{\mathrm{KL}}\left(\phi\left(\mathbf{y},\mathbf{z}\right)||g\left(\mathbf{y},\mathbf{z}\right)\right) := \int\int \phi\left(\mathbf{y},\mathbf{z}\right)\left[\ln\frac{\phi\left(\mathbf{y},\mathbf{z}\right)}{g\left(\mathbf{y},\mathbf{z}\right)}\right] d\mathbf{y}\, d\mathbf{z}$$
$$= \underbrace{\int\int \phi\left(\mathbf{z}|\mathbf{y}\right)\left[\phi\left(\mathbf{y}\right)\ln\frac{\phi\left(\mathbf{y}\right)}{g\left(\mathbf{y}\right)}\right] d\mathbf{y}\, d\mathbf{z}}_{(1)}$$
$$+ \underbrace{\int\int \phi\left(\mathbf{y}\right)\left[\phi\left(\mathbf{z}|\mathbf{y}\right)\ln\frac{\phi\left(\mathbf{z}|\mathbf{y}\right)}{g\left(\mathbf{z}|\mathbf{y}\right)}\right] d\mathbf{z}\, d\mathbf{y}}_{(2)}.$$

For the term (1) in the above expression we may continue as follows: $\int\int \phi\left(\mathbf{z}|\mathbf{y}\right)\left[\phi\left(\mathbf{y}\right)\ln\frac{\phi(\mathbf{y})}{g(\mathbf{y})}\right] d\mathbf{y}\, d\mathbf{z} = \int \phi\left(\mathbf{z}|\mathbf{y}\right) d\mathbf{z}\left[\int \phi\left(\mathbf{y}\right)\ln\frac{\phi(\mathbf{y})}{g(\mathbf{y})} d\mathbf{y}\right] = \mathcal{D}_{\mathrm{KL}}\left(\phi\left(\mathbf{y}\right)||g\left(\mathbf{y}\right)\right)$, where we used Fubini's theorem, the fact that the term on the first line in square brackets is independent on $\mathbf{Z}$ and the fact that $\int \phi\left(\mathbf{z}|\mathbf{y}\right) d\mathbf{z} = 1$. By using again Fubini's theorem, for the term (2) in the above expression instead we have: $\int\int \phi\left(\mathbf{y}\right)\left[\phi\left(\mathbf{z}|\mathbf{y}\right)\ln\frac{\phi(\mathbf{z}|\mathbf{y})}{g(\mathbf{z}|\mathbf{y})}\right] d\mathbf{z}\, d\mathbf{y} = \mathbb{E}_{\phi}\left[\mathcal{D}_{\mathrm{KL}}\left(\phi(\mathbf{z}|\mathbf{Y})||\, g(\mathbf{z}|\mathbf{Y})\right)\right]$, thus proving the result. □

A.2 Proof of Lemma 1

The proof is organized in 3 steps. In **Step 1** we show that the optimization problem in (6) is a convex optimization problem (COP) and we then devise its augmented Lagrangian. In **Step 2** we explicit the Karush-Kuhn-Tucker (KKT) conditions and verify that these are satisfied by the solution in (9). Recall that, for a COP, KKT conditions are necessary and sufficient(Boyd and Vandenberghe, 2004, Chapter 5). Finally, in **Step 3**, we compute the minimum of the cost function corresponding to the optimal solution.

Step 1. We start with observing that the cost function $\mathcal{L}\left(f(\mathbf{z})\right)$ in (6) can be conveniently rewritten as $\mathcal{L}\left(f(\mathbf{z})\right) = \int l\left(f(\mathbf{z})\right) d\mathbf{z}$, with $l\left(f(\mathbf{z})\right) := f(\mathbf{z})\left[\ln\left(\frac{f(\mathbf{z})}{g(\mathbf{z})}\right) + \alpha(\mathbf{z})\right]$. Clearly, $\mathcal{L}(\cdot)$ is twice differentiable and we now prove that it is also a strictly convex functional in f. We do this by showing that its second variation is positive definite on the space of integrable functions and we explicit the first and the second variation of $\mathcal{L}\left(f(\mathbf{z})\right)$, i.e. $\delta\mathcal{L}(f,\delta f)$ and $\delta^2\mathcal{L}(f,\delta f)$, in terms of the first and second derivative of $l\left(f(\mathbf{z})\right)$ with respect to $f(\mathbf{z})$ (i.e. $\frac{\partial l(f(\mathbf{z}))}{\partial f}$ and $\frac{\partial^2 l(f(\mathbf{z}))}{\partial f^2}$, respectively). By computing $\delta\mathcal{L}(f,\delta f)$ we get $\delta\mathcal{L}\left(f,\delta f\right) = \int \frac{\partial l(f(\mathbf{z}))}{\partial f}\delta f\, d\mathbf{z}$, with $\frac{\partial l(f(\mathbf{z}))}{\partial f} = \ln f(\mathbf{z}) + (\alpha(\mathbf{z}) + 1 - \ln g(\mathbf{z}))$. This leads to the following expression for the second variation of $\mathcal{L}(f,\delta f)$

$$\delta^2\mathcal{L}\left(f,\delta f\right) = \int \delta f\left(\frac{\partial^2 l\left(f(\mathbf{z})\right)}{\partial f^2}\right)\delta f\, d\mathbf{z} \qquad \text{(A.1)}$$

To show convexity of $\mathcal{L}$ it then suffices to observe that, since $f(\mathbf{z})$ is non-negative on its support, then the quantity under the integral in (A.1) is positive. In turn, this implies that $\delta^2\mathcal{L}\left(f,\delta f\right)$ is strictly positive for any measurable, non-zero variation δf (see also (Kirk, 2004, Chapter 4) for a detailed discussion). Hence, in order to conclude that the problem in (6) is a COP, it suffices to observe that the constraints in (8) are linear in $f(\mathbf{z})$. The augmented Lagrangian of the COP in (6) is:

$$\mathcal{L}_{aug}\left(f(\mathbf{z}),\boldsymbol{\lambda}\right) := \mathcal{L}\left(f(\mathbf{z})\right) + \sum_{j\in\mathcal{E}_0\cup\mathcal{I}} \lambda_j\, c_j\left[f(\mathbf{z})\right], \qquad \text{(A.2)}$$

where $\boldsymbol{\lambda} := [\lambda_0, \lambda_1, \ldots, \lambda_{n_e+n_l}]^T$ is the column vector stacking all the Lagrange multipliers.

Step 2. We showed that the problem in (6) is a COP and hence the KKT conditions are necessary and sufficient optimality conditions. That is, in order to be a unique minimizer of the problem, the candidate function $f(\mathbf{z})$ must satisfy the conditions made explicit in Table A.1. We now impose the stationarity condition (see Table A.1) and first note that the augmented Lagrangian (A.2) can be written as: $\mathcal{L}_{aug}\left(f(\mathbf{z}),\boldsymbol{\lambda}\right) :=$

Primal feasibility:	$c_j\left[f(\mathbf{z})\right]=0,$	$\forall j\in\mathcal{E}_0$
	$c_j\left[f(\mathbf{z})\right]\leq 0,$	$\forall j\in\mathcal{I}$
Dual feasibility:	$\lambda_j\geq 0,$	$\forall j\in\mathcal{I}$
Complementary slackness:	$\lambda_j\, c_j\left[f(\mathbf{z})\right]=0,$	$\forall j\in\mathcal{I}$
Stationarity:	$\delta\mathcal{L}_{aug}\left(f,\delta f,\boldsymbol{\lambda}\right)=0,$	$\forall\,\delta f$

Table A.1
KKT conditions for the problem in (6).

$\int f(\mathbf{z})\ \left[\ln\left(\frac{f(\mathbf{z})}{g(\mathbf{z})}\right)+\alpha(\mathbf{z})\right]\,d\mathbf{z}+\langle\boldsymbol{\lambda},\int f(\mathbf{z})\,\mathbf{h}\left(\mathbf{z}\right)d\mathbf{z}-\mathbf{H}\rangle$. Hence:

$$\mathcal{L}_{aug}\left(f(\mathbf{z}),\boldsymbol{\lambda}\right)=\int\tilde{l}\left(f(\mathbf{z}),\boldsymbol{\lambda}\right)d\mathbf{z}-\langle\boldsymbol{\lambda},\mathbf{H}\rangle,\qquad(A.3)$$

where the quantity under the integral is given by $\tilde{l}\left(f(\mathbf{z}),\boldsymbol{\lambda}\right):=f(\mathbf{z})\ \left[\ln\left(\frac{f(\mathbf{z})}{g(\mathbf{z})}\right)+\alpha(\mathbf{z})+\langle\boldsymbol{\lambda},\mathbf{h}(\mathbf{z})\rangle\right]$. By computing the first variation of (A.3) with respect to δf we obtain $\delta\mathcal{L}_{aug}\left(f,\delta f,\boldsymbol{\lambda}\right)=\int\frac{\partial\tilde{l}(f(\mathbf{z}),\boldsymbol{\lambda})}{\partial f}\delta f\,d\mathbf{z}$, and thus, by imposing the stationarity condition (i.e. $\delta\mathcal{L}_{aug}\left(f,\delta f,\boldsymbol{\lambda}\right)=0,\ \forall\delta f$), we get $\frac{\partial\tilde{l}(f(\mathbf{z}),\boldsymbol{\lambda})}{\partial f}=0$. That is, $\frac{\partial\tilde{l}(f(\mathbf{z}),\boldsymbol{\lambda})}{\partial f}=\ln\left(\frac{f(\mathbf{z})}{g(\mathbf{z})}\right)+\alpha\left(\mathbf{z}\right)+\langle\boldsymbol{\lambda},\mathbf{h}(\mathbf{z})\rangle+1=0$, from which it immediately follows that all the optimal solution candidates must be of the form

$$\hat{f}^{*}\left(\mathbf{z},\boldsymbol{\lambda}\right):=g(\mathbf{z})e^{-\{1+\alpha(\mathbf{z})+\langle\boldsymbol{\lambda},\mathbf{h}(\mathbf{z})\rangle\}}.\qquad(A.4)$$

In the above expression, the notation $\hat{f}^{*}\left(\mathbf{z},\boldsymbol{\lambda}\right)$ was introduced to stress that the optimal solution candidate is a function of the Lagrange multipliers. These can be found by solving the following dual problem

$$\begin{aligned}\boldsymbol{\lambda}^{*}\in\arg\max_{\boldsymbol{\lambda}}\ &\mathcal{L}^{D}\left(\boldsymbol{\lambda}\right)\\ \text{s.t.:}\ &\lambda_j\text{ free },\ \forall j\in\mathcal{E}_0,\\ &\lambda_j\geq 0,\ \ \forall j\in\mathcal{I}\end{aligned}\qquad(A.5)$$

choosing $\boldsymbol{\lambda}^{*}$ so that $\hat{f}^{*}\left(\mathbf{z},\boldsymbol{\lambda}^{*}\right)$ is primal feasible (see (Ben-Tal et al., 1988)). In the problem, $\mathcal{L}^{D}\left(\boldsymbol{\lambda}\right)$ is the Lagrange dual function $\mathcal{L}^{D}\left(\boldsymbol{\lambda}\right):=\inf_{f(\mathbf{z})\geq 0}\mathcal{L}_{aug}\left(f(\mathbf{z}),\boldsymbol{\lambda}\right)$. Note that the vector $\boldsymbol{\lambda}^{*}$ must satisfy the dual feasibility condition. Now, Assumption 1 implies strong duality (see Remark 8) and hence the complementary slackness condition (see Table A.1) is also fulfilled. Additionally, $\mathcal{L}_{aug}\left(f(\mathbf{z}),\boldsymbol{\lambda}\right)$ is strictly convex in $f(\mathbf{z})$ and hence $\inf_{f(\mathbf{z})\geq 0}\mathcal{L}_{aug}\left(f(\mathbf{z}),\boldsymbol{\lambda}\right)=\mathcal{L}_{aug}\left(\hat{f}^{*}\left(\mathbf{z},\boldsymbol{\lambda}\right),\boldsymbol{\lambda}\right)$. Thus:

$$\begin{aligned}\mathcal{L}^{D}\left(\boldsymbol{\lambda}\right)&=\mathcal{L}_{aug}\left(\hat{f}^{*}\left(\mathbf{z},\boldsymbol{\lambda}\right),\boldsymbol{\lambda}\right)=-\int\hat{f}^{*}\left(\mathbf{z},\boldsymbol{\lambda}\right)d\mathbf{z}-\langle\boldsymbol{\lambda},\mathbf{H}\rangle\\&=-\int g\left(\mathbf{z}\right)\ e^{-\{1+\alpha(\mathbf{z})+\langle\boldsymbol{\lambda},\mathbf{h}(\mathbf{z})\rangle\}}d\mathbf{z}-\langle\boldsymbol{\lambda},\mathbf{H}\rangle.\end{aligned}\qquad(A.6)$$

Note now that the last equivalence gives (12) in the statement of the lemma and hence the problem in (11). Moreover, the complementary slackness condition on the pair of optimizers $f^{*}(\mathbf{z}),\boldsymbol{\lambda}^{*}$ implies, for a COP, that there is no duality gap. That is, $\mathcal{L}^{D}\left(\boldsymbol{\lambda}^{*}\right)=\mathcal{L}\left(f^{*}(\mathbf{z})\right)$. In turn, this means that the Lagrange multipliers associated to inactive inequality constraints must be equal to 0, while all the Lagrange multipliers associated to active inequality constraints must be non-negative. Therefore, the optimal solution of the COP in (6) is given by $\hat{f}^{*}\left(\mathbf{z},\boldsymbol{\lambda}^{*}\right)=f^{*}(\mathbf{z})=g(\mathbf{z})\,e^{-\left\{1+\alpha(\mathbf{z})+\sum_{j\in\mathcal{I}_a(f^{*}(\mathbf{z}))}\lambda_j^{*}h_j(\mathbf{z})\right\}}$, which was obtained by taking into account that only the Lagrange multipliers associated to the active constraints are non-zero. The above expression is equal to (9) where we highlighted the role of λ_0 as a normalization constant. This concludes the proof of **(R1)**.

Step 3. Finally, since there is no duality gap, the minimum value of the primal problem (i.e. the COP in (6)) can be obtained from (A.6). This leads to $\mathcal{L}\left(f^{*}(\mathbf{z})\right)=-\left(1+\sum_{j\in\mathcal{I}_a(f^{*}(\mathbf{z}))}\lambda_j^{*}\,H_j\right)$, and thus completes the proof. □

References

Abbeel, P., Ng, A. Y., 2004. Apprenticeship learning via inverse reinforcement learning. In: Proceedings of the Twenty-First International Conference on Machine Learning. ICML '04. Association for Computing Machinery, New York, NY, USA, p. 1.

Argall, B. D., Chernova, S., Veloso, M., Browning, B., 2009. A survey of robot learning from demonstration. Robotics and Autonomous Systems 57 (5), 469 – 483.

Bae, I., Moon, J., Seo, J., 2019. Toward a comfortable driving experience for a self-driving shuttle bus. Electronics 8 (9), 943.

Baggio, G., Katewa, V., Pasqualetti, F., 2019. Data-driven minimum-energy controls for linear systems. IEEE Control Systems Letters 3 (3), 589–594.

Basseville, M., 2013. Divergence measures for statistical data processing—an annotated bibliography. Signal Processing 93 (4), 621–633.

Ben-Tal, A., Teboulle, M., Charnes, A., 1988. The role of duality in optimization problems involving entropy functionals with applications to information theory. Journal of optimization theory and applications 58, 209–223.

Bertsekas, D., 2021. Multiagent reinforcement learning: Rollout and policy iteration. IEEE/CAA Journal of Automatica Sinica 8 (2), 249–272.

Bot, R., Grad, S.-M., Wanka, G., 2005. Duality for optimization problems with entropy-like objective functions. Journal of Information and Optimization Sciences 22, 415–441.

Boyd, S. P., Vandenberghe, L., 2004. Convex optimization. Cambridge University Press.

Bryson, A. E., June 1996. Optimal control-1950 to 1985. IEEE Control Systems Magazine 16 (3), 26–33.

Censor, Y., Elfving, T., 1982. New methods for linear inequalities. Linear Algebra and Its Applications 42, 199–211.

Colin, K., Bombois, X., Bako, L., Morelli, F., 2020. Data informativity for the open-loop identification of mimo systems in the prediction error framework. Automatica 117, 109000.

Coulson, J., Lygeros, J., Dörfler, F., 2019a. Data-enabled predictive control: In the shallows of the DeePC. In: 2019 18th European Control Conference (ECC). pp. 307–312.

Coulson, J., Lygeros, J., Dörfler, F., 2019b. Regularized and distributionally robust data-enabled predictive control. In: 2019 IEEE 58th Conference on Decision and Control (CDC). pp. 2696–2701.

Cover, T. M., Thomas, J. A., 2006. Elements of Information Theory (Wiley Series in Telecommunications and Signal Processing). Wiley-Interscience, USA.

De Persis, C., Tesi, P., 2020. Formulas for data-driven control: Stabilization, optimality, and robustness. IEEE Transactions on Automatic Control 65 (3), 909–924.

Deng, J., Gagliardi, D., Del Re, L., 2019. Microscopic driving behavior modelling at highway entrances using bayesian network. In: 2019 American Control Conference (ACC). pp. 977–982.

Duffin, R. J., Dantzig, G. B., Fan, K., 1956. Linear inequalities and related systems. Princeton University Press.

Edwards, A., Sahni, H., Schroecker, Y., Isbell, C., 09–15 Jun 2019. Imitating latent policies from observation. In: Proceedings of the 36th International Conference on Machine Learning. Vol. 97 of Proceedings of Machine Learning Research. PMLR, pp. 1755–1763.

Englert, P., Vien, N. A., Toussaint, M., 2017. Inverse KKT: Learning cost functions of manipulation tasks from demonstrations. The International Journal of Robotics Research 36 (13-14), 1474–1488.

Fan, K., 1968. On infinite systems of linear inequalities. Journal of Mathematical Analysis and Applications 21 (3), 475–478.

Fan, K., 1975. Two applications of a consistency theorem for systems of linear inequalities. Linear Algebra and its Applications 11 (2), 171–180.

Gagliardi, D., Russo, G., 2020. On the synthesis of control policies from example datasets. In: 21st IFAC World Congress (see `https://arxiv.org/abs/2001.04428`).

Garrabe, E., Russo, G., 2022. On the design of autonomous agents from multiple data sources. IEEE Control Systems Letters 6, 698–703.

Georgiou, T., Lindquist, A., 2003. Kullback-Leibler approximation of spectral density functions. IEEE Transactions on Information Theory 49 (11), 2910–2917.

Gonçalves da Silva, G. R., Bazanella, A. S., Lorenzini, C., Campestrini, L., 2019. Data-driven LQR control design. IEEE Control Systems Letters 3 (1), 180–185.

Griggs, W., Ordóñez-Hurtado, R., Russo, G., Shorten, R., 2019. A vehicle-in-the-loop emulation platform for demonstrating intelligent transportation systems. In: Control Strategies for Advanced Driver Assistance Systems and Autonomous Driving Functions. Springer, pp. 133–154.

Guan, P., Raginsky, M., Willett, R. M., 2014. Online Markov Decision processes with Kullback–Leibler control cost. IEEE Transactions on Automatic Control 59 (6), 1423–1438.

Guy, T. V., Derakhshan, S. F., Štěch, J., 2018. Lazy fully probabilistic design: Application potential. In: Belardinelli, F., Argente, E. (Eds.), Multi-Agent Systems and Agreement Technologies. Springer International Publishing, Cham, pp. 281–291.

Hanawal, M., Liu, H., Zhu, H., Paschalidis, I., 2019. Learning policies for Markov Decision Processes from data. IEEE Transactions on Automatic Control 64, 2298–2309.

Herzallah, R., 2015. Fully probabilistic control for stochastic nonlinear control systems with input dependent noise. Neural networks 63, 199–207.

Hiebert, K. L., 1980. Solving systems of linear equations and inequalities. SIAM Journal on Numerical Analysis 17 (3), 447–464.

Hou, Z., Xu, J.-X., 2009. On data-driven control theory: the state of the art and perspective. Acta Automatica Sinica 35, 650–667.

Hou, Z.-S., Wang, Z., 2013. From model-based control to data-driven control: Survey, classification and perspective. Information Sciences 235, 3 – 35.

Kappen, H., Gomez, Opper, M., 2012. Optimal control as a graphical model inference problem. Machine Learning 87, 159–182.

Karlin, S., Studden, W. J., 1966. Optimal experimental designs. The Annals of Mathematical Statistics 37 (4), 783–815.

Kárný, M., 1996. Towards fully probabilistic control design. Automatica 32 (12), 1719–1722.

Kárný, M., Guy, T. V., 2006. Fully probabilistic control design. Systems & Control Letters 55 (4), 259–265.

Keel, L. H., Bhattacharyya, S. P., 2008. Controller synthesis free of analytical models: Three term controllers. IEEE Transactions on Automatic Control 53 (6), 1353–1369.

Kirk, D. E., 2004. Optimal control theory: an introduction. Courier Corporation.

Kullback, S., Leibler, R., 1951. On information and sufficiency. Annals of Mathematical Statistics 22, 79–87.

Kárný, M., Kroupa, T., 2012. Axiomatisation of fully probabilistic design. Information Sciences 186 (1), 105 – 113.

Markovsky, I., Rapisarda, P., 2007. On the linear quadratic data-driven control. In: 2007 European Control Conference (ECC). pp. 5313–5318.

McKinnon, C. D., Schoellig, A. P., 2019. Learn fast, forget slow: Safe predictive learning control for sys-

tems with unknown and changing dynamics performing repetitive tasks. IEEE Robotics and Automation Letters 4 (2), 2180–2187.

Moser, D., Ramezani, Z., Gagliardi, D., Zhou, J., del Re, L., 2017. Risk functions oriented autonomous overtaking. In: 2017 11th Asian Control Conference (ASCC). pp. 1017–1022.

Moser, D., Waschl, H., Schmied, R., Efendic, H., del Re, L., 2015. Short term prediction of a vehicle's velocity trajectory using ITS. SAE International Journal of Passenger Cars-Electronic and Electrical Systems 8 (2015-01-0295), 364–370.

Nakka, Y. K., Liu, A., Shi, G., Anandkumar, A., Yue, Y., Chung, S.-J., 2021. Chance-constrained trajectory optimization for safe exploration and learning of nonlinear systems. IEEE Robotics and Automation Letters 6 (2), 389–396.

Nguyen, N. A., Moser, D., Schrangl, P., del Re, L., Jones, S., 2017. Autonomous overtaking using stochastic model predictive control. In: 2017 11th Asian Control Conference (ASCC). pp. 1005–1010.

OpenStreetMap contributors, 2017. Planet dump retrieved from https://planet.osm.org . `https://www.openstreetmap.org` .

Pavon, M., Ferrante, A., 2006. On the Georgiou-Lindquist approach to constrained Kullback-Leibler approximation of spectral densities. IEEE Transactions on Automatic Control 51 (4), 639–644.

Pegueroles, B. G., Russo, G., June 2019. On robust stability of fully probabilistic control with respect to data-driven model uncertainties. In: 2019 18th European Control Conference (ECC). pp. 2460–2465.

Peterka, V., 1981. Bayesian approach to system identification. Elsevier, pp. 239–304.

Ramachandran, D., Amir, E., 2007. Bayesian inverse reinforcement learning. In: Proceedings of the 20th International Joint Conference on Artifical Intelligence. IJCAI'07. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, pp. 2586–2591.

Ratliff, N. D., Bagnell, J. A., Zinkevich, M. A., 2006. Maximum margin planning. In: Proceedings of the 23rd International Conference on Machine Learning. ICML '06. ACM, New York, NY, USA, pp. 729–736.

Ratliff, N. D., Silver, D., Bagnell, J. A., Jul 2009. Learning to search: Functional gradient techniques for imitation learning. Autonomous Robots 27 (1), 25–53.

Rockafeller, R., 1976. Duality and stablity in extremum problems involving convex functions. Pacific Journal of Mathematics 21, 167–186.

Rosolia, U., Borrelli, F., 2018. Learning model predictive control for iterative tasks. A data-driven control framework. IEEE Transactions on Automatic Control 63 (7), 1883–1896.

Russo, G., 2021. On the crowdsourcing of behaviors for autonomous agents. IEEE Control Systems Letters 5 (4), 1321–1326.

Salvador, J. R., delaPena, D. M., Alamo, T., Bemporad, A., 2018. Data-based predictive control via direct weight optimization. IFAC-PapersOnLine 51 (20), 356 – 361, 6th IFAC Conference on Nonlinear Model Predictive Control NMPC 2018.

Singh, M., Vishnoi, N. K., 2014. Entropy, optimization and counting. In: Proceedings of the Forty-Sixth Annual ACM Symposium on Theory of Computing. STOC '14. Association for Computing Machinery, New York, NY, USA, p. 50–59.

Tanaskovic, M., Fagiano, L., Novara, C., Morari, M., 2017. Data-driven control of nonlinear systems: An on-line direct approach. Automatica 75, 1 – 10.

Todorov, E., 2007. Linearly-solvable Markov decision problems. In: Schölkopf, B., Platt, J., Hoffman, T. (Eds.), Advances in Neural Information Processing Systems. Vol. 19. MIT Press.

Todorov, E., 2009. Efficient computation of optimal actions. Proceedings of the National Academy of Sciences 106 (28), 11478–11483.

van Waarde, H. J., 2021. Beyond persistent excitation: Online experiment design for data-driven modeling and control. IEEE Control Systems Letters, 1–1.

van Waarde, H. J., De Persis, C., Camlibel, M. K., Tesi, P., 2020. Willems' fundamental lemma for state-space systems and its extension to multiple datasets. IEEE Control Systems Letters 4 (3), 602–607.

Van Waarde, H. J., Eising, J., Trentelman, H. L., Camlibel, M. K., 2020. Data informativity: a new perspective on data-driven analysis and control. IEEE Transactions on Automatic Control 65 (11), 4753–4768.

Vitus, M. P., Tomlin, C. J., 2013. A probabilistic approach to planning and control in autonomous urban driving. In: 52nd IEEE Conference on Decision and Control. pp. 2459–2464.

Wabersich, K. P., Zeilinger, M. N., June 2018. Scalable synthesis of safety certificates from data with application to learning-based control. In: 2018 European Control Conference (ECC). pp. 1691–1697.

Xu, T., Paschalidis, I. C., June 2019. Learning models for writing better doctor prescriptions. In: 2019 18th European Control Conference (ECC). pp. 2454–2459.

Zhu, B., Baggio, G., 2019. On the existence of a solution to a spectral estimation problem à la Byrnes–Georgiou–Lindquist. IEEE Transactions on Automatic Control 64 (2), 820–825.

Zhu, H., Liu, H., Ataei, A., Munk, Y., Daniel, T., Paschalidis, I. C., 01 2020. Learning from animals: How to navigate complex terrains. PLOS Computational Biology 16 (1), 1–17.

Ziebart, B. D., Maas, A., Bagnell, J. A., Dey, A. K., 2008. Maximum entropy inverse reinforcement learning. In: Proc. AAAI. pp. 1433–1438.

Ziegler, J. G., Nichols, N. B., 1942. Optimum Settings for Automatic Controllers. Transactions of the ASME 64, 759–768.