

# A Probabilistic Framework for Moving-Horizon Estimation: Stability and Privacy Guarantees

Vishaal Krishnan

Sonia Martínez \*

## Abstract

This work proposes a unifying probabilistic framework for the design of robustly asymptotically stable moving-horizon estimators (MHE) for discrete-time nonlinear systems, and a mechanism to incorporate differential privacy in moving-horizon estimation. We begin with an investigation of the classical notion of strong local observability of nonlinear systems and its relationship to optimization-based state estimation. We then present a general moving-horizon estimation framework for strongly locally observable systems, as an iterative minimization scheme in the space of probability measures. This framework allows for the minimization of the estimation cost with respect to different metrics. In particular, we consider two variants, which we name  $W_2$ -MHE and KL-MHE, where the minimization scheme uses the 2-Wasserstein distance and the KL-divergence, respectively. The  $W_2$ -MHE yields a gradient-based estimator whereas the KL-MHE yields a particle filter, for which we investigate asymptotic stability and robustness properties. Stability results for these moving-horizon estimators are derived in the probabilistic setting, against the backdrop of the classical notion of strong local observability which, to the best of our knowledge, differentiates it from other previous works. We then propose a mechanism to encode differential privacy of the measurements used by the estimator via an entropy regularization of the MHE objective functional. In particular, we find sufficient bounds on the regularization parameter to achieve the desired level of differential privacy. Numerical simulations demonstrate the performance of these estimators.

## 1 Introduction

Moving-horizon estimation (MHE) is an optimization-based state estimation method that uses the most recent measurements within a moving-time horizon to recursively update state estimates. In principle, its optimization-based formulation enables it to handle nonlinearities and state constraints much more effectively than other known methods. This, coupled with the adoption of increasingly powerful, inexpensive computing platforms has brought new impetus to the adoption of moving-horizon estimation in various data-driven applications. In many cases, data is acquired from particular individuals or users, which introduces new ethical concerns about data collection and manipulation, highlighting an increasing need for data privacy. Such is the case in home monitoring and traffic estimation (with vehicle GPS data) applications, to name a few. Motivated by this, here we design and analyze a new class of moving-horizon estimation filters that can guarantee the differential-privacy of the data.

---

\*The authors are with the Department of Mechanical and Aerospace Engineering, University of California at San Diego, La Jolla CA 92093 USA (email: v6krishn@ucsd.edu; soniamd@ucsd.edu).

The origins of MHE can be traced back to the limited memory optimal filters introduced in [19]. Theoretical investigations on MHE have broadly been directed at their asymptotic stability [2, 29, 33] and robustness [18, 20, 24] properties. These properties have primarily been built upon underlying assumptions of input/output-to-state (IOSS) stability, which is adopted as the notion of detectability, wherein the norm of the state is bounded given the sequences of inputs and outputs. However, alternative foundations for the stability results in other classical notions of observability, such as strong observability [25], have remained unexplored. The connection between nonlinear observability theory and estimation problems runs deep, see [22] and more recently [32], and it is worthwhile to explore this connection in the context of optimization-based estimation methods such as moving-horizon estimation.

The problem of state estimation is fundamentally about dealing with uncertainty, manifested as uncertainty in the initial conditions and/or in the evolution of the system in the presence of unknown disturbances. This is appropriately formulated in the space of probability measures over the state space of the system. Recent advances in gradient flows in the space of probability measures [5], [30], and the corresponding discrete-time movement-minimizing schemes [28] present powerful theoretical tools that can be applied to recursive optimization-based estimation methods such as moving-horizon estimation, and can serve as a unifying framework for their design and analysis.

Another important consideration in the MHE problem is the cost of computation. The problem formulation more commonly involves solving an optimization problem at every time instant, with the state estimate and disturbances as decision variables in the optimization, where the dimension of the problem scales with the size of the horizon. This approach, in general, tends to be computationally intensive, which poses a hurdle for implementation in real-time. This has motivated the search for fast MHE that implement one or more iterations of the optimization at every time instant. Recently, in [3], [4], the authors develop such a method for noiseless systems and provide theoretical guarantees on convergence. However, these works assume the convexity of the cost function, which is restrictive for general nonlinear systems, and not well connected to notions of observability. None of these works has considered the additional question of privacy.

Differential privacy [11] has emerged over the past decade as a benchmark in data privacy. The typical setting assumes independence between the records in static databases; however, basic existing mechanisms fail to provide guarantees when correlations exist between the records in the database. This is the case when data is employed by a state estimation process whose output is then released: there is a dynamic system from which a time series of sensor measurements is obtained, and the measurement data and the released estimates are correlated.

In [8, 9], the authors generalize the definition of differential privacy to include general notions of distance between datasets and design differentially private mechanisms for Bayesian inference. In [23, 31], the authors investigate privacy-preserving mechanisms for the case where correlations exist between database records. Privacy-preserving mechanisms for functions and functional data were investigated in [15]. The work [27] studies the problem of differentially-private state estimation, introducing the formal notion of differential privacy into the framework of Kalman filter design for dynamic systems. The authors of [13] consider the problem of optimal state estimation for linear discrete-time systems with measurements corrupted by Laplacian noise. A finite-dimensional distributed convex optimization is consid-

ered in [26], where differential privacy is achieved by perturbation of the objective function. We refer the reader to [7] for a broad overview of the systems and control-theoretic perspective on differential privacy.

*Contributions:* The contributions of this work are two-fold: establishing the robust asymptotic stability of the proposed moving-horizon estimator in a probabilistic framework, founded on the notion of strong local observability; and incorporating differential privacy in moving-horizon estimation. We begin with the well-studied notion of strong local observability of nonlinear, discrete-time systems and investigate its relationship to the optimization-based state estimation problem. To handle uncertain initial conditions and the possible non-uniqueness of solutions to the estimation problem, we adopt a generalized problem formulation over the space of probability measures over the state space. More precisely, we define the MHE as a proximal gradient descent in the space of probability measures, with a non-convex, time-varying cost function. This probabilistic setting serves as a unifying framework for moving-horizon estimation and allows us to develop different classes of moving-horizon estimators by simply varying the metric used to define the proximal operator, and to obtain implementable filters by Monte Carlo methods. We then consider the Wasserstein metric and the KL-divergence, which yield the more familiar MHE and a particle filter, respectively. Following this, we present an analysis of the convergence and robustness properties of these estimators in the probabilistic setting, under assumptions of strong local observability. Further, we modify the optimization problem (in the space of probability measures) by an entropy regularization to derive conditions that guarantee a desired level of differential privacy for these filters.

*Paper organization:* The rest of the paper is organized as follows. In Section 2, we introduce the notation and mathematical preliminaries used in the paper. We present the optimization-based state estimation problem in Section 4, where Section 4.1 deals with the Full Information Estimation (FIE) problem and the Moving-horizon Estimation (MHE) problem is introduced in Section 4.2. We present the MHE method based on proximal gradient descent with the Wasserstein metric in Section 5, and with the KL-divergence in Section 6. In Section 7, we address the differential privacy considerations for the moving-horizon estimators designed. The results from numerical experiments are presented in Section 8, with the conclusions in Section 9.

## 2 Notation and preliminaries

In this section, we introduce the notation and mathematical preliminaries relevant to this paper.

Let  $\|\cdot\| : \mathbb{R}^d \rightarrow \mathbb{R}_{\geq 0}$  denote the Euclidean norm on  $\mathbb{R}^d$  and  $|\cdot| : \mathbb{R} \rightarrow \mathbb{R}_{\geq 0}$  the absolute value function. We denote by  $\nabla = \left( \frac{\partial}{\partial x_1}, \dots, \frac{\partial}{\partial x_n} \right)$  the gradient operator in  $\mathbb{R}^d$ . For any  $x \in \mathcal{X} \subset \mathbb{R}^d$ , we let  $\mu \in \mathcal{P}(\mathcal{X})$  be an absolutely continuous probability measure on  $\mathcal{X} \subset \mathbb{R}^d$ . We denote by  $\rho$  the corresponding density function, where  $d\mu = \rho d\text{vol}$ , with  $\text{vol}$  being the Lebesgue measure. For  $M \subseteq \mathcal{X}$ , let the distance  $d(x, M)$  of a point  $x \in \mathcal{X}$  to the set  $M$  be given by  $d(x, M) = \inf_{y \in M} \|x - y\|$ . We denote by  $\langle p, q \rangle$  the inner product of functions  $p, q : \mathcal{X} \rightarrow \mathbb{R}$  with respect to the Lebesgue measure  $\text{vol}$ , given by  $\langle p, q \rangle = \int_{\mathcal{X}} pq d\text{vol}$ . Let  $F : \mathcal{P}(\mathcal{X}) \rightarrow \mathbb{R}$  be a smooth real-valued function on the space of probability measures on  $\mathcal{X} \subset \mathbb{R}^d$ . We denote by  $\frac{\delta F}{\delta \mu}(x)$  the derivative of  $F$  with respect to  $\mu$ , see [12], such that a perturbation  $\delta\mu$

of the measure results in a perturbation  $\delta F = \int_{\mathcal{X}} \frac{\delta F}{\delta \mu} d(\delta \mu)$ . Given a map  $\mathcal{T} : \mathcal{X} \rightarrow \mathcal{Y}$  and a measure  $\mu \in \mathcal{P}(\mathcal{X})$ , in the space of probability measures  $\mathcal{P}(\mathcal{X})$ , we let  $\nu = \mathcal{T}_{\#}\mu$  denote the pushforward measure of  $\mu$  by  $\mathcal{T}$ , where for a measurable set  $\mathcal{B} \subset \mathcal{T}(\mathcal{X})$ , we have  $\nu(\mathcal{B}) = \mathcal{T}_{\#}\mu(\mathcal{B}) = \mu(\mathcal{T}^{-1}(\mathcal{B}))$ . Moreover, we denote by  $\mathbb{E}_{\mu}$  the expectation operator w.r.t. the measure  $\mu$ .

We now introduce the notion of  $l$ -smoothness that underlies the results on convergence of gradient descent methods.

**Definition 1. ( $l$ -smoothness).** A function  $p : \mathcal{X} \rightarrow \mathbb{R}$  is called  $l$ -smooth (or Lipschitz differentiable) if for any  $x, y \in \mathcal{X}$ , we have  $|\nabla p(y) - \nabla p(x)| \leq l\|y - x\|$ .

The following lemma [6] can be easily verified for  $l$ -smooth functions:

**Lemma 1. ( $l$ -smooth functions).** For an  $l$ -smooth function  $p : \mathcal{X} \rightarrow \mathbb{R}$  and any  $x, y \in \mathcal{X}$ , we have  $|p(y) - p(x) - \langle \nabla p(x), y - x \rangle| \leq \frac{l}{2}\|y - x\|^2$ . •

We now define the proximal operator on  $\mathcal{X}$  with respect to a function  $F : \mathcal{X} \rightarrow \mathbb{R}$ , as follows:

$$\text{prox}_F(x) = \arg \min_{\tilde{x} \in \mathcal{X}} \frac{1}{2}\|\tilde{x} - x\|^2 + F(\tilde{x}).$$

The notion of observability used in this paper is intricately related to solutions of inverse problems, with an associated notion of well-posedness that is introduced below:

**Definition 2. (Well posedness [21]).** Let  $\mathcal{X}$  and  $\mathcal{Y}$  be normed spaces, and  $P : \mathcal{X} \rightarrow \mathcal{Y}$  a mapping. The equation  $P(x) = y$  is called well-posed if:

1. *Existence:* For every  $y \in \mathcal{Y}$ , there is (at least one)  $x \in \mathcal{X}$  such that  $P(x) = y$ .
2. *Uniqueness:* For every  $y \in \mathcal{Y}$ , there is at most  $x \in \mathcal{X}$  such that  $P(x) = y$ .
3. *Stability:* The solution  $x$  depends continuously on  $y$ , that is, for any sequence  $\{x_i\} \subset \mathcal{X}$  such that  $P(x_i) \rightarrow P(x)$ , it follows that  $x_i \rightarrow x$ .

We now introduce the notion of lower semicontinuity of set-valued maps, which underlies some of the results on optimization-based state estimation in this paper.

**Definition 3. (Lower semicontinuity of set-valued maps).** A point-to-set mapping  $H : \mathcal{Z} \subset \mathbb{R} \rightrightarrows \mathbb{R}^d$  is lower semicontinuous at a point  $\alpha \in \mathcal{Z}$  if for any  $x \in H(\alpha)$  and sequences  $\{\alpha_i\} \subseteq \mathcal{Z}$ ,  $\{x_i\} \subseteq \mathbb{R}^d$  with  $\{\alpha_i\} \rightarrow \alpha$ ,  $\{x_i\} \rightarrow x$  such that  $x_i \in H(\alpha_i)$  for all  $i$ , it holds that  $x \in H(\alpha)$ . If  $H$  is lower semicontinuous at every  $\alpha \in \mathcal{Z}$ , then  $H$  is said to be lower semicontinuous on  $\mathcal{Z}$ .

We now define some notions of distance in the space of probability measures. Let  $\mu_1, \mu_2 \in \mathcal{P}(\mathcal{X})$  be two absolutely continuous probability measures on  $\mathcal{X}$ , with  $\rho_1, \rho_2$  being the corresponding density functions. Also, let  $\Pi(\mu_1, \mu_2) \subset \mathcal{P}(\mathcal{X} \times \mathcal{X})$  be the space of joint probability measures that have  $\mu_1$  and  $\mu_2$  as their marginals. The 2-Wasserstein distance  $W_2(\mu_1, \mu_2)$  between  $\mu_1$  and  $\mu_2$  is given by:

$$W_2^2(\mu_1, \mu_2) = \inf_{\pi \in \Pi(\mu_1, \mu_2)} \int_{\mathcal{X} \times \mathcal{X}} \|x - y\|^2 d\pi(x, y).$$

In what follows, we let  $\frac{\delta W_2^2(\mu_1, \mu_2)}{\delta \mu_1} = \phi_1$ , where  $\phi_1$  is the so-called the Kantorovich potential [30] associated with the transport from  $\mu_1$  to  $\mu_2$ .

The KL-divergence from  $\mu_1$  to  $\mu_2$  is given by:

$$D_{\text{KL}}(\mu_1 || \mu_2) = \int_{\mathcal{X}} \log \left( \frac{d\mu_1(x)}{d\mu_2(x)} \right) d\mu_1(x) = \int_{\mathcal{X}} \rho_1(x) \log \left( \frac{\rho_1(x)}{\rho_2(x)} \right) d\text{vol}(x).$$

The max-divergence between  $\mu_1$  and  $\mu_2$  is defined as:

$$D_{\text{max}}(\mu_1, \mu_2) = \sup_{x \in \mathcal{X}} \left| \log \left( \frac{\rho_1(x)}{\rho_2(x)} \right) \right|.$$

We refer the reader to [14] for a detailed overview of the relations between the various metrics and divergences in probability spaces.

We define an estimator  $\mathcal{E} : \mathcal{Y} \rightarrow \mathcal{P}(\mathcal{X})$  as a function that accepts as input data  $y$  from the metric space  $\mathcal{Y}$  and releases as output  $\mathcal{E}[y]$ , a probability measure over the space  $\mathcal{X}$ .

**Definition 4. (Differential privacy).** *Given  $\delta$ , an estimator  $\mathcal{E}$  is  $\epsilon$ -differentially private if for any two  $\delta$ -adjacent measurements  $y_1, y_2 \in \mathcal{Y}$  (that is  $d_{\mathcal{Y}}(y_1, y_2) \leq \delta$ ), and any measurable  $A \subseteq \mathcal{X}$ , we have  $\mathcal{E}[y_1](A) \leq e^\epsilon \mathcal{E}[y_2](A)$ .*

Note that the condition  $d_{\mathcal{Y}}(y_1^m, y_2^m) \leq \delta$  is a generalization of the notion of adjacency to arbitrary metric spaces that we adopt in this paper. We now have the following lemma on the connection between the notions of differential privacy and max-divergence introduced above:

**Lemma 2. (Differential privacy and max-divergence).** *An estimator  $\mathcal{E}$  is  $\epsilon$ -differentially private iff  $D_{\text{max}}(\mathcal{E}[y_1], \mathcal{E}[y_2]) \leq \epsilon$  for any  $y_1, y_2 \in \mathcal{Y}$  with  $d_{\mathcal{Y}}(y_1, y_2) \leq \delta$ .*

*Proof.* Clearly, if for any  $y_1, y_2 \in \mathcal{Y}$  with  $d_{\mathcal{Y}}(y_1, y_2) \leq \delta$  we have  $D_{\text{max}}(\mathcal{E}[y_1], \mathcal{E}[y_2]) \leq \epsilon$ , then:

$$\epsilon \geq D_{\text{max}}(\mathcal{E}[y_1], \mathcal{E}[y_2]) = \sup_{x \in \mathcal{X}} \left| \log \left( \frac{\rho_1(x)}{\rho_2(x)} \right) \right| \geq \left| \log \left( \frac{\rho_1(x)}{\rho_2(x)} \right) \right|.$$

This implies that for any  $x \in \mathcal{X}$ , we have  $\rho_1(x) \leq e^\epsilon \rho_2(x)$ . Now, for any  $A \subseteq \mathcal{X}$ , we have  $\mathcal{E}[y_1^m](A) = \int_A \rho_1(x) d\text{vol} \leq \int_A e^\epsilon \rho_2(x) d\text{vol} = e^\epsilon \int_A \rho_2(x) d\text{vol} = e^\epsilon \mathcal{E}[y_2^m](A)$ , which implies that  $\mathcal{E}$  is  $\epsilon$ -differentially private. The forward implication can be easily verified.  $\square$

Thus,  $\epsilon$ -differential privacy essentially imposes an upper bound on the sensitivity of the estimate generated by  $\mathcal{E}$  (in the sense of the max-divergence  $D_{\text{max}}$ ), to the measurement.

### 3 Observability notions

In this paper, we consider systems of the form:

$$\Omega : \begin{cases} x_{k+1} = f(x_k, w_k), \\ y_k = h(x_k) + v_k, \end{cases} \quad (1)$$

where  $f : \mathbb{X} \times \mathbb{W} \rightarrow \mathbb{X}$  and  $h : \mathbb{X} \rightarrow \mathbb{Y}$ ,  $w_k \in \mathbb{W}$  is the process noise,  $v_k \in \mathbb{V}$  is the measurement noise at time instant  $k$ , and  $\mathbb{X} \subset \mathbb{R}^{d_x}$ ,  $\mathbb{Y} \subset \mathbb{R}^{d_y}$ ,  $\mathbb{W} \subset \mathbb{R}^{d_w}$ , and  $\mathbb{V} \subset \mathbb{R}^{d_v}$ .

**Assumption 1. (Lipschitz continuity).** The functions  $f$  and  $h$  are Lipschitz continuous, with  $\|f(x_1, w_1) - f(x_2, w_2)\| \leq c_f^{(1)} \|x_1 - x_2\| + c_f^{(2)} \|w_1 - w_2\|$  and  $\|h(x_1) - h(x_2)\| \leq c_h \|x_1 - x_2\|$ .

**Assumption 2. (Noise characteristics).** The noise sequences  $\{w_k\}_{k \in \mathbb{N}}$  and  $\{v_k\}_{k \in \mathbb{N}}$  are i.i.d samples from distributions  $\omega$  and  $\nu$  (with supports in  $\mathbb{W}$  and  $\mathbb{V}$ ). The sets  $\mathbb{W}$  and  $\mathbb{V}$  are bounded, with  $|w_k| \leq W$  and  $|v_k| \leq V$ . Moreover, we assume that  $\mathbb{E}_\omega[w_k] = 0$  and  $\mathbb{E}_\nu[v_k] = 0$ .

We also introduce the following autonomous system corresponding to (1):

$$\Sigma : \begin{cases} x_{k+1} = f(x_k, 0) = f_0(x_k), \\ y_k = h(x_k). \end{cases} \quad (2)$$

With a slight abuse of notation, for any  $x \in \mathbb{X}$ , we let  $\Sigma_T(x) = (h(x), h \circ f_0(x), \dots, h \circ f_0^T(x))$ , the sequence of outputs over a horizon of length  $T + 1$  for the system (2) from the state  $x \in \mathbb{X}$ . Similarly, for the system (1), we let  $\Omega(x, \mathbf{w}_{i:j}) = (h(x), h \circ f(x, w_i), \dots, h \circ f(\dots f(f(x, w_i), w_{i+1}), \dots, w_j))$ , for some sequence of process noise samples  $\{w_k\}$ , where  $\mathbf{w}_{i:j} = (w_i, \dots, w_j)$ .

The theoretical results in the moving-horizon estimation literature have largely been derived in the setting of input/output-to-state (IOSS) stability, as in [18, 20, 29] to name a few, which is a notion of norm-observability, see [17], wherein the norm of the state is bounded using the sequences of inputs and outputs. However, there are other classical notions of observability based on the notion of distinguishability, which generalize the approach taken to linear systems. For a detailed treatment, we refer the reader to [25] and [1]. In this paper, we explore the connection between the classical notion of strong local observability and moving-horizon estimation.

We now introduce the notion of strong local observability used in this paper:

**Definition 5. (Strong local observability).** The system  $\Sigma$  defined in (2) is called strongly locally observable if there exists a  $T_0 \in \mathbb{N}$  such that for any given  $x \in \mathbb{X}$  and  $T \geq T_0$ , we have that  $\Sigma_T^{-1} \circ \Sigma_T(x)$  is a set of isolated points. Moreover, for all  $x \in \mathbb{X}$  and  $T_1, T_2 \geq T_0$ , we have that  $\Sigma_{T_1}^{-1} \circ \Sigma_{T_1}(x) = \Sigma_{T_2}^{-1} \circ \Sigma_{T_2}(x)$ . We call  $T_0$  the minimum horizon length of  $\Sigma$ .

The above definition is equivalent to the definitions contained in [1, 25], which has been restated it in a manner suitable for the optimization-based estimation framework considered here. As seen from the above definition, strong observability is based on a distinguishability notion, and when it holds globally (i.e.,  $\Sigma_T^{-1} \circ \Sigma_T = id$  for all  $T \geq T_0$ ) it is equivalent to the notion of uniform observability, as established in [16].

For systems with process noise, of the form  $\Omega$  in (1), we introduce the notion of almost sure strong local observability.

**Definition 6. (Almost sure strong local observability).** The system  $\Omega$  defined in (1) is called almost surely strongly locally observable if there exists a  $T^w \in \mathbb{N}$  such that, given a process noise sequence  $\mathbf{w}_{0:T-1} \in \mathbb{W}^T$ , for  $T \geq T^w$ , any  $\mathbf{y}_{0:T} = \Omega_{\mathbf{w}_{0:T-1}}(x) \in \mathbb{Y}^{T+1}$ , and  $T \geq T^w$ , we have that  $\Omega_{\mathbf{w}_{0:T-1}}^{-1}(\mathbf{y}_{0:T})$  is a set of isolated points almost surely. More precisely, the set of noise sequences  $\mathbf{w}_{0:T-1}$  for which  $\Omega_{\mathbf{w}_{0:T-1}}^{-1}(\mathbf{y}_{0:T})$  is not a set of isolated points, is of measure zero. Moreover, we call  $T^w$  the minimum horizon length of  $\Omega$ .

We now present a fundamental result that characterizes strong local observability via a rank condition.

**Lemma 3. (Observability rank condition [25]).** *The system  $\Sigma$  is locally strongly observable with minimum horizon length  $T_0$  if and only if  $\text{Rank}(\nabla\Sigma_T(x)) = \dim(\mathbb{X})$  for all  $T \geq T_0$  and  $x \in \mathbb{X}$ . The system  $\Omega$  is almost surely locally strongly observable with minimum horizon length  $T^w$  if and only if  $\text{Rank}(\nabla\Omega_{\mathbf{w}_{0:T-1}}(x)) = \dim(\mathbb{X})$  almost surely for all  $T \geq T^w$  and  $x \in \mathbb{X}$ .* •

We now present an example to illustrate these concepts.

**Example 1.** *Consider a system with the state space  $\mathbb{X} = (0, \infty)$ , with  $x_{k+1} = f_0(x_k)$  and  $y_k = h(x_k)$ , such that:*

$$f_0(x) = \begin{cases} 3x, & \text{for } x \in (0, a\pi - \epsilon], \\ \gamma(x) & \text{for } x \in (a\pi - \epsilon, a\pi + \epsilon], \\ 2x + a\pi, & \text{for } x \in (a\pi + \epsilon, \infty), \end{cases}$$

for some  $a \in \mathbb{N}$ ,  $\epsilon$  small and a smooth function  $\gamma$  such that  $\gamma(a\pi - \epsilon) = 3(a\pi - \epsilon)$  and  $\gamma(a\pi + \epsilon) = 2(a\pi + \epsilon) + a\pi$ . Moreover, let the output  $h(x) = \sin x$ . We note that  $\nabla h(x) = \cos x$  which implies that  $\nabla h((2m+1)\pi/2) = 0$  for all  $m \in \mathbb{N}$ . Applying Lemma 3 for this system, we can infer that for  $a = 2$ , we get that the minimum horizon length  $T_0 = 3$ . This is because the system becomes strongly locally observable at  $x = \pi/2$  only over a horizon of length  $T_0 = 3$ , that is  $\nabla\Sigma_k(\pi/2) = \mathbf{0}_{k+1}$  for  $k \in \{0, 1, 2\}$ . This is a case of a one-dimensional system which is strongly locally observable with a minimum horizon of length  $T_0 = 3$ . With larger values of  $a$ , the minimum horizon length is further increased. •

We make the following assumption in the rest of the paper:

**Assumption 3. (Strong local observability).**

1. The system  $\Sigma$  in (2) is strongly locally observable with minimum horizon length  $T_0$ .
2. The system  $\Omega$  in (1) is almost surely strongly locally observable with minimum horizon length  $T^w$ .

## 4 Optimization-based state estimation

We now begin by addressing the state estimation problem for the autonomous system  $\Sigma$ , and develop a recursive moving-horizon estimator for it.

### 4.1 Full-Information Estimation (FIE)

Let  $\{y_k\}_{k \in \{0\} \cup \mathbb{N}}$  be a sequence of measurements generated by the system  $\Sigma$ . Let  $\{0, \dots, T\}$  be a time horizon such that  $T \geq T_0$ , the minimum horizon length of the system  $\Sigma$ , and denote  $\mathbf{y}_{0:T} = (y_0, \dots, y_T)$ . The problem of estimation essentially aims at characterizing  $\Sigma_T^{-1}(\mathbf{y}_{0:T})$ , which is an inverse problem, and optimal estimation formulates this problem

as an optimization. Assumptions 1, and 3, on Lipschitz continuity and strong local observability, respectively, ensure that the inverse problem is locally well-posed as in Definition 2.

To formulate the inverse problem as an optimization, consider a convex function  $J_T(\mathbf{y}_{0:T}, \cdot) : \mathbb{Y}^{T+1} \rightarrow \mathbb{R}_{\geq 0}$  such that  $J_T(\mathbf{y}_{0:T}, \xi) = 0$  if and only if  $\xi = \mathbf{y}_{0:T}$ . Moreover, we let  $\lim_{T \rightarrow \infty} J_T(\mathbf{y}_{0:T}, \Sigma_T(x)) = \infty$  if  $x \notin \Sigma_T^{-1}(\mathbf{y}_{0:T})$  for  $T \geq T_0$ . Now, the problem of interest becomes:

$$x_0 \in \arg \min_{x \in \mathbb{X}} J_T(\mathbf{y}_{0:T}, \Sigma_T(x)). \quad (3)$$

In the above,  $\mathbf{y}_{0:T}$  is the data in the estimation problem, which is given. Since the objective is to solve the original inverse problem, and we would like to use gradient descent-based methods, we would like for every local minimizer of  $J_T(\mathbf{y}_{0:T}, \Sigma_T(x))$  to belong to the set  $\Sigma_T^{-1}(\mathbf{y}_{0:T})$ , or, in other words, that every local minimizer is also global. We therefore make the following additional assumption on the system  $\Sigma$  and the choice of  $J_T$ . For a conciseness of notation, in the following assumption and lemma, we let  $J_T(\cdot) = J_T(\mathbf{y}_{0:T}, \cdot)$ , suppressing the data  $\mathbf{y}_{0:T}$  in the notation where useful, and is understood from context.

**Assumption 4. (Lower semicontinuity of sublevel sets).** *We assume that, for all  $T \geq T_0$ , the convex function  $J_T : \mathbb{Y}^{T+1} \rightarrow \mathbb{R}$  is such that the set-valued map  $\mathcal{S}_{\mathbb{X}}(\alpha) = \Sigma_T^{-1}(\mathcal{S}_{\mathbb{Y}^{T+1}}^{J_T}(\alpha) \cap \Sigma_T(\mathbb{X}))$  is lower semicontinuous, where  $\mathcal{S}_{\mathbb{Y}^{T+1}}^{J_T}(\alpha) = \{\xi \in \mathbb{Y}^{T+1} | J_T(\xi) \leq \alpha\}$ .*

The above assumption ensures that the function  $J_T(\mathbf{y}_{0:T}, \Sigma_T(\cdot))$  satisfies the condition for the local minimizers to be global (Theorem 1 from [34]). The following lemma provides a sufficient condition for it to hold.

**Lemma 4. (Second-order sufficient condition for lower semicontinuity).** *Assumption 4 holds if for any  $x \in \mathbb{X}$  such that  $\nabla(J_T(\mathbf{y}_{0:T}, \Sigma_T(x))) = 0$  we have  $J_T(\mathbf{y}_{0:T}, \Sigma_T(x)) = 0$ , or the following condition holds when  $J_T(\mathbf{y}_{0:T}, \Sigma_T(x)) \neq 0$  for any  $v \in \mathbb{R}^{d_x}$ ,  $v \neq 0$ :*

$$\frac{\left\langle \nabla^2_{\Sigma_T}[v, v](x), \nabla J_T \Big|_{\Sigma_T(x)} \right\rangle}{\|\nabla_{\Sigma_T}[v]\|^2} \leq -\lambda_{\max} \left( \text{Hess } J_T \Big|_{\Sigma_T(x)} \right),$$

where  $\text{Hess } J_T$  is the Hessian of  $J_T$ . •

The final inequality in Lemma 4 merely states that those critical points at which the cost function does not reach the global minimum value are local maximizers.

We are now ready to present the following theorem that establishes the equivalence between the inverse problem of characterizing the set  $\Sigma_T^{-1}(\mathbf{y}_{0:T})$  and the optimization (3).

**Theorem 1. (Inverse as minimizer).** *For a convex  $J_T(\mathbf{y}_{0:T}, \cdot) : \mathbb{Y}^{T+1} \rightarrow \mathbb{R}_{\geq 0}$  such that  $J_T(\mathbf{y}_{0:T}, \xi) = 0$  if and only if  $\xi = \mathbf{y}_{0:T}$  for any  $\mathbf{y}_{0:T} \in \mathbb{Y}^{T+1}$ , under Assumptions 3 and 4, and any  $T \geq T_0$ , it holds that  $z \in \Sigma_T^{-1}(\mathbf{y}_{0:T})$  if and only if  $z$  is a minimizer of  $J_T(\mathbf{y}_{0:T}, \Sigma_T(\cdot))$ .*

*Proof.* If  $z \in \Sigma_T^{-1}(\mathbf{y}_{0:T})$ , we have that  $h \circ f_0^k(z) = y_k$  for all  $k \in \{0, \dots, T\}$ . It now follows that  $J_T(\mathbf{y}_{0:T}, \Sigma_T(z)) = 0$ . Since,  $J_T(\mathbf{y}_{0:T}, \Sigma_T(z)) \geq 0$  by definition, we infer that  $z$  is a global minimizer of  $J_T(\mathbf{y}_{0:T}, \Sigma_T(\cdot))$ .

Suppose that  $z$  is a local minimizer of  $J_T(\mathbf{y}_{0:T}, \Sigma_T(\cdot))$ . By Assumption 4 and Theorem 1 in [34], we get that the local minima of  $J_T(\mathbf{y}_{0:T}, \Sigma_T(\cdot))$  are also global, which implies that  $J_T(\mathbf{y}_{0:T}, \Sigma_T(z)) = 0$ , and therefore  $\Sigma_T(z) = \mathbf{y}_{0:T}$ .  $\square$

Theorem 1 suggests that the state estimates for the system  $\Sigma$  can be obtained by minimizing  $J_T(\mathbf{y}_{0:T}, \Sigma_T(\cdot))$  over a horizon of length  $T \geq T_0$ . This is also called the full information estimation (FIE) problem in the optimal state estimation literature [18, 29], as it works with the entire sequence of output measurements over the horizon  $\{0, \dots, T\}$ .

Now, from Assumption 3 and Theorem 1, we have that  $\Sigma_T^{-1}(\mathbf{y}_{0:T})$  is a set of isolated points which are minimizers of  $J_T(\mathbf{y}_{0:T}, \Sigma_T(\cdot))$ . It then follows that  $\Sigma_T^{-1}(\mathbf{y}_{0:T})$  is the set of stable fixed points of the negative gradient vector field of  $J_T(\mathbf{y}_{0:T}, \Sigma_T(\cdot))$ . We let  $\mathcal{C}_0$  be the basin of attraction of this set. Moreover, we note that  $f^k(\Sigma_T^{-1}(\mathbf{y}_{0:T}))$  is the set of stable fixed points of the negative gradient vector field of  $J_T(\mathbf{y}_{k:k+T}, f^k \circ \Sigma_T(\cdot))$ , and we let  $\mathcal{C}_k$  be the basin of attraction of  $\Sigma_T^{-1}(\mathbf{y}_{k:k+T})$ . We have used above the fact that  $\Sigma_T^{-1}(\mathbf{y}_{k:k+T}) = f_0^k(\Sigma_T^{-1}(\mathbf{y}_{0:T}))$ , which follows from the definition of strong local observability.

We now lift the FIE problem (3) to the space of probability measures over  $\mathbb{X}$ , as a minimization in expectation of the estimation objective function:

$$\mu_0 \in \arg \min_{\mu \in \mathcal{P}(\mathbb{X})} \mathbb{E}_\mu [J_T(\mathbf{y}_{0:T}, \Sigma_T(\cdot))]. \quad (4)$$

The above formulation allows us to capture information about the (probably many) optimal estimates through a probability measure  $\mu_0$ , and help encode distributional constraints, which will be considered in a forthcoming publication.

In the following, we develop recursive moving-horizon estimators that generate sequences  $\{\mu_k\}_{k \in \mathbb{N}}$  of probability measures in  $\mathcal{P}(\mathbb{X})$  as estimates. We then obtain practically implementable estimators using Monte Carlo methods to sample from the measures  $\mu_k$ .

## 4.2 Moving-Horizon Estimation (MHE)

In the previous section, we presented a formulation of the full information estimation (FIE) problem for the autonomous system  $\Sigma$ , which uses the entire measurement sequence over a horizon of length  $T \geq T_0$ . However, the minimum horizon length  $T_0$  may be large, which would make the estimation computationally intensive. Moreover, we would like to progressively assimilate the incoming measurements online. We therefore adopt a moving-horizon estimation method which, at any time instant  $k+N$ , uses the output measurements from the horizon  $\{k+1, \dots, k+N\}$  (of length  $N < T_0$ ), and the state estimate at the time instant  $k-1$ , to obtain the state estimate at instant  $k$ , recursively.

We let  $G_k^N(z) = J_{N-1}(\mathbf{y}_{k+1:k+N}, \Sigma_N(z))$  be the objective function over the horizon  $\{k+1, \dots, k+N\}$ , at the time instant  $k+N$ , where  $\mathbf{y}_{k+1:k+N} = (y_{k+1}, \dots, y_{k+N})$ .

**Assumption 5. (Moving-horizon cost).** *We make the following assumptions on the cost function  $G_k^N$ :*

1. the cost  $G_k^N$  is  $l$ -smooth,

2. it holds that  $|G_{k+1}^N(f_0(z)) - G_k^N(z)| \leq L\|\nabla G_k^N(z)\|^2$ ,
3. the previous constants are such that  $lL \leq \frac{1}{2}$ ,
4. for any two  $\delta$ -adjacent measurements  $\mathbf{y}, \tilde{\mathbf{y}} \in \mathbb{Y}^{T+1}$ , such that  $\|\mathbf{y} - \tilde{\mathbf{y}}\| \leq \delta$  and with corresponding costs  $G_k^N$  and  $\tilde{G}_k^N$ , for  $k \in \{0, \dots, T\}$  and  $N \leq T - k$ , we have  $\|\nabla(G_k^N - \tilde{G}_k^N)(x)\| \leq l\delta$  for all  $x \in \mathbb{X}$ .

We now formulate the general moving-horizon estimation method as follows:

$$\begin{aligned} \mu_k \in \arg \min_{\mu \in \mathcal{P}(\mathbb{X})} D(\mu, f_{0\#}\mu_{k-1}) + \eta \mathbb{E}_\mu [G_k^N], \\ \text{given } \mu_0 \in \mathcal{P}(\mathbb{X}), \end{aligned} \tag{5}$$

where  $D : \mathcal{P}(\mathbb{X}) \times \mathcal{P}(\mathbb{X}) \rightarrow \mathbb{R}_{\geq 0}$  is a placeholder for a metric, divergence or transport cost on  $\mathcal{P}(\mathbb{X})$ . We obtain implementable observers from the above formulation by sampling from the measures, by Monte Carlo methods. As discussed in the ensuing sections, using the 2-Wasserstein distance  $W_2$  yields the more familiar MHE formulation, whereas with the KL-divergence we obtain a moving-horizon particle filter. Hence, this formulation is proposed as a unifying probabilistic framework for moving-horizon estimation, where different estimators are generated by different choices of  $D$ .

We now introduce the following asymptotic stability notion for estimators that will be used in investigating the properties of the estimators we design.

**Definition 7. (Asymptotic stability of state estimator).** We call an estimator of the form (5) an asymptotically stable observer for the system  $\Sigma$  if the sequence of estimates  $\{\mu_k\}_{k \in \mathbb{N}}$  is such that  $\lim_{k \rightarrow \infty} \mu_k(\Sigma_T^{-1}(\mathbf{y}_{k:k+T})) = 1$  for  $T \geq T_0$ .

## 5 A $W_2$ -Moving-Horizon Estimator

In this section, we derive a moving-horizon estimator, which we refer to as the  $W_2$ -MHE, to generate a sequence of probability distributions  $\{\mu_k\}_{k \in \mathbb{N}}$ . This is based on the one-step minimization scheme of [30] in  $\mathcal{P}(\mathbb{X})$  w.r.t. the Wasserstein metric  $W_2$ , which we extend to the moving-horizon setting. For every  $k > 0$ , consider:

$$\begin{aligned} \mu_k \in \arg \min_{\mu \in \mathcal{P}(\mathbb{X})} \frac{1}{2} W_2^2(\mu, f_{0\#}\mu_{k-1}) + \eta \mathbb{E}_\mu [G_k^N], \\ \text{given } \mu_0 \in \mathcal{P}(\mathbb{X}). \end{aligned} \tag{6}$$

We let  $\mathcal{K}_k$  be the support of  $\mu_k$ , with  $\mathcal{K}_0 \subseteq \mathcal{C}_0$ , where  $\mathcal{C}_0$  is as defined earlier in Section 4.1.

### 5.1 Sample update scheme for $W_2$ -MHE

We now derive a sample update scheme for  $W_2$ -MHE, which also yields an implementable filter for the  $W_2$ -MHE formulation.

We note that any local minimizer  $\mu_k$  of (6) is a critical point of the objective functional and therefore, it satisfies:

$$c = \frac{\delta}{\delta\mu} \left[ \left( \frac{1}{2} W_2^2(\mu, f_{0\#\mu_{k-1}}) + \eta \mathbb{E}_\mu [G_k^N] \right) \right] \Big|_{\mu=\mu_k} = \phi_k + \eta G_k^N,$$

where  $\phi_k$  is the Kantorovich potential [30] associated with the transport from  $\mu_k$  to  $f_{0\#\mu_{k-1}}$ , and  $c$  is a constant (from the constraint  $\int_{\mathbb{X}} d\mu(x) = 1$ , for  $\mu \in \mathcal{P}(\mathbb{X})$ , due to which the first variation is defined up to an additive constant). From the above equation, we now obtain:

$$\nabla\phi_k(x) + \eta\nabla G_k^N(x) = 0.$$

The gradient of the Kantorovich potential  $\phi_k$  defines the deterministic optimal transport map  $T_k$  (note that this notation is not to be confused with that of the time horizon  $T$ ) w.r.t. the  $W_2$ -distance from  $\mu_k$  to  $f_{0\#\mu_{k-1}}$ , which determines  $\nabla\phi_k(x) = x - T_k^{-1}(x)$  (where  $\mu_k = T_{k\#}f_{0\#\mu_{k-1}}$ ). We therefore get:

$$x = T_k^{-1}(x) - \eta\nabla G_k^N(x). \quad (7)$$

The above equation allows us to design an implementable filter for the  $W_2$ -MHE (6). We let  $z_k \sim \mu_k$ , that is,  $z_k \in \mathcal{K}_k$  is sampled from the distribution  $\mu_k$ . From (7), it holds that  $z_k = T_k^{-1}(z_k) - \eta\nabla G_k^N(z_k)$ . Since  $(T_k^{-1})_{\#\mu_k} = f_{0\#\mu_{k-1}}$ , we let  $T_k^{-1}(z_k) = f_0(z_{k-1})$ , a sample of the distribution  $f_{0\#\mu_{k-1}}$ , and we obtain the following recursive estimator:

$$z_k = f_0(z_{k-1}) - \eta\nabla G_k^N(z_k), \quad k > 0. \quad (8)$$

We now note that the estimate  $z_k$  in (8) corresponds to a critical point of the following minimizing movement scheme:

$$z_k \in \arg \min_z \frac{1}{2} \|z - f_0(z_{k-1})\|^2 + \eta G_k^N(z), \quad k > 0, \quad (9)$$

$$z_0 \sim \mu_0 \in \mathcal{P}(\mathbb{X}).$$

**Lemma 5. (Strong convexity).** *For  $\eta < l^{-1}$ , the objective function in (9) is strongly convex, and therefore  $\text{prox}_{\eta G_k^N}(f_0(x))$  is a singleton for any  $x \in \mathbb{X}$ .*

*Proof.* Let  $\Theta(z) = \frac{1}{2} \|z - f_0(\tilde{z})\|^2 + \eta G_k^N(z)$ . We have that  $\nabla\Theta(z_1) - \nabla\Theta(z_2) = z_1 - z_2 + \eta(\nabla G_k^N(z_1) - \nabla G_k^N(z_2))$ . It now follows that  $\langle \nabla\Theta(z_1) - \nabla\Theta(z_2), z_1 - z_2 \rangle = \|z_1 - z_2\|^2 + \eta \langle \nabla G_k^N(z_1) - \nabla G_k^N(z_2), z_1 - z_2 \rangle$ . From Assumption 5-(1), on the moving-horizon cost, we now get that  $\langle \nabla\Theta(z_1) - \nabla\Theta(z_2), z_1 - z_2 \rangle \geq (1 - \eta l) \|z_1 - z_2\|^2$ , and since  $\eta l < 1$ , we infer that  $\Theta$  is strongly convex, and therefore has a unique minimizer. Thus,  $\text{prox}_{\eta G_k^N}(f_0(\tilde{z})) = \arg \min_z \Theta(z)$  is a singleton.  $\square$

We note that the minimization (9) defines a proximal mapping w.r.t. the Euclidean metric, which we represent in a compact form using the proximal operator as:

$$z_k = \text{prox}_{\eta G_k^N}(f_0(z_{k-1})), \quad k > 0, \quad (10)$$

$$z_0 \sim \mu_0 \in \mathcal{P}(\mathbb{X}),$$

where  $\text{supp}(\mu_0) = \mathcal{K}_0 \subseteq \mathcal{C}_0$ .

## 5.2 Asymptotic stability of $W_2$ -MHE

We present the asymptotic stability result for  $W_2$ -MHE in this section, before which we introduce the following assumption on positive invariance of the discrete-time dynamics defined by the map  $\text{prox}_{\eta G_k^N} \circ f$ .

**Assumption 6. (Positive invariance).** *We assume that there exists  $\alpha > (1 - \sqrt{1 - 2l\bar{L}})l^{-1}$  such that for all  $\eta \in (0, \alpha)$ , we have  $\text{prox}_{\eta G_k^N}(f(\mathcal{C}_{k-1})) \subseteq \mathcal{C}_k$ .*

The above assumption ensures that under the discrete-time dynamics defined by the map  $\text{prox}_{\eta G_k^N} \circ f$ , any sequence starting in the basin of attraction  $\mathcal{C}_0$  of  $\Sigma_T^{-1}(\mathbf{y}_{0:T})$  remains within the basins of attraction  $\mathcal{C}_k$  of  $\Sigma_T^{-1}(\mathbf{y}_{k:k+T})$  at the subsequent instants of time  $k \in \mathbb{N}$ .

We are now ready to present the asymptotic stability result for  $W_2$ -MHE:

**Theorem 2. (Asymptotic stability of  $W_2$ -MHE).** *The estimator (6), under Assumptions 3 to 6, with a constant step size  $\eta \in \left(\frac{1 - \sqrt{1 - 2l\bar{L}}}{l}, \min\left\{\alpha, \frac{1}{l}\right\}\right)$ , is an asymptotically stable observer for the system  $\Sigma$ .*

*Proof.* By Assumption 5-(1), on the moving-horizon cost, and Lemma 1, we have:

$$|G_k^N(f_0(z_{k-1})) - G_k^N(z_k) - \langle \nabla G_k^N(z_k), f_0(z_{k-1}) - z_k \rangle| \leq \frac{l}{2} \|f_0(z_{k-1}) - z_k\|^2.$$

Substituting from (8) into the above, we get:

$$|G_k^N(f_0(z_{k-1})) - G_k^N(z_k) - \eta \|\nabla G_k^N(z_k)\|^2| \leq \eta^2 \frac{l}{2} \|\nabla G_k^N(z_k)\|^2.$$

It now follows that:

$$G_k^N(z_k) \leq G_k^N(f_0(z_{k-1})) - \eta \left(1 - \frac{l}{2}\eta\right) \|\nabla G_k^N(z_k)\|^2.$$

From Assumption 5-(2), on the moving-horizon cost, we have:

$$G_k^N(z_k) \leq G_{k-1}^N(z_{k-1}) + L \|\nabla G_{k-1}^N(z_{k-1})\|^2 - \eta \left(1 - \frac{l}{2}\eta\right) \|\nabla G_k^N(z_k)\|^2.$$

Summing the above inequality from  $k = 1$  to  $K$ , we get:

$$\eta \left(1 - \frac{l}{2}\eta\right) \sum_{k=1}^K \|\nabla G_k^N(z_k)\|^2 - L \sum_{k=1}^K \|\nabla G_{k-1}^N(z_{k-1})\|^2 \leq G_0^N(z_0) - G_K^N(z_K).$$

From here, we obtain:

$$\begin{aligned} \left[\eta \left(1 - \frac{l}{2}\eta\right) - L\right] \sum_{k=1}^K \|\nabla G_k^N(z_k)\|^2 &\leq G_0^N(z_0) - G_K^N(z_K) + L \|\nabla G_0^N(z_0)\|^2 \\ &\leq G_0^N(z_0) + L \|\nabla G_0^N(z_0)\|^2. \end{aligned}$$

Since  $\eta \in \left( \frac{1 - \sqrt{1 - 2lL}}{l}, \frac{1}{l} \right)$ , we have that  $\eta(1 - \frac{l}{2}\eta) - L > 0$  and therefore, taking limits in the previous inequality, we deduce that the series is summable. The latter implies that  $\lim_{k \rightarrow \infty} \nabla G_k^N(z_k) = 0$ , and from (8), we have that  $\lim_{k \rightarrow \infty} \|z_k - f(z_{k-1})\| = 0$ .

It now follows, by definition, from the above that  $\lim_{k \rightarrow \infty} \nabla G_k^{T+1}(z_k) = \lim_{k \rightarrow \infty} \nabla (J_T(\mathbf{y}_{k:k+T}, \Sigma_T(z_k))) = 0$ , over a horizon of length  $T + 1$  (with  $T \geq T_0$ ). We now have that the initial condition  $z_0 \in \mathcal{K}_0 \subseteq \mathcal{C}_0$  and Assumption 6 ensure that  $z_k \in \mathcal{C}_k$ , the basin of attraction of  $f^k(\Sigma_T^{-1}(\mathbf{y}_{0:T}))$  and from the fact that  $\lim_{k \rightarrow \infty} \nabla (J_T(\mathbf{y}_{k:k+T}, \Sigma_T(z_k))) = 0$ , we infer that  $\{z_k\}$  converges to the local minima of  $J_T(\mathbf{y}_{k:k+T}, \Sigma_T(\cdot))$ . By Theorem 1, it now follows that  $\{z_k\}$  converges to the set  $\Sigma_T^{-1}(\mathbf{y}_{k:k+T})$ . Therefore  $\lim_{k \rightarrow \infty} d(z_k, \Sigma_T^{-1}(\mathbf{y}_{k:k+T})) = 0$ .

Moreover, since  $\lim_{k \rightarrow \infty} d(z_k, \Sigma_T^{-1}(\mathbf{y}_{k:k+T})) = 0$  for all  $z_0 \in \mathcal{K}_0$ , it follows that  $\lim_{k \rightarrow \infty} \mathcal{K}_k = \Sigma_T^{-1}(\mathbf{y}_{k:k+T})$ . We know that  $\text{supp}(\mu_k) = \mathcal{K}_k$ , and therefore we get that  $\lim_{k \rightarrow \infty} \mu_k(\Sigma_T^{-1}(\mathbf{y}_{k:k+T})) = 1$ .  $\square$

### 5.3 Robustness of $W_2$ -MHE

We now characterize the performance of the estimator (6) on the system  $\Omega$  in (1). Since the true process and measurement noise sequences remain unknown, we are interested in the robustness properties of the estimator (11), in the form of an upper bound by the norms of the disturbance sequences on the estimation error.

We begin by constructing a reference estimator that recursively generates the estimate sequence, given the true disturbance sequences  $\{w_k\}_{k \in \mathbb{N}}$  and  $\{v_k\}_{k \in \mathbb{N}}$ , as follows:

$$\begin{aligned} \bar{\mu}_k \in \arg \min_{\mu \in \mathcal{P}(\mathbb{X})} \frac{1}{2} W_2^2(\mu, f_{0\#} \bar{\mu}_{k-1}) + \eta \mathbb{E}_\mu [\bar{G}_k^N], \\ \text{given } \bar{\mu}_0 \in \mathcal{P}(\mathbb{X}). \end{aligned} \quad (11)$$

where, we employ for conciseness  $\mathbf{w} \equiv \mathbf{w}_{k:k+N-1} = (w_k, \dots, w_{k+N-1})$  and  $\mathbf{v} \equiv \mathbf{v}_{k+1:k+N} = (v_{k+1}, \dots, v_{k+N})$ , so that  $\bar{G}_k^N(z) \equiv \bar{G}_k^N(z, \mathbf{w}, \mathbf{v}) = J_{N-1}(\mathbf{y}_{k+1:k+N}, \Omega_{\mathbf{w}_{k:k+N-1}}(z) + \mathbf{v}_{k+1:k+N})$ . Note that  $G_k^N = \bar{G}_k^N|_{\mathbf{w}=0, \mathbf{v}=0}$ . We let  $\bar{\mathcal{K}}_k$  be the support of  $\bar{\mu}_k$ , with  $\bar{\mathcal{K}}_0 \subseteq \bar{\mathcal{C}}_0$ , where the definition of  $\bar{\mathcal{C}}_k$  is similar to that of  $\mathcal{C}_k$  but taking the noise  $\{w_k\}$  and  $\{v_k\}$  into account.

**Assumption 7. (*l-Smoothness w.r.t. disturbances*).** We assume that  $\|\nabla G_k^N(z) - \nabla \bar{G}_k^N(z)\| \leq l_w \|\mathbf{w}_{k:k+N-1}, \mathbf{v}_{k+1:k+N}\|$  for all  $z \in \mathbb{X}$ .

Following the proof of Theorem 2, under the same set of underlying assumptions, we infer that the reference estimator (11) is almost surely an asymptotically stable observer for the system  $\Omega$ , given a particular realization of the disturbances  $\{w_k\}_{k \in \mathbb{N}}$  and  $\{v_k\}_{k \in \mathbb{N}}$ .

We now present the following theorem on the robustness of the estimator (6), characterized by a bound on the error in the estimates generated by (6) with respect to the estimates generated by the reference estimator (11):

**Theorem 3. (*Robustness of  $W_2$ -MHE*).** Under Assumptions 1, 3, 5, and 7, given the estimate sequences  $\{\mu_k\}_{k \in \mathbb{N}}$  generated by (6) and  $\{\bar{\mu}_k\}_{k \in \mathbb{N}}$  generated by the reference estimator (11), with  $\mu_0 = \bar{\mu}_0$ , we have  $W_2(\mu_k, \bar{\mu}_k) \leq \frac{c_f^{(2)}}{c_f^{(1)}} W C_k + \frac{\eta l_w \sqrt{N}}{c_f^{(1)}} (W + V) C_k$ , for all  $k \in \mathbb{N}$ ,

where  $C_k = \sum_{\ell=1}^k \left( \frac{c_f^{(1)}}{1-\eta} \right)^\ell$ .

*Proof.* The estimator (11) yields the following reference recursive scheme:

$$\bar{z}_k = f(\bar{z}_{k-1}, w_{k-1}) - \eta \nabla \bar{G}_k^N(\bar{z}_k), \quad (12)$$

where the above is derived similarly to the noiseless case. Let  $\{z_k\}_{k \in \mathbb{N}}$  and  $\{\bar{z}_k\}_{k \in \mathbb{N}}$  be the estimate sequences generated by (8) and (12) respectively, with  $z_0 = \bar{z}_0$ , for which we have:

$$\begin{aligned} \|z_k - \bar{z}_k\| &= \|f_0(z_{k-1}) - f(\bar{z}_{k-1}, w_{k-1}) - \eta \nabla G_k^N(z_k) + \eta \nabla \bar{G}_k^N(\bar{z}_k)\| \\ &= \|f_0(z_{k-1}) - f_0(\bar{z}_{k-1}) + f_0(\bar{z}_{k-1}) - f(\bar{z}_{k-1}, w_{k-1}) \\ &\quad - \eta \nabla G_k^N(z_k) + \eta \nabla G_k^N(\bar{z}_k) - \eta \nabla G_k^N(\bar{z}_k) + \eta \nabla \bar{G}_k^N(\bar{z}_k)\| \\ &\leq c_f^{(1)} \|z_{k-1} - \bar{z}_{k-1}\| + c_f^{(2)} \|w_{k-1}\| + \eta l \|z_k - \bar{z}_k\| \\ &\quad + l_w \eta \|(\mathbf{w}_{k:k+N-1}, \mathbf{v}_{k+1:k+N})\|, \end{aligned}$$

where the final inequality follows from Assumptions 1, 5, and 7, on the several Lipschitz properties of  $f$  the gradient of  $G_k^N$ , and  $\bar{G}_k^N$ , respectively. Further, since  $\eta l < 1$ , we obtain from the above that:

$$\begin{aligned} \|z_k - \bar{z}_k\| &\leq \left( \frac{1}{1 - \eta l} \right) \left( c_f^{(1)} \|z_{k-1} - \bar{z}_{k-1}\| + c_f^{(2)} \|w_{k-1}\| \right. \\ &\quad \left. + \eta l_w \|(\mathbf{w}_{k:k+N-1}, \mathbf{v}_{k+1:k+N})\| \right) \\ &\leq \left( \frac{c_f^{(1)}}{1 - \eta l} \right)^k \|z_0 - \bar{z}_0\| + \frac{c_f^{(2)}}{c_f^{(1)}} \sum_{\ell=1}^k \left( \frac{c_f^{(1)}}{1 - \eta l} \right)^\ell \|w_{k-\ell}\| \\ &\quad + \frac{\eta l_w}{c_f^{(1)}} \sum_{\ell=1}^k \left( \frac{c_f^{(1)}}{1 - \eta l} \right)^\ell \|(\mathbf{w}_{k-\ell+1:k-\ell+N}, \mathbf{v}_{k-\ell+2:k-\ell+N+1})\| \\ &\leq \frac{c_f^{(2)}}{c_f^{(1)}} W C_k + \frac{\eta l_w \sqrt{N}}{c_f^{(1)}} (W + V) C_k. \end{aligned}$$

We note that if  $\frac{c_f^{(1)}}{1 - \eta l} < 1$ , we have that  $\lim_{k \rightarrow \infty} C_k = \frac{c_f^{(1)}}{1 - \eta l - c_f^{(1)}}$  is finite, and therefore,  $\|z_k - \bar{z}_k\|$  is bounded as  $k \rightarrow \infty$ . We note here that even when  $z_0 \neq \bar{z}_0$ , the effect of this initial discrepancy vanishes as  $k \rightarrow \infty$ .

Now, let  $T_k : \mathcal{K}_k \rightarrow \bar{\mathcal{K}}_k$  be a map such that for sequences  $\{z_k\}$  and  $\{\bar{z}_k\}$  generated by (8) and (12) respectively, with  $z_0 = \bar{z}_0$ , we have  $T_k(z_k) = \bar{z}_k$ . It then follows that  $T_{k\#}\mu_k = \bar{\mu}_k$ . Now, from the above, and by definition of the 2-Wasserstein distance, we have:

$$\begin{aligned} W_2(\mu_k, \bar{\mu}_k) &\leq \left( \int_{z \in \mathcal{K}_k} \|z - T_k(z)\|^2 d\mu_k(z) \right)^{\frac{1}{2}} \\ &\leq \left( \int_{z \in \mathcal{K}_k} \left| \frac{c_f^{(2)}}{c_f^{(1)}} W C_k + \frac{\eta l_w \sqrt{N}}{c_f^{(1)}} (W + V) C_k \right|^2 d\mu_k(z) \right)^{\frac{1}{2}} \\ &\leq \frac{c_f^{(2)}}{c_f^{(1)}} W C_k + \frac{\eta l_w \sqrt{N}}{c_f^{(1)}} (W + V) C_k. \end{aligned}$$

□

## 6 A KL-Moving-Horizon Estimator

In this section, we derive a moving-horizon estimator, which we refer to as KL-MHE, to generate a sequence of probability distributions  $\{\mu_k\}_{k \in \mathbb{N}}$ . Using the KL-divergence  $D_{\text{KL}}$  as the choice of divergence in the moving-horizon formulation (5), we obtain:

$$\begin{aligned} \mu_k &\in \arg \min_{\mu \in \mathcal{P}(\mathcal{X})} D_{\text{KL}}(\mu \| f_{0\#} \mu_{k-1}) + \eta \mathbb{E}_\mu [G_k^N], \\ \text{given } \mu_0 &\in \mathcal{P}(\mathcal{X}). \end{aligned} \quad (13)$$

We note that any local minimizer  $\mu_k$  of (13) is a critical point of the objective functional, and, therefore, it satisfies:

$$c = \frac{\delta}{\delta \mu} [D_{\text{KL}}(\mu \| f_{0\#} \mu_{k-1}) + \eta \mathbb{E}_\mu [G_k^N]] \Big|_{\mu = \mu_k},$$

where  $c$  is a constant (from the constraint  $\int_{\mathcal{X}} d\mu(x) = 1$ , for  $\mu \in \mathcal{P}(\mathcal{X})$ , due to which the first variation is defined up to an additive constant). From the above, we get:

$$c = \log \left( \frac{\rho_k}{f_{0\#} \rho_{k-1}} \right) (x) + \eta G_k^N(x),$$

where for any  $\ell \in \{0, 1, \dots\}$ ,  $\rho_\ell$  is the density function corresponding to the measure  $\mu_\ell$ . Therefore, the corresponding recursive update scheme for the density function is given by:

$$\rho_k(x) = c_k (f_{0\#} \rho_{k-1}(x)) \exp(-\eta G_k^N(x)), \quad (14)$$

where  $c_k$  is the normalization constant. We note that the above is a particle filter formulation, with the horizon cost  $G_k^N$  defining the weighting function. Implementable filters are obtained by a Sequential Monte Carlo method, see [10]. We now present the asymptotic stability result for KL-MHE:

**Theorem 4. (Asymptotic stability of KL-MHE).** *The estimator (13), under Assumptions 1 to 4, is an asymptotically stable observer for the system  $\Sigma$ .*

*Proof.* We know that for any map  $\mathcal{T}$  and measure  $\mu$ , we have that  $d\mathcal{T}_\# \mu(x) = d\mu(\mathcal{T}^{-1}(x))$ . It then follows from (14) that:

$$\rho_k(x) = c_k \rho_{k-1}(f_0^{-1}(x)) \exp(-\eta G_k^N(x)).$$

We now rewrite the above as:

$$\rho_k(f_0(x)) = c_k \rho_{k-1}(x) \exp(-\eta G_k^N(f_0(x))).$$

Repeating the above process  $k$  times, we obtain:

$$\rho_k(f_0^k(x)) = C_k \rho_0(x) \exp \left( -\eta \sum_{\ell=1}^k G_\ell^N(f_0^\ell(x)) \right),$$

where  $C_k = c_k c_{k-1} \dots c_1$  is the normalization constant. If  $x \notin \Sigma_T^{-1}(\mathbf{y}_{0:T})$ , we have that  $\lim_{k \rightarrow \infty} \rho_k(f_0^k(x)) = 0$ , since  $\sum_{\ell=1}^k G_\ell^N(f_0^\ell(x)) \rightarrow \infty$  as  $k \rightarrow \infty$  for all  $x \notin \Sigma_T^{-1}(\mathbf{y}_{0:T})$  (by definition of the cost function, the sum diverges over an infinitely long horizon). Thus, we get:

$$\lim_{k \rightarrow \infty} \mu_k \left( f_0^k \left( \Sigma_T^{-1}(\mathbf{y}_{0:T}) \right) \right) = \lim_{k \rightarrow \infty} \mu_k \left( \Sigma_T^{-1}(\mathbf{y}_{k:k+T}) \right) = 1.$$

□

## 7 Differential privacy

In this section, we discuss the mechanism for encoding the desired level of differential privacy in moving-horizon estimators. We then apply this mechanism to the two estimators presented in the previous sections, the  $W_2$ -MHE and KL-MHE. We conclude the section with a discussion on differential privacy of the estimators over a time horizon. Our aim here is to guarantee differential privacy of the measurement data  $\mathbf{y}_{0:T}$ , when the estimate sequence  $\{\mu_k\}$  is released (made public). We consider the class of scenarios where an adversary can access the released estimates, while the measurement data itself is not accessible to the adversary. Our goal in incorporating differential privacy in estimation is to ensure that the adversary is not able to distinguish (in the sense of  $\epsilon$ -differential privacy) between measurement sequences that are  $\delta$ -adjacent, using the released estimates, which is an underlying risk when the estimates are directly released without such a consideration.

Given the framework (5), we encode differential privacy by an entropic regularization of the estimation objective function, as follows:

$$\mu_k \in \arg \min_{\mu \in \mathcal{P}(\mathcal{X})} \left[ s_k D(\mu, f_{0\#} \mu_{k-1}) + s_k \eta \mathbb{E}_\mu [G_k^N] - (1 - s_k) S^{\mathcal{K}_k}(\mu) \right], \quad (15)$$

$$\text{given } \mu_0 \in \mathcal{P}(\mathcal{X}),$$

where  $s_k \in [0, 1]$  is a tunable time-dependent parameter and  $\mathcal{K}_k$  is the support of  $f_{0\#} \mu_{k-1}$  (with  $\mathcal{K}_0$  being the support of  $\mu_0$ ). Moreover,  $S^A(\mu) = \int_A \rho \log(\rho) \text{dvol}$ , where  $A \subset \mathcal{X}$  and  $d\mu = \rho \text{dvol}$ . We note that when  $s_k = 1$ , the above formulation reduces to (5) and when  $s_k = 0$ , it is equivalent to an entropy maximization problem, yielding a uniform distribution over the set  $f_0(\mathcal{K}_{k-1})$  as the solution. Clearly, the uniform distribution is insensitive to the measurements, and therefore offers maximum privacy, while being of no value to the estimation objective. The ensuing analysis in this section is directed at determining upper bounds on the parameter sequence  $\{s_k\}_{k \in \mathbb{N}}$  such that the MHE offers  $\epsilon$ -differential privacy. We rewrite the optimization problem (15) for  $s_k \in (0, 1]$  as follows:

$$\mu_k \in \arg \min_{\mu \in \mathcal{P}(\mathcal{X})} \left[ D(\mu, f_{0\#} \mu_{k-1}) + \eta \mathbb{E}_\mu [G_k^N] - \left( \frac{1 - s_k}{s_k} \right) S^{\mathcal{K}_k}(\mu) \right], \quad (16)$$

$$\text{given } \mu_0 \in \mathcal{P}(\mathcal{X}),$$

Let  $\mathbf{y}, \tilde{\mathbf{y}} \in \mathbb{Y}^{T+N+1}$  be two  $\delta$ -adjacent measurement sequences as in Definition 4, over a horizon  $\{0, \dots, T+N\}$ , such that  $\|\mathbf{y} - \tilde{\mathbf{y}}\| \leq \delta$  and let  $\{\mu_k\}_{k \in \mathbb{N}}$  and  $\{\tilde{\mu}_k\}_{k \in \mathbb{N}}$  be the sequences of estimates derived from (16). In the following, we determine conditions on  $\{s_k\}_{k \in \mathbb{N}}$  that guarantee differential privacy for each of the estimators derived in previous sections.

## 7.1 Differentially private $W_2$ -MHE

We now design a differentially private  $W_2$ -moving-horizon estimator. We begin by considering:

$$\mu_k \in \arg \min_{\mu \in \mathcal{P}(\mathbb{X})} \left[ \frac{1}{2} W_2^2(\mu, f_{0\#} \mu_{k-1}) + \eta \mathbb{E}_\mu [G_k^N] - \left( \frac{1 - s_k}{s_k} \right) S^{\mathcal{K}_k}(\mu) \right], \quad (17)$$

given  $\mu_0 \in \mathcal{P}(\mathbb{X})$ ,

for  $s_k \in (0, 1]$ .

The following theorem provides a sufficient upper bound on  $s_T$  such that the entropy-regularized  $W_2$ -MHE in (17) is  $\epsilon_T$ -differentially private at a time instant  $T$ .

**Theorem 5. (Sensitivity of  $W_2$ -MHE).** *Given two  $\delta$ -adjacent measurement sequences  $\mathbf{y}, \tilde{\mathbf{y}} \in \mathbb{Y}^{T+N+1}$ , under Assumption 5, we have that the estimates generated by (17) satisfy  $D_{\max}(\mu_T, \tilde{\mu}_T) \leq \epsilon_T$  if  $s_T \leq \epsilon_T \left( \epsilon_T + c_f^T \text{diam}(\mathcal{K}_0) \left( \eta l \delta + c_f^T \text{diam}(\mathcal{K}_0) q(\delta) \right) \right)^{-1}$ , where  $q : \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}_{\geq 0}$  is a class- $\mathcal{K}$  function that satisfies  $q(0) = 0$ .*

*Proof.* Let  $G_k^N$  and  $\tilde{G}_k^N$  be the estimation objective functions at time instant  $k$ , corresponding to the measurement sequences  $\mathbf{y}$  and  $\tilde{\mathbf{y}}$  respectively, and let  $\mu_k$  and  $\tilde{\mu}_k$  be the respective estimated probability measures, with  $\rho_k, \tilde{\rho}_k$  the corresponding density functions. From (17), we get that for all  $k \in \{0, \dots, T\}$ ,  $\mu_k$ , being the local minimizer is also a critical point of the objective functional. We therefore obtain:

$$\phi_k(x) + G_k^N(x) + \left( \frac{1 - s_k}{s_k} \right) \log(\rho_k(x)) = c,$$

where  $\phi_k$  is the Kantorovich potential associated with the transport from  $\mu_k$  to  $f_{0\#} \mu_{k-1}$  and  $c$  is a constant. It now follows that:

$$\nabla \phi_k(x) + \nabla G_k^N(x) + \left( \frac{1 - s_k}{s_k} \right) \nabla \log(\rho_k(x)) = 0.$$

Similarly, we have:

$$\nabla \tilde{\phi}_k(x) + \nabla \tilde{G}_k^N(x) + \left( \frac{1 - s_k}{s_k} \right) \nabla \log(\tilde{\rho}_k(x)) = 0.$$

Taking the difference between the above two equations:

$$\nabla \left[ \log \left( \frac{\rho_k}{\tilde{\rho}_k} \right) \right] (x) = - \left( \frac{s_k}{1 - s_k} \right) \left[ \nabla(\phi_k - \tilde{\phi}_k)(x) + \nabla(G_k^N - \tilde{G}_k^N)(x) \right].$$

We have that  $\nabla\phi_k(x) = x - T_k^{-1}(x)$ , where  $\mu_k = T_{k\#}(f_{0\#}\mu_{k-1})$ . This implies that  $\nabla(\phi_k - \tilde{\phi}_k)(x) = -(T_k^{-1}(x) - \tilde{T}_k^{-1}(x))$ . However,  $T_k^{-1}(x), \tilde{T}_k^{-1}(x) \in f_0(\mathcal{K}_{k-1}) = f_0^k(\mathcal{K}_0)$ , and therefore  $\|\nabla(\phi_k - \tilde{\phi}_k)(x)\| \leq c_f^k \text{diam}(\mathcal{K}_0)q(\delta)$ , for all  $x \in f_0^k(\mathcal{K}_0)$  and some class- $\mathcal{K}$  function  $q$ . We let  $q$  characterize the dependence of  $\phi$  on the measurement sequence, and we get that  $\|\nabla(\phi_k - \tilde{\phi}_k)(x)\| = 0$  for all  $x \in \mathbb{X}$ , when  $\delta = 0$ . Moreover, by Assumption 5, we get  $\|\nabla(G_k^N - \tilde{G}_k^N)(x)\| \leq l\delta$ . Therefore, we obtain:

$$\left\| \nabla \left[ \log \left( \frac{\rho_k}{\tilde{\rho}_k} \right) \right] \right\| \leq \left( \frac{s_k}{1-s_k} \right) \left( c_f^k \text{diam}(\mathcal{K}_0)q(\delta) + l\delta \right). \quad (18)$$

We also have that for any  $x \in f_0^k(\mathcal{K}_0)$ :

$$\log \left( \frac{\rho_k}{\tilde{\rho}_k} \right) (x) = \log \left( \frac{\rho_k}{\tilde{\rho}_k} \right) (\bar{x}) + \int_0^1 \nabla \left[ \log \left( \frac{\rho_k}{\tilde{\rho}_k} \right) \right] (\gamma(t)) \cdot \dot{\gamma}(t) dt, \quad (19)$$

where  $\gamma(0) = \bar{x}$  and  $\gamma(1) = x$ . Since  $\rho_k$  and  $\tilde{\rho}_k$  are continuous, with  $\int_{f_0^k(\mathcal{K}_0)} (\rho_k - \tilde{\rho}_k) = 0$  (since  $\int_{f_0^k(\mathcal{K}_0)} \rho_k = \int_{f_0^k(\mathcal{K}_0)} \tilde{\rho}_k = 1$ ), there exists an  $\bar{x} \in f_0^k(\mathcal{K}_0)$  such that  $\rho_k(\bar{x}) = \tilde{\rho}_k(\bar{x})$ , which implies that  $\log \left( \frac{\rho_k}{\tilde{\rho}_k} \right) (\bar{x}) = 0$ . From (18) and (19), for a straight line segment  $\gamma$ , we therefore obtain:

$$\left| \log \left( \frac{\rho_k}{\tilde{\rho}_k} \right) (x) \right| \leq \left( \frac{s_k}{1-s_k} \right) \left( c_f^k \text{diam}(\mathcal{K}_0)q(\delta) + l\delta \right) \times c_f^k \text{diam}(\mathcal{K}_0),$$

where we have used the fact that  $\int_0^1 |\dot{\gamma}(t)| dt = \|x - \bar{x}\| \leq \text{diam}(f_0^k(\mathcal{K}_0)) \leq c_f^k \text{diam}(\mathcal{K}_0)$ . Thus, for  $k = T$ , we let:

$$\left| \log \left( \frac{\rho_T}{\tilde{\rho}_T} \right) (x) \right| \leq \left( \frac{s_T}{1-s_T} \right) \left( c_f^T \text{diam}(\mathcal{K}_0)q(\delta) + l\delta \right) \times c_f^T \text{diam}(\mathcal{K}_0) \leq \epsilon_T,$$

from which we obtain that:

$$s_T \leq \frac{\epsilon_T}{\left( \epsilon_T + c_f^T \text{diam}(\mathcal{K}_0) \left( \eta l \delta + c_f^T \text{diam}(\mathcal{K}_0)q(\delta) \right) \right)},$$

and since  $\left| \log \left( \frac{\rho_T}{\tilde{\rho}_T} \right) (x) \right| \leq \epsilon_T$  for all  $x \in f_0^T(\mathcal{K}_0)$ , we have that  $\sup_{x \in f_0^T(\mathcal{K}_0)} \left| \log \left( \frac{\rho_T}{\tilde{\rho}_T} \right) \right| = D_{\max}(\mu_T, \tilde{\mu}_T) \leq \epsilon_T$ .  $\square$

As noted earlier, Theorem 5 provides a sufficient upper bound on  $s_T$  for differential privacy of the estimate at  $T$ . The goal, however, is to guarantee the desired level of differential privacy over a time horizon  $\{0, \dots, T\}$ . The key issue here is that the recursive update scheme of the estimator introduces a dependence between the estimates at different time instants. This essentially means that imposing an upper bound on sensitivity for the marginal distributions  $\mu_k$  individually, without regard to the dependence between these distributions, may not be sufficient. Therefore, to guarantee the desired level of differential privacy over the time horizon, we must impose an upper bound on the sensitivity of the joint distribution  $\sigma \in \mathcal{P}(\mathbb{X}^{T+1})$ , where the estimates  $\mu_k$  are the marginals of  $\sigma$  over  $\mathbb{X}$ .

The following theorem provides a sufficient upper bound on  $\{s_k\}_{k=1}^T$  such that the entropy-regularized  $W_2$ -MHE in (17) is  $\epsilon$ -differentially private over a time horizon  $\{0, \dots, T\}$ .

**Theorem 6. (Differentially private  $W_2$ -MHE).** Given two  $\delta$ -adjacent measurement sequences  $\mathbf{y}, \tilde{\mathbf{y}} \in \mathbb{Y}^{T+N+1}$ , under Assumption 5, we have that the estimates generated by (17) satisfy  $D_{\max}(\sigma, \tilde{\sigma}) \leq \epsilon$  if  $\sum_{k=1}^T \left(\frac{s_k}{1-s_k}\right) c_f^k \leq \frac{\epsilon}{l\delta \text{diam}(\mathcal{K}_0)}$ .

*Proof.* Let  $G_k^N$  and  $\tilde{G}_k^N$  be the estimation objective functions at time instant  $k$ , corresponding to the measurement sequences  $\mathbf{y}$  and  $\tilde{\mathbf{y}}$  respectively, and let  $\sigma$  and  $\tilde{\sigma}$  be the respective joint probability measures over the horizon  $\{0, \dots, T\}$ . With a slight abuse of notation, we allow  $\sigma$  and  $\tilde{\sigma}$  to also denote the joint density function. We now have:

$$\sigma(x_0, x_1, \dots, x_T) = \rho_0(x_0)\sigma(x_1, \dots, x_T|x_0) = \rho_0(x_0)\rho_1(x_1|x_0)\rho_2(x_2|x_1) \dots \rho_T(x_T|x_{T-1}),$$

where  $\rho_k(x_k|x_{k-1})$  is the marginal density at  $x_k$  at time instant  $k$ , given that the distribution at time instant  $k-1$  is concentrated at  $x_{k-1}$ . Moreover, we note that the  $W_2$ -MHE (17) yields a Markov process, which allows us to express  $\rho_k(x_k|x_{k-1}, \dots, x_0) = \rho_k(x_k|x_{k-1})$ . Now,  $\rho_k(x_k|x_{k-1})$  is the density corresponding to the measure obtained by the following:

$$\mu_k \in \arg \min_{\mu \in \mathcal{P}(\mathcal{X})} \left[ \frac{1}{2} W_2^2(\mu, \partial_{f_0(x_{k-1})}) + \eta \mathbb{E}_\mu [G_k^N] - \left(\frac{1-s_k}{s_k}\right) S^{\mathcal{K}_k}(\mu) \right],$$

where  $\partial_\xi$  is the Dirac measure concentrated at  $\xi$ . From the above, we get that for all  $k \in \{0, \dots, T\}$ ,  $\mu_k$ , being the local minimizer is also a critical point of the objective functional. Applying similar steps to those in the proof of Theorem 5, we obtain:

$$\left| \log \left( \frac{\rho_k}{\tilde{\rho}_k} \right) (x|x_{k-1}) \right| \leq \left( \frac{s_k}{1-s_k} \right) l\delta c_f^k \text{diam}(\mathcal{K}_0).$$

Now, we have:

$$\left| \log \left( \frac{\sigma}{\tilde{\sigma}} \right) (x_0, \dots, x_T) \right| \leq \sum_{k=1}^T \left| \log \left( \frac{\rho}{\tilde{\rho}} \right) (x_k|x_{k-1}) \right| \leq \sum_{k=1}^T \left( \frac{s_k}{1-s_k} \right) l\delta c_f^k \text{diam}(\mathcal{K}_0).$$

By taking

$$l\delta \text{diam}(\mathcal{K}_0) \sum_{k=1}^T \left( \frac{s_k}{1-s_k} \right) c_f^k \leq \epsilon,$$

we obtain the following inequality:

$$\sum_{k=1}^T \left( \frac{s_k}{1-s_k} \right) c_f^k \leq \frac{\epsilon}{l\delta \text{diam}(\mathcal{K}_0)},$$

and that  $D_{\max}(\sigma, \tilde{\sigma}) \leq \epsilon$ . □

We note that for a given  $\epsilon$ , the upper bound on the sequence  $\{s_k\}$  decreases with  $\delta$ . In other words, guaranteeing  $\epsilon$ -differential privacy w.r.t. measurement sequences that are farther apart requires the addition of more noise and a greater loss in estimation accuracy. This is because the weighting on the entropic regularization term in the estimation objective increases when  $s_k$  is reduced. The same is the case when  $\epsilon$  is reduced for a given  $\delta$ , which corresponds to a more stringent privacy requirement.

## 7.2 Differentially private KL-MHE

We now design a differentially private KL-moving-horizon estimator. We begin by considering the entropy-regularized KL-MHE formulation, given by:

$$\mu_k \in \arg \min_{\mu \in \mathcal{P}(\mathcal{X})} \left[ D_{\text{KL}}(\mu \| f_{0\#} \mu_{k-1}) + \eta \mathbb{E}_\mu [G_k^N] - \left( \frac{1 - s_k}{s_k} \right) S^{\mathcal{K}_k}(\mu) \right], \quad (20)$$

given  $\mu_0 \in \mathcal{P}(\mathcal{X})$ ,

for  $s_k \in (0, 1]$ . The corresponding recursive update scheme for (20) is given by:

$$\rho_k(x) = c_k (f_{0\#} \rho_{k-1}(x))^{s_k} e^{-\eta s_k G_k^N(x)}, \quad (21)$$

which will be derived in the proof of Theorem 7 below.

The following theorem provides a sufficient upper bound on  $s_k$  such that the entropy-regularized KL-MHE in (20) is  $\epsilon_T$ -differentially private at a time instant  $T$ , while ignoring the correlations between the estimates  $\mu_k$  across time.

**Theorem 7. (Sensitivity of KL-MHE).** *Given two  $\delta$ -adjacent measurement sequences  $\mathbf{y}, \tilde{\mathbf{y}} \in \mathbb{Y}^{T+N+1}$ , under Assumption 5, we have that the estimates generated by (20) satisfy  $D_{\max}(\mu_T, \tilde{\mu}_T) \leq \epsilon_T$  if  $\sum_{k=1}^T \left( \prod_{i=k}^T s_i \right) \leq \epsilon_T \left( 2\eta \max_{k \in \{0, \dots, T\}} \left( \alpha_k + l c_f^k \delta \text{diam}(\mathcal{K}_0) \right) \right)^{-1}$ , where  $\alpha_k = \min_{\xi \in f_0^k(\mathcal{K}_0)} \left| \left( G_k^N - \tilde{G}_k^N \right) (\xi) \right|$ .*

*Proof.* Let  $G_k^N$  and  $\tilde{G}_k^N$  be the estimation objective functions at time instant  $k$ , corresponding to the measurement sequences  $\mathbf{y}$  and  $\tilde{\mathbf{y}}$  respectively, and let  $\mu_k$  and  $\tilde{\mu}_k$  be the respective estimated probability measures, with  $\rho_k, \tilde{\rho}_k$  the corresponding density functions. From (20), we get that for all  $k \in \{0, \dots, T\}$ ,  $\mu_k$ , being the local minimizer is also a critical point of the objective functional. We therefore obtain:

$$\begin{aligned} & \frac{\delta}{\delta \mu} \left[ D_{\text{KL}}(\mu \| f_{0\#} \mu_{k-1}) + \eta \mathbb{E}_\mu [G_k^N] - \left( \frac{1 - s_k}{s_k} \right) S^{\mathcal{K}_k}(\mu) \right] \Big|_{\mu_k} \\ & = \bar{c}_k, \end{aligned}$$

from which we derive that:

$$\log \left( \frac{\rho_k}{f_{0\#} \rho_{k-1}} \right) (x) + \eta G_k^N(x) + \left( \frac{1 - s_k}{s_k} \right) \log \rho_k(x) = \bar{c}_k.$$

The above equation can be rewritten as follows:

$$\rho_k(x) = c_k (f_{0\#} \rho_{k-1}(x))^{s_k} e^{-\eta s_k G_k^N(x)} = c_k (\rho_{k-1}(f_0^{-1}(x)))^{s_k} e^{-\eta s_k G_k^N(x)},$$

where  $c_k$  is the normalization constant. We therefore obtain:

$$\rho_k(f_0(x)) = c_k (\rho_{k-1}(x))^{s_k} e^{-\eta s_k G_k^N(f_0(x))}.$$

Expanding the above, we get:

$$\rho_T(f_0^T(x)) = C_T (\rho_0(x))^{\prod_{k=1}^T s_k} e^{-\eta \sum_{k=1}^T (\prod_{i=k}^T s_i) G_k^N(f_0^k(x))},$$

where  $C_T = c_1 c_2 \dots c_T$ . Similarly, we have:

$$\tilde{\rho}_T(f_0^T(x)) = \tilde{C}_T (\tilde{\rho}_0(x))^{\prod_{k=1}^T s_k} e^{-\eta \sum_{k=1}^T (\prod_{i=k}^T s_i) \tilde{G}_k^N(f_0^k(x))},$$

where  $\tilde{C}_T = \tilde{c}_1 \tilde{c}_2 \dots \tilde{c}_T$  and  $\rho_0 = \tilde{\rho}_0$ , as we assume that the estimator starts with the same initial  $\mu_0$ . From the above two equations, we obtain:

$$\log \left( \frac{\rho_T}{\tilde{\rho}_T} \right) (f_0^T(x)) = \log \left( \frac{C_T}{\tilde{C}_T} \right) - \eta \sum_{k=1}^T \left( \prod_{i=k}^T s_i \right) (G_k^N - \tilde{G}_k^N) (f_0^k(x)).$$

The max-divergence between  $\mu_T$  and  $\tilde{\mu}_T$  can be upper bounded now by:

$$\begin{aligned} D_{\max}(\mu_T, \tilde{\mu}_T) &= \sup_{x \in \mathcal{K}_0} \left| \log \left( \frac{\rho_T}{\tilde{\rho}_T} \right) (f_0^T(x)) \right| \\ &\leq \left| \log \left( \frac{C_T}{\tilde{C}_T} \right) \right| + \sup_{x \in \mathcal{K}_0} \eta \sum_{k=1}^T \left( \prod_{i=k}^T s_i \right) \times \left| (G_k^N - \tilde{G}_k^N) (f_0^k(x)) \right| \\ &\leq 2 \sup_{x \in \mathcal{K}_0} \eta \sum_{k=1}^T \left( \prod_{i=k}^T s_i \right) \left| (G_k^N - \tilde{G}_k^N) (f_0^k(x)) \right|, \end{aligned}$$

where the final inequality is due to the following (note that we use the fact that  $\rho = \tilde{\rho}$ , as mentioned earlier):

$$\begin{aligned} \left| \log \left( \frac{C_T}{\tilde{C}_T} \right) \right| &= \left| \log \left( \frac{\int_{x \in \mathcal{K}_0} (\rho_0(x))^{\prod_{k=1}^T s_k} e^{-\eta \sum_{k=1}^T (\prod_{i=k}^T s_i) G_k^N(f_0^k(x))} dx}{\int_{x \in \mathcal{K}_0} (\tilde{\rho}_0(x))^{\prod_{k=1}^T s_k} e^{-\eta \sum_{k=1}^T (\prod_{i=k}^T s_i) \tilde{G}_k^N(f_0^k(x))} dx} \right) \right| \\ &\leq \sup_{x \in \mathcal{K}_0} \left| \log \left( \frac{e^{-\eta \sum_{k=1}^T (\prod_{i=k}^T s_i) G_k^N(f_0^k(x))}}{e^{-\eta \sum_{k=1}^T (\prod_{i=k}^T s_i) \tilde{G}_k^N(f_0^k(x))}} \right) \right| \\ &\leq \sup_{x \in \mathcal{K}_0} \eta \sum_{k=1}^T \left( \prod_{i=k}^T s_i \right) \left| (G_k^N - \tilde{G}_k^N) (f_0^k(x)) \right|. \end{aligned}$$

We now have, for all  $k \in \{1, \dots, T\}$ :

$$(G_k^N - \tilde{G}_k^N) (f_0^k(x)) = (G_k^N - \tilde{G}_k^N) (\xi_k) + \int_0^1 \nabla (G_k^N - \tilde{G}_k^N) (\gamma_k(t)) \cdot \dot{\gamma}_k(t) dt,$$

where  $\gamma_k(0) = \xi_k$  and  $\gamma_k(1) = f_0^k(x)$ . From Assumption 5, we have  $\left\| \nabla (G_k^N - \tilde{G}_k^N) (\xi) \right\| \leq l\delta$ . Moreover, let  $\xi_k \in f_0^k(\mathcal{K}_0)$  such that  $\left| (G_k^N - \tilde{G}_k^N) (\xi_k) \right| = \min_{f_0^k(\mathcal{K}_0)} \left| (G_k^N - \tilde{G}_k^N) \right| = \alpha_k$ , and we obtain:

$$\left| (G_k^N - \tilde{G}_k^N) (f_0^k(x)) \right| \leq \alpha_k + l\delta \text{diam}(f_0^k(\mathcal{K}_0)) \leq \alpha_k + l c_f^k \delta \text{diam}(\mathcal{K}_0).$$

This yields the following inequality:

$$\begin{aligned} 2 \sup_{x \in \mathcal{K}_0} \eta \sum_{k=1}^T \left( \prod_{i=k}^T s_i \right) \left| \left( G_k^N - \tilde{G}_k^N \right) (f_0^k(x)) \right| &\leq 2\eta \sum_{k=1}^T \left( \prod_{i=k}^T s_i \right) \left( \alpha_k + lc_f^k \delta \text{diam}(\mathcal{K}_0) \right) \\ &\leq 2\eta \max_k \left( \alpha_k + lc_f^k \delta \text{diam}(\mathcal{K}_0) \right) \sum_{k=1}^T \left( \prod_{i=k}^T s_i \right). \end{aligned}$$

We now let:

$$2\eta \max_k \left( \alpha_k + lc_f^k \delta \text{diam}(\mathcal{K}_0) \right) \sum_{k=1}^T \left( \prod_{i=k}^T s_i \right) \leq \epsilon_T,$$

which yields the bound

$$\sum_{k=1}^T \left( \prod_{i=k}^T s_i \right) \leq \frac{\epsilon_T}{2\eta \max_k \left( \alpha_k + lc_f^k \delta \text{diam}(\mathcal{K}_0) \right)},$$

and we get  $D_{\max}(\mu_T, \tilde{\mu}_T) \leq \epsilon_T$ .  $\square$

We note here that, in practice, with the choice of a sufficiently large domain  $\mathcal{K}_0$ , we can ensure that  $\alpha_k = \min_{\xi \in f_0^k(\mathcal{K}_0)} \left| \left( G_k^N - \tilde{G}_k^N \right) (\xi) \right| = 0$  for all  $k \in \{0, \dots, T\}$ . This is owing to the fact that for a large enough  $\mathcal{K}_0$ , we will have  $\min_{\xi \in f_0^k(\mathcal{K}_0)} \left( G_k^N - \tilde{G}_k^N \right) (\xi) \leq 0 \leq \max_{\xi \in f_0^k(\mathcal{K}_0)} \left( G_k^N - \tilde{G}_k^N \right) (\xi)$ . Moreover, since the function  $G_k^N - \tilde{G}_k^N$  is continuous, there must therefore exist a point  $\xi^*$  such that  $\left( G_k^N - \tilde{G}_k^N \right) (\xi^*) = 0$ .

As with the  $W_2$ -MHE, we now characterize the differential privacy of the KL-MHE over a horizon  $\{0, \dots, T\}$ . We recall that the KL-MHE yields a sequence of distributions  $\{\mu_k\}_{k=0}^T$  over the time horizon. Differential privacy over the horizon requires an upper bound on the sensitivity of the joint distribution  $\sigma$  over the horizon, where  $\mu_k$  is the marginal of  $\sigma$  at the time instant  $k$ . As before, with a slight abuse of notation, letting  $\sigma$  also denote the joint density function, we have:

$$\sigma(x_0, x_1, \dots, x_T) = \rho_0(x_0)\sigma(x_1, \dots, x_T|x_0) = \rho_0(x_0)\rho_1(x_1|x_0)\rho_2(x_2|x_1) \dots \rho_T(x_T|x_{T-1}).$$

From the above, we infer that to estimate the sensitivity of the joint density function, we must estimate the sensitivity of the conditionals  $\rho_k(x_k|x_{k-1})$ . The conditional  $\rho_k(x_k|x_{k-1})$  at any time instant  $k$ , is obtained from the coupling between the marginal distributions  $\mu_k$  and  $\mu_{k-1}$ .

We now obtain an upper bound for the case where the marginals  $\mu_k$  are independently coupled. In other words, we suppose that:

$$\sigma(x_0, x_1, \dots, x_T) = \rho_0(x_0)\sigma(x_1, \dots, x_T|x_0) = \rho_0(x_0)\rho_1(x_1)\rho_2(x_2) \dots \rho_T(x_T). \quad (22)$$

**Theorem 8. (Differentially private KL-MHE).** *Given two  $\delta$ -adjacent measurement sequences  $\mathbf{y}, \tilde{\mathbf{y}} \in \mathbb{Y}^{T+N+1}$ , under Assumption 5 and the independent coupling (22), we have that the estimates generated by (20) satisfy  $D_{\max}(\sigma, \tilde{\sigma}) \leq \epsilon$  if  $\sum_{k=1}^T \sum_{l=1}^k \left( \prod_{i=l}^k s_i \right) \leq \epsilon \left( 2\eta \max_k \left( \alpha_k + lc_f^k \delta \text{diam}(\mathcal{K}_0) \right) \right)^{-1}$ , where  $\alpha_k = \min_{\xi \in f_0^k(\mathcal{K}_0)} \left| \left( G_k^N - \tilde{G}_k^N \right) (\xi) \right|$ .*

*Proof.* Let  $G_k^N$  and  $\tilde{G}_k^N$  be the estimation objective functions at time instant  $k$ , corresponding to the measurement sequences  $\mathbf{y}$  and  $\tilde{\mathbf{y}}$  respectively, and let  $\sigma$  and  $\tilde{\sigma}$  be the respective joint probability measures over the horizon  $\{0, \dots, T\}$ . With a slight abuse of notation, we allow  $\sigma$  and  $\tilde{\sigma}$  to also denote the joint density function. From (22), we get:

$$\log \left( \frac{\sigma}{\tilde{\sigma}} \right) (x_0, \dots, x_T) = \sum_{k=1}^T \log \left( \frac{\rho_k}{\tilde{\rho}_k} \right) (x_k),$$

which implies that:

$$D_{\max}(\sigma, \tilde{\sigma}) \leq \sum_{k=1}^T D_{\max}(\mu_k, \tilde{\mu}_k).$$

From the proof of Theorem 7 on the sensitivity of KL-MHE, we further get:

$$D_{\max}(\sigma, \tilde{\sigma}) \leq \sum_{k=1}^T D_{\max}(\mu_k, \tilde{\mu}_k) \leq 2\eta \max_k \left( \alpha_k + lc_f^k \delta \text{diam}(\mathcal{K}_0) \right) \sum_{k=1}^T \sum_{l=1}^k \left( \prod_{i=l}^k s_i \right).$$

Therefore, it holds that  $D_{\max}(\sigma, \tilde{\sigma}) \leq \epsilon$  if:

$$\sum_{k=1}^T \sum_{l=1}^k \left( \prod_{i=l}^k s_i \right) \leq \frac{\epsilon}{2\eta \max_k \left( \alpha_k + lc_f^k \delta \text{diam}(\mathcal{K}_0) \right)}.$$

□

## 8 Simulation results

In this section, we present results from numerical simulations of the estimators studied in this paper. The simulations were performed in MATLAB (version R2017a) on a 2.5 GHz Intel Core i5 processor.

We considered the following nonlinear discrete-time system:

$$\begin{aligned} x_1(k+1) &= x_1(k) + \tau x_2(k), \\ x_2(k+1) &= x_2(k) - \tau \frac{x_1(k)}{1 + |x_1(k)|^2 + |x_2(k)|^2} + w_k, \\ y(k) &= x_1(k) + v_k, \end{aligned}$$

with  $\tau = 0.1$ ,  $w_k$  and  $v_k$  are i.i.d disturbances, sampled uniformly from the intervals  $[-0.1, 0.1]$  and  $[-0.15, 0.15]$  respectively, and a quadratic estimation objective function  $J_T(\mathbf{y}_{0:T}^{(1)}, \mathbf{y}_{0:T}^{(2)}) = \|\mathbf{y}_{0:T}^{(1)} - \mathbf{y}_{0:T}^{(2)}\|^2$ .

We first present the simulation results for  $W_2$ -MHE. We ran 30 trials of the estimator (9) on the same measurement sequence, with randomly generated initial conditions and over a time horizon of length  $T = 100$ . The length of the moving-horizon was chosen to be  $N = 10$ . Figure 1 contains the plots of the mean of the estimates along with the true states. The root mean squared error (RMSE) for the mean state estimate sequences were found to be  $z_1^{\text{RMSE}} =$

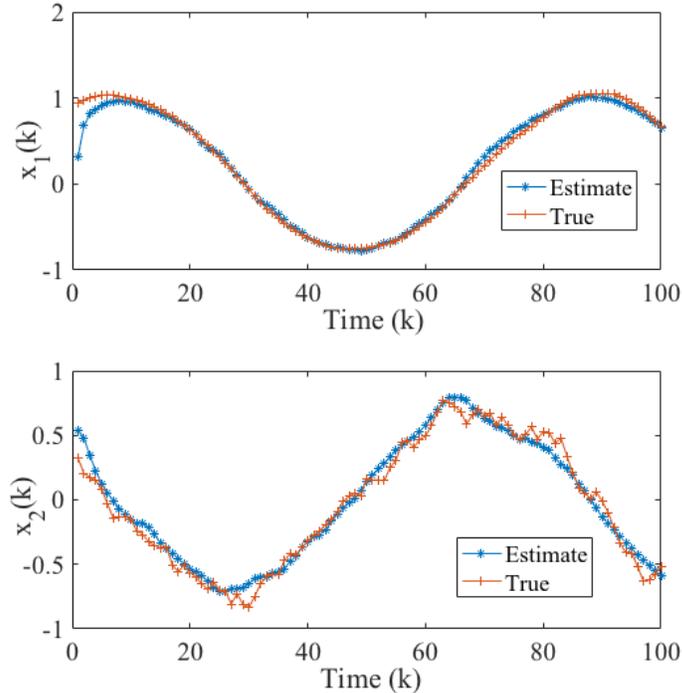


Figure 1: Mean state estimates from 30 trials of  $W_2$ -MHE

0.0856 and  $z_2^{\text{RMSE}} = 0.0846$  for the estimates of  $x_1$  and  $x_2$ , respectively. The average time for computing the state estimate through the minimization (9) using the *fminunc* function in MATLAB was observed to be  $t_{\text{comp}} = 0.012 \pm 0.02s$ .

We then implemented the estimator (13) with 30 samples, over a time horizon of length  $T = 100$ . The length of the moving-horizon was chosen to be  $N = 10$ . Figure 2 contains the plots of the mean of the estimates along with the true states. The root mean squared error (RMSE) for the mean state estimate sequences were found to be  $z_1^{\text{RMSE}} = 0.1073$  and  $z_2^{\text{RMSE}} = 0.1144$  for the estimates of  $x_1$  and  $x_2$ , respectively. The average run-time for the minimization (13) by a resampling method was observed to be  $t_{\text{comp}} = (4.8 \pm 0.4) \times 10^{-4}s$ .

In simulation, with 30 samples, we find that the  $W_2$ -MHE performs better with respect to the root mean squared error, while the KL-MHE is much faster. The performance of the KL-MHE is determined by the richness of the sample set and effectiveness of the resampling procedure, choices that depend on context and experience. In this manuscript, we did not attempt to investigate improvements in performance with respect to these choices. The performance of  $W_2$ -MHE does not necessarily improve with the richness of the sample set, but for systems for which  $\Sigma_T^{-1}(\mathbf{y}_{0:T})$  is not a singleton, a richer sample set allows for a more complete characterization of the set of feasible estimates.

Figure 3 illustrates the typical trade-off between accuracy and privacy in moving-horizon estimation. We considered constant weights  $s_k = s$  for the entropic regularization terms in (17) and (20). The values of  $s$  were chosen such that they satisfied the bounds specified in

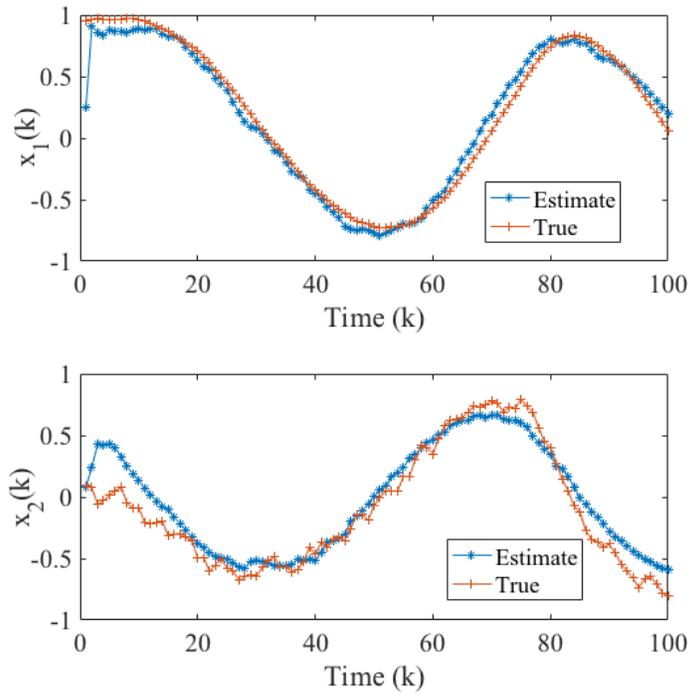


Figure 2: Mean state estimates from KL-MHE with 30 samples

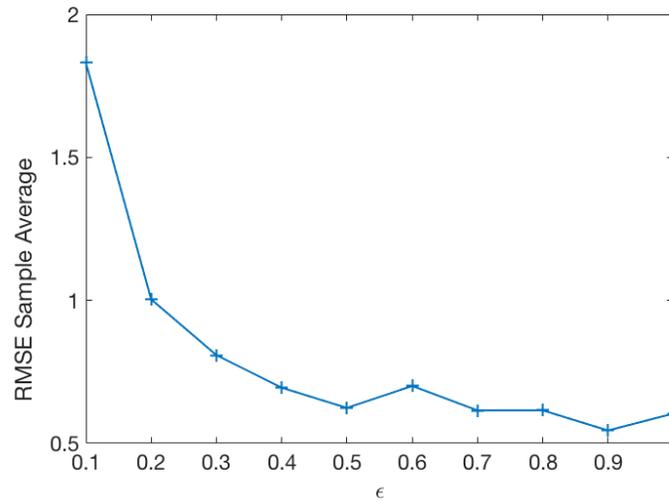


Figure 3: RMSE in estimates of state  $x_1$  for  $W_2$ -MHE, averaged over 30 samples for different values of  $\epsilon$

Theorems 6 and 8 for  $\epsilon$ -differential privacy of the estimators over the horizon. In Figure 3, we plot the RMSE (for the estimates of the state  $x_1$ ) for  $W_2$ -MHE, averaged over the 30 samples, specifying the accuracy, for different values of  $\epsilon$ , the privacy parameter. We recall that a higher value of  $\epsilon$  indicates a less stringent privacy requirement. We notice that the accuracy of the estimators improves with an increase in the privacy parameter.

## 9 Conclusions

In this work, we laid out a unifying probabilistic framework for moving-horizon estimation. We clearly established the connection between the classical notion of strong local observability and the stability of moving-horizon estimation, for nonlinear discrete-time systems. We then proposed a differentially private mechanism based on entropic regularization and derived conditions under which  $\epsilon$ -differential privacy is guaranteed at any given time instant and over time horizons. As an extension to this work, we intend to include distributional constraints in the moving-horizon estimation framework. An important consideration in the estimation problem, in addition to the asymptotic stability, is the rate of convergence of the observer. It is of interest to obtain convergence rate bounds for the moving-horizon estimators proposed in this paper, and to compare their performance for various choices of the metric (or divergence) in the unifying formulation, which will be undertaken in our future work.

## References

- [1] F. Albertini and D. D’Alessandro. Observability and forwardbackward observability of discrete-time nonlinear systems. *Mathematics of Control, Signals and Systems*, 15(4):275290, 2002.
- [2] A. Alessandri, M. Baglietto, and G. Battistelli. Moving-horizon state estimation for nonlinear discrete-time systems: New stability results and approximation schemes. *Automatica*, 44(7):17531765, 2008.
- [3] A. Alessandri and M. Gaggero. Moving-horizon estimation for discrete-time linear and nonlinear systems using the gradient and newton methods. In *IEEE Int. Conf. on Decision and Control*, page 29062911, 2016.
- [4] A. Alessandri and M. Gaggero. Fast moving horizon state estimation for discrete-time systems using single and multi iteration descent methods. *IEEE Transactions on Automatic Control*, 62(9):44994511, 2017.
- [5] L. Ambrosio, N. Gigli, and G. Savaré. *Gradient flows: in metric spaces and in the space of probability measures*. Springer, 2008.
- [6] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.
- [7] J. Cortés, G. E. Dullerud, S. Han, J. Le Ny, S. Mitra, and G. J. Pappas. Differential privacy in control and network systems. In *IEEE Int. Conf. on Decision and Control*, pages 4252–4272, Las Vegas, NV, 2016.

- [8] C. Dimitrakakis, B. Nelson, A. Mitrokotsa, and B. Rubinstein. Robust and private Bayesian inference. In *Int. Conf. on Algorithmic Learning Theory*, page 291305, 2014.
- [9] C. Dimitrakakis, B. Nelson, Z. Zhang, A. Mitrokotsa, and B. Rubinstein. Differential privacy for Bayesian inference through posterior sampling. *Journal of Machine Learning Research*, 18:1–39, 2017.
- [10] A. Doucet, N. D. Freitas, and N. Gordon. Sequential Monte Carlo methods in practice. Springer, 2001.
- [11] C. Dwork and A. Roth. The algorithmic foundations of differential privacy. *Foundations and Trends in Theoretical Computer Science*, 9(34):211407, 2014.
- [12] L. C. Evans. *Partial differential equations*. Graduate studies in mathematics. American Mathematical Society, Providence (R.I.), 1998.
- [13] F. Farokhi, J. Milosevic, and H. Sandberg. Optimal state estimation with measurements corrupted by laplace noise. In *IEEE Int. Conf. on Decision and Control*, page 302307, 2016.
- [14] A. L. Gibbs and F. E. Su. On choosing and bounding probability metrics. *International Statistical Review*, 70(3):419435, 2002.
- [15] R. Hall, A. Rinaldo, and L. Wasserman. Differential privacy for functions and functional data. *Journal of Machine Learning Research*, 14:703727, 2013.
- [16] S. Hanba. Further results on the uniform observability of discrete-time nonlinear systems. *IEEE Transactions on Automatic Control*, 55(4):1034–1038, 2010.
- [17] J. Hespanha, D. Liberzon, D. Angeli, and E. Sontag. Nonlinear norm-observability notions and stability of switched systems. *IEEE Transactions on Automatic Control*, 50(2):154168, 2005.
- [18] W. Hu. Robust stability of optimization-based state estimation. *arXiv preprint arXiv:1702.01903*, 2017.
- [19] A. Jazwinski. Limited memory optimal filtering. *IEEE Transactions on Automatic Control*, 13(5):558563, 1968.
- [20] L. Ji, J. Rawlings, W. Hu, A. Wynn, and M. Diehl. Robust stability of moving horizon estimation under bounded disturbances. *IEEE Transactions on Automatic Control*, 61(11):35093514, 2016.
- [21] A. Kirsch. *An introduction to the mathematical theory of inverse problems*, volume 120. Springer, 2011.
- [22] T. S. Lee, K. P. Dunn, and C. B. Chang. On observability and unbiased estimation of nonlinear systems. In *System Modeling and Optimization*, page 258266. 1982.
- [23] C. Liu, S. Chakraborty, and P. Mittal. Dependence makes you vulnerable: Differential privacy under dependent tuples. In *Network and Distributed System Security Symposium (NDSS)*, volume 16, page 2124, 2016.

- [24] M. Miller. Nonlinear moving horizon estimation in the presence of bounded disturbances. *Automatica*, 79:306314, 2017.
- [25] H. Nijmeijer. Observability of autonomous discrete time non-linear systems: a geometric approach. *International Journal of Control*, 36(5):867874, 1982.
- [26] E. Nozari, P. Tallapragada, and J. Cortés. Differentially private distributed convex optimization via functional perturbation. *IEEE Transactions on Control of Network Systems*, 5(1):395–408, 2018.
- [27] J. Le Ny and G. Pappas. Differentially private filtering. *IEEE Transactions on Automatic Control*, 59(2):341354, 2014.
- [28] G. Peyré. Entropic approximation of Wasserstein gradient flows. *SIAM Journal on Imaging Sciences*, 8(4):23232351, 2015.
- [29] C. Rao, J. Rawlings, and D. Mayne. Constrained state estimation for nonlinear discrete-time systems: Stability and moving horizon approximations. *IEEE Transactions on Automatic Control*, 48(2):246258, 2003.
- [30] F. Santambrogio. {Euclidean, metric, and Wasserstein} gradient flows: an overview. *Bulletin of Mathematical Sciences*, 7(1):87154, 2017.
- [31] S. Song, Y. Wang, and K. Chaudhuri. Pufferfish privacy mechanisms for correlated data. In *ACM Int. Conf. on Management of Data*, page 12911306. ACM, 2017.
- [32] J. Tsiniias and C. Kitsos. Observability and state estimation for a class of nonlinear systems. *arXiv preprint arXiv:1803.08386*, 2018.
- [33] A. Wynn, M. Vukov, and M. Diehl. Convergence guarantees for moving horizon estimation based on the real-time iteration scheme. *IEEE Transactions on Automatic Control*, 59(8):22152221, 2014.
- [34] I. Zang and M. Avriel. On functions whose local minima are global. *Journal of Optimization Theory & Applications*, 16(3-4):183190, 1975.