A Global Lipschitz Stability Perspective for Understanding Approximate Approaches in Bayesian Sequential Learning

Liliang Wang liliangw@umich.edu
Department of Aerospace Engineering
University of Michigan
Ann Arbor, MI 48109, USA

Alex Gorodetsky goroda@umich.edu
Department of Aerospace Engineering
University of Michigan
Ann Arbor, MI 48109, USA

Editor:

Abstract

We establish a general, non-asymptotic error analysis framework for understanding the effects of incremental approximations made by practical approaches for Bayesian sequential learning (BSL) on their long-term inference performance. Our setting covers inverse problems, state estimation, and parameter-state estimation. In these settings, we bound the difference—termed the learning error—between the unknown true posterior and the approximate posterior computed by these approaches, using three widely used distribution metrics: total variation, Hellinger, and Wasserstein distances. This framework builds on our establishment of the global Lipschitz stability of the posterior with respect to the prior across these settings. To the best of our knowledge, this is the first work to establish such global Lipschitz stability under the Hellinger and Wasserstein distances and the first general error analysis framework for approximate BSL methods.

Our framework offers two sets of upper bounds on the learning error. The first set demonstrates the stability of general approximate BSL methods with respect to the incremental approximation process, while the second set is estimable in many practical scenarios.

Furthermore, as an initial step toward understanding the phenomenon of learning error decay, which is sometimes observed, we identify sufficient conditions under which data assimilation leads to learning error reduction.

Keywords: Bayesian sequential learning, approximate methods, stability of Bayesian learning, Wasserstein distance, online variational inference

1 Introduction

Approximations are at the heart of Bayesian learning approaches addressing real-world problems: in most practical scenarios, exact Bayesian solutions are intractable and approximations are made. In this paper, we study approximate methods for Bayesian sequential learning (BSL), in the context of inverse problems (IP), state estimation (SE), and parameter-state estimation (PS). When data arrive and are assimilated incrementally in these settings, due to either large-scale data or a dynamical system, BSL seeks to obtain a sequence of posteriors $P_k, k = 1, 2, \cdots$, by recursively updating an initial prior distribution

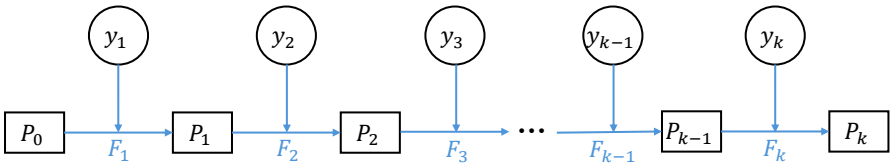


Figure 1: Process of BSL. At each step k, BSL updates the prior P_{k-1} using the data y_k to obtain the posterior P_k. The symbol F_k denotes the map from a prior to its corresponding posterior at step k.

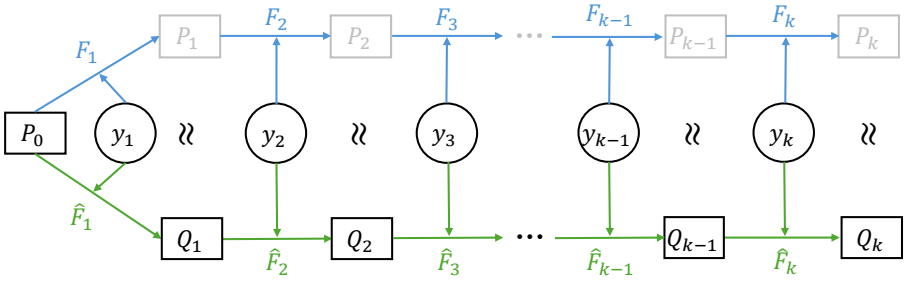


Figure 2: Procedure of approximate approaches for BSL. The exact posterior P_k is typically intractable. The approximate approaches for BSL instead compute an approximate posterior Q_k at each step k, by using the previous approximation Q_{k-1} as the prior and approximately assimilating the new data y_k. Here, Q_k is obtained from Q_{k-1} via $Q_k = \hat{F}_k(Q_{k-1})$. The distance between P_k and Q_k, $d(P_k, Q_k)$, is referred to as the learning error.

P_0. This process is illustrated in Figure 1. At each step k, BSL updates the prior P_{k-1}, which is the posterior of the previous step, using the newly received data y_k, to obtain the posterior P_k. This update can be expressed as $P_k = F_k(P_{k-1})$. Here, F_k denotes the prior-to-posterior map, which varies across the three problem contexts. In IP, SE, and PS, the posterior P_k at step k represents, respectively, the distribution of some unknown parameter X given all available data $\mathcal{Y}_{1:k} = (y_1, \ldots, y_k)$, the distribution of a time-evolving state X_k conditioned on $\mathcal{Y}_{1:k}$, and the joint distribution of X_k and system parameters W given $\mathcal{Y}_{1:k}$.

In practice, the exact posterior P_k is typically intractable to compute. Therefore, practical BSL approaches generate approximate solutions. More specifically, at each step k, they compute an approximate posterior Q_k, by approximately assimilating the new data y_k, using the previous approximate posterior Q_{k-1} as the prior. In other words, Q_k is the result of approximately applying F_k onto Q_{k-1}. We denote this approximate map by $\hat{F}_k$, i.e., $Q_k = \hat{F}_k(Q_{k-1})$. This procedure is shown in Figure 2. Let d denote a metric for probability measures, such as the total variation (TV) or Hellinger distance. We define the distance between the exact solution P_k and the approximate solution Q_k, denoted by $d(P_k, Q_k)$, as the *learning error*. As a result, the prior error at step k is precisely the learning error at previous step $k-1$.

It is important to note that Q_k is not the exact posterior obtained by using Q_{k-1} as the prior and assimilating the data y_k at step k. Let Q_k^* denote this exact posterior, i.e., $Q_k^* = F_k(Q_{k-1})$. As Q_k^* is normally intractable to obtain, approximate methods for BSL

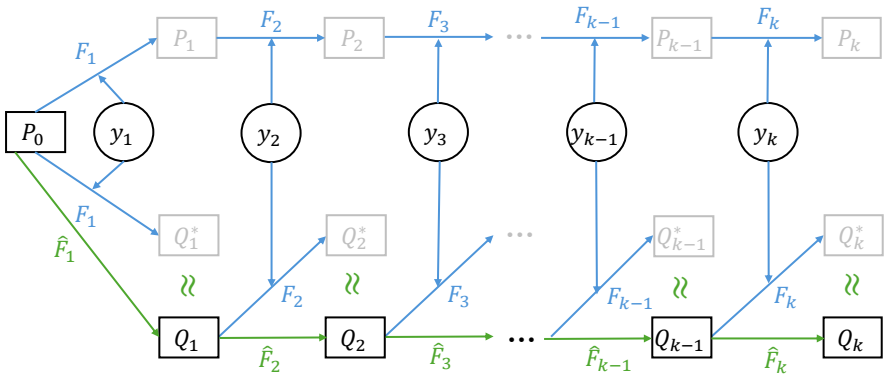


Figure 3: Break-down procedure of approximate approaches in BSL. The exact posterior and approximate posterior at step k are denoted by P_k and Q_k, respectively. When the previous approximate posterior, Q_{k-1}, is used as the prior, the corresponding exact posterior is denoted by $Q_k^* = F_k(Q_{k-1})$. As Q_k^* is normally infeasible to compute, approximate approaches for BSL construct $Q_k = \hat{F}_k(Q_{k-1})$ to approximate Q_k^*. This process is referred to as the incremental approximation process, and the distance $d(Q_k, Q_k^*)$ is called the incremental approximation error.

instead approximate it by Q_k. This procedure is illustrated in Figure 3. The process of approximating Q_k^* with Q_k is referred to as the *incremental approximation process* at step k. We call the distance between Q_k and Q_k^*, $d(Q_k, Q_k^*)$, the *incremental approximation error* of step k.

In recent years, a wide range of methods have been developed for approximately solving BSL problems. These include online/sequential variational inference (VI) methods (e.g., Campbell et al. (2021); Dowling et al. (2023); Neri et al. (2020); Wang and Gorodetsky (2025)) and optimal transport-based sequential learning approaches (e.g., Al-Jarrah et al. (2024, 2025); Grashorn et al. (2024)). Approximate BSL methods also encompass classical Bayesian filtering techniques such as the Gaussian filter (Ito and Xiong (2000)) and sequential importance sampling (particle filter) (e.g., Freitas and Neil (2001); Ning and Ionides (2023)). The Gaussian filter category includes Kalman filter variants, such as the unscented Kalman filter (UKF) and ensemble Kalman filter (EnKF) (Säkkä (2013)).

Despite the extensive development of approximate approaches for BSL, rigorous analysis of these approaches remains limited. Among the available studies, most focus on asymptotic properties of specific approaches (e.g., Crisan et al. (2020); Le Gland et al. (2009); Crisan and Doucet (2002)). Non-asymptotic analysis, by contrast, has been developed to a lesser extent. Existing contributions, such as Elvira et al. (2021), Zhao and Cui (2024), Rebeschini and van Handel (2015), Carrillo et al. (2024), and Chagneux et al. (2024), typically provide results tailored to very specific methods. There is a gap in the literature: a general, method-agnostic framework for non-asymptotic analysis of approximate approaches for BSL is still lacking, and general properties of these methods are rarely studied.

This work aims to bridge this gap by providing a non-asymptotic error analysis framework that applies broadly to various approximate approaches for BSL. This framework enables us to understand how the incremental approximation errors affect the long-term

learning performance of approximate BSL methods. It can also serve as a foundational tool for further analysis and evaluation of a specific method when integrated with method-specific theoretical results. For instance, existing accuracy analysis for offline VI (e.g., Katsevich and Rigollet (2024); Burt et al. (2020); Zhang and Gao (2020)) can be incorporated to assess the incremental approximation error (Q_k, Q_k^*) when applying our framework to sequential VI methods.

Error analysis of general approximate approaches for BSL requires understanding how the prior error, $d(P_{k-1}, Q_{k-1})$, propagates to the posterior error, $d(P_k, Q_k^*)$. In practical scenarios, the prior error can be substantial. As a result, existing local stability results for the prior-to-posterior map F_k, which require the prior error to be sufficiently small, often fail to capture the relationship between the prior and posterior errors in practical BSL settings.

Moreover, since the exact previous posterior P_{k-1} is typically intractable, our knowledge of it is highly limited. It is often impossible to identify a class of distributions which P_{k-1} belongs to, let alone a class that contains both P_{k-1} and Q_{k-1}. Additionally, because our analysis targets general approximate BSL methods, we do not restrict Q_{k-1} to lie within any specific class. Consequently, the existing results of global robustness of posterior, which rely on assumptions that the priors belong to a certain class, are not applicable in this context.

As elaborated in Section 1.2.1, most existing results on the stability of the posterior with respect to the prior are either limited to local stability or rely on restrictive assumptions that the prior belongs to a specific class. Therefore, to establish meaningful error bounds for general approximate approaches in BSL, it is essential to answer the following fundamental question:

QUESTION 1 (global stability of posterior): Is the posterior globally stable with respect to the prior, irrespective of the prior class?

We then use the answer to derive results addressing the following questions:

QUESTION 2 (stability w.r.t. the incremental approximation process): Is the overall learning error $d(P_k, Q_k)$ of general approximate approaches for BSL guaranteed to be upper bounded, provided that the incurred incremental approximation errors $d(Q_j, Q_j^*), j \leq k$ are upper bounded, regardless of how large those bounds may be? Furthermore, given another sequence of approximate posteriors $\tilde{Q}_j, j \leq k$, is the distance between the two approximate posteriors Q_k and $\tilde{Q}_k$, namely $d(Q_k, \tilde{Q}_k)$, also guaranteed to be bounded if their respective incremental approximation errors $d(Q_j, Q_j^*)$ and $d(\tilde{Q}_j, \tilde{Q}_j^*), j \leq k$ are bounded?

QUESTION 3 (accuracy analysis): Does there exist a finite upper bound for the learning error $d(P_k, Q_k)$ which is independent of the true posteriors $P_j, j \leq k$, and determined solely by the approximate posteriors $Q_j, j \leq k$ and incremental approximation errors $d(Q_j, Q_j^*), j \leq k$?

In practical applications, it is often observed that the learning error $d(P_k, Q_k)$ decreases as the number of update steps k increases. Existing analysis of this phenomenon typically focus on specific algorithms, such as bootstrap particle filter and mean-field EnKF. To gain theoretical insight into this behavior for general BSL approaches, we pose the following question:

QUESTION 4 (learning error decay): Under what conditions does the learning error of *general* approximate methods for BSL decay as more data are assimilated?

1.1 Contributions

We answer these questions through three sets of results:

1. (**Global stability of Bayesian inference for arbitrary priors**) We establish global Lipschitz stability of the posterior with respect to the prior, i.e., the pointwise global Lipschitz continuity of the prior-to-posterior map F_k under suitable assumptions in Theorem 14. The formal definition of pointwise globally Lipschitz continuous functions can be found in Definition 1. The assumptions required by Theorem 14 under the TV and Hellinger distances are very mild and hold in most problems. To the best of our knowledge, this is the first work to prove pointwise *global* Lipschitz continuity of the prior-to-posterior map under the Hellinger and Wasserstein distances. For the TV distance, for which global Lipschitz stability bounds already exist, our derived bound improves the state-of-the-art one by 50% under same assumptions.

2. (**Stability and accuracy analyses**) Building on Theorem 14, we establish two sets of upper bounds on the learning error $d(P_k, Q_k)$ in Theorem 16. The first set is linear in the incremental approximation errors $d(Q_j, Q_j^*), j \leq k$, with the approximate posteriors $Q_j, j \leq k$ entering the bounds only through the incremental approximation errors. This result shows that, under mild assumptions, the overall learning error remains bounded as long as the incremental approximation errors are bounded. Moreover, under the same assumptions, the first set of bounds also implies that the distance between two approximate posteriors, Q_k and $\tilde{Q}_k$, is bounded, provided that both sequences of incremental approximation errors, $d(Q_j, Q_j^*)$ and $d(\tilde{Q}_j, \tilde{Q}_j^*), j \leq k$, are bounded. The second set of bounds depends only on the approximate posteriors and incremental approximation errors, and is independent of the exact posteriors $P_j, j \leq k$, making it computable in many practical applications.

3. (**Learning error reduction by data assimilation**) In Theorems 19—23, we provide sufficient conditions on the system and the distributions P_{k-1} and Q_{k-1} under which $d(P_k, Q_k^*) \leq d(P_{k-1}, Q_{k-1})$ holds. Combining these theorems with the triangle inequality $d(P_k, Q_k) \leq d(P_k, Q_k^*) + d(Q_k^*, Q_k)$ enables us to derive sufficient conditions on the system, posteriors and approximate posteriors that guarantee $d(P_k, Q_k) \leq d(P_{k-1}, Q_{k-1})$—without assuming that the approximate posteriors belong to any specific class or satisfy certain structural properties. These results offer an initial step toward explaining the phenomenon of learning error decay in general BSL methods.

   In addition to the above, the contributions of this work also include:

4. (**Analysis under the Wasserstein distance**) We provide an error analysis under the Wasserstein distance for problems involving dynamical systems, including both state estimation and parameter-state estimation. To the best of our knowledge, this work presents the first error analysis under the Wasserstein distance for BSL in parameter-state estimation. Additionally, there are very few existing studies on learning error in state estimation under the Wasserstein distance. Compared to other commonly used probability metrics, the Wasserstein distance has two key advantages. First, it captures not only the magnitude of differences between distributions but

also how these differences are spatially distributed, making it well-suited for comparing distributions with different supports (was; Catalano et al. (2021); Rousseau and Scricciolo (2024)). Second, it enables meaningful comparisons between continuous and discrete distributions (Catalano et al. (2021); was), which is especially useful for analyzing Monte Carlo-based BSL methods.

5. (**Practical insights**) Our theoretical results offer guidance for choosing the initial prior and approximate posteriors in practical BSL applications. Specifically, under certain conditions, choosing a well-designed alternative of the true initial prior can enhance learning accuracy in state estimation problems. Additionally, under the same conditions, selecting an approximate posterior Q_k with a larger immediate incremental approximation error $d(Q_k^*, Q_k)$ may be preferable if it facilitates more accurate incremental approximations in later stages of the learning process, relative to a choice that minimizes the immediate error but complicates the incremental approximation process over time.

1.2 Related work

This paper makes two main contributions: (i) establishing the global Lipschitz stability of the posterior with respect to the prior in Bayesian learning, and (ii) developing a general, non-asymptotic error analysis framework for approximate approaches to BSL. In this section, we review related literature on posterior stability in Bayesian learning and on non-asymptotic analyses of approximate BSL methods.

1.2.1 Posteriors stability in Bayesian learning

This section reviews the literature on posterior stability with respect to the prior in inverse and state estimation problems.

Posterior stability in inverse problems In this work, we establish the global Lipschitz stability of the posterior w.r.t. the prior, as formalized in Theorem 14. Specifically, for inverse problems, the theorem shows that for every prior $\mu \in \bar{\mathcal{P}}_k(\mathcal{X})$, the distance between the posteriors $F_k(\mu)$ and $F_k(\mu')$ is bounded by

$$d(F_k(\mu), F_k(\mu')) \leq K(\mu) d(\mu, \mu'), \quad \forall \mu' \in \bar{\mathcal{P}}_k(\mathcal{X}), \tag{1}$$

where $\bar{\mathcal{P}}_k(\mathcal{X})$ denotes the set of all prior probability measures of which the corresponding posteriors are well-defined. The formal definition of $\bar{\mathcal{P}}_k(\mathcal{X})$ can be found in Definition 7. Here $\mathcal{X}$ represents the space of the unknown variable X. Equation (1) establishes the pointwise global Lipschitz continuity of the prior-to-posterior map F_k.

Over the past four decades, considerable work has been devoted to understanding the posterior stability w.r.t. the prior in Bayesian inverse problems. Broadly, the literature falls into three categories: global robustness analysis, local sensitivity analysis, and local stability analysis.

Global robustness analysis investigates the range of values that a functional of the posterior, $\phi(F_k(\mu))$, may take as the prior μ varies within a specified class Γ. Specifically, it computes

$$\underline{\phi} = \inf_{\mu \in \Gamma} \phi(F_k(\mu)), \quad \overline{\phi} = \sup_{\mu \in \Gamma} \phi(F_k(\mu)),$$

and treats the interval $(\underline{\phi}, \overline{\phi})$ as a robustness range of possible outputs of $\phi(F_k(\mu))$ (Berger and Moreno (1994)). The prior class Γ can be parametric (e.g., normal distribution with fixed mean and bounded covariance (Leamer (1982)) or non-parametric (e.g., Sivaganesan (1994); Bose (1994); Ruggeri (1990)).

Local sensitivity analysis (e.g., Gustafson and L. (1995); Sedighi et al. (2021)) studies the local sensitivity of the posterior w.r.t. the prior. This sensitivity is defined as:

$$s := \lim_{\epsilon \to 0} \frac{d(F_k(\mu), F_k(\mu_\epsilon))}{d(\mu, \mu_\epsilon)},$$

where μ_ϵ is the prior obtained by applying the perturbation ϵ to the base prior μ.

Posterior stability w.r.t. the prior, without restricting the prior to a particular class, has been studied in different works (e.g., Basu (2000, 1998); Sprungk (2020); Garbuno-Inigo et al. (2023)). For example, Garbuno-Inigo et al. (2023) proves a relation between the posterior difference and the prior difference under a family of integral probability metrics (IPM). Specifically, it shows:

$$d_c(F_k(\mu), F_k(\mu')) \leq K'(\mu, \mu') d_{c_{y_k}}(\mu, \mu'), \qquad \forall \mu, \mu' \in \mathcal{P}_1(\mathcal{X}; c_{y_k}), \tag{2}$$

where d_c and $d_{c_{y_k}}$ are two metrics within the IPM family, and the space $\mathcal{P}_1(\mathcal{X}; c_{y_k})$ is a set depending on the data y_k. Because K' depends on both priors μ and μ', this result does not establish global pointwise Lipschitz continuity of F_k.

The works in Basu (2000, 1998) establish the local Lipschitz stability, showing that if the likelihood is bounded above, then for each prior $\mu \in \bar{\mathcal{P}}_k(\mathcal{X})$, there exists $\delta > 0$ such that

$$d(F_k(\mu), F_k(\mu')) \leq \tilde{K}(\mu) d(\mu, \mu'), \quad if\ \mu' \in \bar{\mathcal{P}}_k(\mathcal{X})\ and\ d(\mu, \mu') < \delta. \tag{3}$$

The study in Sprungk (2020) proves the pointwise global Lipschitz continuity of F_k under the TV distance, and the pointwise local Lipschitz continuity of F_k under the Hellinger and 1-Wasserstein distances. Under the same assumptions, our Theorem 14 yields an upper bound on the posterior distance $d(F_k(\mu), F_k(\mu'))$ under the TV distance that is half the bound established in Sprungk (2020). A detailed comparison between our results and those in Sprungk (2020) is provided in Appendix A.1.

To conclude, prior work on global posterior robustness typically requires that the priors belong to a specific class. In contrast, most existing work that do not impose such restrictions on priors only establish local posterior stability with respect to the prior.

It is important to note that the Lipschitz constants in posterior stability results typically depend on the evidence terms associated with at least one of the priors. The evidence, also known as the marginal likelihood, is a functional of the prior. Specifically, the evidence of a prior μ, $Z_k(\mu)$, is defined as

$$Z_k(\mu) = \mathbb{E}_{X \sim \mu}[p(y \mid X)],$$

where y denotes the data and $p(y \mid X)$ is the likelihood. The evidence $Z_k(\mu)$ represents the expected value of the data's probability density under the prior distribution μ.

In our result, the Lipschitz constant $K(\mu)$ in Equation (1) depends on the evidence $Z_k(\mu)$. To the best of our knowledge, the Lipschitz constants appearing in the existing

results on the Lipschitz stability of the posterior either explicitly depend on the evidence terms of one or both priors (e.g., Basu (2000, 1998); Sprungk (2020)), or can be easily shown to be larger than another Lipschitz constant that is a functional of the priors' evidence terms under their respective assumptions (e.g., Garbuno-Inigo et al. (2023)). For instance, the Lipschitz constant $\tilde{K}$ in Equation (3) is directly expressed as a functional of $Z_k(\mu)$ in Basu (2000, 1998) and Sprungk (2020). Furthermore, the proof of Equation (2) in Garbuno-Inigo et al. (2023) shows the existence of another Lipschitz constant—depending on both $Z_k(\mu)$ and $Z_k(\mu')$—that is strictly smaller than the Lipschitz constant $K'(\mu, \mu')$ stated in their result.

Posterior stability in state estimation problems For state estimation problems, the global Lipschitz continuity of F_k under the TV distance has been shown in Rebesschini and van Handel (2015), Law et al. (2015), and Sanz-Alonso et al. (2023) under a strong assumption that the likelihood is bounded above and below by positive constants. This assumption is stronger than ours in Theorem 14. Under this assumption, our Theorem 14 also establishes global Lipschitz continuity of F_k and provides an upper bound on the posterior distance $d(F_k(\mu), F_k(\mu'))$ that is $\frac{1}{2}$ of the bound given in these works. See Appendix A.2 for a detailed derivation.

1.2.2 Existing error analyses for approximate approaches for BSL

This section reviews existing non-asymptotic error analyses for approximate approaches to BSL. As previously noted, such analyses are relatively underexplored. The existing results generally exhibit two limitations:

- Most of them are restricted to specific methods and can not generalize to other algorithms;
- Many rely on strong assumptions.

For instance, Chagneux et al. (2024) derives upper bounds on the error between the expectation of a specific type of functional under the approximate smoothing distribution and the expectation of the same functional under the true, unknown smoothing distribution in state estimation problems. The approximate smoothing distribution is the output of a particular sequential VI method based on the backward factorization of the smoothing distributions. The assumptions imposed in Chagneux et al. (2024) include the following condition:

$$\sigma_1 \leq p(y_{k+1} \mid X_{k+1})p(X_{k+1} \mid X_k) \leq \sigma_2, \quad \forall X_k, X_{k+1} \tag{4}$$

where the data y_{k+1} is given, $p(X_{k+1} \mid X_k)$ denotes the transition probability density from the state X_{k+1} to the state X_k (i.e., the state transition model), and $p(y_{k+1} \mid X_{k+1})$ denotes the probability density of y_{k+1} given X_{k+1} (i.e., the observation model). The condition $\sigma_1 \leq p(y_{k+1} \mid X_{k+1})p(X_{k+1} \mid X_k)$ is quite restrictive. For example, it is not satisfied in common settings where both the state transition and observation models are Gaussian.

The work in Elvira et al. (2021) analyzes the learning error of a block-adaptive particle filter that periodically updates the number of particles. The study in Sanz-Alonso et al. (2023) provides learning error bounds for the bootstrap particle filter. Both analyses are

based on the assumption that the observation model is bounded both above and below by positive constants—an assumption that is quite strong.

In Zhao and Cui (2024), the authors present an error analysis for a tensor train-based sequential learning method in parameter-state estimation problems. Although the assumptions made in Zhao and Cui (2024) are mild, the analysis is tailored to a specific procedure used by this tensor train-based method and is not able to provide meaningful results for other approaches.

1.3 Organization of the paper

The remainder of this paper is organized as follows. Section 2 introduces the necessary notation, definitions, and the problems considered. Section 3 presents our assumptions and main theorems. Section 4 provides several discussions of the main theorems. In Section 5, we apply our results to a specific class of BSL methods—online VI. Section 6 presents an illustrative numerical example. Finally, we conclude the paper in Section 7.

2 Background

In this section, we present the necessary notation, provide the definition of pointwise globally Lipschitz continuous functions, introduce three types of distances for probability measures, and formally describe the problems considered in this paper.

2.1 Notation

Let $\mathbb{R}$ denote the set of real numbers, and $\mathbb{R}_+$ the set of positive real numbers. For $a, b \in \mathbb{R}$, we let $a \wedge b$ and $a \vee b$ denote the minimum and maximum of a and b, respectively. The symbol $d_{\mathcal{Z}}$ denotes a metric for the space $\mathcal{Z}$ and $\mathcal{B}(\mathcal{Z})$ denotes the Borel σ-algebra on $\mathcal{Z}$. If a measure μ is absolutely continuous w.r.t. a measure ν, we write $\mu \ll \nu$.

Random variables are denoted by uppercase letters. The realization of a random variable is denoted by the corresponding lowercase letter (e.g., y for Y). Let expectation and probability be denoted as $\mathbb{E}$ and $\mathbb{P}$, respectively. If the probability measure $\mathbb{P}$ is absolutely continuous w.r.t. the Lebesgue measure, its probability density is denoted by p. The conditional probability density $p(X = x \mid Y = y)$ is denoted by $p(x \mid y)$ for brevity. The symbol $\mathcal{N}(m, R)$ denotes a normal distribution with mean m and covariance R. The value of the probability density function (pdf) of this normal distribution evaluated at z is denoted by $p_{\mathcal{N}}(z \mid m, R)$.

We use d_{TV}, d_H and W_1 to denote the total variation distance, Hellinger distance and 1-Wasserstein distance, respectively.

2.2 Definitions

This section provides definitions of pointwise globally Lipschitz continuous functions and the distances for probability measures.

2.2.1 Pointwise globally Lipschitz continuous functions

The definition of pointwise globally Lipschitz continuous functions is as follows.

Definition 1 (Pointwise globally Lipschitz continuous functions) *Let $(\mathcal{Z}, d_{\mathcal{Z}})$ and $(\tilde{\mathcal{Z}}, d_{\tilde{\mathcal{Z}}})$ be two metric spaces. A map $f : \mathcal{Z} \to \tilde{\mathcal{Z}}$ is called pointwise globally Lipschitz continuous if for every $z \in \mathcal{Z}$, there exists a constant $K(z) \in \mathbb{R}_+$ such that*

$$d_{\tilde{\mathcal{Z}}}\big(f(z), f(z')\big) \leq K(z) d_{\mathcal{Z}}(z, z'), \quad \forall z' \in \mathcal{Z}\,. \tag{5}$$

For every $z \in \mathcal{Z}$, a constant $K(z) \in \mathbb{R}_{\geq 0}$ that satisfies Equation (5) is called a pointwise Lipschitz constant of f at z.

Remark 2 *Function f is called pointwise locally Lipschitz continuous if the inequality in Equation (5) only holds for $z' \in B(z, r(z))$, where $B(z, r(z))$ denotes an open ball centered at z with radius $r(z)$ (Aggarwal and Cobzaș (2023)).*

2.2.2 Distances for probability measures

In this section, we introduce three types of distances that quantify the difference between probability measures: total variation distance, Hellinger distance, and Wasserstein distance.

Let μ and μ' be two probability measures defined on the measurable space $(\mathcal{Z}, \mathcal{B}(\mathcal{Z}))$. Let $\mathcal{P}(\mathcal{Z})$ denote the set of all probability measures defined on the space $(\mathcal{Z}, \mathcal{B}(\mathcal{Z}))$. Assume that μ and μ' are both absolutely continuous with respect to the σ-finite measure ν (such ν always exists). The total variation distance and Hellinger distance are defined as follows.

Definition 3 (Total variation distance) *The total variation distance between two probability measures $\mu, \mu' \in \mathcal{P}(\mathcal{Z})$ is defined by*

$$d_{TV}(\mu, \mu') := \sup_{A \in \mathcal{B}(\mathcal{Z})} \big|\mu(A) - \mu'(A)\big| = \frac{1}{2} \int_{\mathcal{Z}} \left| \frac{d\mu}{d\nu}(z) - \frac{d\mu'}{d\nu}(z) \right| \nu(dz).$$

Definition 4 (Hellinger distance) *The Hellinger distance between two probability measures $\mu, \mu' \in \mathcal{P}(\mathcal{Z})$ is defined by*

$$d_H(\mu, \mu') := \sqrt{\frac{1}{2} \int_{\mathcal{Z}} \left(\sqrt{\frac{d\mu}{d\nu}(z)} - \sqrt{\frac{d\mu'}{d\nu}(z)} \right)^2 \nu(dz)}.$$

The definitions of the Wasserstein distance and Wasserstein space are provided below.

Definition 5 (Wasserstein distance and Wasserstein space) *Let $(\mathcal{Z}, d_{\mathcal{Z}})$ be a Polish metric space. For two probability measures $\mu, \mu' \in \mathcal{P}_q(\mathcal{Z})$, the Wasserstein distance of order q between μ and μ' is defined as*

$$W_q(\mu, \mu') := \inf_{\gamma \in \Gamma(\mu, \mu')} \left(\mathbb{E}_{(Z, Z') \sim \gamma} [d_{\mathcal{Z}}^q(Z, Z')] \right)^{\frac{1}{q}},$$

where $\Gamma(\mu, \mu')$ denotes the set of all couplings of μ and μ', and $\mathcal{P}_q(\mathcal{Z})$ is a space defined as:

$$\mathcal{P}_q(\mathcal{Z}) := \left\{ \mu \in \mathcal{P}(\mathcal{Z}) : \int_{\mathcal{Z}} d_{\mathcal{Z}}(z_0, z)^q \mu(dz) < \infty \right\},$$

where $z_0 \in \mathcal{Z}$. The space $\mathcal{P}_q$ is called the Wasserstein space of order q (Vallani (2009)). The distance W_q is referred to as the q-Wasserstein distance. In this paper, we focus on the 1-Wasserstein distance, W_1, which is well-defined on the space $\mathcal{P}_1$.

2.3 Problem statement

This section formally introduces the three problems considered in this paper: inverse problems, state estimation in dynamical systems, and parameter-state estimation in dynamical systems.

2.3.1 Inverse problems

In inverse problems, the objective is to learn an unknown variable X using the received data $\mathcal{Y}_{1:N} = (y_1, y_2, \cdots, y_N)$. From the Bayesian perspective, X is treated as a random variable. Let $\mathcal{X}$ denote the space for the realizations of X, equipped with a metric $d_{\mathcal{X}}$. We assume $(\mathcal{X}, d_{\mathcal{X}})$ is a Polish metric space. Each data point y_i can be viewed as a realization of the observable Y. We assume the conditional distribution of Y given X admits a density $p(y \mid x) = h(y, x)$ w.r.t. the Lebesgue measure. Given the data y_k, the likelihood model $h(y_k, x)$, as a function of x, is assumed to be measurable.

At each update step k, the prior refers to the distribution of X before assimilating the data y_k, and the posterior refers to the distribution of X after incorporating y_k. Given a prior μ, by Bayes' rule, the correpsonding posterior $F_k\mu$ (also denoted by $F_k(\mu)$) is given by

$$F_k\mu(dx) = \frac{h(y_k, x)\mu(dx)}{\int_{\mathcal{X}} h(y_k, x)\mu(dx)}.$$

If the prior μ satisfies

$$0 < \int_{\mathcal{X}} h(y_k, x)\mu(dx) < \infty,$$

then the posterior $F_k\mu$ is well-defined.

2.3.2 State estimation

State estimation requires the solution of a sequence of inverse problems, interspersed with propagation of the measure through the dynamics.

Consider the following discrete-time Markov dynamical system with partially observable states:

$$\begin{aligned} p(x_k \mid x_{0:k-1}) &= p(x_k \mid x_{k-1}) = T_k(x_k, x_{k-1}), \\ p(y_k \mid x_{0:k}, \mathcal{Y}_{1:k-1}) &= p(y_k \mid x_k) = h_k(y_k, x_k), \end{aligned} \tag{6}$$

where x_k and y_k represent the realizations of the hidden state X_k and observation Y_k, respectively. Here $x_{0:k}$ represents the sequence $(x_0, x_1, \cdots, x_k)$. The state space $\mathcal{X}$ is assumed to be identical across all time steps k and forms a Polish metric space with a metric $d_{\mathcal{X}}$. Given the data y_k, the observation model $h_k(y_k, x_k)$, as a function of x_k, is assumed to be measurable. We also assume that the state transition model $T_k(x_k, x_{k-1})$ is measurable.

At each time step k, the prior denotes the distribution of the state X_{k-1} before incorporating the data y_k, and the posterior represents the distribution of the state X_k after assimilating y_k. Let the prior μ be an arbitrary probability measure. By Bayes' rule, the corresponding posterior $F_k\mu$ is given by

$$F_k\mu(dx_k) = \frac{h_k(y_k, x_k)\left(\int_{\mathcal{X}} T_k(x_k, x_{k-1})\mu(dx_{k-1})\right)dx_k}{\int_{\mathcal{X}} h_k(y_k, x_k)\left(\int_{\mathcal{X}} T_k(x_k, x_{k-1})\mu(dx_{k-1})\right)dx_k}.$$

If μ satisfies

$$\int_{\mathcal{X}} T_k(x_k, x_{k-1})\mu(dx_{k-1}) < \infty, \quad \forall x_k \in \mathcal{X},$$

and

$$0 < \int_{\mathcal{X}} h_k(y_k, x_k) \left(\int_{\mathcal{X}} T_k(x_k, x_{k-1})\mu(dx_{k-1}) \right) dx_k < \infty,$$

then the posterior $F_k\mu$ is well-defined.

State estimation problems assume fully known models. The next subsection introduces parameter-state estimation, where model parameters are unknown.

2.3.3 Parameter-state estimation

Consider the following discrete-time Markov dynamical system with unknown system parameters:

$$\begin{aligned} p(x_k \mid x_{0:k-1}, w) &= p(x_k \mid x_{k-1}, w) = T_k(x_k, x_{k-1}, w), \\ p(y_k \mid x_{0:k}, \mathcal{Y}_{1:k-1}, w) &= p(y_k \mid x_k, w) = h_k(y_k, x_k, w), \end{aligned} \tag{7}$$

where w represents the realization of the system parameters W. We assume that the state space $\mathcal{X}$ is time-invariant. The state space $\mathcal{X}$ and parameter space $\mathcal{W}$ form a Polish metric space $(\mathcal{X} \times \mathcal{W}, d_{\mathcal{X}\times\mathcal{W}})$ with a metric $d_{\mathcal{X}\times\mathcal{W}}$. Assume the state transition model T_k is measurable. Given the data y_k, the observation model $h(y_k, x_k, w)$, as a function of x_k and w, is also assumed to be measurable.

At each time step k, the prior refers to the joint distribution of the state X_{k-1} and system parameters W before assimilating the data y_k, and the posterior refers to the joint distribution the state X_k and W after assimilating y_k. Suppose that the prior μ is absolutely continuous with respect to the Lebesgue measure, with density π. Then, by Bayes' rule, the correpsonding posterior $F_k\mu$ is given by

$$F_k\mu(dx_k, dw) = \frac{h_k(y_k, x_k, w) \left(\int_{\mathcal{X}} T_k(x_k, x_{k-1}, w)\pi(x_{k-1}, w)dx_{k-1} \right) dx_k dw}{\int_{\mathcal{X}\times\mathcal{W}} h_k(y_k, x_k, w) \left(\int_{\mathcal{X}} T_k(x_k, x_{k-1}, w)\pi(x_{k-1}, w)dx_{k-1} \right) dx_k dw}.$$

If μ satisfies:

$$\int_{\mathcal{X}} T_k(x_k, x_{k-1}, w)\pi(x_{k-1}, w)dx_{k-1} < \infty, \quad \forall x_k \in \mathcal{X}, w \in \mathcal{W},$$

and

$$0 < \int_{\mathcal{X}\times\mathcal{W}} h_k(y_k, x_k, w) \left(\int_{\mathcal{X}} T_k(x_k, x_{k-1}, w)\pi(x_{k-1}, w)dx_{k-1} \right) dx_k dw < \infty,$$

then the posterior $F_k\mu$ is well-defined.

Remark 6 *Systems 6 and 7 encompass both autonomous and non-autonomous systems. If external inputs are present, they are assumed to be incorporated into the state transition model T_k.*

2.4 Unified Bayesian learning framework for different problems

In this section, we present a unified framework for Bayesian learning in the three problem settings introduced earlier: inverse problems, state estimation, and parameter-state estimation.

Throughout this paper, let $\bar{\mathcal{X}}$ represent the space $\mathcal{X}$ in inverse and state estimation problems, and $\mathcal{X} \times \mathcal{W}$ in parameter-state estimation problems. Let $\mathcal{P}(\bar{\mathcal{X}})$ denote the set of all probability measures defined on the measurable space $(\bar{\mathcal{X}}, \mathcal{B}(\bar{\mathcal{X}}))$. Similarly, let $\mathcal{M}(\bar{\mathcal{X}})$ denote the set of all σ-finite measures defined on $(\bar{\mathcal{X}}, \mathcal{B}(\bar{\mathcal{X}}))$. Let λ denote the Lebesgue measure defined on the measurable space $(\mathcal{X} \times \mathcal{W}, \mathcal{B}(\mathcal{X} \times \mathcal{W}))$.

We refer to the prior distributions at step k of which the corresponding posteriors are well-defined as *admissible priors*. We then define $\bar{\mathcal{P}}_k(\mathcal{X})$ as the set of all admissible prior probability measures in inverse and state estimation problems. For parameter-state estimation problems, $\bar{\mathcal{P}}_k(\mathcal{X} \times \mathcal{W})$ is defined as the set of admissible priors that are absolutely continuous with respect to the Lebesgue measure. The formal definition of $\bar{\mathcal{P}}_k(\bar{\mathcal{X}})$ is provided below.

Definition 7 (Admissible prior space $\bar{\mathcal{P}}_k(\bar{\mathcal{X}})$) *At step k, given the data y_k, the space $\bar{\mathcal{P}}_k(\bar{\mathcal{X}})$ is defined as follows:*

- *in inverse problems:* $\bar{\mathcal{P}}_k(\mathcal{X}) := \{\mu \in \mathcal{P}(\mathcal{X}) : 0 < \int_{\mathcal{X}} h(y_k, x)\mu(dx) < \infty\}$,
- *in state estimation problems:*

$$\bar{\mathcal{P}}_k(\mathcal{X}) := \Bigg\{\mu \in \mathcal{P}(\mathcal{X}) : \int_{\mathcal{X}} T_k(x_k, x_{k-1})\mu(dx_{k-1}) < \infty \quad \forall x_k \in \mathcal{X}, \\ 0 < \int_{\mathcal{X}} h_k(y_k, x_k) \left(\int_{\mathcal{X}} T_k(x_k, x_{k-1})\mu(dx_{k-1})\right) dx_k < \infty\Bigg\},$$

- *in parameter-state estimation problems:*

$$\bar{\mathcal{P}}_k(\mathcal{X} \times \mathcal{W}) := \Bigg\{\mu \in \mathcal{P}(\mathcal{X} \times \mathcal{W}) : \mu \ll \lambda, \\ \int_{\mathcal{X}} T_k(x_k, x_{k-1}, w)\frac{d\mu}{d\lambda}(x_{k-1}, w)dx_{k-1} < \infty \quad \forall x_k \in \mathcal{X}, w \in \mathcal{W}, \\ 0 < \int_{\mathcal{X}\times\mathcal{W}} h_k(y_k, x_k, w) \left(\int_{\mathcal{X}} T_k(x_k, x_{k-1}, w)\frac{d\mu}{d\lambda}(x_{k-1}, w)dx_{k-1}\right) dx_k dw < \infty\Bigg\}.$$

Remark 8 *The space $\bar{\mathcal{P}}_k(\bar{\mathcal{X}})$ is determined by the system. For instance, in inverse problems, given the data y_k, if the likelihood model h satisfies $h(y_k, x) > 0$ for all $x \in \mathcal{X}$ and $\sup_{x\in\mathcal{X}} h(y_k, x) < \infty$, then we have $\bar{\mathcal{P}}_k(\mathcal{X}) = \mathcal{P}(\mathcal{X})$.*

Next, we introduce the formal definition of the prior-to-posterior map F_k and evidence Z_k.

Definition 9 (Prior-to-posterior map F_k and evidence Z_k) *Let $F_k : \bar{\mathcal{P}}_k(\bar{\mathcal{X}}) \to \mathcal{P}(\bar{\mathcal{X}})$ and $\tilde{F}_k : \bar{\mathcal{P}}_k(\bar{\mathcal{X}}) \to \mathcal{M}(\bar{\mathcal{X}})$ be two functions. For a probability measure $\mu \in \bar{\mathcal{P}}_k(\bar{\mathcal{X}})$, let $F_k\mu$*

and $\tilde{F}_k\mu$ denote the function applications $F_k(\mu)$ and $\tilde{F}_k(\mu)$, respectively. At step k, given the data y_k, the probability measure $F_k\mu$ is defined by

$$F_k\mu(A) := \frac{\tilde{F}_k\mu(A)}{Z_k(\mu)}, \quad \forall A \in \mathcal{B}(\bar{\mathcal{X}}), \tag{8}$$

where the evidence Z_k, which is a real-valued function, is defined as

$$Z_k(\mu) := \int_{\bar{\mathcal{X}}} \tilde{F}_k\mu(d\bar{x}). \tag{9}$$

The measure $\tilde{F}_k\mu$ is defined as follows:

- *in inverse problems: $\tilde{F}_k\mu(dx) := h(y_k, x)\mu(dx)$,*
- *in state estimation problems:*

$$\tilde{F}_k\mu(dx) := h_k(y_k, x)\left(\int_{\mathcal{X}} T_k(x, x_{k-1})\mu(dx_{k-1})\right)dx,$$

- *in parameter-state estimation problems:*

$$\tilde{F}_k\mu(dx, dw) := h_k(y_k, x, w)\left(\int_{\mathcal{X}} T_k(x, x_{k-1}, w)\pi(x_{k-1}, w)dx_{k-1}\right)dxdw,$$

  where π is the density of μ with respect to the Lebesgue measure.

As illustrated in Figure 3, given the data y_k, P_k and Q_k^* are obtained by $P_k = F_k(P_{k-1})$ and $Q_k^* = F_k(Q_{k-1})$, respectively. The approximate posterior Q_k is computed through $Q_k = \hat{F}_k(Q_{k-1})$.

3 Main results

In this section, we present the main theorems of this paper.

3.1 Assumptions

In this section, we present the assumptions necessary for establishing the theorems in the following sections. To distinguish the assumptions by context, we label them with "IP" for inverse problems, "SE" for state estimation problems, and "PS" for parameter-state estimation problems. We begin with the assumptions for inverse problems.

Assumption IP.1 (Upper-bounded likelihood) *Given the data y_k, the likelihood model $h(y_k, x)$ in the inverse problems defined in Section 2.3.1 satisfies*

$$C_h(y_k) := \sup_{x\in\mathcal{X}} h(y_k, x) < \infty. \tag{10}$$

Remark 10 *For a given data y_k, $C_h(y_k)$ is a constant. This assumption is quite mild and holds for most inverse problems. For instance, in the common setting where $Y = H(X)+V$, with H being a forward map and V an additive noise, this assumption is satisfied when the pdf of V is upper bounded—such as when V follows a normal distribution.*

The following assumption is required for theorems related to the 1-Wasserstein distance for inverse problems.

Assumption IP.2 (Lipschitz continuous likelihood) *Given the data y_k, for inverse problems, the likelihood model $h(y_k, x)$, as a function of $x \in \mathcal{X}$, is (globally) Lipschitz continuous. Let $\|h\|_{Lip}(y_k)$ denote its best Lipschitz constant.*

Next, we present the assumptions required for state estimation problems.

Assumption SE.1 *At time step k, given the data y_k, the state transition model T_k and observation model h_k in system 6 satisfy*

$$C_{Th}(y_k; k) := \sup_{x_{k-1}\in\mathcal{X}} \int_{\mathcal{X}} h_k(y_k, x) T_k(x, x_{k-1}) dx < \infty. \tag{11}$$

Remark 11 *Assumption SE.1 is a weaker condition than assuming $h_k(y_k, x)$ is uniformly upper bounded for $x \in \mathcal{X}$, as the latter directly implies Equation* (11)*.*

The following assumption is used for theorems involving the 1-Wasserstein distance for state estimation problems.

Assumption SE.2 *At time step k, given the data y_k, the state transition model T_k and observation model h_k in system 6 satisfy*

$$T_{Lip}(x_k; k) := \sup_{x,x'\in\mathcal{X}, x\neq x'} \frac{|T_k(x_k, x) - T_k(x_k, x')|}{d_{\mathcal{X}}(x, x')} < \infty, \quad \forall x_k \in \mathcal{X},$$

and

$$C^*_{Th}(y_k; k) := \int_{\mathcal{X}} h_k(y_k, x_k) T_{Lip}(x_k; k) dx_k < \infty. \tag{12}$$

Lastly, we state the assumptions required for parameter-state estimation problems.

Assumption PS.1 *At time step k, given the data y_k, the state transition model T_k and observation model h_k in system 7 satisfy*

$$\tilde{C}_{Th}(y_k; k) := \sup_{x_{k-1}\in\mathcal{X}, w\in\mathcal{W}} \int_{\mathcal{X}} h_k(y_k, x, w) T_k(x, x_{k-1}, w) dx < \infty. \tag{13}$$

Remark 12 *Assumption PS.1 is mild and sets a weaker condition than that the observation model h_k satisfies $\sup_{x\in\mathcal{X}, w\in\mathcal{W}} h_k(y_k, x, w) < \infty$.*

The following assumptions are required for theorems related to the 1-Wasserstein distance for parameter-state estimation problems.

Assumption PS.2 *At time step k, given the data y_k, the state transition model T_k and observation model h_k in system 7 satisfy*

$$\tilde{f}_{Lip}(x_k;k) := \sup_{(x,w),(x',w')\in\mathcal{X}\times\mathcal{W},(x,w)\neq(x',w')} \frac{|\tilde{f}_k(y_k,x_k,x,w)-\tilde{f}_k(y_k,x_k,x',w')|}{d_{\mathcal{X}\times\mathcal{W}}((x,w),(x',w'))} < \infty, \quad \forall x_k\in\mathcal{X},$$

where $\tilde{f}_k(y_k,x_k,x,w) := h_k(y_k,x_k,w)T_k(x_k,x,w), \quad \forall x\in\mathcal{X}, w\in\mathcal{W},$
and

$$\tilde{C}^*_{Th}(y_k;k) := \int_{\mathcal{X}} \tilde{f}_{Lip}(x_k;k)dx_k < \infty. \tag{14}$$

Assumption PS.3 *The metric space $(\mathcal{X}\times\mathcal{W}, d_{\mathcal{X}\times\mathcal{W}})$ satisfies*

$$d_{\mathcal{X}\times\mathcal{W}}((x,w),(x,w')) \le d_{\mathcal{X}\times\mathcal{W}}((\tilde{x},w),(x',w')), \quad \forall x,\tilde{x},x'\in\mathcal{X},\ w,w'\in\mathcal{W}.$$

Remark 13 *Assumption PS.3 is generally not restrictive as it is actually the property shared by many metrics. For instance, the Euclidean distance satisfies this assumption.*

An additional assumption required for establishing results under the 1-Wasserstein distance for all problems is as follows.

Assumption IP-SE-PS.4 (Bounded metric space) *The metric space $(\bar{\mathcal{X}}, d_{\bar{\mathcal{X}}})$ satisfies*

$$\sup_{\bar{x},\bar{x}'\in\bar{\mathcal{X}}} d_{\bar{\mathcal{X}}}(\bar{x},\bar{x}') = D < \infty,$$

where $\bar{\mathcal{X}}$ represents $\mathcal{X}$ in inverse problems and state estimation problems, and represents $\mathcal{X}\times\mathcal{W}$ in parameter-state estimation problems, as introduced in Section 2.4.

3.2 Pointwise global Lipschitz continuity of the prior-to-posterior map

Theorem 14 establishes the pointwise global Lipschitz continuity of the prior-to-posterior map F_k.

Theorem 14 (Pointwise global Lipschitz continuity of prior-to-posterior map F_k) *Given the data y_k, the prior-to-posterior map $F_k : (\bar{\bar{\mathcal{P}}}_k(\bar{\mathcal{X}}), d) \to (\mathcal{P}(\bar{\mathcal{X}}), d)$ is pointwise globally Lipschitz continuous, i.e., for every $\mu\in\bar{\bar{\mathcal{P}}}_k(\bar{\mathcal{X}})$, there exists a constant $K(\mu;y_k)$ such that*

$$d(F_k(\mu),F_k(\mu')) \le K(\mu;y_k)d(\mu,\mu'), \quad \forall\mu'\in\bar{\bar{\mathcal{P}}}_k(\bar{\mathcal{X}}), \tag{15}$$

- *in **inverse problems***
  - *when d is the total variation distance or the Hellinger distance, and $\bar{\bar{\mathcal{P}}}_k(\bar{\mathcal{X}})$ represents $\bar{\mathcal{P}}_k(\mathcal{X})$, under Assumption IP.1; or*

Table 1: Pointwise global Lipschitz constant $K(\mu; y_k)$ of the prior-to-posterior map F_k for different problems and distances of probability measures.

d / Problem	Inverse problems	State estimation	Parameter-state estimation
d_{TV}	$\frac{C_h(y_k)}{Z_k(\mu)}$	$\frac{C_{Th}(y_k;k)}{Z_k(\mu)}$	$\frac{\tilde{C}_{Th}(y_k;k)}{Z_k(\mu)}$
d_H	$2\sqrt{\frac{C_h(y_k)}{Z_k(\mu)}}$	$2\sqrt{\frac{C_{Th}(y_k;k)}{Z_k(\mu)}}$	$2\sqrt{\frac{\tilde{C}_{Th}(y_k;k)}{Z_k(\mu)}}$
W_1	$\frac{2D\|h\|_{\mathrm{Lip}}(y_k)+C_h(y_k)}{Z_k(\mu)}$	$\frac{2DC^*_{Th}(y_k;k)}{Z_k(\mu)}$	$\frac{2D\tilde{C}^*_{Th}(y_k;k)+\tilde{C}_{Th}(y_k;k)}{Z_k(\mu)}$

  - *when d is the 1-Wasserstein distance, and $\bar{\bar{\mathcal{P}}}_k(\bar{\mathcal{X}})$ represents $\bar{\mathcal{P}}_k(\mathcal{X}) \bigcap \mathcal{P}_1(\mathcal{X})$, under Assumptions IP.1, IP.2, and IP-SE-PS.4;*

- *in **state estimation problems***

  - *when d is the total variation distance or the Hellinger distance, and $\bar{\bar{\mathcal{P}}}_k(\bar{\mathcal{X}})$ represents $\bar{\mathcal{P}}_k(\mathcal{X})$, under Assumption SE.1; or*
  - *when d is the 1-Wasserstein distance, and $\bar{\bar{\mathcal{P}}}_k(\bar{\mathcal{X}})$ represents $\bar{\mathcal{P}}_k(\mathcal{X}) \bigcap \mathcal{P}_1(\mathcal{X})$, under Assumptions SE.2 and IP-SE-PS.4;*

- *in **parameter-state estimation problems***

  - *when d is the total variation distance or the Hellinger distance, and $\bar{\bar{\mathcal{P}}}_k(\bar{\mathcal{X}})$ represents $\bar{\mathcal{P}}_k(\mathcal{X} \times \mathcal{W})$, under Assumption PS.1; or*
  - *when d is the 1-Wasserstein distance, and $\bar{\bar{\mathcal{P}}}_k(\bar{\mathcal{X}})$ is $\bar{\mathcal{P}}_k(\mathcal{X} \times \mathcal{W}) \bigcap \mathcal{P}_1(\mathcal{X} \times \mathcal{W})$, under Assumptions PS.1, PS.2, PS.3, and IP-SE-PS.4.*

The pointwise global Lipschitz constants $K(\mu; y_k)$ associated with different distances and various problems are summarized in Table 1.

Proof See Appendix B. ■

Remark 15 *Note that the posterior $F_k(\mu')$ depends on the evidence $Z_k(\mu')$, which typically differs from the evidence $Z_k(\mu)$. As a result, without imposing very strong assumptions, it is challenging to establish a global bound on the posterior distance $d(F_k(\mu), F_k(\mu'))$ that holds uniformly for all $\mu' \in \bar{\bar{\mathcal{P}}}_k(\bar{\mathcal{X}})$ and is independent of $Z_k(\mu')$. To the best of our knowledge, Theorem 14 provides the first such global bound under the Hellinger and Wasserstein distances.*

The Lipschitz constant $K(\mu; y_k)$ depends only on one of the priors, μ, and is independent of the other prior, μ'. Moreover, inequality (15) holds for all $\mu, \mu' \in \bar{\bar{\mathcal{P}}}_k(\bar{\mathcal{X}})$.

Given that $P_k = F_k(P_{k-1})$ and $Q_k^* = F_k(Q_{k-1})$, Theorem 14 implies that, regardless of how different the priors P_{k-1} and Q_{k-1} are, the propagated learning error satisfies

$$d(P_k, Q_k^*) \leq K(P_{k-1}; y_k) d(P_{k-1}, Q_{k-1}), \tag{16}$$

and

$$d(P_k, Q_k^*) \leq K(Q_{k-1}; y_k) d(P_{k-1}, Q_{k-1}). \tag{17}$$

As discussed in Section 1.2.1, for the Hellinger and Wasserstein distances, the existing stability results for Bayesian learning either (i) hold with a Lipschitz constant depending on both priors (e.g.,Garbuno-Inigo et al. (2023)), or (ii) establish only the pointwise local Lipschitz continuity of F_k (e.g., Basu (2000, 1998); Sprungk (2020)). In case (i), the bound takes the form

$$d(P_k, Q_k^*) \leq K'(P_{k-1}, Q_{k-1}; y_k) d(P_{k-1}, Q_{k-1}). \tag{18}$$

In case (ii), the bound on $d(P_k, Q_k^*)$ is valid only when $d(P_{k-1}, Q_{k-1})$ is sufficiently small. In practical settings, however, the prior error $d(P_{k-1}, Q_{k-1})$ can be large, and such results often fail to yield bounds on $d(P_k, Q_k^*)$.

Combining inequality (16) with the triangle inequality $d(P_k, Q_k) \leq d(P_k, Q_k^*) + d(Q_k^*, Q_k)$, we obtain

$$d(P_k, Q_k) \leq K(P_{k-1}; y_k) d(P_{k-1}, Q_{k-1}) + d(Q_k^*, Q_k). \tag{19}$$

Recursively applying inequality (19) from step k down to 1 yields an upper bound on the learning error $d(P_k, Q_k)$. The approximate posteriors $Q_j, j \leq k$ only enter this bound through the incremental approximation errors $d(Q_j^*, Q_j), j \leq k$. This bound shows that if the incremental approximation errors $d(Q_j, Q_j^*)$ remain bounded for all $j \leq k$, then the overall learning error $d(P_k, Q_k)$ is also bounded. Further discussion of this bound is provided in Sections 3.3 and 4.1.

Similarly, applying inequality (17) yields

$$d(P_k, Q_k) \leq K(Q_{k-1}; y_k) d(P_{k-1}, Q_{k-1}) + d(Q_k^*, Q_k), \tag{20}$$

and the recursive application of inequality (20) from k to 1 leads to a second bound on $d(P_k, Q_k)$ that depends only on the approximate posteriors $Q_j, j \leq k$ and the incremental approximation errors. This bound can be computed in many practical scenarios. The learning error bounds derived from inequalities (19) and (20) are presented in detail in Section 3.3.

In contrast, learning error bounds derived from inequality (18) depend on the incremental approximation errors as well as both the exact and approximate posteriors. These bounds are generally not computable, as the exact posteriors are typically intractable. Moreover, they do not guarantee that the learning error is bounded when the incremental approximation errors are bounded. Given bounded incremental approximation errors, these learning error bounds themselves may still be unbounded, due to their dependence on the approximate posteriors (the condition that the incremental approximation errors $d(Q_j^*, Q_j), j \leq k$ are bounded does not guarantee that a functional of the approximate posteriors $Q_j, j \leq k$ is bounded).

These results underscore the necessity of Theorem 14 for the establishment of a broadly applicable and practical analysis framework for approximate approaches of BSL.

3.3 Learning error bounds of approximate approaches for BSL

Theorem 14 uncovers the relation between the propagated error $d(P_k, Q_k^*)$ and the prior error $d(P_{k-1}, Q_{k-1})$. As illustrated in Figure 3, this result can be combined with the triangle

Table 2: Definition of function $C(\mathcal{Y}_{j+1:k}, p(\mathcal{Y}_{j+1:k} \mid \mathcal{Y}_{1:j}), j)$

d / Problem	Inverse problems	State estimation	Parameter-state estimation
d_{TV}	$\frac{\prod_{i=j+1}^{k} C_h(y_i)}{p(\mathcal{Y}_{j+1:k}\|\mathcal{Y}_{1:j})}$	$\frac{\prod_{i=j+1}^{k} C_{Th}(y_i;i)}{p(\mathcal{Y}_{j+1:k}\|\mathcal{Y}_{1:j})}$	$\frac{\prod_{i=j+1}^{k} \tilde{C}_{Th}(y_i;i)}{p(\mathcal{Y}_{j+1:k}\|\mathcal{Y}_{1:j})}$
d_H	$\frac{2^{k-j}\prod_{i=j+1}^{k}\sqrt{C_h(y_i)}}{\sqrt{p(\mathcal{Y}_{j+1:k}\|\mathcal{Y}_{1:j})}}$	$\frac{2^{k-j}\prod_{i=j+1}^{k}\sqrt{C_{Th}(y_i;i)}}{\sqrt{p(\mathcal{Y}_{j+1:k}\|\mathcal{Y}_{1:j})}}$	$\frac{2^{k-j}\prod_{i=j+1}^{k}\sqrt{\tilde{C}_{Th}(y_i;i)}}{\sqrt{p(\mathcal{Y}_{j+1:k}\|\mathcal{Y}_{1:j})}}$
W_1	$\frac{\prod_{i=j+1}^{k}(2D\|h\|_{\mathrm{Lip}}(y_i)+C_h(y_i))}{p(\mathcal{Y}_{j+1:k}\|\mathcal{Y}_{1:j})}$	$\frac{(2D)^{k-j}\prod_{i=j+1}^{k} C^*_{Th}(y_i;i)}{p(\mathcal{Y}_{j+1:k}\|\mathcal{Y}_{1:j})}$	$\frac{\prod_{i=j+1}^{k}(2D\tilde{C}^*_{Th}(y_i;i)+\tilde{C}_{Th}(y_i;i))}{p(\mathcal{Y}_{j+1:k}\|\mathcal{Y}_{1:j})}$

Table 3: Definition of function $C'(\mathcal{Y}_{j+1:k}, Q_{j:k-1}, j)$

d / Problem	Inverse problems	State estimation	Parameter-state estimation
d_{TV}	$\prod_{i=j+1}^{k} \frac{C_h(y_i)}{Z_i(Q_{i-1})}$	$\prod_{i=j+1}^{k} \frac{C_{Th}(y_i;i)}{Z_i(Q_{i-1})}$	$\prod_{i=j+1}^{k} \frac{\tilde{C}_{Th}(y_i;i)}{Z_i(Q_{i-1})}$
d_H	$2^{k-j}\prod_{i=j+1}^{k} \frac{\sqrt{C_h(y_i)}}{\sqrt{Z_i(Q_{i-1})}}$	$2^{k-j}\prod_{i=j+1}^{k} \frac{\sqrt{C_{Th}(y_i;i)}}{\sqrt{Z_i(Q_{i-1})}}$	$2^{k-j}\prod_{i=j+1}^{k} \frac{\sqrt{\tilde{C}_{Th}(y_i;i)}}{\sqrt{Z_i(Q_{i-1})}}$
W_1	$\prod_{i=j+1}^{k} \frac{2D\|h\|_{\mathrm{Lip}}(y_i)+C_h(y_i)}{Z_i(Q_{i-1})}$	$(2D)^{k-j}\prod_{i=j+1}^{k} \frac{C^*_{Th}(y_i;i)}{Z_i(Q_{i-1})}$	$\prod_{i=j+1}^{k} \frac{2D\tilde{C}^*_{Th}(y_i;i)+\tilde{C}_{Th}(y_i;i)}{Z_i(Q_{i-1})}$

inequality $d(P_i, Q_i) \leq d(P_i, Q_i^*) + d(Q_i^*, Q_i)$ and applied recursively from $i = k$ down to $i = 1$ to derive upper bounds on the overall learning error $d(P_k, Q_k)$.

Let $\mathcal{Y}_{m:n}$ denote the data received from step m to step n, i.e., $(y_m, y_{m+1}, \cdots, y_n)$. Theorem 16 presents two sets of upper bounds on the overall learning error $d(P_k, Q_k)$, with each set containing one bound under the TV distance, one under the Hellinger distance, and one under the 1-Wasserstein distance. The first set of bounds is provided in Equation (21), while the second set is given in Equation (22).

Theorem 16 *Given a sequence of data $\mathcal{Y}_{1:k}$, suppose the assumptions in Theorem 14 hold for all steps up to step k. Assume $P_i \in \bar{\mathcal{P}}_{i+1}(\bar{\mathcal{X}})$ for all $i \in [0, k-1]$ and $Q_i \in \bar{\mathcal{P}}_{i+1}(\bar{\mathcal{X}})$ for all $i \in [1, k-1]$. For any step $k \geq 2$, the distance between the true posterior P_k and the approximate posterior Q_k is bounded by*

$$d(P_k, Q_k) \leq \sum_{j=1}^{k-1} C(\mathcal{Y}_{j+1:k}, p(\mathcal{Y}_{j+1:k} \mid \mathcal{Y}_{1:j}), j) d(Q_j^*, Q_j) + d(Q_k^*, Q_k), \tag{21}$$

and

$$d(P_k, Q_k) \leq \sum_{j=1}^{k-1} C'(\mathcal{Y}_{j+1:k}, Q_{j:k-1}, j) d(Q_j^*, Q_j) + d(Q_k^*, Q_k), \tag{22}$$

where $Q_{j:k-1}$ represents $(Q_j, Q_{j+1}, \cdots, Q_{k-1})$. The definitions of the real-valued functions C and C' are provided in Table 2 and Table 3, respectively.

Proof See Appendix C. ■

Remark 17 *We note that, the bound provided in Equation* (21) *is a linear function of the incremental approximation errors. This can significantly facilitate further stability analysis.*

Remark 18 *For inverse problems and state estimation problems, Theorem 16 holds for approximate posteriors from arbitrary distribution families, without requiring them to be absolutely continuous with respect to the Lebesgue measure. This makes Theorem 16 applicable to a broad class of BSL methods, including those based on the Monte Carlo technique.*

As shown in Equation (21), the first set of upper bounds depends only on the incremental approximation errors $d(Q_j^*, Q_j), j \leq k$ and the constant $C(\mathcal{Y}_{j+1:k}, p(\mathcal{Y}_{j+1:k} \mid \mathcal{Y}_{1:j}), j)$. For a given problem, the probability density $p(\mathcal{Y}_{j+1:k} \mid \mathcal{Y}_{1:j})$ is fixed, and therefore the constant $C(\mathcal{Y}_{j+1:k}, p(\mathcal{Y}_{j+1:k} \mid \mathcal{Y}_{1:j}), j)$ is also fixed. These bounds thus demonstrate that, as long as the incremental approximation errors, $d(Q_j^*, Q_j), j \leq k$, remain bounded, the overall learning error $d(P_k, Q_k)$ is guaranteed to be bounded. Furthermore, if we are given another sequence of approximate posteriors $\tilde{Q}_j, j \leq k$, Equation (21), combined with the triangle inequality $d(Q_k, \tilde{Q}_k) \leq d(Q_k, P_k) + d(\tilde{Q}_k, P_k)$, shows that the distance between two approximate posteriors Q_k and $\tilde{Q}_k$ is bounded, as long as both sequences of incremental approximation errors, $d(Q_j^*, Q_j)$ and $d(\tilde{Q}_j^*, \tilde{Q}_j), j \leq k$, are bounded. A detailed discussion of the stability of approximate BSL methods with respect to the incremental approximation process, which is demonstrated by Equation (21), is provided in Section 4.1. However, the probability density $p(\mathcal{Y}_{j+1:k} \mid \mathcal{Y}_{1:j})$ in the first set of bounds is normally intractable. As a result, while these bounds provide a stability guarantee, they are generally not computable in practice.

In contrast, the second set of upper bounds, given in Equation (22), is independent of $p(\mathcal{Y}_{j+1:k} \mid \mathcal{Y}_{1:j})$. Computing these bounds only requires the approximate posteriors $Q_j, 1 \leq j \leq k-1$, which are the learning results, and the incremental approximate errors $d(Q_j^*, Q_j), j \leq k$. This makes the second set of bounds estimable in many real-world applications. Further discussion on the estimation of these bounds is provided in Section 4.2.

In practical applications, it can sometimes be observed that the learning error $d(P_k, Q_k)$ decreases as the number of update steps k increases. However, this phenomenon cannot be explained by the results presented so far in this paper. Although Theorem 14, as discussed in Section 1.2.1, yields tighter upper bounds on $d(P_k, Q_k^*)$ than those obtainable from existing results, it can be easily proved that these bounds are still larger than $d(P_{k-1}, Q_{k-1})$. Consequently, this leads to an increase in the derived upper bounds on $d(P_k, Q_k)$ as k grows. In the following section, we will attempt to understand these observations by providing sufficient conditions under which the learning error decreases after data assimilation.

3.4 Error reduction by data assimilation

Understanding the phenomenon of learning error decay requires understanding what leads to $d(P_k, Q_k) \leq d(P_{k-1}, Q_{k-1})$. Recall that $Q_k = \hat{F}_k(Q_{k-1})$. If the map $\hat{F}_k$ exhibits any specific structural properties, these properties can be used to help establish sufficient conditions ensuring learning error reduction. However, since our analysis targets general approximate

Table 4: Definitions of variable $\bar{x}$, function g, and space $\bar{\mathcal{M}}(\bar{\mathcal{X}})$ in different problems. The symbol $\mathcal{M}(\mathcal{X})$ denotes the set of all σ-finite measures defined on the space $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$, and λ is the Lebesgue measure defined on the space $(\mathcal{X} \times \mathcal{W}, \mathcal{B}(\mathcal{X} \times \mathcal{W}))$.

Problem	$\bar{x}$	$g(\bar{x}, y_k)$	$\bar{\mathcal{M}}(\bar{\mathcal{X}})$
Inverse problems	x	$h(y_k, x)$	$\mathcal{M}(\mathcal{X})$
State estimation	x_{k-1}	$\int_{\mathcal{X}} h_k(y_k, x_k) T_k(x_k, x_{k-1}) dx_k$	$\mathcal{M}(\mathcal{X})$
Parameter-state estimation	(x_{k-1}, w)	$\int_{\mathcal{X}} h_k(y_k, x_k, w) T_k(x_k, x_{k-1}, w) dx_k$	$\{\lambda\}$

methods for BSL, we do not impose any structural assumptions on $\hat{F}_k$. Instead, we leverage the structure of approximate BSL methods and establish sufficient conditions on the system, P_{k-1}, and Q_{k-1}, under which the inequality $d(P_k, Q_k^*) \leq d(P_{k-1}, Q_{k-1})$ holds. If these conditions are met and the incremental approximation error satisfies $d(Q_k, Q_k^*) \leq d(P_{k-1}, Q_{k-1}) - d(P_k, Q_k^*)$, then the learning error at step k is smaller than that at step $k-1$.

To facilitate this analysis, we introduce a unified notation involving a variable $\bar{x}$, a system-dependent function g, and a space of reference measures, $\bar{\mathcal{M}}(\bar{\mathcal{X}})$, as summarized in Table 4. We first present sufficient conditions that guarantee error reduction under the TV distance for all three learning problems considered in this paper.

Theorem 19 *Given the data y_k, assume that $P_{k-1} \in \bar{\mathcal{P}}_k(\bar{\mathcal{X}})$ and $Q_{k-1} \in \bar{\mathcal{P}}_k(\bar{\mathcal{X}})$. Suppose that there exists a measure $\nu \in \bar{\mathcal{M}}(\bar{\mathcal{X}})$ such that $P_{k-1} \ll \nu, Q_{k-1} \ll \nu$, and*

$$\begin{aligned}\int_{\bar{\mathcal{X}}^*} g(y_k, \bar{x}) \left| \frac{dP_{k-1}}{d\nu}(\bar{x}) - \frac{dQ_{k-1}}{d\nu}(\bar{x}) \right| \nu(d\bar{x}) \\ \leq \int_{\bar{\mathcal{X}}^*} g(y_k, \bar{x}) \nu(d\bar{x}) \int_{\bar{\mathcal{X}}^*} \left| \frac{dP_{k-1}}{d\nu}(\bar{x}) - \frac{dQ_{k-1}}{d\nu}(\bar{x}) \right| \nu(d\bar{x}),\end{aligned}$$

and

$$\int_{\bar{\mathcal{X}}^*} g(y_k, \bar{x}) \nu(d\bar{x}) \leq \int_{\bar{\mathcal{X}}} g(y_k, \bar{x}) P_{k-1}(d\bar{x}) \vee \int_{\bar{\mathcal{X}}} g(y_k, \bar{x}) Q_{k-1}(d\bar{x}),$$

where

$$\bar{\mathcal{X}}^* := \begin{cases} \{\bar{x} \in \bar{\mathcal{X}} : \frac{dP_{k-1}}{d\nu}(\bar{x}) \geq \frac{dQ_{k-1}}{d\nu}(\bar{x})\}, & \text{if } Z_k(P_{k-1}) \geq Z_k(Q_{k-1}), \\ \{\bar{x} \in \bar{\mathcal{X}} : \frac{dP_{k-1}}{d\nu}(\bar{x}) \leq \frac{dQ_{k-1}}{d\nu}(\bar{x})\}, & \text{otherwise.} \end{cases}$$

Then the distributions P_k and Q_k^* satisfy

$$d_{TV}(P_k, Q_k^*) \leq d_{TV}(P_{k-1}, Q_{k-1}).$$

Proof See Appendix D.1. ■

We next establish conditions under which error reduction occurs under the Hellinger distance. The first such condition is given below.

Assumption ER.H.1 *Given the data y_k, there exists a measure $\nu \in \bar{\mathcal{M}}(\bar{\mathcal{X}})$ such that $P_{k-1} \ll \nu, Q_{k-1} \ll \nu$, and*

$$\int_{\bar{\mathcal{X}}} g(y_k, \bar{x}) \sqrt{\frac{dP_{k-1}}{d\nu}(\bar{x}) \frac{dQ_{k-1}}{d\nu}(\bar{x})} \nu(d\bar{x}) \geq \int_{\bar{\mathcal{X}}} g(y_k, \bar{x}) \nu(d\bar{x}) \int_{\bar{\mathcal{X}}} \sqrt{\frac{dP_{k-1}}{d\nu}(\bar{x}) \frac{dQ_{k-1}}{d\nu}(\bar{x})} \nu(d\bar{x}),$$

and

$$\int_{\bar{\mathcal{X}}} g(y_k, \bar{x}) \nu(d\bar{x}) \geq \sqrt{\int_{\bar{\mathcal{X}}} g(y_k, \bar{x}) P_{k-1}(d\bar{x}) \int_{\bar{\mathcal{X}}} g(y_k, \bar{x}) Q_{k-1}(d\bar{x})}.$$

An alternative sufficient condition for Hellinger error reduction is provided in the following assumption.

Assumption ER.H.2 *Given the data y_k, there exists a measure $\nu \in \bar{\mathcal{M}}(\bar{\mathcal{X}})$ such that $P_{k-1} \ll \nu, Q_{k-1} \ll \nu$, and*

$$\begin{aligned} &\int_{\bar{\mathcal{X}}} g(y_k, \bar{x}) \left(\sqrt{\frac{dP_{k-1}}{d\nu}(\bar{x})} - \sqrt{\frac{dQ_{k-1}}{d\nu}(\bar{x})} \right)^2 \nu(d\bar{x}) \\ &\quad \leq \int_{\bar{\mathcal{X}}} g(y_k, \bar{x}) \nu(d\bar{x}) \int_{\bar{\mathcal{X}}} \left(\sqrt{\frac{dP_{k-1}}{d\nu}(\bar{x})} - \sqrt{\frac{dQ_{k-1}}{d\nu}(\bar{x})} \right)^2 \nu(d\bar{x}), \end{aligned}$$

and

$$\int_{\bar{\mathcal{X}}} g(y_k, \bar{x}) \nu(d\bar{x}) \leq \sqrt{\int_{\bar{\mathcal{X}}} g(y_k, \bar{x}) P_{k-1}(d\bar{x}) \int_{\bar{\mathcal{X}}} g(y_k, \bar{x}) Q_{k-1}(d\bar{x})}.$$

Theorem 20 *Assume that $P_{k-1} \in \bar{\mathcal{P}}_k(\bar{\mathcal{X}})$ and $Q_{k-1} \in \bar{\mathcal{P}}_k(\bar{\mathcal{X}})$. Suppose **either** Assumption ER.H.1 or Assumption ER.H.2 holds. Then the distributions P_k and Q_k^* satisfy*

$$d_H(P_k, Q_k^*) \leq d_H(P_{k-1}, Q_{k-1}).$$

Proof See Appendix D.2. ■

Remark 21 *Theorems 19 and 20 characterize a scenario where we may have $d(P_k, Q_k^*) \leq d(P_{k-1}, Q_{k-1})$, with d being the TV or Hellinger distance.*

This scenario arises when P_{k-1} and Q_{k-1} are relatively similar where the function g takes large values, and both assign high average probability to those regions. Additionally, P_{k-1} and Q_{k-1} should differ more significantly in regions where g is small but both assign, on average, low probability to such regions. It should be noted that both P_{k-1} and Q_{k-1} assigning low average probability to low-g regions does not preclude them from being substantially different in those regions.

Let the symbol "$\otimes$" denote the product of two independent probability measures. Specifically, let ν_1 and ν_2 be two probability measures defined on measurable spaces $(\mathcal{Z}_1, \mathcal{B}(\mathcal{Z}_1))$ and $(\mathcal{Z}_2, \mathcal{B}(\mathcal{Z}_2))$, respectively. The probability measure $\nu_1 \otimes \nu_2$ is defined as

$$\nu_1 \otimes \nu_2(A, B) = \nu_1(A)\nu_2(B), \quad \forall A \in \mathcal{B}(\mathcal{Z}_1), B \in \mathcal{B}(\mathcal{Z}_2).$$

We now present sufficient conditions for error reduction under the 1-Wasserstein distance in inverse problems.

Theorem 22 *In inverse problems, given the data y_k, suppose that $P_{k-1} \in \bar{\mathcal{P}}_k(\mathcal{X}) \bigcap \mathcal{P}_1(\mathcal{X})$ and $Q_{k-1} \in \bar{\mathcal{P}}_k(\mathcal{X}) \bigcap \mathcal{P}_1(\mathcal{X})$. Assume the likelihood model h and the distributions P_{k-1} and Q_{k-1} satisfy*

$$\frac{\mathbb{E}_{(X,X')\sim P_{k-1}\bigotimes Q_{k-1}}[d_{\mathcal{X}}(X,X')h(y_k,X)h(y_k,X')]}{\mathbb{E}_{X\sim P_{k-1}}[h(y_k,X)]\,\mathbb{E}_{X\sim Q_{k-1}}[h(y_k,X)]} \leq \sup_{x_0\in\mathcal{X}}\left|\mathbb{E}_{X\sim P_{k-1}}[d_{\mathcal{X}}(X,x_0)] - \mathbb{E}_{X\sim Q_{k-1}}[d_{\mathcal{X}}(X,x_0)]\right|.$$

Then the distributions P_k and Q_k^* satisfy $W_1(P_k, Q_k^*) \leq W_1(P_{k-1}, Q_{k-1})$.

Proof See Appendix D.3. ■

Finally, we outline the sufficient conditions for error reduction under the 1-Wasserstein distance in state estimation and parameter-state estimation problems.

Theorem 23 *Given the data y_k, assume that the distributions P_{k-1} and Q_{k-1} belong to the space $\bar{\mathcal{P}}_k(\bar{\mathcal{X}}) \bigcap \mathcal{P}_1(\bar{\mathcal{X}})$. Suppose the state transition model T_k, the observation model h_k, and the distributions P_{k-1} and Q_{k-1} satisfy*

$$\mathbb{E}_{(\bar{X},\bar{X}')\sim P_k^-\bigotimes Q_k^-}[d_{\bar{\mathcal{X}}}(\bar{X},\bar{X}')h_k(y_k,\bar{X})h_k(y_k,\bar{X}')] \leq \left(\int_{\bar{\mathcal{X}}} h_k(y_k,\bar{x})d\bar{x}\right)^2 \mathbb{E}_{(\bar{X},\bar{X}')\sim P_k^-\bigotimes Q_k^-}[d_{\bar{\mathcal{X}}}(\bar{X},\bar{X}')],$$

and

$$\mathbb{E}_{(\bar{X},\bar{X}')\sim P_k^-\bigotimes Q_k^-}[d_{\bar{\mathcal{X}}}(\bar{X},\bar{X}')] \leq \sup_{\bar{x}_0\in\bar{\mathcal{X}}}\left|\mathbb{E}_{\bar{X}\sim P_{k-1}}[d_{\bar{\mathcal{X}}}(\bar{X},\bar{x}_0)] - \mathbb{E}_{\bar{X}\sim Q_{k-1}}[d_{\bar{\mathcal{X}}}(\bar{X},\bar{x}_0)]\right|,$$

and

$$\left(\int_{\bar{\mathcal{X}}} h_k(y_k,\bar{x})d\bar{x}\right)^2 \leq \mathbb{E}_{\bar{X}\sim P_k^-}[h_k(y_k,\bar{X})]\,\mathbb{E}_{\bar{X}\sim Q_k^-}[h_k(y_k,\bar{X})],$$

where P_k^- and Q_k^- are probability measures with Lebesgue densities p_k^- and q_k^-, respectively, defined as:

- *in state estimation problems,*

$$p_k^-(x) := \mathbb{E}_{X\sim P_{k-1}}[T_k(x,X)], \quad q_k^-(x) := \mathbb{E}_{X\sim Q_{k-1}}[T_k(x,X)];$$

- *in parameter-state estimation problems,*

$$p_k^-(x,w) := \int_{\mathcal{X}} T_k(x,x_{k-1},w)p_{k-1}(x_{k-1},w)dx_{k-1},$$
$$q_k^-(x,w) := \int_{\mathcal{X}} T_k(x,x_{k-1},w)q_{k-1}(x_{k-1},w)dx_{k-1},$$

where p_{k-1} and q_{k-1} are the Lebesgue densities of P_{k-1} and Q_{k-1}, respectively.

Then, the distributions P_k and Q_k^* satisfy

$$W_1(P_k, Q_k^*) \leq W_1(P_{k-1}, Q_{k-1}).$$

Proof See Appendix D.4. ■

Remark 24 *The results presented in Theorems 19, 20, 22, and 23 apply to any pair of distributions P_{k-1} and Q_{k-1} that satisfy the sufficient conditions specified therein.*

4 Discussion

In this section, we provide discussions of the main theorems. These discussions are carried out under the assumption that the conditions of Theorems 14 and 16 are satisfied.

4.1 Stability of approximate approaches in BSL with respect to the incremental approximation process

In this section, we discuss the stability of approximate approaches for BSL with respect to the incremental approximation process, based on Theorems 16 and 14.

4.1.1 Bounded input, bounded output stability of learning error evolution

The evolution of the learning error $d(Q_k, P_k)$ over the step k can be viewed as a dynamical system with input $d(Q_k, Q_k^*)$. According to Theorems 14 and 16, $d(Q_k, P_k)$ is upper bounded by

$$\begin{aligned} d(Q_k, P_k) &\leq \tilde{C}(y_k, p(y_k \mid \mathcal{Y}_{1:k-1}), k) d(Q_{k-1}, P_{k-1}) + d(Q_k, Q_k^*) \\ &\leq \sum_{i=1}^{k-1} C(\mathcal{Y}_{i+1:k}, p(\mathcal{Y}_{i+1:k} \mid \mathcal{Y}_{1:i}), i) d(Q_i, Q_i^*) + d(Q_k, Q_k^*), \end{aligned}$$

where $\tilde{C}(y_k, p(y_k \mid \mathcal{Y}_{1:k-1}), k)$ is a constant given y_k, $p(y_k \mid \mathcal{Y}_{1:k-1})$, and k. Function C is defined in Table 2. This implies that the learning error system exhibits bounded input, bounded output (BIBO) stability: starting from a distribution Q_{k_0} at an arbitrary step $k_0 < k$, if the input $d(Q_i, Q_i^*)$ remains upper bounded over steps $i \in (k_0, k]$, then the output $d(Q_k, P_k)$ is upper bounded.

4.1.2 Distance between different outputs of an approximate BSL approach

Consider two different sequences of approximate posteriors Q_i and $\tilde{Q}_i, i \leq k$. By Theorem 16 and the triangle inequality satisfied by the distance d, the distance between Q_k and $\tilde{Q}_k$ is bounded by:

$$\begin{aligned} d(Q_k, \tilde{Q}_k) &\leq d(Q_k, P_k) + d(\tilde{Q}_k, P_k) \\ &\leq \sum_{i=1}^{k-1} C(\mathcal{Y}_{i+1:k}, p(\mathcal{Y}_{i+1:k} \mid \mathcal{Y}_{1:i}), i) \left(d(Q_i, Q_i^*) + d(\tilde{Q}_i, \tilde{Q}_i^*) \right) + d(Q_k, Q_k^*) + d(\tilde{Q}_k, \tilde{Q}_k^*). \end{aligned} \tag{23}$$

Given the data $\mathcal{Y}_{1:k}$, the coefficients $C(\mathcal{Y}_{i+1:k}, p(\mathcal{Y}_{i+1:k} \mid \mathcal{Y}_{1:i}), i)$ are fixed for each $i \leq k-1$. Inequality (23) thus indicates that, as long as both sequences of incremental approximation errors, $d(Q_i, Q_i^*)$ and $d(\tilde{Q}_i, \tilde{Q}_i^*), i \leq k$ are upper bounded, the distance between the approximate posteriors Q_k and $\tilde{Q}_k$—representing two different outputs of an approximate BSL method—remains bounded.

4.2 Estimation of learning error bounds

As shown in Equation (22), estimating the learning error bound $d(P_k, Q_k)$ requires evaluating the evidence terms $Z_i(Q_{i-1})$ for $i = 2, 3, \cdots, k$, where Q_{i-1} are the outputs of the learning methods. Since each $Z_i(Q_{i-1})$ involves only a single data point y_i, computing these terms is generally much easier than evaluating the offline learning evidence, which involves the entire batch of data.

In practice, $Z_i(Q_{i-1})$ can often be estimated, for instance, using Monte Carlo methods with samples from Q_{i-1}. This is particularly convenient for methods that represent the approximate posteriors with empirical distributions. While the ability to estimate $Z_i(Q_{i-1})$ implies that one could, in principle, construct Q_i^*, practical BSL methods do not typically do so. Setting $Q_i = Q_i^*$ would lead to increasingly complex representations of Q_i over time, significantly increasing computational costs and reducing inference speed-both of which are critical considerations in BSL applications. For example, one may want to retain a Gaussian form for Q_i throughout the learning process.

In addition to evidence terms, estimating the learning error bounds under the TV and Hellinger distances requires evaluating $C_h(y_i)$ for inverse problems, $C_{Th}(y_i; i)$ for state estimation problems, and $\tilde{C}_{Th}(y_i; i)$ for parameter-state estimation problems.

Given the data y_i, it is usually straightforward to compute $C_h(y_i)$, which is the upper bound of the likelihood. Although $C_{Th}(y_i; i)$ and $\tilde{C}_{Th}(y_i; i)$ may be more difficult to compute directly, they are upper bounded by their respective observation model bounds. These bounds can therefore be used as conservative estimates for $C_{Th}(y_i; i)$ and $\tilde{C}_{Th}(y_i; i)$, although this may result in looser learning error bounds.

In summary, the learning error bounds under the TV and Hellinger distances are estimable in many practical settings. However, estimating the learning error bound under the 1-Wasserstein distance using Equation (22) is generally more challenging, as it requires computing $\|h\|_{\text{Lip}}(y_i)$, $C_{Th}^*(y_i; i)$, and $\tilde{C}_{Th}^*(y_i; i)$, with the latter two being particularly difficult to evaluate in practice. Nevertheless, the inequality $W_1(P_k, Q_k) \leq D d_{TV}(P_k, Q_k)$ (Sprungk (2020); Gibbs and Edward (2002)), where D is defined in Assumption IP-SE-PS.4, provides an alternative way to upper bound 1-Wasserstein learning error using the estimated TV bound.

Finally, it is worth noticing that the bound in Equation (22) can be computed recursively, which is especially suitable for sequential learning applications.

4.3 Inaccurate initial prior

Based on Theorem 14, Theorem 16 can be easily extended to the scenarios in which the true initial prior P_0 is unknown, and an estimated initial prior P_0' is used for inference. In

this case, if we have $P_0' \in \bar{\mathcal{P}}_1(\bar{\mathcal{X}})$, the learning error is bounded by

$$d(Q_k, P_k) \leq \sum_{i=1}^{k-1} C(\mathcal{Y}_{i+1:k}, p(\mathcal{Y}_{i+1:k} \mid \mathcal{Y}_{1:i}), i) d(Q_i, Q_i^*) + d(Q_k, Q_k^*) \\ + C(\mathcal{Y}_{2:k}, p(\mathcal{Y}_{2:k} \mid y_1), 1) \hat{C}(y_1, p(y_1)) d(P_0', P_0),$$

where $Q_1^* = F_1(P_0')$ and $Q_i^* = F_i(Q_{i-1})$ for $i \geq 2$. Here $\hat{C}(y_1, p(y_1))$ is a constant depending on data y_1 and probability density $p(y_1)$. This upper bound grows linearly with the initial prior error $d(P_0', P_0)$. Assuming the incremental approximation errors $d(Q_i, Q_i^*), i \leq k$ are bounded, the learning error remains bounded as long as the initial prior error $d(P_0', P_0)$ is bounded.

4.4 Choice of initial prior and approximate posteriors

In this section, we combine Theorems 14 and 16 with existing results on long-term posterior behavior to provide insights into the choice of initial prior and approximate posteriors in state estimation problems.

Let $F_{m:n}$ denote the composition of the prior-to-posterior maps from step m to n:

$$F_{m:n} := F_n \circ F_{n-1} \circ \cdots \circ F_m, \qquad m < n,$$

where "$\circ$" denotes function composition. For instance, $F_n \circ F_{n-1}(\mu) := F_n(F_{n-1}(\mu))$ for a probability measure μ. Using this notation, the true posterior at step k can be written as $P_k = F_{1:k}(P_0)$ and $P_k = F_{j+1:k}(P_j)$ for any $j < k$. According to Le Gland and Oudjane (2004), given the data $\mathcal{Y}_{j:k}$, if the product of state transition and observation models, $T_i(x_i, x_{i-1})h_i(y_i, x_i)$, is bounded above and below by positive constants for all $i \in [j, k]$, then $d_{TV}(F_{j:k}(\mu), F_{j:k}(\mu'))$ can become arbitrarily small for any μ and μ', when $k - j$ is sufficiently large.

4.4.1 Choice of initial prior

Let P_0' denote an alternative initial prior. Suppose we use P_0' as the true initial prior and approximate the corresponding posterior $P_k' = F_{1:k}(P_0')$ with Q_k at each time step k. Assume $P_i' \in \bar{\mathcal{P}}_{i+1}(\bar{\mathcal{X}})$ for all $i \in [0, k-1]$. Then the learning error under the TV distance satisfies:

$$\begin{aligned} d_{TV}(Q_k, P_k) &\leq d_{TV}(Q_k, P_k') + d_{TV}(P_k', P_k) \\ &\leq \sum_{i=1}^{k-1} C(\mathcal{Y}_{i+1:k}, p'(\mathcal{Y}_{i+1:k} \mid \mathcal{Y}_{1:i}), i) d_{TV}(Q_i, Q_i^*) + d_{TV}(Q_k, Q_k^*) + d_{TV}(P_k', P_k), \end{aligned} \tag{24}$$

where $Q_1^* = F_1(P_0')$ and $Q_i^* = F_i(Q_{i-1})$ for $i \geq 2$. Here the probability density $p'(\mathcal{Y}_{i+1:k} \mid \mathcal{Y}_{1:i})$ is computed under P_0'. The constant C is defined in Table 2 and is proportional to $1/p'(\mathcal{Y}_{i+1:k} \mid \mathcal{Y}_{1:i})$.

If $T_i(x_i, x_{i-1})h_i(y_i, x_i)$ are bounded below and above by positive constants for all $i \in [1, k]$, then as k increases, $d(P_k', P_k)$ will become negligible and the upper bound given in Equation (24) will be dominated by the first two terms.

Comparing Equations (24) and (21), we see that the bound in Equation (24) can be significantly smaller if P_0' leads to either lower incremental approximation errors or higher probability densities $p'(\mathcal{Y}_{i+1:k} \mid \mathcal{Y}_{1:i})$. For instance, smaller incremental approximation errors may arise when P_0' yields a simpler optimization problem in practice. This highlights the following insight: *when the product of the state transition and observation models are bounded above and below, using an alternative, well-chosen initial prior P_0', even if the true initial prior P_0 is available, may improve learning accuracy.*

4.4.2 Choice of approximate posterior Q

By applying Theorem 14 and triangle inequality iteratively, we obtain the following upper bound on the learning error:

$$\begin{aligned} &d(Q_k, P_k) \leq d(Q_k, F_{j+1:k}(Q_j)) + d(F_{j+1:k}(Q_j), P_k) \\ &\leq \sum_{i=j+1}^{k-1} C(\mathcal{Y}_{i+1:k}, p_q(\mathcal{Y}_{i+1:k} \mid \mathcal{Y}_{1:i}), i) d(Q_i, Q_i^*) + d(Q_k, Q_k^*) + d(F_{j+1:k}(Q_j), P_k), \end{aligned} \tag{25}$$

where $p_q(\mathcal{Y}_{i+1:k} \mid \mathcal{Y}_{1:i})$ is the probability density computed using Q_j as the true posterior at step j. Function C is defined in Table 2. Since $P_k = F_{j+1:k}(P_j)$, the bound in Equation (25) will be gradually dominated by the first two terms as k grows, provided that the product of the state transition and observation models is bounded above and below by positive constants.

When selecting Q_j at an earlier step j, we may compare two candidates. One offers a smaller immediate incremental approximation error $d(Q_j, Q_j^*)$. The other has a larger immediate incremental approximation error but may facilitate more accurate approximations at future steps, resulting in smaller incremental approximation errors $d(Q_i, Q_i^*), i > j$. This can occur, for instance, if the candidate belongs to a simpler distribution family that is easier to propagate in subsequent steps.

In such cases, inequality (25) suggests that, in the case where the product of state transition and observation models are bounded above and below, the second choice may yield a lower learning error over time (note that the summation in the upper bound given in inequality (25) starts from $j + 1$).

5 Application example

In this section, we consider a specific class of algorithms for BSL: online variational inference for joint state and parameter estimation. Online VI has emerged in recent years as one of the most popular and effective approaches for sequential inference. However, theoretical analysis of their learning error remains sparse. We leverage Theorem 16 to derive learning error bounds for these methods in specific system settings.

Online VI approaches for joint state and parameter estimation can be categorized into two types:

- Type 1 methods (e.g., Neri et al. (2020); Wang and Gorodetsky (2025)) approximate the joint posterior of state X_k and parameter W, solving the parameter-state estimation problem considered in this paper.

- Type 2 methods (e.g., Campbell et al. (2021); Dowling et al. (2023)) estimate the posterior distribution of state X_k, and compute a point estimate $\hat{w}_k$ of the system parameter.

We refer to these as Type 1 and Type 2 methods, respectively.

5.1 Type 1 online VI methods

Consider the following discrete-time Markov system:

$$\begin{aligned} p(x_k \mid x_{0:k-1}, w) &= p(x_k \mid x_{k-1}, w) = T_k(x_k, x_{k-1}, w), \\ p(y_k \mid x_{0:k}, \mathcal{Y}_{1:k-1}, w) &= p(y_k \mid x_k, w) = p_{\mathcal{N}}(y_k \mid \Phi_k(x_k, w), \Gamma), \end{aligned} \tag{26}$$

where $y_k \in \mathbb{R}^r$ and $p_{\mathcal{N}}(y_k \mid \Phi_k(x_k, w), \Gamma)$ denotes the value of the probability density function the normal distribution with mean $\Phi_k(x_k, w)$ and covariance Γ evaluated at y_k. Assume the initial distribution $P_0(X_0, W)$ is absolutely continuous w.r.t. the Lebesgue measure.

Standard Type 1 methods (e.g., Neri et al. (2020)) compute the approximate posterior Q_k by maximizing the evidence lower bound (ELBO). Let q_k and q_{k-1} denote the pdfs of Q_k and Q_{k-1}, respectively. The ELBO at step k is given by:

$$\mathcal{L}_k(Q_k) := \mathbb{E}_{q_k(x_k,w)}[\log\left(p_{\mathcal{N}}(y_k \mid H_k(x_k, w), \Gamma) q_k^{*-}(x_k, w)\right)] - \mathbb{E}_{q_k(x_k,w)}[\log q_k(x_k, w)],$$

where $q_k^{*-}(x_k, w) := \int_{\mathcal{X}} T_k(x_k, x_{k-1}, w) q_{k-1}(x_{k-1}, w) dx_{k-1}$. The following corollary applies Theorem 16 to bound the learning error $d(P_k, Q_k)$ for standard Type 1 methods applied to system (26).

Corollary 25 *Suppose the ELBO satisfies $\mathcal{L}_i(Q_i) \geq \epsilon_i$[1] for all $i \in [1, k]$. Assume $P_i \in \bar{\mathcal{P}}_{i+1}(\mathcal{X} \times \mathcal{W})$ for all $i \in [0, k-1]$ and $Q_i \in \bar{\mathcal{P}}_{i+1}(\mathcal{X} \times \mathcal{W})$ for all $i \in [1, k-1]$. Also assume that Q_i is absolutely continuous with respect to the Lebesgue measure for all $i \in [1, k]$. Then the learning error $d(P_k, Q_k)$ is bounded as:*

$$\begin{aligned} d(P_k, Q_k) \leq \sum_{j=1}^{k-1} C_{VI}(\mathcal{Y}_{j+1:k}, p(\mathcal{Y}_{j+1:k} \mid \mathcal{Y}_{1:j}), j) \sqrt{-\frac{r}{2}\log(2\pi) - \frac{1}{2}\log(\det(\Gamma)) - \epsilon_j} \\ + \alpha \sqrt{-\frac{r}{2}\log(2\pi) - \frac{1}{2}\log(\det(\Gamma)) - \epsilon_k}, \end{aligned} \tag{27}$$

and

$$\begin{aligned} d(P_k, Q_k) \leq \sum_{j=1}^{k-1} C'_{VI}(\mathcal{Y}_{j+1:k}, Q_{j:k-1}, j) \sqrt{-\frac{r}{2}\log(2\pi) - \frac{1}{2}\log(\det(\Gamma)) - \epsilon_j} \\ + \alpha \sqrt{-\frac{r}{2}\log(2\pi) - \frac{1}{2}\log(\det(\Gamma)) - \epsilon_k}, \end{aligned} \tag{28}$$

for d being the total variation or Hellinger distance.

1. Note that neither the ELBO $\mathcal{L}_i(Q_i)$ nor the threshold ϵ_i is necessarily positive.

Table 5: Definitions of C_{VI}, C'_{VI}, and α for different distances of probability measures.

Distance d	$C_{VI}(\mathcal{Y}_{j+1:k}, p(\mathcal{Y}_{j+1:k} \mid \mathcal{Y}_{1:j}), j)$	$C'_{VI}(\mathcal{Y}_{j+1:k}, Q_{j:k-1}, j)$	α
d_{TV}	$\frac{(2\pi)^{-\frac{r(k-j)}{2}} \det(\Gamma)^{-\frac{k-j}{2}}}{\sqrt{2} p(\mathcal{Y}_{j+1:k} \mid \mathcal{Y}_{1:j})}$	$\frac{(2\pi)^{-\frac{r(k-j)}{2}} \det(\Gamma)^{-\frac{k-j}{2}}}{\sqrt{2} \prod_{i=j+1}^{k} Z_i(Q_{i-1})}$	$\frac{1}{\sqrt{2}}$
d_H	$\frac{2^{(k-j)}(2\pi)^{-\frac{r(k-j)}{4}} \det(\Gamma)^{-\frac{k-j}{4}}}{\sqrt{2p(\mathcal{Y}_{j+1:k} \mid \mathcal{Y}_{1:j})}}$	$\frac{2^{(k-j)}(2\pi)^{-\frac{r(k-j)}{4}} \det(\Gamma)^{-\frac{k-j}{4}}}{\sqrt{2} \prod_{i=j+1}^{k} \sqrt{Z_i(Q_{i-1})}}$	$\frac{1}{\sqrt{2}}$
W_1	$\frac{D(2\pi)^{-\frac{r(k-j)}{2}} \det(\Gamma)^{-\frac{k-j}{2}}}{\sqrt{2} p(\mathcal{Y}_{j+1:k} \mid \mathcal{Y}_{1:j})}$	$\frac{D(2\pi)^{-\frac{r(k-j)}{2}} \det(\Gamma)^{-\frac{k-j}{2}}}{\sqrt{2} \prod_{i=j+1}^{k} Z_i(Q_{i-1})}$	$\frac{D}{\sqrt{2}}$

If in addition the metric space $(\mathcal{X} \times \mathcal{W}, d_{\mathcal{X} \times \mathcal{W}})$ satisfies

$$\sup_{(x,w),(x',w') \in \mathcal{X} \times \mathcal{W}} d_{\mathcal{X} \times \mathcal{W}}((x,w),(x',w')) = D < \infty,$$

then inequalities (27) *and* (28) *also hold with d being the* 1*-Wasserstein distance.*

The definitions of functions C_{VI}, C'_{VI}, and the constant α for each distance are summarized in Table 5.

Proof See Appendix E.1. ■

5.2 Type 2 online VI methods

Type 2 methods solve a different problem: they assume the existence of a true but unknown system parameter $\bar{w}$ and aim to estimate the true posterior P_k, defined as the distribution of the state X_k given all received data $\mathcal{Y}_{1:k}$ under the underlying system with parameter $\bar{w}$, and compute a point estimate $\hat{w}_k$ of the system parameter.

Some Type 2 methods, such as Dowling et al. (2023), assume the observation model is known and only the state transition model is unknown. We adopt this setting and consider the following system:

$$\begin{aligned} p(x_k \mid x_{0:k-1}, w) = p(x_k \mid x_{k-1}, w) = T_k(x_k, x_{k-1}, w), \\ p(y_k \mid x_{0:k}, \mathcal{Y}_{1:k-1}, w) = p(y_k \mid x_k) = p_{\mathcal{N}}(y_k \mid \Phi_k(x_k), \Gamma), \end{aligned} \tag{29}$$

where $y_k \in \mathbb{R}^r$. Assume the initial distribution $P_0(X_0)$ is absolutely continuous w.r.t. the Lebesgue measure. Let q_k denote the pdf of Q_k. Standard Type 2 methods obtain Q_k and $\hat{w}_k$ by maximizing the following ELBO:

$$\mathcal{L}_k(Q_k, \hat{w}_k) := \mathbb{E}_{q_k(x_k)}[\log \left(p_{\mathcal{N}}(y_k \mid \Phi_k(x_k), \Gamma) q^{*-}_{k,\hat{w}_k}(x_k) \right)] - \mathbb{E}_{q_k(x_k)}[\log q_k(x_k)],$$

where $q^{*-}_{k,\hat{w}_k}(x_k) := \int_{\mathcal{X}} T_k(x_k, x_{k-1}, \hat{w}_k) Q_{k-1}(dx_{k-1})$.

By applying Theorem 16 to the state estimation setting considered here, we obtain Corollary 26, which provides an upper bound on the state estimation error $d(P_k, Q_k)$.

Table 6: Definitions of function β for different distances of probability measures. The function $Z_{k,\hat{w}_k}$is defined as $Z_{k,\hat{w}_k}(Q_{k-1}) := \int_{\mathcal{X}} p_{\mathcal{N}}(y_k \mid H_k(x), \Gamma) \left(\int_{\mathcal{X}} T_k(x, x_{k-1}, \hat{w}_k) Q_{k-1}(dx_{k-1})\right) dx$.

Function β	d_{TV}, W_1	d_H
$\beta(Q_{k-1}, y_k, \hat{w}_k, \|\hat{w}_k - \bar{w}\|_m, k)$	$\frac{\sqrt{2}\tilde{C}_{VI}(y_k;k)\|\hat{w}_k-\bar{w}\|_m}{Z_{k,\hat{w}_k}(Q_{k-1})}$	$2\sqrt{\frac{\tilde{C}_{VI}(y_k;k)\|\hat{w}_k-\bar{w}\|_m}{Z_{k,\hat{w}_k}(Q_{k-1})}}$

Assumption VI.1 *At time step k, assume the state transition and observation model of system* (29) *satisfies*

$$g_{Lip}(x;k) := \sup_{\forall x_{k-1}\in\mathcal{X}, w, w'\in\mathcal{W}, w\neq w'} \frac{|T_k(x, x_{k-1}, w) - T_k(x, x_{k-1}, w')|}{\|w - w'\|_m} < \infty, \qquad \forall x \in \mathcal{X},$$

where $\|\cdot\|_m$ is a metric on the parameter space $\mathcal{W}$,
and

$$\tilde{C}_{VI}(y_k;k) := \int_{\mathcal{X}} p_{\mathcal{N}}(y_k \mid \Phi_k(x), \Gamma) g_{Lip}(x;k) dx < \infty.$$

Corollary 26 *Suppose Assumption VI.1 holds for all time steps up to k. Assume the ELBO satisfies $\mathcal{L}_i(Q_i, \hat{w}_i) \geq {\epsilon_i}^2$ for all $i \in [1, k]$. Assume $P_i \in \bar{\mathcal{P}}_{i+1}(\mathcal{X})$[3] for all $i \in [0, k-1]$ and $Q_i \in \bar{\mathcal{P}}_{i+1}(\mathcal{X})$ for all $i \in [1, k-1]$. Also assume that Q_i is absolutely continuous with respect to the Lebesgue measure for all $i \in [1, k]$. Then the state estimation error satisfies*

$$\begin{aligned} d(P_k, Q_k) \leq \sum_{j=1}^{k-1} C'_{VI}(\mathcal{Y}_{j+1:k}, Q_{j:k-1}, j) \Bigg(& \sqrt{-\frac{r}{2}\log(2\pi) - \frac{1}{2}\log(\det(\Gamma)) - \epsilon_j} \\ & + \beta\left(Q_{j-1}, y_j, \hat{w}_j, \|\hat{w}_j - \bar{w}\|_m, j\right)) \\ & + \alpha\left(\sqrt{-\frac{r}{2}\log(2\pi) - \frac{1}{2}\log(\det(\Gamma)) - \epsilon_k} + \beta\left(Q_{k-1}, y_k, \hat{w}_k, \|\hat{w}_k - \bar{w}\|_m, k\right)\right), \end{aligned} \tag{30}$$

with d being the total variation or Hellinger distance.

Assuming further that the metric space $(\mathcal{X}, d_{\mathcal{X}})$ satisfies $\sup_{x,x'\in\mathcal{X}} d_{\mathcal{X}}(x, x') = D < \infty$, then inequality (30) *holds with d being the 1-Wasserstein distance. The functions C'_{VI} and β are defined in Tables 5 and 6, respectively.*

Proof See Appendix E.2. ■

6 Illustrative numerical example

In this section, we present a numerical example to validate the upper bounds established in Theorem 14 for inverse problems and to investigate the following questions:

2. Neither the ELBO $\mathcal{L}_i(Q_i, \hat{w}_i)$ nor ϵ_i is necessarily positive.
3. The set $\bar{\mathcal{P}}_{i+1}(\mathcal{X})$ corresponds to the system with the true parameter $\bar{w}$.

- How tight are the derived bounds?
- How does the tightness of the derived bound on the posterior distance $d(F_k(\mu), F_k(\mu'))$ change when the priors μ and μ' vary?
- As the priors μ and μ' vary, do $d(F_k(\mu), F_k(\mu'))$ and the derived bound on it exhibit similar trends?

As will be shown, the derived bounds can be remarkably tight. To minimize the influence of computational and estimation errors on the validation of the theorem, we consider a simple one-dimensional inverse problem with a linear Gaussian likelihood model and Gaussian initial prior distributions. Under this setting, both the posterior distributions and the Hellinger distance between them admit closed-form expressions. The model is chosen to be one-dimensional to minimize the estimation error of the TV distance.

Consider the following model:

$$Y = aX + \eta, \quad \eta \sim \mathcal{N}(0, 3),$$

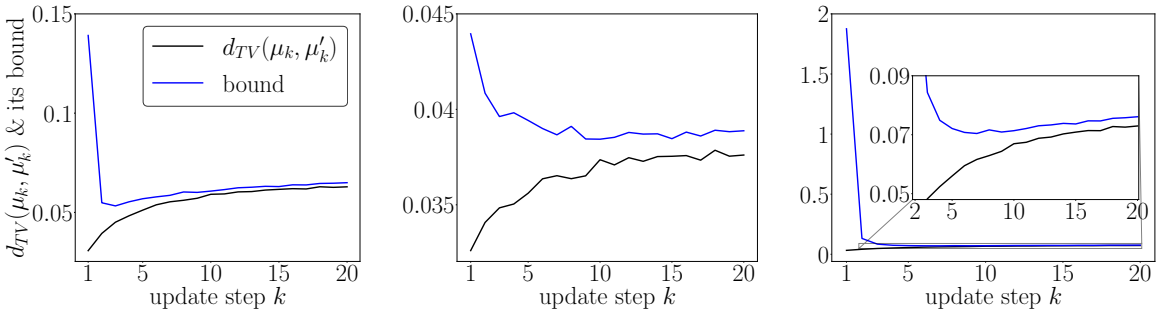
where $a = 1.1$. We generate a single observation $Y = y$ and evaluate three cases with three different sets of initial distributions μ_0 and μ_0'. The first two sets are $\mu_0 = \mathcal{N}(-10, 5)$ and $\mu_0' = \mathcal{N}(8, 5)$, and $\mu_0 = \mathcal{N}(0, 1)$ and $\mu_0' = \mathcal{N}(2, 1)$, representing cases where the initial distributions are significantly different and relatively close, respectively. The third set of initial distributions consists of Gaussian distributions whose means are uniformly sampled from the interval $[-10, 10]$ and variances uniformly sampled from the interval $[0, 5]$. In this case, the variances of the two initial distributions are not necessarily equal. We refer to the cases using the first, second, and third sets of initial distributions as Case 1, Case 2, and Case 3, respectively. We update the corresponding posteriors sequentially over 20 iterations. For each update step k, the updates follow $\mu_k = F_k(\mu_{k-1})$ and $\mu_k' = F_k(\mu_{k-1}')$. The same data y is used at every iteration, i.e., $y_k = y$ for all k.

In this example, the assumptions in Theorem 14 for the TV and Hellinger distances are satisfied. The TV distance between μ_k and μ_k', $d_{TV}(\mu_k, \mu_k')$, and the derived bound on it are shown in Figure 4. The Hellinger distance between μ_k and μ_k', $d_H(\mu_k, \mu_k')$, and its corresponding bound are shown in Figure 5.

Figures 4 and 5 validate the results under the TV and Hellinger distances for inverse problems in Theorem 14. Both the TV and Hellinger bounds become progressively tighter over update steps as more data are assimilated. Notably, the TV bound can be quite sharp. Additionally, while the Hellinger bound tracks the actual distance closely as both decay, the TV bound initially decreases and then increases as $d_{TV}(\mu_k, \mu_k')$ steadily rises.

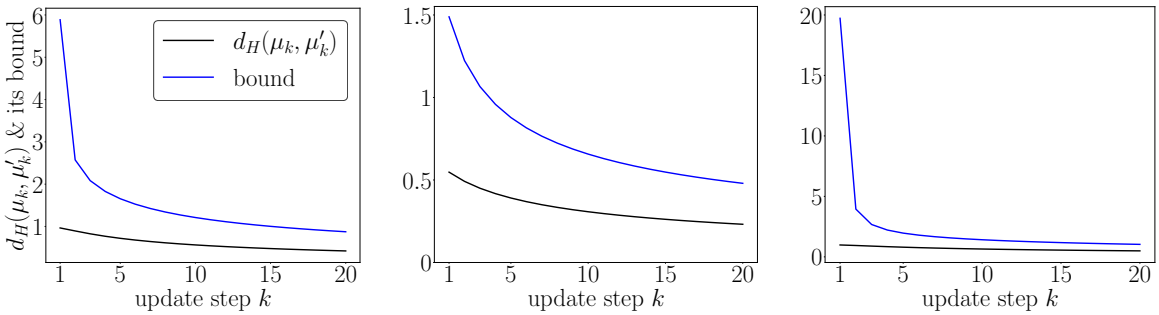
7 Conclusions

In this work, we have presented a rigorous theoretical analysis of general approximate approaches for Bayesian sequential learning (BSL) in the context of inverse problems, state estimation, and parameter-state estimation. By establishing pointwise global Lipschitz continuity of the prior-to-posterior map under the total variation, Hellinger, and 1-Wasserstein distances, we derived two sets of upper bounds on the learning error associated with these methods. The derived bounds demonstrate the stability of these approaches with respect



(a) Case 1 (b) Case 2 (c) Case 3

Figure 4: TV distance between posteriors, $d_{TV}(\mu_k, \mu'_k)$, and its upper bound provided by Theorem 14. Because of the symmetry of the TV distance, Theorem 14 yields the bound $\frac{C_h(y)}{Z_k(\mu_{k-1}) \vee Z_k(\mu'_{k-1})} d_{TV}(\mu_{k-1}, \mu'_{k-1})$. A zoom-in view of steps 2 through 20 is shown in panel (c). The posterior distance is always below the derived bound. As more data are assimilated, the bound becomes increasingly tight.



(a) Case 1 (b) Case 2 (c) Case 3

Figure 5: Hellinger distance between posteriors, $d_H(\mu_k, \mu'_k)$, and its upper bound from Theorem 14. Because the Hellinger distance is symmetric, Theorem 14 provides the bound $2\sqrt{\frac{C_h(y)}{Z_k(\mu_{k-1}) \vee Z_k(\mu'_{k-1})}} d_H(\mu_{k-1}, \mu'_{k-1})$. The posterior distance remains below the derived bound throughout. When $d_H(\mu_k, \mu'_k)$ decreases, the derived bound also declines.

to the incremental approximation process and offer estimable guidance in real-world applications.

Our results mark the first establishment of global Lipschitz stability of the posterior with respect to the prior under the Hellinger and Wasserstein distances, and provide the first general error analysis framework for approximate BSL methods. Moreover, we identified sufficient conditions under which learning error decays, a phenomenon often observed empirically but not previously explored theoretically for general approximate methods in BSL.

To demonstrate the applicability of our framework, we apply it to two different online variational inference methods in the context of joint state and parameter learning.

Our theoretical results yield practical insights. In particular, for state estimation problems, under certain conditions, using a well-designed alternative initial prior—rather than the true one—can lead to improved learning accuracy. Moreover, under the same conditions, our results suggest that selecting an approximate posterior Q_k with a larger immediate incremental approximation error $d(Q_k, Q_k^*)$ may be advantageous if it enables more accurate incremental approximations in subsequent steps, compared to a choice that minimizes the immediate incremental approximation error but leads to less accurate incremental approximations over time.

Acknowledgments and Disclosure of Funding

This work was supported by an NSF CAREER Award, grant number CMMI-2238913.

Appendix A. Comparison with the results in existing literature

A.1 Comparison with the results in Sprungk (2020)

We compare the posterior distance bounds provided by Theorem 14 for inverse problems with those presented in Sprungk (2020). When the assumptions required by the results in Sprungk (2020) are satisfied, the assumptions for Theorem 14 are also satisfied. Under these *same* assumptions, the corresponding upper bounds are summarized in Table 7 below.

Table 7: Upper bounds on the distance between posteriors, $d(F_k(\mu), F_k(\mu'))$. The Lipschitz constant $\|h\|_{\text{Lip}}(y_k)$, metric space diameter D, and evidence function Z_k are defined in Assumption IP.1, Assumption IP-SE-PS.4, and Definition 2.4, respectively. Under the assumptions in Sprungk (2020), it can be readily shown that we have $Z_k(\mu) \vee Z_k(\mu') \leq 1$.

Distance	Ours	Sprungk (2020)
Total variation	$\frac{1}{\boldsymbol{Z_k(\mu)\vee Z_k(\mu')}}\boldsymbol{d_{TV}(\mu,\mu')}$	$\frac{2}{Z_k(\mu)\vee Z_k(\mu')}d_{TV}(\mu,\mu')$
Hellinger	$\boldsymbol{2\sqrt{\frac{1}{Z_k(\mu)\vee Z_k(\mu')}}d_H(\mu,\mu')}$	$\frac{2}{Z_k(\mu)\wedge Z_k(\mu')}d_H(\mu,\mu')$
1-Wasserstein	$\frac{1+2D\|h\|_{\text{Lip}}(y_k)}{Z_k(\mu)\vee Z_k(\mu')}W_1(\mu,\mu')$	$\frac{\left(1+D\|h\|_{\text{Lip}}(y_k)\right)^2}{\left(Z_k(\mu)\wedge Z_k(\mu')\right)^2}W_1(\mu,\mu')$, $\tilde{K}(\mu,\mu')W_1(\mu,\mu')$

where

$$\tilde{K}(\mu,\mu') = (1+D\|h\|_{\text{Lip}}(y_k))\left(\frac{1+\|h\|_{\text{Lip}}(y_k)\frac{|\mu|_{\mathcal{P}_1}}{Z_k(\mu)}}{Z_k(\mu')} \wedge \frac{1+\|h\|_{\text{Lip}}(y_k)\frac{|\mu'|_{\mathcal{P}_1}}{Z_k(\mu')}}{Z_k(\mu)}\right),$$

with $|\mu|_{\mathcal{P}_1}$ defined under the metric space $(\mathcal{X}, d_{\mathcal{X}})$ as

$$|\mu|_{\mathcal{P}^1} := \inf_{x_0\in\mathcal{X}} \int_{\mathcal{X}} d_{\mathcal{X}}(x,x_0)\mu(dx).$$

Table 7 shows that our bounds are tighter than those presented in Sprungk (2020) under the total variation and Hellinger distances. As we have $Z_k(\mu) \vee Z_k(\mu') \leq 1$, the ratio between our bound for the Hellinger distance between the posteriors and that in Sprungk (2020) is less than $\frac{Z_k(\mu)\wedge Z_k(\mu')}{Z_k(\mu)\vee Z_k(\mu')}$. This ratio can be particularly small when the priors differ significantly. For 1-Wasserstein distance, Sprungk (2020) presents two upper bounds. It can be readily proven that our bound is tighter than $\frac{\left(1+D\|h\|_{\text{Lip}}(y_k)\right)^2}{\left(Z_k(\mu)\wedge Z_k(\mu')\right)^2}W_1(\mu,\mu')$. However, for two arbitrary priors μ and μ', a direct comparison between our bound and $\tilde{K}(\mu,\mu')W_1(\mu,\mu')$ is not possible; the relative tightness depends on the specific choice of μ and μ'. It is noteworthy that the bounds in Sprungk (2020) only establish pointwise *local* Lipschitz continuity of the prior-to-posterior map F_k under the Hellinger or 1-Wasserstein distance. In contrast, our bounds enable the establishment of pointwise *global* Lipschitz continuity of F_k under all three distance metrics. It is also noteworthy that although both results prove the global Lipschitz stability of posterior with respect to prior under total variation distance, our upper bound is tighter by a factor of $1/2$.

A.2 Comparison with results in Rebesschini and van Handel (2015); Law et al. (2015); Sanz-Alonso et al. (2023)

For state estimation problems, Rebesschini and van Handel (2015); Law et al. (2015); Sanz-Alonso et al. (2023) prove that given the data y_k, if there exists $\kappa \in (0,1)$ such that the observation model $h_k(y_k, x)$ satisfies

$$\kappa \leq \alpha h_k(y_k, x) \leq \frac{1}{\kappa}, \quad \forall x \in \mathcal{X}, \tag{31}$$

where $\alpha > 0$ is a scaling factor, then the prior-to-posterior map F_k satisfies

$$d_{TV}(F_k(\mu), F_k(\mu')) \leq \frac{2}{\kappa^2} d_{TV}(\mu, \mu'), \quad \forall \mu, \mu' \in \bar{\mathcal{P}}_k(\mathcal{X}), \tag{32}$$

This result establishes global Lipschitz stability under the assumption that the observation model $h_k(y_k, x)$ is bounded above and below by positive constants. While the observation function is often upper bounded in practice, the lower boundedness assumption is much stronger and does not hold in many problems. For example, consider a common setting: a scalar system with observation $Y = aX + V$, where V is a Gaussian noise. In this case, the likelihood is not bounded below by a positive constant.

Next, we show that under the assumption in Equation (31), the upper bound provided by Theorem 14 is half of that given in Equation (32). When Equation (31) holds, the assumption for Theorem 14 under the total variation distance for state estimation problems is satisfied. Theorem 14 yields:

$$d_{TV}(F_k(\mu), F_k(\mu')) \leq \frac{C_{Th}(y_k; k)}{Z_k(\mu) \vee Z_k(\mu')} d_{TV}(\mu, \mu'), \quad \forall \mu, \mu' \in \bar{\mathcal{P}}_k(\mathcal{X}), \tag{33}$$

Under the assumption in Equation (31), $C_{Th}(y_k; k)$ and $Z_k(\mu)$ satisfy

$$C_{Th}(y_k; k) := \sup_{x_{k-1} \in \mathcal{X}} \int_{\mathcal{X}} h_k(y_k, x) T_k(x, x_{k-1}) dx \leq \sup_{x \in \mathcal{X}} h_k(y_k, x) \leq \frac{1}{\alpha \kappa}, \tag{34}$$

and

$$Z_k(\mu) = \int_{\mathcal{X}} h(y_k, x) F_k^-(\mu)(dx) \geq \frac{\kappa}{\alpha} \int_{\mathcal{X}} F_k^-(\mu)(dx) = \frac{\kappa}{\alpha}, \tag{35}$$

where $F_k^-(\mu)$ represents the probability measure obtained by propagating the prior μ through the dynamics. Substituting Equations (34) and (35) into Equation (33) gives

$$d_{TV}(F_k(\mu), F_k(\mu')) \leq \frac{1}{\kappa^2} d_{TV}(\mu, \mu'), \quad \forall \mu, \mu' \in \bar{\mathcal{P}}_k(\mathcal{X}). \tag{36}$$

Thus, the bound given in Equation (36) is exactly $\frac{1}{2}$ of the bound provided by Equation (32).

Appendix B. Proof of Theorem 14

Theorem 14 is a conclusion of Propositions 27, 28, 29, 30, 31, 32, 49, 50, and 51.

B.1 Proof of Theorem 14 under the total variation distance

Theorem 14 under the total variation distance is a conclusion of Propositions 27—29.

Proposition 27 *In inverse problems, given the data y_k, under Assumption IP.1, $\forall \mu, \mu' \in \bar{\mathcal{P}}_k(\mathcal{X})$, the posteriors $F_k(\mu)$ and $F_k(\mu')$ satisfy*

$$d_{TV}(F_k(\mu), F_k(\mu')) \leq \frac{C_h(y_k)}{Z_k(\mu)} d_{TV}(\mu, \mu').$$

Proof Let ν be a σ-finite measure defined on the measurable space $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$. Assume that $\mu \ll \nu$ and $\mu' \ll \nu$. As we have $F_k(\mu) \ll \mu$ and $F_k(\mu') \ll \mu'$, $F_k(\mu)$ and $F_k(\mu')$ are both absolutely continuous w.r.t. ν.

By the definiton of total variation distance, we have

$$\begin{aligned}
&d_{TV}(F_k(\mu), F_k(\mu')) \\
&\quad = \frac{1}{2}\int_{\mathcal{X}} \left| \frac{h(y_k,x)}{Z_k(\mu)} \cdot \frac{d\mu}{d\nu}(x) - \frac{h(y_k,x)}{Z_k(\mu')} \cdot \frac{d\mu'}{d\nu}(x) \right| \nu(dx) \\
&\quad \leq \frac{1}{2}\int_{\mathcal{X}} \left| \frac{h(y_k,x)}{Z_k(\mu)} \cdot \frac{d\mu}{d\nu}(x) - \frac{h(y_k,x)}{Z_k(\mu)} \cdot \frac{d\mu'}{d\nu}(x) \right| \nu(dx) \\
&\qquad + \frac{1}{2}\int_{\mathcal{X}} \left| \frac{h(y_k,x)}{Z_k(\mu)} \cdot \frac{d\mu'}{d\nu}(x) - \frac{h(y_k,x)}{Z_k(\mu')} \cdot \frac{d\mu'}{d\nu}(x) \right| \nu(dx). \qquad (37)
\end{aligned}$$

The first term on the right-hand side in Equation (37) can be expressed as:

$$\begin{aligned}
&\frac{1}{2}\int_{\mathcal{X}} \left| \frac{h(y_k,x)}{Z_k(\mu)} \cdot \frac{d\mu}{d\nu}(x) - \frac{h(y_k,x)}{Z_k(\mu)} \cdot \frac{d\mu'}{d\nu}(x) \right| \nu(dx) \\
&\qquad = \frac{1}{2Z_k(\mu)} \int_{\mathcal{X}} h(y_k,x) \left| \frac{d\mu}{d\nu}(x) - \frac{d\mu'}{d\nu}(x) \right| \nu(dx).
\end{aligned}$$

We can write the second term on the right-hand side in Equation (37) as:

$$\begin{aligned}
&\frac{1}{2}\int_{\mathcal{X}} \left| \frac{h(y_k,x)}{Z_k(\mu)} \cdot \frac{d\mu'}{d\nu}(x) - \frac{h(y_k,x)}{Z_k(\mu')} \cdot \frac{d\mu'}{d\nu}(x) \right| \nu(dx) \\
&\qquad = \frac{1}{2} \left| \frac{1}{Z_k(\mu)} - \frac{1}{Z_k(\mu')} \right| \int_{\mathcal{X}} h(y_k,x) \cdot \frac{d\mu'}{d\nu}(x) \nu(dx).
\end{aligned}$$

Note that

$$\int_{\mathcal{X}} h(y_k,x) \cdot \frac{d\mu'}{d\nu}(x) \nu(dx) = Z_k(\mu').$$

Therefore the second term in Equation (37) can be reformulated as:

$$\frac{1}{2}\int_{\mathcal{X}} \left| \frac{h(y_k,x)}{Z_k(\mu)} \cdot \frac{d\mu'}{d\nu}(x) - \frac{h(y_k,x)}{Z_k(\mu')} \cdot \frac{d\mu'}{d\nu}(x) \right| \nu(dx) = \frac{|Z_k(\mu) - Z_k(\mu')|}{2Z_k(\mu)}.$$

It follows that

$$d_{TV}(F_k(\mu), F_k(\mu')) \leq \frac{1}{2Z_k(\mu)} \left(\int_{\mathcal{X}} h(y_k, x) \left| \frac{d\mu}{d\nu}(x) - \frac{d\mu'}{d\nu}(x) \right| \nu(dx) + \left| Z_k(\mu) - Z_k(\mu') \right| \right). \tag{38}$$

Similarly, we can prove that

$$d_{TV}(F_k(\mu), F_k(\mu')) \leq \frac{1}{2Z_k(\mu')} \left(\int_{\mathcal{X}} h(y_k, x) \left| \frac{d\mu}{d\nu}(x) - \frac{d\mu'}{d\nu}(x) \right| \nu(dx) + \left| Z_k(\mu) - Z_k(\mu') \right| \right). \tag{39}$$

Combining Equations (38) and (39) gives

$$d_{TV}(F_k(\mu), F_k(\mu')) \leq \frac{1}{2(Z_k(\mu) \vee Z_k(\mu'))} \left(\int_{\mathcal{X}} h(y_k, x) \left| \frac{d\mu}{d\nu}(x) - \frac{d\mu'}{d\nu}(x) \right| \nu(dx) + \left| Z_k(\mu) - Z_k(\mu') \right| \right). \tag{40}$$

Define a space $\mathcal{X}^*$ as:

$$\mathcal{X}^* := \begin{cases} \{x \in \mathcal{X} : \frac{d\mu}{d\nu}(x) \geq \frac{d\mu'}{d\nu}(x)\}, & \text{if } Z_k(\mu) \geq Z_k(\mu'), \\ \{x \in \mathcal{X} : \frac{d\mu}{d\nu}(x) \leq \frac{d\mu'}{d\nu}(x)\}, & \text{otherwise.} \end{cases}$$

Suppose we have $Z_k(\mu) \geq Z_k(\mu')$. Then the space $\mathcal{X}^*$ is given by

$$\mathcal{X}^* = \{x \in \mathcal{X} : \frac{d\mu}{d\nu}(x) - \frac{d\mu'}{d\nu}(x) \geq 0\}.$$

Equation (40) becomes

$$d_{TV}(F_k(\mu), F_k(\mu')) \leq \frac{1}{2(Z_k(\mu) \vee Z_k(\mu'))} \left(\int_{\mathcal{X}} h(y_k, x) \left| \frac{d\mu}{d\nu}(x) - \frac{d\mu'}{d\nu}(x) \right| \nu(dx) + Z_k(\mu) - Z_k(\mu') \right). \tag{41}$$

According to Radon-Nikodym theorem, $\frac{d\mu}{d\nu}$ and $\frac{d\mu'}{d\nu}$, as Radon-Nikodym derivatives, are $(\mathcal{B}(\mathcal{X}), \mathcal{B}(\mathbb{R}))$-measurable. Therefore, the function $\frac{d\mu}{d\nu} - \frac{d\mu'}{d\nu}$ is also $(\mathcal{B}(\mathcal{X}), \mathcal{B}(\mathbb{R}))$-measurable. Since $[0, +\infty)$ belongs to $\mathcal{B}(\mathbb{R})$, the set $\mathcal{X}^*$ belongs to $\mathcal{B}(\mathcal{X})$. Then the term $Z_k(\mu) - Z_k(\mu')$ can be written as

$$\begin{aligned} Z_k(\mu) - Z_k(\mu') = & \int_{\mathcal{X}^*} h(y_k, x) \left(\frac{d\mu}{d\nu}(x) - \frac{d\mu'}{d\nu}(x) \right) \nu(dx) \\ & + \int_{\mathcal{X} \setminus \mathcal{X}^*} h(y_k, x) \left(\frac{d\mu}{d\nu}(x) - \frac{d\mu'}{d\nu}(x) \right) \nu(dx). \end{aligned} \tag{42}$$

The term $\int_{\mathcal{X}} h(y_k,x)\left|\frac{d\mu}{d\nu}(x)-\frac{d\mu'}{d\nu}(x)\right|\nu(dx)$ in Equation (41) can be expressed as

$$\int_{\mathcal{X}} h(y_k,x)\left|\frac{d\mu}{d\nu}(x)-\frac{d\mu'}{d\nu}(x)\right|\nu(dx) = \int_{\mathcal{X}^*} h(y_k,x)\left(\frac{d\mu}{d\nu}(x)-\frac{d\mu'}{d\nu}(x)\right)\nu(dx) + \int_{\mathcal{X}\setminus\mathcal{X}^*} h(y_k,x)\left(\frac{d\mu'}{d\nu}(x)-\frac{d\mu}{d\nu}(x)\right)\nu(dx). \tag{43}$$

Substitute Equation (42) and (43) into Equation (41). We obtain

$$d_{TV}(F_k(\mu),F_k(\mu')) \leq \frac{1}{Z_k(\mu)\vee Z_k(\mu')}\int_{\mathcal{X}^*} h(y_k,x)\left|\frac{d\mu}{d\nu}(x)-\frac{d\mu'}{d\nu}(x)\right|\nu(dx). \tag{44}$$

We can write the total variation distance between the two priors μ and μ' as

$$d_{TV}(\mu,\mu') = \frac{1}{2}\left(\int_{\mathcal{X}^*}\left(\frac{d\mu}{d\nu}(x)-\frac{d\mu'}{d\nu}(x)\right)\nu(dx) + \int_{\mathcal{X}\setminus\mathcal{X}^*}\left(\frac{d\mu'}{d\nu}(x)-\frac{d\mu}{d\nu}(x)\right)\nu(dx)\right).$$

Note that

$$\int_{\mathcal{X}}\frac{d\mu}{d\nu}(x)\nu(dx) = 1 = \int_{\mathcal{X}}\frac{d\mu'}{d\nu}(x)\nu(dx).$$

Therefore, we have

$$\int_{\mathcal{X}^*}\frac{d\mu}{d\nu}(x)\nu(dx) + \int_{\mathcal{X}\setminus\mathcal{X}^*}\frac{d\mu}{d\nu}(x)\nu(dx) = \int_{\mathcal{X}^*}\frac{d\mu'}{d\nu}(x)\nu(dx) + \int_{\mathcal{X}\setminus\mathcal{X}^*}\frac{d\mu'}{d\nu}(x)\nu(dx).$$

It follows that

$$\int_{\mathcal{X}^*}\left(\frac{d\mu}{d\nu}(x)-\frac{d\mu'}{d\nu}(x)\right)\nu(dx) = \int_{\mathcal{X}\setminus\mathcal{X}^*}\left(\frac{d\mu'}{d\nu}(x)-\frac{d\mu}{d\nu}(x)\right)\nu(dx),$$

and we can re-write $d_{TV}(\mu,\mu')$ as

$$d_{TV}(\mu,\mu') = \int_{\mathcal{X}^*}\left|\frac{d\mu}{d\nu}(x)-\frac{d\mu'}{d\nu}(x)\right|\nu(dx). \tag{45}$$

Similarly, we can prove that Equations (44) and (45) hold when $Z_k(\mu) < Z_k(\mu')$.
Given Equations (44) and (45), under Assumption IP.1, we have

$$\begin{aligned} d_{TV}(F_k(\mu),F_k(\mu')) &\leq \frac{C_h(y_k)}{Z_k(\mu)\vee Z_k(\mu')}\int_{\mathcal{X}^*}\left(\frac{d\mu}{d\nu}(x)-\frac{d\mu'}{d\nu}(x)\right)\nu(dx) \\ &= \frac{C_h(y_k)}{Z_k(\mu)\vee Z_k(\mu')}d_{TV}(\mu,\mu') \\ &\leq \frac{C_h(y_k)}{Z_k(\mu)}d_{TV}(\mu,\mu') \end{aligned}$$

■

Proposition 28 *In state estimation problems, given the data y_k, under Assumption SE.1, $\forall \mu, \mu' \in \bar{\mathcal{P}}_k(\mathcal{X})$, the posteriors $F_k(\mu)$ and $F_k(\mu')$ satisfy*

$$d_{TV}(F_k(\mu), F_k(\mu')) \leq \frac{C_{Th}(y_k;k)}{Z_k(\mu)} d_{TV}(\mu, \mu').$$

Proof Let ν be a σ-finite measure defined on the measurable space $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$ such that $\mu \ll \nu$ and $\mu' \ll \nu$. Let π and π' denote the Radon-Nikodym derivatives $\frac{d\mu}{d\nu}$ and $\frac{d\mu'}{d\nu}$, respectively. By the definition of F_k for state estimation problems, $d_{TV}(F_k(\mu), F_k(\mu'))$ can be written as

$$\begin{aligned}
d_{TV}(F_k(\mu), F_k(\mu')) &= \frac{1}{2}\int_{\mathcal{X}} \left| \frac{h_k(y_k,x)}{Z_k(\mu)} \int_{\mathcal{X}} T_k(x, x_{k-1}) \mu(dx_{k-1}) \right. \\
&\quad \left. + \frac{h_k(y_k,x)}{Z_k(\mu)} \int_{\mathcal{X}} T_k(x, x_{k-1}) \mu'(dx_{k-1}) - \frac{h_k(y_k,x)}{Z_k(\mu')} \int_{\mathcal{X}} T_k(x, x_{k-1}) \mu'(dx_{k-1}) \right| dx \\
&\leq \frac{1}{2}\int_{\mathcal{X}} \left| \frac{h_k(y_k,x)}{Z_k(\mu)} \int_{\mathcal{X}} T_k(x, x_{k-1}) \mu(dx_{k-1}) - \frac{h_k(y_k,x)}{Z_k(\mu)} \int_{\mathcal{X}} T_k(x, x_{k-1}) \mu'(dx_{k-1}) \right| dx \\
&\quad + \frac{1}{2}\int_{\mathcal{X}} \left| \frac{h_k(y_k,x)}{Z_k(\mu)} \int_{\mathcal{X}} T_k(x, x_{k-1}) \mu'(dx_{k-1}) \right. \\
&\quad \left. - \frac{h_k(y_k,x)}{Z_k(\mu')} \int_{\mathcal{X}} T_k(x, x_{k-1}) \mu'(dx_{k-1}) \right| dx \\
&= \frac{1}{2Z_k(\mu)} \int_{\mathcal{X}} \left| h_k(y_k,x) \int_{\mathcal{X}} T_k(x, x_{k-1}) \left(\mu(dx_{k-1}) - \mu'(dx_{k-1}) \right) \right| dx \\
&\quad + \frac{1}{2} \left| \frac{1}{Z_k(\mu)} - \frac{1}{Z_k(\mu')} \right| \int_{\mathcal{X}} h_k(y_k,x) \left(\int_{\mathcal{X}} T_k(x, x_{k-1}) \mu'(dx_{k-1}) \right) dx. \qquad (46)
\end{aligned}$$

The first term on the right-hand side of Equation (46) satisfies

$$\begin{aligned}
&\frac{1}{2Z_k(\mu)} \int_{\mathcal{X}} \left| h_k(y_k,x) \int_{\mathcal{X}} T_k(x, x_{k-1}) \left(\mu(dx_{k-1}) - \mu'(dx_{k-1}) \right) \right| dx \\
&\quad \leq \frac{1}{2Z_k(\mu)} \int_{\mathcal{X}} h_k(y_k,x) \left(\int_{\mathcal{X}} T_k(x, x_{k-1}) \left| \pi(x_{k-1}) - \pi'(x_{k-1}) \right| \nu(dx_{k-1}) \right) dx \\
&\quad = \frac{1}{2Z_k(\mu)} \int_{\mathcal{X}} \left(\int_{\mathcal{X}} h_k(y_k,x) T_k(x, x_{k-1}) dx \right) \left| \pi(x_{k-1}) - \pi'(x_{k-1}) \right| \nu(dx_{k-1}), \qquad (47)
\end{aligned}$$

where first equality is obtained using Tonelli's theorem.
Note that we have

$$\int_{\mathcal{X}} h_k(y_k,x) \left(\int_{\mathcal{X}} T_k(x, x_{k-1}) \mu'(dx_{k-1}) \right) dx = Z_k(\mu').$$

Hence the second term on the right-hand side of Equation (46) can be written as

$$\begin{aligned}
&\frac{1}{2} \left| \frac{1}{Z_k(\mu)} - \frac{1}{Z_k(\mu')} \right| \int_{\mathcal{X}} h_k(y_k,x) \left(\int_{\mathcal{X}} T_k(x, x_{k-1}) \mu'(dx_{k-1}) \right) dx \\
&\quad = \frac{|Z_k(\mu) - Z_k(\mu')|}{2Z_k(\mu)Z_k(\mu')} Z_k(\mu') = \frac{|Z_k(\mu) - Z_k(\mu')|}{2Z_k(\mu)}. \qquad (48)
\end{aligned}$$

Plug Equations (47) and (48) into (46). We obtain

$$\begin{aligned} &d_{TV}(F_k(\mu), F_k(\mu')) \\ &\quad \leq \frac{1}{2Z_k(\mu)} \left(\int_{\mathcal{X}} \left(\int_{\mathcal{X}} h_k(y_k, x) T_k(x, x_{k-1}) dx \right) \left| \pi(x_{k-1}) - \pi'(x_{k-1}) \right| \nu(dx_{k-1}) \right. \\ &\qquad \left. + |Z_k(\mu) - Z_k(\mu')| \right). \end{aligned} \tag{49}$$

Similarly, we can show that $d_{TV}(F_k(\mu), F_k(\mu'))$ also satisfies

$$\begin{aligned} &d_{TV}(F_k(\mu), F_k(\mu')) \\ &\quad \leq \frac{1}{2Z_k(\mu')} \left(\int_{\mathcal{X}} \left(\int_{\mathcal{X}} h_k(y_k, x) T_k(x, x_{k-1}) dx \right) \left| \pi(x_{k-1}) - \pi'(x_{k-1}) \right| \nu(dx_{k-1}) \right. \\ &\qquad \left. + |Z_k(\mu) - Z_k(\mu')| \right). \end{aligned} \tag{50}$$

Combining Equation (49) and (50) gives

$$\begin{aligned} d_{TV}(F_k(\mu), F_k(\mu')) \leq \frac{1}{2(Z_k(\mu) \vee Z_k(\mu'))} &\left(\int_{\mathcal{X}} \left(\int_{\mathcal{X}} h_k(y_k, x) T_k(x, x_{k-1}) dx \right) \right. \\ &\left. \cdot \left| \pi(x_{k-1}) - \pi'(x_{k-1}) \right| \nu(dx_{k-1}) + |Z_k(\mu) - Z_k(\mu')| \right). \end{aligned} \tag{51}$$

By Tonelli's theorem, we can write the evidence terms $Z_k(\mu)$ and $Z_k(\mu')$ as:

$$\begin{aligned} Z_k(\mu) &= \int_{\mathcal{X}} h_k(y_k, x) \left(\int_{\mathcal{X}} T_k(x, x_{k-1}) \pi(x_{k-1}) \nu(dx_{k-1}) \right) dx_k \\ &= \int_{\mathcal{X}} \left(\int_{\mathcal{X}} h_k(y_k, x) T_k(x, x_{k-1}) dx \right) \pi(x_{k-1}) \nu(dx_{k-1}), \end{aligned} \tag{52}$$

and

$$\begin{aligned} Z_k(\mu') &= \int_{\mathcal{X}} h_k(y_k, x) \left(\int_{\mathcal{X}} T_k(x, x_{k-1}) \pi'(x_{k-1}) \nu(dx_{k-1}) \right) dx \\ &= \int_{\mathcal{X}} \left(\int_{\mathcal{X}} h_k(y_k, x) T_k(x, x_{k-1}) dx \right) \pi'(x_{k-1}) \nu(dx_{k-1}). \end{aligned} \tag{53}$$

Define a space $\mathcal{X}^*$ as:

$$\mathcal{X}^* := \begin{cases} \{x \in \mathcal{X} : \pi(x) \geq \pi'(x)\}, & \text{if } Z_k(\mu) \geq Z_k(\mu'), \\ \{x \in \mathcal{X} : \pi(x) \leq \pi'(x)\}, & \text{otherwise.} \end{cases}$$

Following an argument analogous to that in the proof of Proposition 27, we obtain

$$\begin{aligned} d_{TV}(F_k(\mu), F_k(\mu')) \leq \frac{1}{Z_k(\mu) \vee Z_k(\mu')} &\int_{\mathcal{X}^*} \left(\int_{\mathcal{X}} h_k(y_k, x) T_k(x, x_{k-1}) dx \right) \\ &\cdot |\pi(x_{k-1}) - \pi'(x_{k-1})| \nu(dx_{k-1}), \end{aligned} \tag{54}$$

and

$$d_{TV}(\mu, \mu') = \int_{\mathcal{X}^*} |\pi(x_{k-1}) - \pi'(x_{k-1})| \nu(dx_{k-1}). \tag{55}$$

Combining Equations (54) and (55) with Assumption SE.1 gives

$$\begin{aligned} d_{TV}(F_k(\mu), F_k(\mu')) \leq & \frac{C_{Th}(y_k;k)}{Z_k(\mu) \vee Z_k(\mu')} d_{TV}(\mu, \mu') \\ \leq & \frac{C_{Th}(y_k;k)}{Z_k(\mu)} d_{TV}(\mu, \mu'). \end{aligned}$$

■

Proposition 29 *In parameter-state estimation problems, given the data y_k, under Assumption PS.1, $\forall \mu, \mu' \in \bar{\mathcal{P}}_k(\mathcal{X} \times \mathcal{W})$, the posteriors $F_k(\mu)$ and $F_k(\mu')$ satisfy*

$$d_{TV}(F_k(\mu), F_k(\mu')) \leq \frac{\tilde{C}_{Th}(y_k;k)}{Z_k(\mu)} d_{TV}(\mu, \mu').$$

Proof Let π and π' denote the densities of μ and μ' with respect to the Lebesgue measure, respectively. By the definition of F_k for parameter-state estimation problems, $d_{TV}(F_k(\mu), F_k(\mu'))$ can be written as

$$\begin{aligned} & d_{TV}(F_k(\mu), F_k(\mu')) \\ & = \frac{1}{2} \int_{\mathcal{X}\times\mathcal{W}} \left| \frac{h_k(y_k, x, w)}{Z_k(\mu)} \int_{\mathcal{X}} T_k(x, x_{k-1}, w) \pi(x_{k-1}, w) dx_{k-1} \right. \\ & \qquad \left. - \frac{h_k(y_k, x, w)}{Z_k(\mu')} \int_{\mathcal{X}} T_k(x, x_{k-1}, w) \pi'(x_{k-1}, w) dx_{k-1} \right| dx dw \\ & \leq \frac{1}{2} \int_{\mathcal{X}\times\mathcal{W}} \left| \frac{h_k(y_k, x, w)}{Z_k(\mu)} \int_{\mathcal{X}} T_k(x, x_{k-1}, w) \pi(x_{k-1}, w) dx_{k-1} \right. \\ & \qquad \left. - \frac{h_k(y_k, x, w)}{Z_k(\mu)} \int_{\mathcal{X}} T_k(x, x_{k-1}, w) \pi'(x_{k-1}, w) dx_{k-1} \right| dx dw \\ & \qquad + \frac{1}{2} \int_{\mathcal{X}\times\mathcal{W}} \left| \frac{h_k(y_k, x, w)}{Z_k(\mu)} \int_{\mathcal{X}} T_k(x, x_{k-1}, w) \pi'(x_{k-1}, w) dx_{k-1} \right. \\ & \qquad \left. - \frac{h_k(y_k, x, w)}{Z_k(\mu')} \int_{\mathcal{X}} T_k(x, x_{k-1}, w) \pi'(x_{k-1}, w) dx_{k-1} \right| dx dw \end{aligned} \tag{56}$$

The first integral on the right-hand side of Equation (56) satisfies

$$\begin{aligned} & \frac{1}{2} \int_{\mathcal{X}\times\mathcal{W}} \left| \frac{h_k(y_k, x, w)}{Z_k(\mu)} \int_{\mathcal{X}} T_k(x, x_{k-1}, w) \pi(x_{k-1}, w) dx_{k-1} \right. \\ & \qquad \left. - \frac{h_k(y_k, x, w)}{Z_k(\mu)} \int_{\mathcal{X}} T_k(x, x_{k-1}, w) \pi'(x_{k-1}, w) dx_{k-1} \right| dx dw \\ & \leq \frac{1}{2 Z_k(\mu)} \int_{\mathcal{X}\times\mathcal{W}} h_k(y_k, x, w) \\ & \qquad \cdot \left(\int_{\mathcal{X}} T_k(x, x_{k-1}, w) \left| \pi(x_{k-1}, w) - \pi'(x_{k-1}, w) \right| dx_{k-1} \right) dx dw. \end{aligned}$$

According Tonelli's theorem, we have

$$\begin{aligned}&\int_{\mathcal{X}\times\mathcal{W}} h_k(y_k,x,w)\left(\int_{\mathcal{X}} T_k(x,x_{k-1},w)\Big|\pi(x_{k-1},w)-\pi'(x_{k-1},w)\Big|dx_{k-1}\right)dxdw\\ &\quad=\int_{\mathcal{X}\times\mathcal{W}}\left(\int_{\mathcal{X}} h_k(y_k,x,w)T_k(x,x_{k-1},w)dx\right)\Big|\pi(x_{k-1},w)-\pi'(x_{k-1},w)\Big|dx_{k-1}dw.\end{aligned}$$

The second integral on the right-hand side of Equation (56) can be written as

$$\begin{aligned}&\frac{1}{2}\int_{\mathcal{X}\times\mathcal{W}}\left|\frac{h_k(y_k,x,w)}{Z_k(\mu)}\int_{\mathcal{X}} T_k(x,x_{k-1},w)\pi'(x_{k-1},w)dx_{k-1}\right.\\ &\qquad\left.-\frac{h_k(y_k,x,w)}{Z_k(\mu')}\int_{\mathcal{X}} T_k(x,x_{k-1},w)\pi'(x_{k-1},w)dx_{k-1}\right|dxdw\\ &=\frac{1}{2}\left|\frac{1}{Z_k(\mu)}-\frac{1}{Z_k(\mu')}\right|\\ &\qquad\cdot\int_{\mathcal{X}\times\mathcal{W}} h_k(y_k,x,w)\left(\int_{\mathcal{X}} T_k(x,x_{k-1},w)\pi'(x_{k-1},w)dx_{k-1}\right)dxdw\\ &=\frac{|Z_k(\mu)-Z_k(\mu')|}{2Z_k(\mu)Z_k(\mu')}Z_k(\mu')=\frac{|Z_k(\mu)-Z_k(\mu')|}{2Z_k(\mu)}.\end{aligned}$$

By Tonelli's theorem, we can write the evidence term $Z_k(\mu)$ as:

$$\begin{aligned}Z_k(\mu)&=\int_{\mathcal{X}\times\mathcal{W}} h_k(y_k,x,w)\left(\int_{\mathcal{X}} T_k(x,x_{k-1},w)\pi(x_{k-1},w)dx_{k-1}\right)dxdw\\ &=\int_{\mathcal{X}\times\mathcal{W}}\left(\int_{\mathcal{X}} h_k(y_k,x,w)T_k(x,x_{k-1},w)dx\right)\pi(x_{k-1},w)dx_{k-1}dw.\end{aligned}\tag{57}$$

Similarly, the evidence term $Z_k(\mu')$ can be expressed as:

$$Z_k(\mu')=\int_{\mathcal{X}\times\mathcal{W}}\left(\int_{\mathcal{X}} h_k(y_k,x,w)T_k(x,x_{k-1},w)dx\right)\pi'(x_{k-1},w)dx_{k-1}dw.\tag{58}$$

Define a space $\mathcal{S}^*$ as:

$$\mathcal{S}^*:=\begin{cases}\{(x,w)\in\mathcal{X}\times\mathcal{W}:\pi(x,w)\geq\pi'(x,w)\}, & \text{if } Z_k(\mu)\geq Z_k(\mu'),\\ \{(x,w)\in\mathcal{X}\times\mathcal{W}:\pi(x,w)\leq\pi'(x,w)\}, & \text{otherwise.}\end{cases}$$

Analogous to the proof of Proposition 27, the following can be proved:

$$\begin{aligned}d_{TV}(F_k(\mu),F_k(\mu'))\leq\frac{1}{Z_k(\mu)\vee Z_k(\mu')}&\int_{\mathcal{S}^*}\left(\int_{\mathcal{X}} h_k(y_k,x,w)T_k(x,x_{k-1},w)dx\right)\\ &\cdot|\pi(x_{k-1},w)-\pi'(x_{k-1},w)|dx_{k-1}dw,\end{aligned}\tag{59}$$

and

$$d_{TV}(\mu,\mu')=\int_{\mathcal{S}^*}|\pi(x_{k-1},w)-\pi'(x_{k-1},w)|dx_{k-1}dw.\tag{60}$$

Combining Equations (59) and (60) with Assumption PS.1 yields:

$$\begin{aligned} d_{TV}(F_k(\mu), F_k(\mu')) \leq & \frac{\tilde{C}_{Th}(y_k;k)}{Z_k(\mu) \vee Z_k(\mu')} d_{TV}(\mu, \mu') \\ \leq & \frac{\tilde{C}_{Th}(y_k;k)}{Z_k(\mu)} d_{TV}(\mu, \mu'). \end{aligned}$$

■

B.2 Proof of Theorem 14 under Hellinger distance

Theorem 14 under the Hellinger distance is a conclusion of Propositions 30—32.

Proposition 30 *In inverse problems, given the data y_k, under Assumption IP.1, $\forall \mu, \mu' \in \bar{\mathcal{P}}_k(\mathcal{X})$, the posteriors $F_k(\mu)$ and $F_k(\mu')$ satisfy*

$$d_H(F_k(\mu), F_k(\mu')) \leq 2\sqrt{\frac{C_h(y_k)}{Z_k(\mu)}} d_H(\mu, \mu').$$

Proposition 31 *In state estimation problems, given the data y_k, under Assumption SE.1, $\forall \mu, \mu' \in \bar{\mathcal{P}}_k(\mathcal{X})$, the posteriors $F_k(\mu)$ and $F_k(\mu')$ satisfy*

$$d_H(F_k(\mu), F_k(\mu')) \leq 2\sqrt{\frac{C_{Th}(y_k;k)}{Z_k(\mu)}} d_H(\mu, \mu').$$

Proposition 32 *In parameter-state estimation problems, given the data y_k, under Assumption PS.1, $\forall \mu, \mu' \in \bar{\mathcal{P}}_k(\mathcal{X} \times \mathcal{W})$, the posteriors $F_k(\mu)$ and $F_k(\mu')$ satisfy*

$$d_H(F_k(\mu), F_k(\mu')) \leq 2\sqrt{\frac{\tilde{C}_{Th}(y_k;k)}{Z_k(\mu)}} d_H(\mu, \mu').$$

To prove Propositions 30— 32, we first present a supporting proposition, Proposition 35, which aids in the proof of Propositions 30— 32.

B.2.1 Supporting proposition for the proof of Theorem 14 under the Hellinger distance: Proposition 35

Proposition 35 is used to prove Propositions 30— 32. In this section, we provide Proposition 35 and its proof.

Before presenting Proposition 35, we first define distances for scaled probability measures. Let $\tilde{\mathcal{P}}(\mathcal{Z})$ denote the set of all scaled probability measures defined on $(\mathcal{Z}, \mathcal{B}(\mathcal{Z}))$, i.e.,

$$\tilde{\mathcal{P}}(\mathcal{Z}) = \{\tilde{\mu} \mid \tilde{\mu}(\epsilon) = c\mu(\epsilon), c \in \mathbb{R}_+, c < +\infty, \mu \in \mathcal{P}(\mathcal{Z}), \epsilon \in \mathcal{B}(\mathcal{Z})\}.$$

Let $\tilde{\mu}$ and $\tilde{\mu}'$ be two measures in $\tilde{\mathcal{P}}(\mathcal{Z})$. Suppose both $\tilde{\mu}$ and $\tilde{\mu}'$ are absolutely continuous with respect to a common reference measure ν (such ν always exists). The distances $\tilde{d}_{TV}$ and $\tilde{d}_H$ between $\tilde{\mu}$ and $\tilde{\mu}'$ are defined as follows.

Definition 33 (Distances for scaled probability measures) *Distance $\tilde{d}_{TV}$ between $\tilde{\mu}$ and $\tilde{\mu}'$ is given by*

$$\tilde{d}_{TV}(\tilde{\mu}, \tilde{\mu}') := \frac{1}{2}\int_{\mathcal{Z}} \left| \frac{d\tilde{\mu}}{d\nu}(z) - \frac{d\tilde{\mu}'}{d\nu}(z) \right| \nu(dz),$$

and the distance $\tilde{d}_H$ between $\tilde{\mu}$ and $\tilde{\mu}'$ is given by

$$\tilde{d}_H(\tilde{\mu}, \tilde{\mu}') := \sqrt{\frac{1}{2}\int_{\mathcal{Z}} \left(\sqrt{\frac{d\tilde{\mu}}{d\nu}(z)} - \sqrt{\frac{d\tilde{\mu}'}{d\nu}(z)} \right)^2 \nu(dz)}.$$

Note that the values of $\tilde{d}_{TV}(\tilde{\mu}, \tilde{\mu}')$ and $\tilde{d}_H(\tilde{\mu}, \tilde{\mu}')$ are invariant to the choice of the reference measure ν. The total variation distance d_{TV} and Hellinger distance d_H can be viewed as specific instances of $\tilde{d}_{TV}$ and $\tilde{d}_H$, respectively, with the inputs restricted to probability measures. Building on Definition 33, we now present the following lemma.

Lemma 34 *If $(\mathcal{Z}, d_{\mathcal{Z}})$ is a metric space, then $(\tilde{\mathcal{P}}(\mathcal{Z}), \tilde{d}_{TV})$ and $(\tilde{\mathcal{P}}(\mathcal{Z}), \tilde{d}_H)$ are metric spaces.*

Let $\mathcal{Z}$ be a set and $d_{\mathcal{Z}} : \mathcal{Z} \times \mathcal{Z} \to \mathbb{R}$ be a function. The pair $(\mathcal{Z}, d_{\mathcal{Z}})$ is a metric space if and only if the following conditions are satisfied for any $z1, z2, z3 \in \mathcal{Z}$.

1. $d_{\mathcal{Z}}(z_1, z_2) = 0 \Leftrightarrow z_1 = z_2$.
2. **Non-negativity**: $d_{\mathcal{Z}}(z_1, z_2) \geq 0$.
3. **Symmetry**: $d_{\mathcal{Z}}(z_1, z_2) = d_{\mathcal{Z}}(z_2, z_1)$.
4. **Triangle inequality**: $d_{\mathcal{Z}}(z_1, z_3) \leq d_{\mathcal{Z}}(z_1, z_2) + d_{\mathcal{Z}}(z_2, z_3)$.

It is straightforward to prove that $(\tilde{\mathcal{P}}(\mathcal{Z}), \tilde{d}_{TV})$ is a metric space and the pair $(\tilde{\mathcal{P}}(\mathcal{Z}), \tilde{d}_H)$ satisfies the first three requirements for being a metric space. The corresponding proofs are omitted for brevity.

Below, we provide a proof of the triangle inequality of $\tilde{d}_H$ in space $\tilde{\mathcal{P}}(\mathcal{Z})$.

Proof Assume $\mu_1, \mu_2, \mu_3 \in \tilde{\mathcal{P}}(\mathcal{Z})$. Let the measure ν satisfies $\mu_1 \ll \nu$, $\mu_2 \ll \nu$, and $\mu_3 \ll \nu$ — note that such ν can always exist. Denote the Radon-Nikodym derivatives $\frac{d\mu_1}{d\nu}$, $\frac{d\mu_2}{d\nu}$, and $\frac{d\mu_3}{d\nu}$ by g_1, g_2, and g_3, respectively. By the definition of the distance $\tilde{d}_H$, $\tilde{d}_H(\mu_1, \mu_3)$ satisfies

$$\tilde{d}_H^2(\mu_1, \mu_3) = \frac{1}{2}\int_{\mathcal{Z}} \left(\sqrt{g_1(z)} - \sqrt{g_3(z)} \right)^2 \nu(dz) = \frac{1}{2}\int_{\mathcal{Z}} \left| \sqrt{g_1(z)} - \sqrt{g_3(z)} \right|^2 \nu(dz). \quad (61)$$

The square of $\tilde{d}_H(\mu_1,\mu_2)+\tilde{d}_H(\mu_2,\mu_3)$ can be written as

$$\begin{aligned}
&\left(\sqrt{\frac{1}{2}\int_{\mathcal{Z}}\left(\sqrt{g_1(z)}-\sqrt{g_2(z)}\right)^2\nu(dz)}+\sqrt{\frac{1}{2}\int_{\mathcal{Z}}\left(\sqrt{g_2(z)}-\sqrt{g_3(z)}\right)^2\nu(dz)}\right)^2\\
&\quad=\frac{1}{2}\int_{\mathcal{Z}}\left(\sqrt{g_1(z)}-\sqrt{g_2(z)}\right)^2\nu(dz)+\frac{1}{2}\int_{\mathcal{Z}}\left(\sqrt{g_2(z)}-\sqrt{g_3(z)}\right)^2\nu(dz)\\
&\qquad+\sqrt{\int_{\mathcal{Z}}\left(\sqrt{g_1(z)}-\sqrt{g_2(z)}\right)^2\nu(dz)}\sqrt{\int_{\mathcal{Z}}\left(\sqrt{g_2(z)}-\sqrt{g_3(z)}\right)^2\nu(dz)}. \qquad (62)
\end{aligned}$$

By Hölder's inequality, the third term in Equation (62) satisfies

$$\begin{aligned}
&\sqrt{\int_{\mathcal{Z}}\left(\sqrt{g_1(z)}-\sqrt{g_2(z)}\right)^2\nu(dz)}\sqrt{\int_{\mathcal{Z}}\left(\sqrt{g_2(z)}-\sqrt{g_3(z)}\right)^2\nu(dz)}\\
&\quad\geq\int_{\mathcal{Z}}\left|\left(\sqrt{g_1(z)}-\sqrt{g_2(z)}\right)\left(\sqrt{g_2(z)}-\sqrt{g_3(z)}\right)\right|\nu(dz)\\
&\quad=\int_{\mathcal{Z}}\left|\sqrt{g_1(z)}-\sqrt{g_2(z)}\right|\left|\sqrt{g_2(z)}-\sqrt{g_3(z)}\right|\nu(dz) \qquad (63)
\end{aligned}$$

Plugging Equation (63) into Equation (62) gives

$$\begin{aligned}
&\left(\sqrt{\int_{\mathcal{Z}}\left(\sqrt{g_1(z)}-\sqrt{g_2(z)}\right)^2\nu(dz)}+\sqrt{\int_{\mathcal{Z}}\left(\sqrt{g_2(z)}-\sqrt{g_3(z)}\right)^2\nu(dz)}\right)^2\\
&\quad\geq\frac{1}{2}\int_{\mathcal{Z}}\left(\sqrt{g_1(z)}-\sqrt{g_2(z)}\right)^2\nu(dz)+\frac{1}{2}\int_{\mathcal{Z}}\left(\sqrt{g_2(z)}-\sqrt{g_3(z)}\right)^2\nu(dz)\\
&\qquad+\int_{\mathcal{Z}}\left|\sqrt{g_1(z)}-\sqrt{g_2(z)}\right|\left|\sqrt{g_2(z)}-\sqrt{g_3(z)}\right|\nu(dz)\\
&\quad=\frac{1}{2}\int_{\mathcal{Z}}\left(\left|\sqrt{g_1(z)}-\sqrt{g_2(z)}\right|+\left|\sqrt{g_2(z)}-\sqrt{g_3(z)}\right|\right)^2\nu(dz). \qquad (64)
\end{aligned}$$

According to the triangle inequality of the Euclidean distance on $\mathbb{R}$, for every $z\in\mathcal{Z}$, we have

$$0\leq\left|\sqrt{g_1(z)}-\sqrt{g_3(z)}\right|\leq\left|\sqrt{g_1(z)}-\sqrt{g_2(z)}\right|+\left|\sqrt{g_2(z)}-\sqrt{g_3(z)}\right|. \qquad (65)$$

Combining Equations (61), (64) and (65) completes the proof. ■

Next, we state Proposition 35, which establishes the global Lipschitz continuity of the map $\tilde{F}_k$.

Proposition 35 *(**Global Lipschitz continuity of** $\tilde{F}_k$) Given data y_k, the function $\tilde{F}_k:(\bar{\mathcal{P}}_k(\bar{\mathcal{X}}),d)\to(\tilde{\mathcal{P}}(\bar{\mathcal{X}}),\tilde{d})$ is globally Lipschitz continuous, i.e., $\tilde{F}_k$ satisfies*

$$\tilde{d}(\tilde{F}_k(\mu),\tilde{F}_k(\mu'))\leq\tilde{K}(y_k)d(\mu,\mu'),\quad\forall\mu,\mu'\in\bar{\mathcal{P}}_k(\bar{\mathcal{X}}),$$

when d is d_{TV} and $\tilde{d}$ is $\tilde{d}_{TV}$, or when d is d_H and $\tilde{d}$ is $\tilde{d}_H$, under the assumption for the total variation and Hellinger distance in Theorem 14.

Table 8: (Global) Lipschitz constant $\tilde{K}(y_k)$ of function $\tilde{F}_k$ for different problems and distances of probability measures.

Problem name or distance d and $\tilde{d}$	Inverse problems	State estimation	Parameter-state learning
$(d_{TV}, \tilde{d}_{TV})$	$C_h(y_k)$	$C_{Th}(y_k;k)$	$\tilde{C}_{Th}(y_k;k)$
$(d_H, \tilde{d}_H)$	$\sqrt{C_h(y_k)}$	$\sqrt{C_{Th}(y_k;k)}$	$\sqrt{\tilde{C}_{Th}(y_k;k)}$

The Lipschitz constants $\tilde{K}(y_k)$ corresponding to different distances and different problems are provided in Table 8.

Remark 36 *The conclusion in Proposition 35 under the total variation distance is not used for the proof of Theorem 14. We include it in Proposition 35 for completeness.*

Proposition 35 is a conclusion of Lemmas 37—42.

Lemma 37 *In inverse problems, under Assumption IP.1,* $\forall \mu, \mu' \in \bar{\mathcal{P}}_k(\mathcal{X})$, $\tilde{F}_k(\mu)$ *and* $\tilde{F}_k(\mu')$ *satisfy*

$$\tilde{d}_{TV}(\tilde{F}_k(\mu), \tilde{F}_k(\mu')) \leq C_h(y_k) d_{TV}(\mu, \mu').$$

Proof Let ν be a σ-finite measure defined on the measurable space $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$. Assume that $\mu \ll \nu$ and $\mu' \ll \nu$. According to the definition of $\tilde{F}_k$, $\tilde{F}_k(\mu)$ and $\tilde{F}_k(\mu')$ are equivalent to $F_k(\mu)$ and $F_k(\mu')$, respectively. Since $F_k(\mu)$ and $F_k(\mu')$ satisfy $F_k(\mu) \ll \mu$ and $F_k(\mu') \ll \mu'$, we have $F_k(\mu) \ll \nu, F_k(\mu') \ll \nu$ and $\tilde{F}_k(\mu) \ll \nu, \tilde{F}_k(\mu') \ll \nu$. The Radon–Nikodym derivatives $\frac{d\tilde{F}_k\mu}{d\nu}$ and $\frac{d\tilde{F}_k\mu'}{d\nu}$ can be written as

$$\begin{aligned}
\frac{d\tilde{F}_k\mu}{d\nu}(x) &= \frac{d\tilde{F}_k\mu}{d\mu}(x) \cdot \frac{d\mu}{d\nu}(x) = h(y_k, x) \cdot \frac{d\mu}{d\nu}(x), \\
\frac{d\tilde{F}_k\mu'}{d\nu}(x) &= \frac{d\tilde{F}_k\mu'}{d\mu'}(x) \cdot \frac{d\mu'}{d\nu}(x) = h(y_k, x) \cdot \frac{d\mu'}{d\nu}(x).
\end{aligned}$$

By the definition of distance $\tilde{d}_{TV}$, $\tilde{d}_{TV}(\tilde{F}_k(\mu), \tilde{F}_k(\mu'))$ can be written as

$$\begin{aligned}
\tilde{d}_{TV}(\tilde{F}_k(\mu), \tilde{F}_k(\mu')) =& \frac{1}{2} \int_{\mathcal{X}} \left| h(y_k, x) \frac{d\mu}{d\nu}(x) - h(y_k, x) \frac{d\mu'}{d\nu}(x) \right| \nu(dx) \\
\leq& \frac{1}{2} C_h(y_k) \int_{\mathcal{X}} \left| \frac{d\mu}{d\nu}(x) - \frac{d\mu'}{d\nu}(x) \right| \nu(dx) \\
=& C_h(y_k) d_{TV}(\mu, \mu').
\end{aligned}$$

■

Lemma 38 *In state estimation problems, under Assumption SE.1,* $\forall \mu, \mu' \in \bar{\mathcal{P}}_k(\mathcal{X})$, $\tilde{F}_k(\mu)$ *and* $\tilde{F}_k(\mu')$ *satisfy*

$$\tilde{d}_{TV}(\tilde{F}_k(\mu), \tilde{F}_k(\mu')) \leq C_{Th}(y_k; k) d_{TV}(\mu, \mu').$$

Proof Let ν be a σ-finite measure defined on the measurable space $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$ such that $\mu \ll \nu$ and $\mu' \ll \nu$. Let $\pi(x)$ and $\pi'(x)$ denote the Radon-Nikodym derivatives $\frac{d\mu}{d\nu}(x)$ and $\frac{d\mu'}{d\nu}(x)$, respectively. By the definition of $\tilde{F}_k$ for state estimation problems, we can write $\tilde{d}_{TV}(\tilde{F}_k(\mu), \tilde{F}_k(\mu'))$ as:

$$\begin{aligned}
&\tilde{d}_{TV}(\tilde{F}_k(\mu), \tilde{F}_k(\mu')) \\
&\quad = \frac{1}{2} \int_{\mathcal{X}} \left| h_k(y_k, x) \int_{\mathcal{X}} T_k(x, x_{k-1}) \mu(dx_{k-1}) - h_k(y_k, x) \int_{\mathcal{X}} T_k(x, x_{k-1}) \mu'(dx_{k-1}) \right| dx \\
&\quad = \frac{1}{2} \int_{\mathcal{X}} \left| h_k(y_k, x) \int_{\mathcal{X}} T_k(x, x_{k-1}) \pi(x_{k-1}) \nu(dx_{k-1}) \right. \\
&\qquad\qquad \left. - h_k(y_k, x) \int_{\mathcal{X}} T_k(x, x_{k-1}) \pi'(x_{k-1}) \nu(dx_{k-1}) \right| dx \\
&\quad \leq \frac{1}{2} \int_{\mathcal{X}} h_k(y_k, x) \left(\int_{\mathcal{X}} T_k(x, x_{k-1}) |\pi(x_{k-1}) - \pi'(x_{k-1})| \nu(dx_{k-1}) \right) dx. \qquad (66)
\end{aligned}$$

By Tonelli's theorem, the right-hand side of Equation (66) can be written as

$$\begin{aligned}
&\frac{1}{2} \int_{\mathcal{X}} h_k(y_k, x) \left(\int_{\mathcal{X}} T_k(x, x_{k-1}) |\pi(x_{k-1}) - \pi'(x_{k-1})| \nu(dx_{k-1}) \right) dx \\
&\quad = \frac{1}{2} \int_{\mathcal{X}} \left(\int_{\mathcal{X}} h_k(y_k, x) T_k(x, x_{k-1}) dx \right) |\pi(x_{k-1}) - \pi'(x_{k-1})| \nu(dx_{k-1}) \\
&\quad \leq \frac{1}{2} C_{Th}(y_k; k) \int_{\mathcal{X}} |\pi(x_{k-1}) - \pi'(x_{k-1})| \nu(dx_{k-1}) \\
&\quad = C_{Th}(y_k; k) d_{TV}(\mu, \mu').
\end{aligned}$$

■

Lemma 39 *In parameter-state estimation problems, under Assumption PS.1,* $\forall \mu, \mu' \in \bar{\mathcal{P}}_k(\mathcal{X} \times \mathcal{W})$, $\tilde{F}_k(\mu)$ *and* $\tilde{F}_k(\mu')$ *satisfy*

$$\tilde{d}_{TV}(\tilde{F}_k(\mu), \tilde{F}_k(\mu')) \leq \tilde{C}_{Th}(y_k; k) d_{TV}(\mu, \mu').$$

Proof Let π and π' denote the Lebesgue densities of μ and μ', respectively. By the definition of $\tilde{F}_k$ for parameter-state estimation problems, we can write $\tilde{d}_{TV}(\tilde{F}_k(\mu), \tilde{F}_k(\mu'))$

as

$$\begin{aligned}
&\tilde{d}_{TV}(\tilde{F}_k(\mu), \tilde{F}_k(\mu')) \\
&\quad = \frac{1}{2}\int_{\mathcal{X}\times\mathcal{W}} \Bigg| h_k(y_k, x, w) \int_{\mathcal{X}} T_k(x, x_{k-1}, w)\pi(x_{k-1}, w)dx_{k-1} \\
&\qquad\qquad - h_k(y_k, x, w)\int_{\mathcal{X}} T_k(x, x_{k-1}, w)\pi'(x_{k-1}, w)dx_{k-1}\Bigg| dxdw \\
&\quad = \frac{1}{2}\int_{\mathcal{X}\times\mathcal{W}} h_k(y_k, x, w)\left| \int_{\mathcal{X}} T_k(x, x_{k-1}, w)\left(\pi(x_{k-1}, w) - \pi'(x_{k-1}, w)\right) dx_{k-1}\right| dxdw \\
&\quad \leq \frac{1}{2}\int_{\mathcal{X}\times\mathcal{W}} h_k(y_k, x, w)\left(\int_{\mathcal{X}} T_k(x, x_{k-1}, w)\left|\pi(x_{k-1}, w) - \pi'(x_{k-1}, w)\right| dx_{k-1}\right) dxdw.
\end{aligned}$$

By Tonelli's theorem, we have

$$\begin{aligned}
&\frac{1}{2}\int_{\mathcal{X}\times\mathcal{W}} h_k(y_k, x, w)\left(\int_{\mathcal{X}} T_k(x, x_{k-1}, w)\left|\pi(x_{k-1}, w) - \pi'(x_{k-1}, w)\right| dx_{k-1}\right) dxdw \\
&\quad = \frac{1}{2}\int_{\mathcal{X}\times\mathcal{W}} \left(\int_{\mathcal{X}} h_k(y_k, x, w)T_k(x, x_{k-1}, w)dx\right)\left|\pi(x_{k-1}, w) - \pi'(x_{k-1}, w)\right| dx_{k-1}dw \\
&\quad \leq \tilde{C}_{Th}(y_k; k)d_{TV}(\mu, \mu').
\end{aligned}$$

■

Lemma 40 *In inverse problems, under Assumption IP.1, $\forall \mu, \mu' \in \bar{\mathcal{P}}_k(\mathcal{X})$, $\tilde{F}_k(\mu)$ and $\tilde{F}_k(\mu')$ satisfy*

$$\tilde{d}_H(\tilde{F}_k(\mu), \tilde{F}_k(\mu')) \leq \sqrt{C_h(y_k)}d_H(\mu, \mu').$$

The proof of Lemma 40 is analogous to that of Lemma 37.

Lemma 41 *In state estimation problems, under Assumption SE.1, $\forall \mu, \mu' \in \bar{\mathcal{P}}_k(\mathcal{X})$, $\tilde{F}_k(\mu)$ and $\tilde{F}_k(\mu')$ satisfy*

$$\tilde{d}_H(\tilde{F}_k(\mu), \tilde{F}_k(\mu')) \leq \sqrt{C_{Th}(y_k; k)}d_H(\mu, \mu').$$

Proof Let ν be a σ-finite measure defined on the measurable space $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$ such that $\mu \ll \nu$ and $\mu' \ll \nu$. Let $\pi(x)$ and $\pi'(x)$ denote the Radon-Nikodym derivatives $\frac{d\mu}{d\nu}(x)$ and

$\frac{d\mu'}{d\nu}(x)$, respectively. By the definition of $\tilde{d}_H$ distance, we have

$$\begin{aligned}
&\tilde{d}_H^2(\tilde{F}_k(\mu),\tilde{F}_k(\mu')) \\
&\quad= \frac{1}{2}\int_{\mathcal{X}}\left(\sqrt{h_k(y_k,x)\int_{\mathcal{X}}T_k(x,x_{k-1})\mu(dx_{k-1})}\right. \\
&\qquad\qquad\left.-\sqrt{h_k(y_k,x)\int_{\mathcal{X}}T_k(x,x_{k-1})\mu'(dx_{k-1})}\right)^2 dx \\
&\quad= \frac{1}{2}\int_{\mathcal{X}}h_k(y_k,x)\left(\int_{\mathcal{X}}T_k(x,x_{k-1})\left(\pi(x_{k-1})+\pi'(x_{k-1})\right)\nu(dx_{k-1})\right. \\
&\qquad\left.-2\sqrt{\int_{\mathcal{X}}T_k(x,x_{k-1})\pi(x_{k-1})\nu(dx_{k-1})}\sqrt{\int_{\mathcal{X}}T_k(x,x_{k-1})\pi'(x_{k-1})\nu(dx_{k-1})}\right)dx. \quad (67)
\end{aligned}$$

By Hölder's inequality, we have

$$\begin{aligned}
&\sqrt{\int_{\mathcal{X}}T_k(x,x_{k-1})\pi(x_{k-1})\nu(dx_{k-1})}\sqrt{\int_{\mathcal{X}}T_k(x,x_{k-1})\pi'(x_{k-1})\nu(dx_{k-1})} \\
&\qquad\geq\int_{\mathcal{X}}T_k(x,x_{k-1})\sqrt{\pi(x_{k-1})}\sqrt{\pi'(x_{k-1})}\nu(dx_{k-1}). \quad (68)
\end{aligned}$$

Substituting Equation (68) into Equation (67) gives

$$\begin{aligned}
\tilde{d}_H^2(\tilde{F}_k(\mu),\tilde{F}_k(\mu')) &\leq \frac{1}{2}\int_{\mathcal{X}}h_k(y_k,x)\left(\int_{\mathcal{X}}T_k(x,x_{k-1})\left(\pi(x_{k-1})+\pi'(x_{k-1})\right)\nu(dx_{k-1})\right. \\
&\qquad\left.-2\int_{\mathcal{X}}T_k(x,x_{k-1})\sqrt{\pi(x_{k-1})}\sqrt{\pi'(x_{k-1})}\nu(dx_{k-1})\right)dx \\
&= \frac{1}{2}\int_{\mathcal{X}}h_k(y_k,x)\left(\int_{\mathcal{X}}T_k(x,x_{k-1})\left(\sqrt{\pi(x_{k-1})}-\sqrt{\pi'(x_{k-1})}\right)^2\nu(dx_{k-1})\right)dx. \quad (69)
\end{aligned}$$

According to Tonelli's theorem, the right-hand side of Equation (69) satisfies

$$\begin{aligned}
&\frac{1}{2}\int_{\mathcal{X}}h_k(y_k,x)\left(\int_{\mathcal{X}}T_k(x,x_{k-1})\left(\sqrt{\pi(x_{k-1})}-\sqrt{\pi'(x_{k-1})}\right)^2\nu(dx_{k-1})\right)dx \\
&\quad= \frac{1}{2}\int_{\mathcal{X}}\left(\int_{\mathcal{X}}h_k(y_k,x)T_k(x,x_{k-1})dx\right)\left(\sqrt{\pi(x_{k-1})}-\sqrt{\pi'(x_{k-1})}\right)^2\nu(dx_{k-1}) \\
&\quad\leq \frac{1}{2}C_{Th}(y_k;k)\int_{\mathcal{X}}\left(\sqrt{\pi(x_{k-1})}-\sqrt{\pi'(x_{k-1})}\right)^2\nu(dx_{k-1}) \\
&\quad= C_{Th}(y_k;k)d_H^2(\mu,\mu').
\end{aligned}$$

■

Lemma 42 *In parameter-state estimation problems, under Assumption PS.1, $\forall\mu,\mu'\in\widetilde{\mathcal{P}}_k(\mathcal{X}\times\mathcal{W})$, $\tilde{F}_k(\mu)$ and $\tilde{F}_k(\mu'))$ satisfy*

$$\tilde{d}_H(\tilde{F}_k(\mu),\tilde{F}_k(\mu'))\leq\sqrt{\tilde{C}_{Th}(y_k;k)}d_H(\mu,\mu').$$

Proof Let π and π' denote the Lebesgue densities of μ and μ', respectively. Following an argument analogous to that used in the proof of Lemma 41, we can show that

$$\begin{aligned}&\tilde{d}_H^2(\tilde{F}_k(\mu), \tilde{F}_k(\mu')) \\ \leq\ & \frac{1}{2}\int_{\mathcal{X}\times\mathcal{W}} h_k(y_k, x, w)\left(\int_{\mathcal{X}} T_k(x, x_{k-1}, w)\left(\sqrt{\pi(x_{k-1}, w)} - \sqrt{\pi'(x_{k-1}, w)}\right)^2 dx_{k-1}\right) \\ & \cdot dxdw. \end{aligned} \tag{70}$$

An argument similar to that in Lemma 39 yields

$$\begin{aligned}&\int_{\mathcal{X}\times\mathcal{W}} h_k(y_k, x, w)\left(\int_{\mathcal{X}} T_k(x, x_{k-1}, w)\left(\sqrt{\pi(x_{k-1}, w)} - \sqrt{\pi'(x_{k-1}, w)}\right)^2 dx_{k-1}\right) \\ &\quad \cdot dxdw \\ &= \int_{\mathcal{X}\times\mathcal{W}} \left(\int_{\mathcal{X}} h_k(y_k, x, w) T_k(x, x_{k-1}, w) dx\right)\left(\sqrt{\pi(x_{k-1}, w)} - \sqrt{\pi'(x_{k-1}, w)}\right)^2 \\ &\qquad \cdot dx_{k-1}dw \\ &\leq \tilde{C}_{Th}(y_k; k)\int_{\mathcal{X}\times\mathcal{W}} \left(\sqrt{\pi(x_{k-1}, w)} - \sqrt{\pi'(x_{k-1}, w)}\right)^2 dx_{k-1}dw \\ &= 2\tilde{C}_{Th}(y_k; k) d_H^2(\mu, \mu'). \end{aligned}$$

■

B.2.2 Proofs of Propositions 30—32

Given Proposition 35, to prove Propositions 30—32, we first prove the following lemma.

Lemma 43 *Let $\tilde{\mu}, \tilde{\mu}' \in \tilde{\mathcal{P}}(\mathcal{Z})$ be two measures such that $\forall A \in \mathcal{B}(\mathcal{Z})$,*

$$\tilde{\mu}(A) = k\mu(A), \quad \tilde{\mu}'(A) = k'\mu'(A), \quad \mu, \mu' \in \mathcal{P}(\mathcal{Z}), \quad k, k' \in \mathbb{R}_+.$$

Then $|\sqrt{k} - \sqrt{k'}|$ satisfies

$$\left|\sqrt{k} - \sqrt{k'}\right| \leq \sqrt{2}\tilde{d}_H(\tilde{\mu}, \tilde{\mu}'). \tag{71}$$

The Hellinger distance between the two probability measures μ and μ' satisfies

$$d_H\left(\mu, \mu'\right) \leq \frac{2}{\sqrt{k}}\tilde{d}_H(\tilde{\mu}, \tilde{\mu}'). \tag{72}$$

Proof We first prove the statement $|\sqrt{k} - \sqrt{k'}| \leq \sqrt{2}\tilde{d}_H(\tilde{\mu}, \tilde{\mu}')$. Let ν be a σ-finite measure defined on $(\mathcal{Z}, \mathcal{B}(\mathcal{Z}))$ such that $\mu \ll \nu$ and $\mu' \ll \nu$. Then we have $\tilde{\mu} \ll \nu$ and $\tilde{\mu}' \ll \nu$. Given

the definition of $\tilde{\mu}$ and $\tilde{\mu}'$, we have

$$\int_{\mathcal{Z}} \frac{d\tilde{\mu}}{d\nu}(z)\nu(dz) = \int_{\mathcal{Z}} \tilde{\mu}(dz) = k\int_{\mathcal{Z}} \mu(dz) = k,$$
$$\int_{\mathcal{Z}} \frac{d\tilde{\mu}'}{d\nu}(z)\nu(dz) = \int_{\mathcal{Z}} \tilde{\mu}'(dz) = k'\int_{\mathcal{Z}} \mu'(dz) = k'.$$

Therefore, $(\sqrt{k}-\sqrt{k'})^2$ can be written as

$$\begin{aligned}(\sqrt{k}-\sqrt{k'})^2 &= k + k' - 2\sqrt{k}\sqrt{k'} \\ &= \int_{\mathcal{Z}} \frac{d\tilde{\mu}}{d\nu}(z)\nu(dz) + \int_{\mathcal{Z}} \frac{d\tilde{\mu}'}{d\nu}(z)\nu(dz) - 2\sqrt{\int_{\mathcal{Z}} \frac{d\tilde{\mu}}{d\nu}(z)\nu(dz)}\sqrt{\int_{\mathcal{Z}} \frac{d\tilde{\mu}'}{d\nu}(z)\nu(dz)}. \quad (73)\end{aligned}$$

By Hölder's inequality, the term $\sqrt{\int_{\mathcal{Z}} \frac{d\tilde{\mu}}{d\nu}(z)\nu(dz)}\sqrt{\int_{\mathcal{Z}} \frac{d\tilde{\mu}'}{d\nu}(z)\nu(dz)}$ satisfies

$$\sqrt{\int_{\mathcal{Z}} \frac{d\tilde{\mu}}{d\nu}(z)\nu(dz)}\sqrt{\int_{\mathcal{Z}} \frac{d\tilde{\mu}'}{d\nu}(z)\nu(dz)} \geq \int_{\mathcal{Z}} \sqrt{\frac{d\tilde{\mu}}{d\nu}(z)}\sqrt{\frac{d\tilde{\mu}'}{d\nu}(z)}\nu(dz). \quad (74)$$

Plugging Equation (74) into Equation (73) gives

$$\begin{aligned}(\sqrt{k}-\sqrt{k'})^2 \leq& \int_{\mathcal{Z}} \frac{d\tilde{\mu}}{d\nu}(z)\nu(dz) + \int_{\mathcal{Z}} \frac{d\tilde{\mu}'}{d\nu}(z)\nu(dz) - 2\int_{\mathcal{Z}} \sqrt{\frac{d\tilde{\mu}}{d\nu}(z)}\sqrt{\frac{d\tilde{\mu}'}{d\nu}(z)}\nu(dz) \\ =& \int_{\mathcal{Z}} \left(\sqrt{\frac{d\tilde{\mu}}{d\nu}(z)} - \sqrt{\frac{d\tilde{\mu}'}{d\nu}(z)}\right)^2 \nu(dz) \\ =& 2\tilde{d}_H^2(\tilde{\mu},\tilde{\mu}').\end{aligned}$$

Next, we prove the statement $d_H(\mu,\mu') \le \frac{2}{\sqrt{k}}\tilde{d}_H(\tilde{\mu},\tilde{\mu}')$.
By the definition of the Hellinger distance, $d_H^2(\mu,\mu')$ can be written as

$$
\begin{aligned}
d_H^2(\mu,\mu') &= \frac{1}{2}\int_{\mathcal{Z}}\left(\sqrt{\frac{d\mu}{d\nu}(z)}-\sqrt{\frac{d\mu'}{d\nu}(z)}\right)^2\nu(dz)\\
&= \frac{1}{2}\int_{\mathcal{Z}}\left(\sqrt{\frac{1}{k}}\sqrt{\frac{d\tilde{\mu}}{d\nu}(z)}-\sqrt{\frac{1}{k'}}\sqrt{\frac{d\tilde{\mu}'}{d\nu}(z)}\right)^2\nu(dz)\\
&= \frac{1}{2}\int_{\mathcal{Z}}\left(\sqrt{\frac{1}{k}}\sqrt{\frac{d\tilde{\mu}}{d\nu}(z)}-\sqrt{\frac{1}{k}}\sqrt{\frac{d\tilde{\mu}'}{d\nu}(z)}+\sqrt{\frac{1}{k}}\sqrt{\frac{d\tilde{\mu}'}{d\nu}(z)}-\sqrt{\frac{1}{k'}}\sqrt{\frac{d\tilde{\mu}'}{d\nu}(z)}\right)^2\nu(dz)\\
&\le \int_{\mathcal{Z}}\left(\left(\sqrt{\frac{1}{k}}\sqrt{\frac{d\tilde{\mu}}{d\nu}(z)}-\sqrt{\frac{1}{k}}\sqrt{\frac{d\tilde{\mu}'}{d\nu}(z)}\right)^2\right.\\
&\qquad\left.+\left(\sqrt{\frac{1}{k}}\sqrt{\frac{d\tilde{\mu}'}{d\nu}(z)}-\sqrt{\frac{1}{k'}}\sqrt{\frac{d\tilde{\mu}'}{d\nu}(z)}\right)^2\right)\nu(dz)\\
&= \frac{1}{k}\int_{\mathcal{Z}}\left(\sqrt{\frac{d\tilde{\mu}}{d\nu}(z)}-\sqrt{\frac{d\tilde{\mu}'}{d\nu}(z)}\right)^2\nu(dz)+\left(\sqrt{\frac{1}{k}}-\sqrt{\frac{1}{k'}}\right)^2\int_{\mathcal{Z}}\frac{d\tilde{\mu}'}{d\nu}(z)\nu(dz)\\
&= \frac{2}{k}\tilde{d}_H^2(\tilde{\mu},\tilde{\mu}')+\frac{(\sqrt{k}-\sqrt{k'})^2}{k} \le \frac{4}{k}\tilde{d}_H^2(\tilde{\mu},\tilde{\mu}').
\end{aligned}
$$

■

Lemma 43 is a generalization of Lemma 44 stated below. Lemma 44 is presented in Lemma 1 in Zhao and Cui (2024) and the proofs of Proposition 5 and Theorem 1 in Cui and Dolgov (2022).

Lemma 44 (Zhao and Cui (2024), Cui and Dolgov (2022)) *Let $\tilde{\mu},\tilde{\mu}' \in \tilde{\mathcal{P}}(\mathcal{Z})$ be two measures, both absolutely continuous with respect to the Lebesgue measure. Let $\tilde{\pi}$ and $\tilde{\pi}'$ represent the densities of $\tilde{\mu}$ and $\tilde{\mu}'$ with respect to the Lebesgue measure, respectively. Assume that $\forall z \in \mathcal{Z}$, $\tilde{\pi}$ and $\tilde{\pi}'$ satisfy*

$$\tilde{\pi}(z) = k\pi(z), \quad \tilde{\pi}'(z) = k'\pi'(z), \quad k,k' \in \mathbb{R}_+,$$

where π and π' are the densities of the probability measures μ and μ' with respect to the Lebesgue measure, respectively. Then $|\sqrt{k}-\sqrt{k'}|$ satisfies

$$\left|\sqrt{k}-\sqrt{k'}\right| \le \sqrt{2}\tilde{d}_H(\tilde{\mu},\tilde{\mu}').$$

The Hellinger distance between the two probability measures μ and μ' satisfies

$$d_H(\mu,\mu') \le \frac{2}{\sqrt{k}}\tilde{d}_H(\tilde{\mu},\tilde{\mu}'). \tag{75}$$

As shown in the proof of Lemma 43, the proof of the conclusion $d_H(\mu, \mu') \leq \frac{2}{\sqrt{k}} \tilde{d}_H(\tilde{\mu}, \tilde{\mu}')$ is based on the conclusion $\left|\sqrt{k} - \sqrt{k'}\right| \leq \sqrt{2}\tilde{d}_H(\tilde{\mu}, \tilde{\mu}')$. Our proof of the conclusion $\left|\sqrt{k} - \sqrt{k'}\right| \leq \sqrt{2}\tilde{d}_H(\tilde{\mu}, \tilde{\mu}')$ differs from the proof in Cui and Dolgov (2022) and is more concise. Given this result, our proof of the conclusion $d_H(\mu, \mu') \leq \frac{2}{\sqrt{k}} \tilde{d}_H(\tilde{\mu}, \tilde{\mu}')$ is analogous to the proof in Cui and Dolgov (2022).

Given the definition of F_k, directly combining Proposition 35 and Lemma 43 proves Propositions 30—32. This argument is similar to an argument used in the proof of Theorem 8 in Zhao and Cui (2024).

B.3 Proof of Theorem 14 under 1-Wasserstein distance

Theorem 14 under the 1-Wasserstein distance is a conclusion of Propositions 49—51. To prove Propositions 49—51, we first present Proposition 45, which is a supporting proposition for the proof of Theorem 14 under the 1-Wasserstein distance.

B.3.1 Supporting proposition for the proof of Theorem 14: Proposition 45

Proposition 45 *(**Global Lipschitz continuity of evidence function** Z_k) Given the data y_k, the evidence function $Z_k : (\bar{\mathcal{P}}_k(\bar{\mathcal{X}}) \bigcap \mathcal{P}_1(\bar{\mathcal{X}}), W_1) \to (\mathbb{R}_+, |\cdot|)$ is globally Lipschitz continuous, i.e., Z_k satisfies*

$$|Z_k(\mu) - Z_k(\mu')| \leq K_Z(y_k) W_1(\mu, \mu'), \quad \forall \mu, \mu' \in \bar{\mathcal{P}}_k(\bar{\mathcal{X}}) \bigcap \mathcal{P}_1(\bar{\mathcal{X}}),$$

if

- *in inverse problems, Assumption IP.2 holds (Theorem 15, Sprungk (2020)); or*
- *in state-parameter estimation problems, Assumption SE.2 holds; or*
- *in parameter-state estimation problems, Assumption PS.2 holds.*

The Lipschitz constant $K_Z(y_k)$ is given in Table 9.

Table 9: (Global) Lipschitz constant $K_Z(y_k)$ of evidence function Z_k in different problems.

Inverse problems	State estimation	Parameter-state learning
$\|h\|_{\text{Lip}}(y_k)$	$C^*_{Th}(y_k; k)$	$\widetilde{C}^*_{Th}(y_k; k)$

Proposition 45 is a conclusion of Lemmas 46—48, as shown below.

Lemma 46 (Theorem 15, Sprungk (2020)) *(**Upper-bounded evidence difference in inverse problems**) For inverse problems, given the data y_k, suppose Assumption IP.2 holds. For any $\mu, \mu' \in \bar{\mathcal{P}}_k(\mathcal{X}) \bigcap \mathcal{P}_1(\mathcal{X})$, the difference between the two evidences, $Z_k(\mu)$ and $Z_k(\mu')$, is bounded by:*

$$\left|Z_k(\mu) - Z_k(\mu')\right| \leq \|h\|_{Lip}(y_k) W_1(\mu, \mu').$$

For completeness, we restate Lemma 46 and provide its full proof below.

Proof The difference between the two evidences $Z_k(\mu)$ and $Z_k(\mu')$ can be written as

$$\left|Z_k(\mu) - Z_k(\mu')\right| = \left| \int_{\mathcal{X}} h(y_k, x)\mu(dx) - \int_{\mathcal{X}} h(y_k, x)\mu'(dx) \right|.$$

Given data y_k, define a function $\tilde{h}$ as

$$\tilde{h}(x) := \frac{h(y_k, x)}{\|h\|_{\text{Lip}}(y_k)}.$$

Then $\tilde{h}$ satisfies

$$\begin{aligned}\sup_{(x_1,x_2)\in\mathcal{X}\times\mathcal{X}} \frac{\left|\tilde{h}(x_1) - \tilde{h}(x_2)\right|}{d_{\mathcal{X}}(x_1, x_2)} =& \frac{1}{\|h\|_{\text{Lip}}(y_k)} \sup_{(x_1,x_2)\in\mathcal{X}\times\mathcal{X}} \frac{\left|h(y_k, x_1) - h(y_k, x_2)\right|}{d_{\mathcal{X}}(x_1, x_2)} \\ =& \frac{\|h\|_{\text{Lip}}(y_k)}{\|h\|_{\text{Lip}}(y_k)} = 1. \end{aligned} \tag{76}$$

By the definition of $\tilde{h}$, we can re-write $\left|Z_k(\mu) - Z_k(\mu')\right|$ as

$$\begin{aligned}\left|Z_k(\mu) - Z_k(\mu')\right| &= \|h\|_{\text{Lip}}(y_k) \left| \int_{\mathcal{X}} \tilde{h}(x)\mu(dx) - \int_{\mathcal{X}} \tilde{h}(x)\mu'(dx) \right| \\ &\leq \|h\|_{\text{Lip}}(y_k) \sup_{f:\mathcal{X}\to\mathbb{R}, \|f\|_{\text{Lip}}\leq 1} \left| \int_{\mathcal{X}} f(x)\mu(dx) - \int_{\mathcal{X}} f(x)\mu'(dx) \right| \\ &= \|h\|_{\text{Lip}}(y_k) W_1(\mu, \mu'). \end{aligned} \tag{77}$$

The first inequality in Equation (77) is obtained because it has been shown in Equation (76) that the best Lipschitz constant of function $\tilde{h}$ is 1. ■

By extending the proof of Lemma 46 to state estimation and parameter-state estimation problems, we obtain Lemmas 47 and 48, as stated below.

Lemma 47 ***(Upper-bounded evidence difference in state estimation problems)*** *For state estimation problems, at time step k, given the data y_k, suppose Assumption SE.2 holds. For any $\mu, \mu' \in \bar{\mathcal{P}}_k(\mathcal{X}) \bigcap \mathcal{P}_1(\mathcal{X})$, the difference between the two evidences, $Z_k(\mu)$ and $Z_k(\mu')$, is bounded by:*

$$\left|Z_k(\mu) - Z_k(\mu')\right| \leq C^*_{Th}(y_k; k) W_1(\mu, \mu').$$

Proof Let ν be a σ-finite measure defined on the measurable space $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$ such that $\mu \ll \nu$ and $\mu' \ll \nu$. Let $\pi(x)$ and $\pi'(x)$ denote the Radon-Nikodym derivatives $\frac{d\mu}{d\nu}(x)$ and $\frac{d\mu'}{d\nu}(x)$, respectively. By the definition of evidence function Z_k, we can write $|Z_k(\mu) - Z_k(\mu')|$ as

$$\begin{aligned}|Z_k(\mu) - Z_k(\mu')| &= \left| \int_{\mathcal{X}} \tilde{F}_k \mu(dx) - \int_{\mathcal{X}} \tilde{F}_k \mu'(dx) \right| \\ &= \left| \int_{\mathcal{X}} h_k(y_k, x) \left(\int_{\mathcal{X}} T_k(x, x_{k-1}) \pi(x_{k-1}) \nu(dx_{k-1}) \right) dx \right. \\ &\quad \left. - \int_{\mathcal{X}} h_k(y_k, x) \left(\int_{\mathcal{X}} T_k(x, x_{k-1}) \pi'(x_{k-1}) \nu(dx_{k-1}) \right) dx \right| \end{aligned} \tag{78}$$

According to Tonelli's theorem, the right-hand side of Equation (78) can be written as

$$\begin{aligned}
&\left| \int_{\mathcal{X}} h_k(y_k, x) \left(\int_{\mathcal{X}} T_k(x, x_{k-1}) \pi(x_{k-1}) \nu(dx_{k-1}) \right) dx \right. \\
&\qquad \left. - \int_{\mathcal{X}} h_k(y_k, x) \left(\int_{\mathcal{X}} T_k(x, x_{k-1}) \pi'(x_{k-1}) \nu(dx_{k-1}) \right) dx \right| \\
&= \left| \int_{\mathcal{X}} \left(\int_{\mathcal{X}} h_k(y_k, x) T_k(x, x_{k-1}) dx \right) \pi(x_{k-1}) \nu(dx_{k-1}) \right. \\
&\qquad \left. - \int_{\mathcal{X}} \left(\int_{\mathcal{X}} h_k(y_k, x) T_k(x, x_{k-1}) dx \right) \pi'(x_{k-1}) \nu(dx_{k-1}) \right|. \qquad (79)
\end{aligned}$$

Given the data y_k, the expression $\int_{\mathcal{X}} h_k(y_k, x) T_k(x, x_{k-1}) dx$ is a function of x_{k-1}. We denote this function as g:

$$g(x_{k-1}) := \int_{\mathcal{X}} h_k(y_k, x) T_k(x, x_{k-1}) dx.$$

We now show that g is Lipschitz continuous and its best Lipschitz constant is no greater than $C^*_{Th}(y_k; k)$.
For two arbitrary points $x_{k-1}, x'_{k-1} \in \mathcal{X}$, we have

$$\begin{aligned}
|g(x_{k-1}) - g(x'_{k-1})| =& \left| \int_{\mathcal{X}} h_k(y_k, x) \left(T_k(x, x_{k-1}) - T_k(x, x'_{k-1}) \right) dx \right| \\
\leq& \int_{\mathcal{X}} h_k(y_k, x) \left| T_k(x, x_{k-1}) - T_k(x, x'_{k-1}) \right| dx \\
\leq& d_{\mathcal{X}}(x_{k-1}, x'_{k-1}) \int_{\mathcal{X}} h_k(y_k, x) T_{\mathrm{Lip}}(x; k) dx \\
=& C^*_{Th}(y_k; k) d_{\mathcal{X}}(x_{k-1}, x'_{k-1}).
\end{aligned}$$

It follows that

$$\|g\|_{\mathrm{Lip}} = \sup_{x_{k-1}, x'_{k-1} \in \mathcal{X}, x_{k-1} \neq x'_{k-1}} \frac{|g(x_{k-1}) - g(x'_{k-1})|}{d_{\mathcal{X}}(x_{k-1}, x'_{k-1})} \leq C^*_{Th}(y_k; k).$$

We proceed by employing the same proof strategy used in the proof of Lemma 46. Define a function $\tilde{g}$ as

$$\tilde{g}(x) := \frac{g(x)}{C^*_{Th}(y_k; k)}.$$

Then $\tilde{g}$ satisfies

$$\|\tilde{g}\|_{\mathrm{Lip}} = \sup_{x, x' \in \mathcal{X}, x \neq x'} \frac{|\tilde{g}(x) - \tilde{g}(x')|}{d_{\mathcal{X}}(x, x')} = \frac{1}{C^*_{Th}(y_k; k)} \sup_{x, x' \in \mathcal{X}, x \neq x'} \frac{|g(x) - g(x')|}{d_{\mathcal{X}}(x, x')} \leq 1.$$

The term $|Z_k(\mu) - Z_k(\mu')|$ can be written as:

$$\begin{aligned}
&|Z_k(\mu) - Z_k(\mu')| \\
&\quad = C^*_{Th}(y_k; k) \left| \int_{\mathcal{X}} \tilde{g}(x_{k-1}) \pi(x_{k-1}) \nu(dx_{k-1}) - \int_{\mathcal{X}} \tilde{g}(x_{k-1}) \pi'(x_{k-1}) \nu(dx_{k-1}) \right|.
\end{aligned}$$

As we have $\|\tilde{g}\|_{\text{Lip}} \leq 1$, the following condition is satisfied:

$$\begin{aligned}
&\left| \int_{\mathcal{X}} \tilde{g}(x_{k-1})\pi(x_{k-1})\nu(dx_{k-1}) - \int_{\mathcal{X}} \tilde{g}(x_{k-1})\pi'(x_{k-1})\nu(dx_{k-1}) \right| \\
&\quad \leq \sup_{f:\mathcal{X}\to\mathbb{R}, \|f\|_{\text{Lip}}\leq 1} \left| \int_{\mathcal{X}} f(x_{k-1})\pi(x_{k-1})\nu(dx_{k-1}) - \int_{\mathcal{X}} f(x_{k-1})\pi'(x_{k-1})\nu(dx_{k-1}) \right| \\
&\quad = W_1(\mu, \mu').
\end{aligned}$$

■

Lemma 48 *(**Upper-bounded evidence difference in parameter-state estimation problems**) For parameter-state estimation problems, at time step k, given the data y_k, suppose Assumption PS.2 holds. For any $\mu, \mu' \in \bar{\mathcal{P}}_k(\mathcal{X}\times\mathcal{W}) \bigcap \mathcal{P}_1(\mathcal{X}\times\mathcal{W})$, the difference between the two evidences, $Z_k(\mu)$ and $Z_k(\mu')$, is bounded by:*

$$\left|Z_k(\mu) - Z_k(\mu')\right| \leq \tilde{C}^*_{Th}(y_k; k) W_1(\mu, \mu').$$

Proof Let π and π' be the densities of μ and μ' with respect to the Lebesgue measure, respectively. By the definition of evidence function Z_k, we can write $|Z_k(\mu) - Z_k(\mu')|$ as

$$\begin{aligned}
&\left|Z_k(\mu) - Z_k(\mu')\right| \\
&\quad = \left| \int_{\mathcal{X}\times\mathcal{W}} h_k(y_k, x, w) \left(\int_{\mathcal{X}} T_k(x, x_{k-1}, w)\pi(x_{k-1}, w) dx_{k-1} \right) dx dw \right. \\
&\qquad \left. - \int_{\mathcal{X}\times\mathcal{W}} h_k(y_k, x, w) \left(\int_{\mathcal{X}} T_k(x, x_{k-1}, w)\pi'(x_{k-1}, w) dx_{k-1} \right) dx dw \right| \\
&\quad = \left| \int_{\mathcal{X}\times\mathcal{W}} \left(\int_{\mathcal{X}} h_k(y_k, x, w) T_k(x, x_{k-1}, w) dx \right) \pi(x_{k-1}, w) dx_{k-1} dw \right. \\
&\qquad \left. - \int_{\mathcal{X}\times\mathcal{W}} \left(\int_{\mathcal{X}} h_k(y_k, x, w) T_k(x, x_{k-1}, w) dx \right) \pi'(x_{k-1}, w) dx_{k-1} dw \right|. \qquad (80)
\end{aligned}$$

The second equality in Equation (80) is obtained by Tonelli's thoerem. The expression $\int_{\mathcal{X}} h_k(y_k, x, w) T_k(x, x_{k-1}, w) dx$ in Equation (80) is a function of (x_{k-1}, w). We denote this function as g:

$$g(x_{k-1}, w) := \int_{\mathcal{X}} h_k(y_k, x, w) T_k(x, x_{k-1}, w) dx.$$

We now prove that function g is Lipschitz continuous and its best Lipschitz constant is no greater than $\tilde{C}^*_{Th}(y_k; k)$.

For two arbitrary points (x_{k-1}, w) and (x'_{k-1}, w') in space $\mathcal{X} \times \mathcal{W}$, we have

$$\begin{aligned}
&|g(x_{k-1}, w) - g(x'_{k-1}, w')| \\
&\quad = \left| \int_{\mathcal{X}} h_k(y_k, x, w) T_k(x, x_{k-1}, w) dx - \int_{\mathcal{X}} h_k(y_k, x, w') T_k(x, x'_{k-1}, w') dx \right| \\
&\quad \leq \int_{\mathcal{X}} \left| h_k(y_k, x, w) T_k(x, x_{k-1}, w) - h_k(y_k, x, w') T_k(x, x'_{k-1}, w') \right| dx \\
&\quad \leq d_{\mathcal{X}\times\mathcal{W}}((x_{k-1}, w), (x'_{k-1}, w')) \int_{\mathcal{X}} \tilde{f}_{\text{Lip}}(x; k) dx \\
&\quad = \tilde{C}^*_{Th}(y_k; k) d_{\mathcal{X}\times\mathcal{W}}((x_{k-1}, w), (x'_{k-1}, w')).
\end{aligned}$$

As (x_{k-1}, w) and (x'_{k-1}, w') are arbitrary, we have

$$\sup_{(x_{k-1},w),(x'_{k-1},w') \in \mathcal{X}\times\mathcal{W}, (x_{k-1},w) \neq (x'_{k-1},w')} \frac{|g(x_{k-1}, w) - g(x'_{k-1}, w')|}{d_{\mathcal{X}\times\mathcal{W}}((x_{k-1}, w), (x'_{k-1}, w'))} \leq \tilde{C}^*_{Th}(y_k; k).$$

By following an argument analogous to those used in the proofs of Lemmas 46 and 47, we can show that

$$\begin{aligned}
|Z_k(\mu) - Z_k(\mu')| &\leq \tilde{C}^*_{Th}(y_k; k) \sup_{f:\mathcal{X}\times\mathcal{W}\to\mathbb{R}, \|f\|_{\text{Lip}} \leq 1} \left| \int_{\mathcal{X}\times\mathcal{W}} f(x_{k-1}, w) \pi(x_{k-1}, w) dx_{k-1} dw \right. \\
&\qquad \left. - \int_{\mathcal{X}\times\mathcal{W}} f(x_{k-1}, w) \pi'(x_{k-1}, w) dx_{k-1} dw \right| \\
&= \tilde{C}^*_{Th}(y_k; k) W_1(\mu, \mu').
\end{aligned}$$

■

B.3.2 Theorem 14 under 1-Wasserstein distance: Propositions 49—51 and their proofs

Theorem 14 under the 1-Wasserstein distance is a conclusion of Propositions 49—51.

Proposition 49 *In inverse problems, given the data y_k, under Assumptions IP.1, IP.2, and IP-SE-PS.4, $\forall \mu, \mu' \in \bar{\mathcal{P}}_k(\mathcal{X}) \bigcap \mathcal{P}_1(\mathcal{X})$, the posteriors $F_k(\mu)$ and $F_k(\mu')$ satisfy*

$$W_1(F_k(\mu), F_k(\mu')) \leq \frac{C_h(y_k) + 2D\|h\|_{Lip}(y_k)}{Z_k(\mu)} W_1(\mu, \mu').$$

Proof By Kantorovich-Rubinstein duality, the 1-Wasserstein distance between the two posterior probability measures $F_k(\mu)$ and $F_k(\mu')$ can be written as

$$\begin{aligned}
W_1(F_k(\mu), F_k(\mu')) &= \sup_{f:\mathcal{X}\to\mathbb{R}, \|f\|_{\text{Lip}} \leq 1} \left| \int_{\mathcal{X}} f(x) F_k\mu(dx) - \int_{\mathcal{X}} f(x) F_k\mu'(dx) \right| \\
&= \sup_{f:\mathcal{X}\to\mathbb{R}, \|f\|_{\text{Lip}} \leq 1} \left| \int_{\mathcal{X}} f(x) \frac{h(y_k, x)}{Z_k(\mu)} \mu(dx) - \int_{\mathcal{X}} f(x) \frac{h(y_k, x)}{Z_k(\mu')} \mu'(dx) \right|.
\end{aligned}$$

We begin by following the proof strategy of Theorem 15 in Sprungk (2020), and first show that $W_1(F_k(\mu), F_k(\mu'))$ can be expressed as

$$W_1(F_k(\mu), F_k(\mu')) = \sup_{f:\mathcal{X}\to\mathbb{R}, \|f\|_{\text{Lip}}\leq 1, f(x_0)=0} \left| \int_{\mathcal{X}} f(x) \frac{h(y_k,x)}{Z_k(\mu)} \mu(dx) - \int_{\mathcal{X}} f(x) \frac{h(y_k,x)}{Z_k(\mu')} \mu'(dx) \right|. \tag{81}$$

The proof for Equation (81) is as follows.

Let x_0 denote an arbitrary point in the space $\mathcal{X}$. Define a function f_0 as:

$$f_0(x) := f(x) - f(x_0).$$

Then we have $f_0(x_0) = 0$. For two arbitrary probability measures $\mu_1, \mu_2 \in \mathcal{P}_1(\mathcal{X})$, we have

$$\begin{aligned}
&\left| \int_{\mathcal{X}} f_0(x)\mu_1(dx) - \int_{\mathcal{X}} f_0(x)\mu_2(dx) \right| \\
&\quad = \left| \int_{\mathcal{X}} (f(x) - f(x_0))\mu_1(dx) - \int_{\mathcal{X}} (f(x) - f(x_0))\mu_2(dx) \right| \\
&\quad = \left| \int_{\mathcal{X}} f(x)\mu_1(dx) - \int_{\mathcal{X}} f(x)\mu_2(dx) - f(x_0)\int_{\mathcal{X}} \mu_1(dx) + f(x_0)\int_{\mathcal{X}} \mu_2(dx) \right| \\
&\quad = \left| \int_{\mathcal{X}} f(x)\mu_1(dx) - \int_{\mathcal{X}} f(x)\mu_2(dx) - f(x_0) + f(x_0) \right| \\
&\quad = \left| \int_{\mathcal{X}} f(x)\mu_1(dx) - \int_{\mathcal{X}} f(x)\mu_2(dx) \right|.
\end{aligned}$$

Therefore, $W_1(F_k(\mu), F_k(\mu'))$ can be re-written as

$$W_1(F_k(\mu), F_k(\mu')) = \sup_{f:\mathcal{X}\to\mathbb{R}, \|f\|_{\text{Lip}}\leq 1, f(x_0)=0} \left| \int_{\mathcal{X}} f(x) \frac{h(y_k,x)}{Z_k(\mu)} \mu(dx) - \int_{\mathcal{X}} f(x) \frac{h(y_k,x)}{Z_k(\mu')} \mu'(dx) \right|.$$

It follows that

$$\begin{aligned}
&W_1(F_k(\mu), F_k(\mu')) \\
&\quad = \sup_{f:\mathcal{X}\to\mathbb{R}, \|f\|_{\text{Lip}}\leq 1, f(x_0)=0} \left| \int_{\mathcal{X}} f(x) \frac{h(y_k,x)}{Z_k(\mu)} \mu(dx) - \int_{\mathcal{X}} f(x) \frac{h(y_k,x)}{Z_k(\mu)} \mu'(dx) \right. \\
&\qquad \left. + \int_{\mathcal{X}} f(x) \frac{h(y_k,x)}{Z_k(\mu)} \mu'(dx) - \int_{\mathcal{X}} f(x) \frac{h(y_k,x)}{Z_k(\mu')} \mu'(dx) \right| \\
&\quad \leq \sup_{f:\mathcal{X}\to\mathbb{R}, \|f\|_{\text{Lip}}\leq 1, f(x_0)=0} \left[\left| \int_{\mathcal{X}} f(x) \frac{h(y_k,x)}{Z_k(\mu)} \mu(dx) - \int_{\mathcal{X}} f(x) \frac{h(y_k,x)}{Z_k(\mu)} \mu'(dx) \right| \right. \\
&\qquad \left. + \left| \int_{\mathcal{X}} f(x) \frac{h(y_k,x)}{Z_k(\mu)} \mu'(dx) - \int_{\mathcal{X}} f(x) \frac{h(y_k,x)}{Z_k(\mu')} \mu'(dx) \right| \right] \\
&\quad \leq \sup_{f:\mathcal{X}\to\mathbb{R}, \|f\|_{\text{Lip}}\leq 1, f(x_0)=0} \left| \int_{\mathcal{X}} f(x) \frac{h(y_k,x)}{Z_k(\mu)} \mu(dx) - \int_{\mathcal{X}} f(x) \frac{h(y_k,x)}{Z_k(\mu)} \mu'(dx) \right| \\
&\qquad + \sup_{f:\mathcal{X}\to\mathbb{R}, \|f\|_{\text{Lip}}\leq 1, f(x_0)=0} \left| \int_{\mathcal{X}} f(x) \frac{h(y_k,x)}{Z_k(\mu)} \mu'(dx) - \int_{\mathcal{X}} f(x) \frac{h(y_k,x)}{Z_k(\mu')} \mu'(dx) \right|.
\end{aligned} \tag{82}$$

The first term on the right-hand side in Equation (82) can be written as

$$\begin{aligned}&\sup_{f:\mathcal{X}\to\mathbb{R},\|f\|_{\mathrm{Lip}}\leq 1,f(x_0)=0}\left|\int_{\mathcal{X}} f(x)\frac{h(y_k,x)}{Z_k(\mu)}\mu(dx)-\int_{\mathcal{X}} f(x)\frac{h(y_k,x)}{Z_k(\mu)}\mu'(dx)\right|\\&=\frac{1}{Z_k(\mu)}\sup_{f:\mathcal{X}\to\mathbb{R},\|f\|_{\mathrm{Lip}}\leq 1,f(x_0)=0}\left|\int_{\mathcal{X}} f(x)h(y_k,x)\mu(dx)-\int_{\mathcal{X}} f(x)h(y_k,x)\mu'(dx)\right|.\end{aligned}$$

Next, we use the proof strategy used in the proof of Theorem 15 in Sprungk (2020) to show that

$$\begin{aligned}&\sup_{f:\mathcal{X}\to\mathbb{R},\|f\|_{\mathrm{Lip}}\leq 1,f(x_0)=0}\left|\int_{\mathcal{X}} f(x)h(y_k,x)\mu(dx)-\int_{\mathcal{X}} f(x)h(y_k,x)\mu'(dx)\right|\\&\qquad\leq\left(D\|h\|_{\mathrm{Lip}}(y_k)+C_h(y_k)\right)W_1(\mu,\tilde{\mu}).\end{aligned}$$

The proof is as follows.

Define a function $g:\mathcal{X}\to\mathbb{R}$ as

$$g(x):=f(x)h(y_k,x).$$

Then g satisfies

$$g(x_0)=f(x_0)h(y_k,x_0)=0.$$

Next, we prove that g is Lipschitz continuous and its best Lipschitz constant is no greater than $D\|h\|_{\mathrm{Lip}}(y_k)+C_h(y_k)$.

For two arbitrary points $x_1,x_2\in\mathcal{X}$, $|g(x_1)-g(x_2)|$ satisfies

$$\begin{aligned}&|g(x_1)-g(x_2)|\\&\quad=|f(x_1)h(y_k,x_1)-f(x_2)h(y_k,x_2)|\\&\quad\leq|f(x_1)h(y_k,x_1)-f(x_1)h(y_k,x_2)|+|f(x_1)h(y_k,x_2)-f(x_2)h(y_k,x_2)|\\&\quad=|f(x_1)||h(y_k,x_1)-h(y_k,x_2)|+h(y_k,x_2)|f(x_1)-f(x_2)|\\&\quad\leq|f(x_1)||h(y_k,x_1)-h(y_k,x_2)|+C_h(y_k)|f(x_1)-f(x_2)|.\end{aligned}\tag{83}$$

Given $\|f\|_{\mathrm{Lip}}\leq 1$ and $f(x_0)=0$, we have

$$\begin{aligned}|f(x_1)|=&|f(x_1)-f(x_0)+f(x_0)|\leq|f(x_1)-f(x_0)|+|f(x_0)|\\=&|f(x_1)-f(x_0)|\leq d(x_1,x_0)\leq D,\end{aligned}\tag{84}$$

and

$$|f(x_1)-f(x_2)|\leq d_{\mathcal{X}}(x_1,x_2).\tag{85}$$

The term $|h(y_k,x_1)-h(y_k,x_2)|$ satisfies

$$|h(y_k,x_1)-h(y_k,x_2)|\leq\|h\|_{\mathrm{Lip}}(y_k)d_{\mathcal{X}}(x_1,x_2).\tag{86}$$

By substituting Equation 84, 85 and 86 into Equation 83, we obtain

$$\begin{aligned}|g(x_1)-g(x_2)|\leq&D\|h\|_{\mathrm{Lip}}(y_k)d_{\mathcal{X}}(x_1,x_2)+C_h(y_k)d_{\mathcal{X}}(x_1,x_2)\\=&\left(D\|h\|_{\mathrm{Lip}}(y_k)+C_h(y_k)\right)d_{\mathcal{X}}(x_1,x_2).\end{aligned}$$

It follows that

$$\frac{|g(x_1)-g(x_2)|}{d_{\mathcal{X}}(x_1,x_2)} \leq D\|h\|_{\mathrm{Lip}}(y_k) + C_h(y_k), \quad \forall x_1, x_2 \in \mathcal{X}, x_1 \neq x_2.$$

Therefore, the function g is Lipschitz continuous and we have

$$\|g\|_{\mathrm{Lip}} = \sup_{x_1,x_2\in\mathcal{X},x_1\neq x_2} \frac{|g(x_1)-g(x_2)|}{d_{\mathcal{X}}(x_1,x_2)} \leq D\|h\|_{\mathrm{Lip}}(y_k) + C_h(y_k).$$

Define a function $\tilde{f}$ as

$$\tilde{f}(x) = \frac{g(x)}{D\|h\|_{\mathrm{Lip}}(y_k) + C_h(y_k)}.$$

Then $\tilde{f}$ is Lipschitz continuous and it satisfies

$$\tilde{f}(x_0) = \frac{g(x_0)}{D\|h\|_{\mathrm{Lip}}(y_k) + C_h(y_k)} = 0,$$

and

$$\|\tilde{f}\|_{\mathrm{Lip}} = \sup_{x_1,x_2\in\mathcal{X},x_1\neq x_2} \frac{1}{D\|h\|_{\mathrm{Lip}}(y_k) + C_h(y_k)} \frac{|g(x_1)-g(x_2)|}{d_{\mathcal{X}}(x_1,x_2)} \leq 1.$$

Then we have

$$\begin{aligned}
&\sup_{f:\mathcal{X}\to\mathbb{R},\|f\|_{\mathrm{Lip}}\leq 1, f(x_0)=0} \left| \int_{\mathcal{X}} f(x)h(y_k,x)\mu(dx) - \int_{\mathcal{X}} f(x)h(y_k,x)\mu'(dx) \right| \\
&\quad = (D\|h\|_{\mathrm{Lip}}(y_k) + C_h(y_k)) \sup_{f:\mathcal{X}\to\mathbb{R},\|f\|_{\mathrm{Lip}}\leq 1, f(x_0)=0} \left| \int_{\mathcal{X}} \tilde{f}(x)\mu(dx) - \int_{\mathcal{X}} \tilde{f}(x)\mu'(dx) \right| \\
&\quad = (D\|h\|_{\mathrm{Lip}}(y_k) + C_h(y_k)) \sup_{\tilde{f}\in\tilde{A}} \left| \int_{\mathcal{X}} \tilde{f}(x)\mu(dx) - \int_{\mathcal{X}} \tilde{f}(x)\mu'(dx) \right|,
\end{aligned}$$

where

$$\begin{aligned}
\tilde{A} := \Big\{ \tilde{f} : \mathcal{X} \to \mathbb{R} \mid \|\tilde{f}\|_{\mathrm{Lip}} \leq 1, \tilde{f}(x_0) = 0, \tilde{f}(x) = \frac{h(y_k,x)f(x)}{D\|h\|_{\mathrm{Lip}}(y_k) + C_h(y_k)}, \\
\|f\|_{\mathrm{Lip}} \leq 1, f(x_0) = 0 \Big\}.
\end{aligned}$$

Note that

$$\tilde{A} \subseteq \{ f : \mathcal{X} \to \mathbb{R} \mid \|f\|_{\mathrm{Lip}} \leq 1, f(x_0) = 0 \}.$$

Therefore we have

$$\begin{aligned}
&\sup_{\tilde{f}\in\tilde{A}} \left| \int_{\mathcal{X}} \tilde{f}(x)\mu(dx) - \int_{\mathcal{X}} \tilde{f}(x)\mu'(dx) \right| \\
&\quad \leq \sup_{f:\mathcal{X}\to\mathbb{R},\|f\|_{\mathrm{Lip}}\leq 1, f(x_0)=0} \left| \int_{\mathcal{X}} f(x)\mu(dx) - \int_{\mathcal{X}} f(x)\mu'(dx) \right| \\
&\quad = W_1(\mu,\mu').
\end{aligned}$$

It follows that

$$\sup_{f:\mathcal{X}\to\mathbb{R},\|f\|_{\mathrm{Lip}}\leq 1,f(x_0)=0}\left|\int_{\mathcal{X}} f(x)h(y_k,x)\mu(dx)-\int_{\mathcal{X}} f(x)h(y_k,x)\mu'(dx)\right| \leq \left(D\|h\|_{\mathrm{Lip}}(y_k)+C_h(y_k)\right)W_1(\mu,\mu'). \tag{87}$$

So far, we have derived an upper bound for the first term in Equation (82) using the proof strategy used for Theorem 15 in Sprungk (2020).

Next, we derive an upper bound for the second term in Equation (82). We have

$$\begin{aligned}&\sup_{f:\mathcal{X}\to\mathbb{R},\|f\|_{\mathrm{Lip}}\leq 1,f(x_0)=0}\left|\int_{\mathcal{X}} f(x)\frac{h(y_k,x)}{Z_k(\mu)}\mu'(dx)-\int_{\mathcal{X}} f(x)\frac{h(y_k,x)}{Z_k(\mu')}\mu'(dx)\right| \\ &= \sup_{f:\mathcal{X}\to\mathbb{R},\|f\|_{\mathrm{Lip}}\leq 1,f(x_0)=0}\left|\left(\frac{1}{Z_k(\mu)}-\frac{1}{Z_k(\mu')}\right)\int_{\mathcal{X}} f(x)h(y_k,x)\mu'(dx)\right| \\ &= \frac{\left|Z_k(\mu)-Z_k(\mu')\right|}{Z_k(\mu)Z_k(\mu')}\sup_{f:\mathcal{X}\to\mathbb{R},\|f\|_{\mathrm{Lip}}\leq 1,f(x_0)=0}\left|\int_{\mathcal{X}} f(x)h(y_k,x)\mu'(dx)\right|.\end{aligned} \tag{88}$$

We first note that the expression $\sup_{f:\mathcal{X}\to\mathbb{R},\|f\|_{\mathrm{Lip}}\leq 1,f(x_0)=0}\left|\int_{\mathcal{X}} f(x)h(y_k,x)\mu'(dx)\right|$ satisfies

$$\begin{aligned}&\sup_{f:\mathcal{X}\to\mathbb{R},\|f\|_{\mathrm{Lip}}\leq 1,f(x_0)=0}\left|\int_{\mathcal{X}} f(x)h(y_k,x)\mu'(dx)\right| \\ &\leq \sup_{f:\mathcal{X}\to\mathbb{R},\|f\|_{\mathrm{Lip}}\leq 1,f(x_0)=0}\int_{\mathcal{X}}\Big|f(x)\Big|h(y_k,x)\mu'(dx).\end{aligned}$$

Recall that for any $x\in\mathcal{X}$, the expression $|f(x)|$ satisfies $|f(x)|\leq D$. Therefore we have

$$\begin{aligned}&\sup_{f:\mathcal{X}\to\mathbb{R},\|f\|_{\mathrm{Lip}}\leq 1,f(x_0)=0}\int_{\mathcal{X}}\Big|f(x)\Big|h(y_k,x)\mu'(dx) \\ &\leq D\sup_{f:\mathcal{X}\to\mathbb{R},\|f\|_{\mathrm{Lip}}\leq 1,f(x_0)=0}\int_{\mathcal{X}} h(y_k,x)\mu'(dx) \\ &= D\int_{\mathcal{X}} h(y_k,x)\mu'(dx)=DZ_k(\mu').\end{aligned} \tag{89}$$

Substituting Equation (89) into Equation (88) gives

$$\sup_{f:\mathcal{X}\to\mathbb{R},\|f\|_{\mathrm{Lip}}\leq 1,f(x_0)=0}\left|\int_{\mathcal{X}} f(x)\frac{h(y_k,x)}{Z_k(\mu)}\mu'(dx)-\int_{\mathcal{X}} f(x)\frac{h(y_k,x)}{Z_k(\mu')}\mu'(dx)\right| \leq \frac{\left|Z_k(\mu)-Z_k(\mu')\right|}{Z_k(\mu)}D. \tag{90}$$

By Lemma 46, $\left|Z_k(\mu)-Z_k(\mu')\right|$ is bounded by:

$$\left|Z_k(\mu)-Z_k(\mu')\right|\leq\|h\|_{\mathrm{Lip}}(y_k)W_1(\mu,\mu'). \tag{91}$$

Plug Equation (91) into Equation (90). We obtain

$$\sup_{f:\mathcal{X}\to\mathbb{R},\|f\|_{\text{Lip}}\leq 1,f(x_0)=0}\left|\int_{\mathcal{X}} f(x)\frac{h(y_k,x)}{Z_k(\mu)}\mu'(dx)-\int_{\mathcal{X}} f(x)\frac{h(y_k,x)}{Z_k(\mu')}\mu'(dx)\right| \leq \frac{D\|h\|_{\text{Lip}}(y_k)}{Z_k(\mu)}W_1(\mu,\mu'). \tag{92}$$

Combining Equation (87) with Equation (92) gives

$$W_1(F_k(\mu),F_k(\mu')) \leq \frac{2D\|h\|_{\text{Lip}}(y_k)+C_h(y_k)}{Z_k(\mu)}W_1(\mu,\mu').$$

■

Proposition 50 *In state estimation problems, given the data y_k, under Assumption SE.2 and IP-SE-PS.4, $\forall \mu,\mu' \in \bar{\mathcal{P}}_k(\mathcal{X})\bigcap\mathcal{P}_1(\mathcal{X})$, the posteriors $F_k(\mu)$ and $F_k(\mu')$ satisfy*

$$W_1(F_k(\mu),F_k(\mu')) \leq \frac{2C^*_{Th}(y_k;k)D}{Z_k(\mu)}W_1(\mu,\mu').$$

The high-level proof strategy for Proposition 50 follows that of Proposition 49, with modifications to accommodate the more complex setting of state estimation problems compared to inverse problems.

Proof Analogous to the proof of Proposition 49, we can show that the 1-Wasserstein distance between the two posteriors $F_k(\mu)$ and $F_k(\mu')$ satisfies

$$\begin{aligned} W_1(F_k(\mu),F_k(\mu')) \leq \frac{1}{Z_k(\mu)}&\sup_{f:\mathcal{X}\to\mathbb{R},\|f\|_{\text{Lip}}\leq 1,f(\bar{x})=0}\left|\int_{\mathcal{X}} f(x)\tilde{F}_k\mu(dx)-\int_{\mathcal{X}} f(x)\tilde{F}_k\mu'(dx)\right| \\ &+\frac{|Z_k(\mu)-Z_k(\mu')|}{Z_k(\mu)Z_k(\mu')}\sup_{f:\mathcal{X}\to\mathbb{R},\|f\|_{\text{Lip}}\leq 1,f(\bar{x})=0}\left|\int_{\mathcal{X}} f(x)\tilde{F}_k\mu'(dx)\right|, \end{aligned} \tag{93}$$

where $\bar{x}\in\mathcal{X}$ is an arbitrary point.

We begin by bounding the first term on the right-hand side of Equation (93), which can be expressed as:

$$\begin{aligned} &\sup_{f:\mathcal{X}\to\mathbb{R},\|f\|_{\text{Lip}}\leq 1,f(\bar{x})=0}\left|\int_{\mathcal{X}} f(x)\tilde{F}_k\mu(dx)-\int_{\mathcal{X}} f(x)\tilde{F}_k\mu'(dx)\right| \\ &\quad=\sup_{f:\mathcal{X}\to\mathbb{R},\|f\|_{\text{Lip}}\leq 1,f(\bar{x})=0}\left|\int_{\mathcal{X}} f(x)h_k(y_k,x)\left(\int_{\mathcal{X}} T_k(x,x_{k-1})\mu(dx_{k-1})\right)dx\right. \\ &\qquad\left.-\int_{\mathcal{X}} f(x)h_k(y_k,x)\left(\int_{\mathcal{X}} T_k(x,x_{k-1})\mu'(dx_{k-1})\right)dx\right|. \end{aligned} \tag{94}$$

Analogous to the proof of Proposition 49, we have $|f(x)| \leq D$. It follows that

$$\begin{aligned}
&\int_{\mathcal{X}\times\mathcal{X}} \left| f(x)h_k(y_k,x)T_k(x,x_{k-1}) \right| \mu(dx_{k-1})dx \\
&\quad = \int_{\mathcal{X}\times\mathcal{X}} \left|f(x)\right| h_k(y_k,x)T_k(x,x_{k-1})\mu(dx_{k-1})dx \\
&\quad = \int_{\mathcal{X}} \left|f(x)\right| h_k(y_k,x) \left(\int_{\mathcal{X}} T_k(x,x_{k-1})\mu(dx_{k-1}) \right) dx \\
&\quad \leq D \int_{\mathcal{X}} h_k(y_k,x) \left(\int_{\mathcal{X}} T_k(x,x_{k-1})\mu(dx_{k-1}) \right) dx \\
&\quad = DZ_k(\mu) < \infty,
\end{aligned}$$

where the second equality is obtained by Tonelli's theorem and the last inequality is obtained because $\mu \in \bar{\mathcal{P}}_k(\mathcal{X})$.
Therefore, the function $f(x)h_k(y_k,x)T_k(x,x_{k-1})$ is integrable and we can use Fubini's theorem to obtain

$$\begin{aligned}
&\int_{\mathcal{X}} f(x)h_k(y_k,x) \left(\int_{\mathcal{X}} T_k(x,x_{k-1})\mu(dx_{k-1}) \right) dx \\
&\quad = \int_{\mathcal{X}} \left(\int_{\mathcal{X}} f(x)h_k(y_k,x)T_k(x,x_{k-1})dx \right) \mu(dx_{k-1}).
\end{aligned} \tag{95}$$

Similarly, we have

$$\begin{aligned}
&\int_{\mathcal{X}} f(x)h_k(y_k,x) \left(\int_{\mathcal{X}} T_k(x,x_{k-1})\mu'(dx_{k-1})) \right) dx \\
&\quad = \int_{\mathcal{X}} \left(\int_{\mathcal{X}} f(x)h_k(y_k,x)T_k(x,x_{k-1})dx \right) \mu'(dx_{k-1}).
\end{aligned} \tag{96}$$

Substituting Equations (95) and (96) into Equation (94) yields:

$$\begin{aligned}
&\sup_{f:\mathcal{X}\to\mathbb{R},\|f\|_{\mathrm{Lip}}\leq 1, f(\bar{x})=0} \left| \int_{\mathcal{X}} f(x)\tilde{F}_k\mu(dx) - \int_{\mathcal{X}} f(x)\tilde{F}_k\mu'(dx) \right| \\
&\quad = \sup_{f:\mathcal{X}\to\mathbb{R},\|f\|_{\mathrm{Lip}}\leq 1, f(\bar{x})=0} \left| \int_{\mathcal{X}} \left(\int_{\mathcal{X}} f(x)h_k(y_k,x)T_k(x,x_{k-1})dx \right) \mu(dx_{k-1}) \right. \\
&\qquad \left. - \int_{\mathcal{X}} \left(\int_{\mathcal{X}} f(x)h_k(y_k,x)T_k(x,x_{k-1})dx \right) \mu'(dx_{k-1}) \right|.
\end{aligned} \tag{97}$$

The expression $\int_{\mathcal{X}} f(x)h_k(y_k,x)T_k(x,x_{k-1})dx$ is a function of x_{k-1}. We denote this function as $\tilde{f}$, i.e.,

$$\tilde{f}(x_{k-1}) := \int_{\mathcal{X}} f(x)h_k(y_k,x)T_k(x,x_{k-1})dx.$$

Next, we prove that $\tilde{f}$ is Lipschitz continuous and its best Lipschitz constant is smaller than or equal to $DC^*_{Th}(y_k;k)$. For two arbitrary points x_{k-1}, $x'_{k-1} \in \mathcal{X}$, we have

$$\begin{aligned}|\tilde{f}(x_{k-1}) - \tilde{f}(x'_{k-1})| =& \left| \int_{\mathcal{X}} f(x)h_k(y_k,x)T_k(x,x_{k-1})dx - \int_{\mathcal{X}} f(x)h_k(y_k,x)T_k(x,x'_{k-1})dx \right| \\ =& \left| \int_{\mathcal{X}} f(x)h_k(y_k,x)\left(T_k(x,x_{k-1}) - T_k(x,x'_{k-1})\right)dx \right| \\ \leq& \int_{\mathcal{X}} \left| f(x) \right| \left| h_k(y_k,x) \right| \left| T_k(x,x_{k-1}) - T_k(x,x'_{k-1}) \right| dx.\end{aligned}$$

As we also have $|f(x)| \leq D$, and

$$|T_k(x,x_{k-1}) - T_k(x,x'_{k-1})| \leq T_{\text{Lip}}(x;k)d_{\mathcal{X}}(x_{k-1},x'_{k-1}),$$

the expression $|\tilde{f}(x_{k-1}) - \tilde{f}(x'_{k-1})|$ satisfies

$$\begin{aligned}|\tilde{f}(x_{k-1}) - \tilde{f}(x'_{k-1})| \leq& Dd(x_{k-1},x'_{k-1}) \int_{\mathcal{X}} h_k(y_k,x)T_{\text{Lip}}(x;k)dx \\ =& DC^*_{Th}(y_k;k)d_{\mathcal{X}}(x_{k-1},x'_{k-1}).\end{aligned}$$

It follows that

$$\sup_{x_{k-1},x'_{k-1}\in\mathcal{X},x_{k-1}\neq x'_{k-1}} \frac{|\tilde{f}(x_{k-1}) - \tilde{f}(x'_{k-1})|}{d_{\mathcal{X}}(x_{k-1},x'_{k-1})} \leq DC^*_{Th}(y_k;k).$$

Define a function g as

$$g(x) := \frac{\tilde{f}(x)}{DC^*_{Th}(y_k;k)}.$$

Then g is Lipschitz continuous with the best Lipschitz constant

$$\|g\|_{\text{Lip}} = \frac{1}{DC^*_{Th}(y_k;k)} \sup_{x,x'\in\mathcal{X},x\neq x'} \frac{|\tilde{f}(x) - \tilde{f}(x')|}{d_{\mathcal{X}}(x,x')} \leq 1.$$

Therefore, we have

$$\begin{aligned}&\sup_{f:\mathcal{X}\to\mathbb{R},\|f\|_{\text{Lip}}\leq 1,f(\bar{x})=0} \left| \int_{\mathcal{X}} f(x)\tilde{F}_k\mu(dx) - \int_{\mathcal{X}} f(x)\tilde{F}_k\mu'(dx) \right| \\ &\quad = DC^*_{Th}(y_k;k) \sup_{g\in\hat{A}} \left| \int_{\mathcal{X}} g(x_{k-1})\mu(dx_{k-1}) - \int_{\mathcal{X}} g(x_{k-1})\mu'(dx_{k-1}) \right|. \end{aligned} \tag{98}$$

where

$$\begin{aligned}\hat{A} =& \Big\{ g : \mathcal{X} \to \mathbb{R} \mid \|g\|_{\text{Lip}} \leq 1, \\ & g(x_{k-1}) = \frac{1}{DC^*_{Th}(y_k;k)} \int_{\mathcal{X}} f(x)h_k(y_k,x)T_k(x,x_{k-1})dx, f \in A \Big\},\end{aligned}$$

with

$$A = \{f : \mathcal{X} \to \mathbb{R} \mid \|f\|_{\text{Lip}} \leq 1, f(\bar{x}) = 0\}.$$

As $\hat{A} \subseteq \{f : \mathcal{X} \to \mathbb{R} \mid \|f\|_{\text{Lip}} \leq 1\}$, we have

$$\begin{aligned}
&\sup_{g \in \hat{A}} \left| \int_{\mathcal{X}} g(x_{k-1}) \mu(dx_{k-1}) - \int_{\mathcal{X}} g(x_{k-1}) \mu'(dx_{k-1}) \right| \\
&\quad \leq \sup_{f:\mathcal{X}\to\mathbb{R}, \|f\|_{\text{Lip}} \leq 1} \left| \int_{\mathcal{X}} f(x_{k-1}) \mu(dx_{k-1}) - \int_{\mathcal{X}} f(x_{k-1}) \mu'(dx_{k-1}) \right| \\
&\quad = W_1(\mu, \mu'). \qquad (99)
\end{aligned}$$

Analogous to the proof of Proposition 49, the second term on the right-hand side of Equation (93) satisfies

$$\frac{|Z_k(\mu) - Z_k(\mu')|}{Z_k(\mu) Z_k(\mu')} \sup_{f:\mathcal{X}\to\mathbb{R}, \|f\|_{\text{Lip}} \leq 1, f(\bar{x})=0} \left| \int_{\mathcal{X}} f(x) \tilde{F}_k \mu'(dx) \right| \leq \frac{D|Z_k(\mu) - Z_k(\mu')|}{Z_k(\mu)}. \qquad (100)$$

By plugging Equations (98), (99), and (100) into Equation (93), we obtain

$$W_1(F_k(\mu), F_k(\mu')) \leq \frac{DC^*_{Th}(y_k; k)}{Z_k(\mu)} W_1(\mu, \mu') + \frac{D|Z_k(\mu) - Z_k(\mu')|}{Z_k(\mu)}.$$

By Lemma 47, the term $|Z_k(\mu) - Z_k(\mu')|$ satisfies

$$|Z_k(\mu) - Z_k(\mu')| \leq C^*_{Th}(y_k; k) W_1(\mu, \mu').$$

It follows that

$$W_1\left(F_k(\mu), F_k(\mu')\right) \leq \frac{2DC^*_{Th}(y_k; k)}{Z_k(\mu)} W_1(\mu, \mu').$$

■

Proposition 51 *In parameter-state estimation problems, given the data y_k, under Assumptions PS.1, PS.2, PS.3 and IP-SE-PS.4, $\forall \mu, \mu' \in \bar{\mathcal{P}}_k(\mathcal{X} \times \mathcal{W}) \bigcap \mathcal{P}_1(\mathcal{X} \times \mathcal{W})$, the posteriors $F_k(\mu)$ and $F_k(\mu')$ satisfy*

$$W_1(F_k(\mu), F_k(\mu')) \leq \frac{2\tilde{C}^*_{Th}(y_k; k) D + \tilde{C}_{Th}(y_k; k)}{Z_k(\mu)} W_1(\mu, \mu').$$

The high-level proof strategy for Proposition 51 are similar to those of Propositions 49 and 50, with modifications to adapt to the increased complexity of parameter-state estimation problems compared to inverse and state estimation problems.

Proof Analogous to the proof of Proposition 49, we can show that the 1-Wasserstein distance between the two posteriors $F_k(\mu)$ and $F_k(\mu')$ satisfies

$$\begin{aligned}
&W_1(F_k(\mu), F_k(\mu')) \\
&\quad \leq \frac{1}{Z_k(\mu)} \sup_{f:\mathcal{X}\times\mathcal{W}\to\mathbb{R}, \|f\|_{\text{Lip}}\leq 1, f(\bar{x},\bar{w})=0} \left| \int_{\mathcal{X}\times\mathcal{W}} f(x,w)\tilde{F}_k\mu(dx,dw) \right. \\
&\qquad \left. - \int_{\mathcal{X}\times\mathcal{W}} f(x,w)\tilde{F}_k\mu'(dx,dw) \right| \\
&\qquad + \frac{|Z_k(\mu) - Z_k(\mu')|}{Z_k(\mu)Z_k(\mu')} \sup_{f:\mathcal{X}\times\mathcal{W}\to\mathbb{R}, \|f\|_{\text{Lip}}\leq 1, f(\bar{x},\bar{w})=0} \left| \int_{\mathcal{X}\times\mathcal{W}} f(x,w)\tilde{F}_k\mu'(dx,dw) \right|.
\end{aligned} \tag{101}$$

where $(\bar{x}, \bar{w})$ is an arbitrary point in the space $\mathcal{X} \times \mathcal{W}$. We first derive an upper bound on the first term on the right-hand side of Equation (101).

Let π and π' denote the densities of μ and μ' with respect to the Lebesgue measure, respectively. We have

$$\begin{aligned}
&\sup_{f\in A} \left| \int_{\mathcal{X}\times\mathcal{W}} f(x,w)\tilde{F}_k\mu(dx,dw) - \int_{\mathcal{X}\times\mathcal{W}} f(x,w)\tilde{F}_k\mu'(dx,dw) \right| \\
&\quad = \sup_{f\in A} \left| \int_{\mathcal{X}\times\mathcal{W}} f(x,w)h_k(y_k,x,w) \left(\int_{\mathcal{X}} T_k(x,x_{k-1},w)\pi(x_{k-1},w)dx_{k-1} \right) dxdw \right. \\
&\qquad \left. - \int_{\mathcal{X}\times\mathcal{W}} f(x,w)h_k(y_k,x,w) \left(\int_{\mathcal{X}} T_k(x,x_{k-1},w)\pi'(x_{k-1},w)dx_{k-1} \right) dxdw \right|,
\end{aligned} \tag{102}$$

where

$$A = \{f : \mathcal{X} \times \mathcal{W} \to \mathbb{R} \mid \|f\|_{\text{Lip}} \leq 1, f(\bar{x}, \bar{w}) = 0\}.$$

Analogous to the proof of Proposition 49, we have

$$|f(x,w)| \leq D, \quad \forall x \in \mathcal{X}, w \in \mathcal{W}\,.$$

It follows that

$$\begin{aligned}
&\int_{\mathcal{X}\times\mathcal{W}\times\mathcal{X}} \left| f(x,w)h_k(y_k,x,w)T_k(x,x_{k-1},w)\pi(x_{k-1},w) \right| dxdwdx_{k-1} \\
&\quad = \int_{\mathcal{X}\times\mathcal{W}\times\mathcal{X}} \left| f(x,w) \right| h_k(y_k,x,w)T_k(x,x_{k-1},w)\pi(x_{k-1},w)dxdwdx_{k-1} \\
&\quad = \int_{\mathcal{X}\times\mathcal{W}} \left| f(x,w) \right| h_k(y_k,x,w) \left(\int_{\mathcal{X}} T_k(x,x_{k-1},w)\pi(x_{k-1},w)dx_{k-1} \right) dxdw \\
&\quad \leq D \int_{\mathcal{X}\times\mathcal{W}} h_k(y_k,x,w) \left(\int_{\mathcal{X}} T_k(x,x_{k-1},w)\pi(x_{k-1},w)dx_{k-1} \right) dxdw \\
&\quad = DZ_k(\mu) < \infty.
\end{aligned} \tag{103}$$

The second equality in Equation (103) is obtained by Tonelli's theorem and the last inequality in Equation (103) is obtained since $\mu \in \bar{\mathcal{P}}_k(\mathcal{X} \times \mathcal{W})$.

Then we can use Fubini's theorem to re-write the first term on the right-hand side of Equation (102) as

$$\int_{\mathcal{X}\times\mathcal{W}} f(x,w)h_k(y_k,x,w)\left(\int_{\mathcal{X}} T_k(x,x_{k-1},w)\pi(x_{k-1},w)dx_{k-1}\right)dxdw$$
$$= \int_{\mathcal{X}\times\mathcal{W}}\left(\int_{\mathcal{X}} f(x,w)h_k(y_k,x,w)T_k(x,x_{k-1},w)dx\right)\pi(x_{k-1},w)dx_{k-1}dw. \tag{104}$$

Similarly, the second term on the right-hand side of Equation (102) can be written as:

$$\int_{\mathcal{X}\times\mathcal{W}} f(x,w)h_k(y_k,x,w)\left(\int_{\mathcal{X}} T_k(x,x_{k-1},w)\pi'(x_{k-1},w)dx_{k-1}\right)dxdw$$
$$= \int_{\mathcal{X}\times\mathcal{W}}\left(\int_{\mathcal{X}} f(x,w)h_k(y_k,x,w)T_k(x,x_{k-1},w)dx\right)\pi'(x_{k-1},w)dx_{k-1}dw. \tag{105}$$

By plugging Equations (104) and (105) into Equation (102), we obtain:

$$\sup_{f\in A}\left|\int_{\mathcal{X}\times\mathcal{W}} f(x,w)\tilde{F}_k\mu(dx,dw) - \int_{\mathcal{X}\times\mathcal{W}} f(x,w)\tilde{F}_k\mu'(dx,dw)\right|$$
$$= \sup_{f\in A}\left|\int_{\mathcal{X}\times\mathcal{W}}\left(\int_{\mathcal{X}} f(x,w)h_k(y_k,x,w)T_k(x,x_{k-1},w)dx\right)\pi(x_{k-1},w)dx_{k-1}dw\right.$$
$$\left. - \int_{\mathcal{X}\times\mathcal{W}}\left(\int_{\mathcal{X}} f(x,w)h_k(y_k,x,w)T_k(x,x_{k-1},w)dx\right)\pi'(x_{k-1},w)dx_{k-1}dw\right|. \tag{106}$$

The expression $\int_{\mathcal{X}} f(x,w)h_k(y_k,x,w)T_k(x,x_{k-1},w)dx$ is a function of (x_{k-1},w), which we denote by $\tilde{f}$:

$$\tilde{f}(x_{k-1},w) := \int_{\mathcal{X}} f(x,w)h_k(y_k,x,w)T_k(x,x_{k-1},w)dx.$$

We now show that $\tilde{f}$ is Lipschitz continuous and its best Lipschitz constant is no greater than $\tilde{C}^*_{Th}(y_k;k)D + \tilde{C}_{Th}(y_k;k)$.

For two arbitrary points (x_{k-1},w) and (x'_{k-1},w') in the space $\mathcal{X}\times\mathcal{W}$, we have

$$\left|\tilde{f}(x_{k-1},w) - \tilde{f}(x'_{k-1},w')\right|$$
$$= \left|\int_{\mathcal{X}} f(x,w)h_k(y_k,x,w)T_k(x,x_{k-1},w)dx - \int_{\mathcal{X}} f(x,w')h_k(y_k,x,w')T_k(x,x'_{k-1},w')dx\right|$$
$$\leq \left|\int_{\mathcal{X}} f(x,w)\left(h_k(y_k,x,w)T_k(x,x_{k-1},w) - h_k(y_k,x,w')T_k(x,x'_{k-1},w')\right)dx\right|$$
$$+ \left|\int_{\mathcal{X}}\left(f(x,w)-f(x,w')\right)h_k(y_k,x,w')T_k(x,x'_{k-1},w')dx\right|. \tag{107}$$

We first show that the first term on the right-hand side of Equation (107) is bounded by the product of a constant and $d_{\mathcal{X}\times\mathcal{W}}((x_{k-1},w),(x'_{k-1},w'))$. We have

$$\left|\int_{\mathcal{X}} f(x,w)\left(h_k(y_k,x,w)T_k(x,x_{k-1},w) - h_k(y_k,x,w')T_k(x,x'_{k-1},w')\right)dx\right|$$
$$\leq \int_{\mathcal{X}}\left|f(x,w)\right|\left|h_k(y_k,x,w)T_k(x,x_{k-1},w) - h_k(y_k,x,w')T_k(x,x'_{k-1},w')\right|dx.$$

Recall that $|f(x,w)| \leq D$ for any $x \in \mathcal{X}, w \in \mathcal{W}$. Hence we have

$$\begin{aligned}
&\int_{\mathcal{X}} \left|f(x,w)\right| \left|h_k(y_k,x,w)T_k(x,x_{k-1},w) - h_k(y_k,x,w')T_k(x,x'_{k-1},w')\right| dx \\
&\quad \leq D \int_{\mathcal{X}} \left|h_k(y_k,x,w)T_k(x,x_{k-1},w) - h_k(y_k,x,w')T_k(x,x'_{k-1},w')\right| dx \\
&\quad \leq D \int_{\mathcal{X}} \tilde{f}_{\text{Lip}}(x;k) d_{\mathcal{X}\times\mathcal{W}}((x_{k-1},w),(x'_{k-1},w'))dx \\
&\quad = D d_{\mathcal{X}\times\mathcal{W}}((x_{k-1},w),(x'_{k-1},w')) \int_{\mathcal{X}} \tilde{f}_{\text{Lip}}(x;k)dx \\
&\quad = D\tilde{C}^*_{Th}(y_k;k) d_{\mathcal{X}\times\mathcal{W}}((x_{k-1},w),(x'_{k-1},w')). \qquad (108)
\end{aligned}$$

Next, we prove that the second term on the right-hand side of Equation (107) is also bounded by the product of a constant and $d_{\mathcal{X}\times\mathcal{W}}((x_{k-1},w),(x'_{k-1},w'))$.

$$\begin{aligned}
&\left| \int_{\mathcal{X}} \left(f(x,w) - f(x,w')\right) h_k(y_k,x,w')T_k(x,x'_{k-1},w')dx \right| \\
&\quad \leq \int_{\mathcal{X}} \left|f(x,w) - f(x,w')\right| h_k(y_k,x,w')T_k(x,x'_{k-1},w')dx \\
&\quad \leq \int_{\mathcal{X}} d_{\mathcal{X}\times\mathcal{W}}((x,w),(x,w'))h_k(y_k,x,w')T_k(x,x'_{k-1},w')dx \\
&\quad \leq \int_{\mathcal{X}} d_{\mathcal{X}\times\mathcal{W}}((x_{k-1},w),(x'_{k-1},w'))h_k(y_k,x,w')T_k(x,x'_{k-1},w')dx \\
&\quad = d_{\mathcal{X}\times\mathcal{W}}((x_{k-1},w),(x'_{k-1},w')) \int_{\mathcal{X}} h_k(y_k,x,w')T_k(x,x'_{k-1},w')dx \\
&\quad \leq \tilde{C}_{Th}(y_k;k) d_{\mathcal{X}\times\mathcal{W}}((x_{k-1},w),(x'_{k-1},w')), \qquad (109)
\end{aligned}$$

where the third inequality is obtained according to Assumption PS.3. Combining Equation (108) with Equation (109) gives

$$\begin{aligned}
\|\tilde{f}\|_{\text{Lip}} &= \sup_{(x_{k-1},w),(x'_{k-1},w')\in\mathcal{X}\times\mathcal{W},(x_{k-1},w)\neq(x'_{k-1},w')} \frac{|\tilde{f}(x_{k-1},w) - \tilde{f}(x'_{k-1},w')|}{d_{\mathcal{X}\times\mathcal{W}}((x_{k-1},w),(x'_{k-1},w'))} \\
&\leq \tilde{C}^*_{Th}(y_k;k)D + \tilde{C}_{Th}(y_k;k).
\end{aligned}$$

Then following an argument analogous to that used in the proof of Proposition 50, we can show that

$$\begin{aligned}
&\sup_{f\in A} \left| \int_{\mathcal{X}\times\mathcal{W}} f(x,w)\tilde{F}_k\mu(dx,dw) - \int_{\mathcal{X}\times\mathcal{W}} f(x,w)\tilde{F}_k\mu'(dx,dw) \right| \\
&\qquad \leq \left(\tilde{C}^*_{Th}(y_k;k)D + \tilde{C}_{Th}(y_k;k)\right) W_1(\mu,\mu'). \qquad (110)
\end{aligned}$$

Similar to the proof of Proposition 49, the second term on the right-hand side of Equation (101) satisfies

$$\frac{|Z_k(\mu) - Z_k(\mu')|}{Z_k(\mu)Z_k(\mu')} \sup_{f:\mathcal{X}\times\mathcal{W}\to\mathbb{R}, \|f\|_{\text{Lip}}\leq 1, f(\bar{x},\bar{w})=0} \left| \int_{\mathcal{X}\times\mathcal{W}} f(x,w)\tilde{F}_k\mu'(dxdw) \right| \leq \frac{D|Z_k(\mu) - Z_k(\mu')|}{Z_k(\mu)}.$$

By Lemma 48, the term $|Z_k(\mu) - Z_k(\mu')|$ satisfies

$$|Z_k(\mu) - Z_k(\mu')| \leq \tilde{C}^*_{Th}(y_k;k)W_1(\mu,\mu').$$

It follows that

$$\frac{|Z_k(\mu) - Z_k(\mu')|}{Z_k(\mu)Z_k(\mu')} \sup_{f:\mathcal{X}\times\mathcal{W}\to\mathbb{R}, \|f\|_{\text{Lip}}\leq 1, f(\bar{x},\bar{w})=0} \left| \int_{\mathcal{X}\times\mathcal{W}} f(x,w)\tilde{F}_k\mu'(dxdw) \right|$$

$$\leq \frac{D\tilde{C}^*_{Th}(y_k;k)}{Z_k(\mu)}W_1(\mu,\mu'). \tag{111}$$

Plugging Equations (110) and (111) into Equation (101) gives

$$W_1(F_k(\mu),F_k(\mu')) \leq \frac{2D\tilde{C}^*_{Th}(y_k;k) + \tilde{C}_{Th}(y_k;k)}{Z_k(\mu)}W_1(\mu,\mu').$$

■

Appendix C. Proofs of Theorem 16

Proof We first prove that Theorem 16 holds for inverse problems under the total variation distance.

By the definitions of the evidence Z_k and the true posterior P_k for inverse problems, we have

$$Z_k(P_{k-1}) = \int_{\mathcal{X}} h(y_k,x)P_{k-1}(dx) = \int_{\mathcal{X}} p(y_k \mid x, \mathcal{Y}_{1:k-1})P_{k-1}(dx) = p(y_k \mid \mathcal{Y}_{1:k-1}),$$

where the third equality is obtained because $P_{k-1}(dx)$ represent the probability of dx given $\mathcal{Y}_{1:k-1}$. The remainder of the proof proceeds by mathematical induction.
Statement: for any integer $k \geq 2$, we have

$$d_{TV}(P_k,Q_k) \leq \sum_{j=1}^{k-1} \frac{\prod_{i=j+1}^{k} C_h(y_i)}{p(\mathcal{Y}_{j+1:k} \mid \mathcal{Y}_{1:j})} d_{TV}(Q^*_j,Q_j) + d_{TV}(Q^*_k,Q_k),$$

and

$$d_{TV}(P_k,Q_k) \leq \sum_{j=1}^{k-1} \left(\prod_{i=j+1}^{k} \frac{C_h(y_i)}{Z_i(Q_{i-1})} \right) d_{TV}(Q^*_j,Q_j) + d_{TV}(Q^*_k,Q_k).$$

Base case: Consider $k = 2$. By the triangle inequality satisfied by the metric d_{TV}, the learning error $d_{TV}(P_2, Q_2)$ can be bounded as

$$d_{TV}(P_2, Q_2) \leq d_{TV}(P_2, Q_2^*) + d_{TV}(Q_2^*, Q_2) = d_{TV}(F_2(P_1), F_2(Q_1)) + d_{TV}(Q_2^*, Q_2).$$

According to Theorem 14, $d_{TV}(F_2(P_1), F_2(Q_1))$ satisfies

$$\begin{aligned} d_{TV}(F_2(P_1), F_2(Q_1)) \leq & \frac{C_h(y_2)}{Z_2(P_1)} d_{TV}(P_1, Q_1), \\ d_{TV}(F_2(P_1), F_2(Q_1)) \leq & \frac{C_h(y_2)}{Z_2(Q_1)} d_{TV}(P_1, Q_1). \end{aligned}$$

As we have

$$Q_1^* = F_1(P_0) = P_1,$$

it follows that

$$d_{TV}(P_1, Q_1) \leq d_{TV}(P_1, Q_1^*) + d_{TV}(Q_1^*, Q_1) = d_{TV}(Q_1^*, Q_1).$$

Then we have

$$\begin{aligned} d_{TV}(P_2, Q_2) \leq & \frac{C_h(y_2)}{Z_2(P_1)} d_{TV}(Q_1^*, Q_1) + d_{TV}(Q_2^*, Q_2) \\ = & \sum_{j=1}^{2-1} \frac{\prod_{i=j+1}^{2} C_h(y_i)}{p(\mathcal{Y}_{j+1:2} \mid \mathcal{Y}_{1:j})} d_{TV}(Q_j^*, Q_j) + d_{TV}(Q_2^*, Q_2), \end{aligned}$$

and

$$\begin{aligned} d_{TV}(P_2, Q_2) \leq & \frac{C_h(y_2)}{Z_2(Q_1)} d_{TV}(Q_1^*, Q_1) + d_{TV}(Q_2^*, Q_2) \\ = & \sum_{j=1}^{2-1} \left(\prod_{i=j+1}^{2} \frac{C_h(y_i)}{Z_i(Q_{i-1})} \right) d_{TV}(Q_j^*, Q_j) + d_{TV}(Q_2^*, Q_2). \end{aligned}$$

Therefore, we conclude that the statement holds for $k = 2$.
Induction step: Assume the statement holds for some integer $k \geq 2$. As d_{TV} satisfies the triangle inequality, the learning error at step $k + 1$ satisfies

$$d_{TV}(P_{k+1}, Q_{k+1}) \leq d_{TV}(P_{k+1}, Q_{k+1}^*) + d_{TV}(Q_{k+1}^*, Q_{k+1}). \tag{112}$$

By Theorem 14, we have

$$\begin{aligned} d_{TV}(P_{k+1}, Q_{k+1}^*) &= d_{TV}(F_{k+1}(P_k), F_{k+1}(Q_k)) \\ &\leq \frac{C_h(y_{k+1})}{Z_{k+1}(P_k)} d_{TV}(P_k, Q_k) = \frac{C_h(y_{k+1})}{p(y_{k+1} \mid \mathcal{Y}_{1:k})} d_{TV}(P_k, Q_k), \end{aligned} \tag{113}$$

and

$$d_{TV}(P_{k+1}, Q_{k+1}^*) = d_{TV}(F_{k+1}(P_k), F_{k+1}(Q_k)) \leq \frac{C_h(y_{k+1})}{Z_{k+1}(Q_k)} d_{TV}(P_k, Q_k).$$

Plugging Equation (113) into Equation (112) gives

$$\begin{aligned}
d_{TV}(P_{k+1}, Q_{k+1}) &\leq \frac{C_h(y_{k+1})}{p(y_{k+1} \mid \mathcal{Y}_{1:k})} d_{TV}(P_k, Q_k) + d_{TV}(Q^*_{k+1}, Q_{k+1}) \\
&\leq \frac{C_h(y_{k+1})}{p(y_{k+1} \mid \mathcal{Y}_{1:k})} \left(\sum_{j=1}^{k-1} \frac{\prod_{i=j+1}^{k} C_h(y_i)}{p(\mathcal{Y}_{j+1:k} \mid \mathcal{Y}_{1:j})} d_{TV}(Q^*_j, Q_j) + d_{TV}(Q^*_k, Q_k) \right) \\
&\qquad + d_{TV}(Q^*_{k+1}, Q_{k+1}) \\
&= \frac{C_h(y_{k+1})}{p(y_{k+1} \mid \mathcal{Y}_{1:k})} \sum_{j=1}^{k-1} \frac{\prod_{i=j+1}^{k} C_h(y_i)}{p(\mathcal{Y}_{j+1:k} \mid \mathcal{Y}_{1:j})} d_{TV}(Q^*_j, Q_j) \\
&\qquad + \frac{C_h(y_{k+1})}{p(y_{k+1} \mid \mathcal{Y}_{1:k})} d_{TV}(Q^*_k, Q_k) + d_{TV}(Q^*_{k+1}, Q_{k+1}) \\
&= \sum_{j=1}^{k-1} \frac{\prod_{i=j+1}^{k+1} C_h(y_i)}{p(\mathcal{Y}_{j+1:k+1} \mid \mathcal{Y}_{1:j})} d_{TV}(Q^*_j, Q_j) + \frac{C_h(y_{k+1})}{p(y_{k+1} \mid \mathcal{Y}_{1:k})} d_{TV}(Q^*_k, Q_k) \\
&\qquad + d_{TV}(Q^*_{k+1}, Q_{k+1}) \\
&= \sum_{j=1}^{k} \frac{\prod_{i=j+1}^{k+1} C_h(y_i)}{p(\mathcal{Y}_{j+1:k+1} \mid \mathcal{Y}_{1:j})} d_{TV}(Q^*_j, Q_j) + d_{TV}(Q^*_{k+1}, Q_{k+1}).
\end{aligned}$$

Similarly, we can prove that

$$d_{TV}(P_{k+1}, Q_{k+1}) \leq \sum_{j=1}^{k} \left(\prod_{i=j+1}^{k+1} \frac{C_h(y_i)}{Z_i(Q_{i-1})} \right) d_{TV}(Q^*_j, Q_j) + d_{TV}(Q^*_{k+1}, Q_{k+1}).$$

Hence, the statement holds for $k+1$.
Conclusion: the statement holds for any $k \geq 2$.

So far, we have shown that Theorem 16 holds for inverse problems under the total variation distance. Analogously, we can prove that Theorem 16 holds for inverse problems under the Hellinger and 1-Wasserstein distances.

For state estimation problems, by the definition of Z_k, we can write $Z_k(P_{k-1})$ as

$$\begin{aligned}
Z_k(P_{k-1}) &= \int_{\mathcal{X}} h_k(y_k, x_k) \left(\int_{\mathcal{X}} T_k(x_k, x_{k-1}) P_{k-1}(dx_{k-1}) \right) dx_k \\
&= \int_{\mathcal{X}} p(y_k \mid x_k) p(x_k \mid \mathcal{Y}_{1:k-1}) dx_k \\
&= p(y_k \mid \mathcal{Y}_{1:k-1}).
\end{aligned} \tag{114}$$

Similarly, for parameter-state estimation problems, the evidence term $Z_k(P_{k-1})$ can be expressed as

$$\begin{aligned}
Z_k(P_{k-1}) &= \int_{\mathcal{X} \times \mathcal{W}} h_k(y_k, x_k, w) \left(\int_{\mathcal{X}} T_k(x_k, x_{k-1}, w) \pi_{k-1}(x_{k-1}, w) dx_{k-1} \right) dx_k dw \\
&= \int_{\mathcal{X} \times \mathcal{W}} p(y_k \mid x_k, w) p(x_k, w \mid \mathcal{Y}_{1:k-1}) dx_k dw \\
&= p(y_k \mid \mathcal{Y}_{1:k-1}),
\end{aligned} \tag{115}$$

where π_{k-1} denotes the density of P_{k-1} with respect to the Lebesgue measure.

Given Equations (114) and (115), the proofs of Theorem 16 for state estimation and parameter-state estimation problems are analogous to the proof of Theorem 16 for inverse problems under the total variation distance.

■

Appendix D. Proofs of error reduction theorems: Theorems 19, 20, 22, and 23

D.1 Proof of Theorem 19

Theorem 19 is a conclusion of Propositions 52—54.

Proposition 52 *For inverse problems, given the data y_k, assume that $P_{k-1} \in \bar{\mathcal{P}}_k(\mathcal{X})$ and $Q_{k-1} \in \bar{\mathcal{P}}_k(\mathcal{X})$. Suppose there exists a σ-finite measure ν defined on $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$ such that $P_{k-1} \ll \nu, Q_{k-1} \ll \nu$, and*

$$\begin{aligned}\int_{\mathcal{X}^*} h(y_k, x)\left|\frac{dP_{k-1}}{d\nu}(x) - \frac{dQ_{k-1}}{d\nu}(x)\right|\nu(dx) \\ \leq \int_{\mathcal{X}^*} h(y_k, x)\nu(dx) \int_{\mathcal{X}^*}\left|\frac{dP_{k-1}}{d\nu}(x) - \frac{dQ_{k-1}}{d\nu}(x)\right|\nu(dx),\end{aligned}$$

and

$$\int_{\mathcal{X}^*} h(y_k, x)\nu(dx) \leq \int_{\mathcal{X}} h(y_k, x)P_{k-1}(dx) \vee \int_{\mathcal{X}} h(y_k, x)Q_{k-1}(dx),$$

where

$$\mathcal{X}^* := \begin{cases}\{x \in \mathcal{X} : \frac{dP_{k-1}}{d\nu}(x) \geq \frac{dQ_{k-1}}{d\nu}(x)\}, & \text{if } Z_k(P_{k-1}) \geq Z_k(Q_{k-1}),\\ \{x \in \mathcal{X} : \frac{dP_{k-1}}{d\nu}(x) \leq \frac{dQ_{k-1}}{d\nu}(x)\}, & \text{otherwise.}\end{cases}$$

Then the posteriors P_k and Q_k^* satisfy

$$d_{TV}(P_k, Q_k^*) \leq d_{TV}(P_{k-1}, Q_{k-1}).$$

Proof Since $P_k = F_k(P_{k-1})$ and $Q_k^* = F_k(Q_{k-1})$, according to Equations (44) and (45), we have

$$d_{TV}(P_k, Q_k^*) \leq \frac{1}{Z_k(P_{k-1}) \vee Z_k(Q_{k-1})}\int_{\mathcal{X}^*} h(y_k, x)\left|\frac{dP_{k-1}}{d\nu}(x) - \frac{dQ_{k-1}}{d\nu}(x)\right|\nu(dx), \quad (116)$$

and

$$d_{TV}(P_{k-1}, Q_{k-1}) = \int_{\mathcal{X}^*}\left|\frac{dP_{k-1}}{d\nu}(x) - \frac{dQ_{k-1}}{d\nu}(x)\right|\nu(dx). \quad (117)$$

If $d_{TV}(P_{k-1}, Q_{k-1}) = 0$, then $P_{k-1} = Q_{k-1}$. It follows that $F_k(P_{k-1}) = F_k(Q_{k-1})$ and $d_{TV}(P_k, Q_k^*) = 0$. If $d_{TV}(P_{k-1}, Q_{k-1}) > 0$, we have

$$\frac{d_{TV}(P_k, Q_k^*)}{d_{TV}(P_{k-1}, Q_{k-1})} \leq \frac{1}{Z_k(P_{k-1}) \vee Z_k(Q_{k-1})}\frac{\int_{\mathcal{X}^*} h(y_k, x)\left|\frac{dP_{k-1}}{d\nu}(x) - \frac{dQ_{k-1}}{d\nu}(x)\right|\nu(dx)}{\int_{\mathcal{X}^*}\left|\frac{dP_{k-1}}{d\nu}(x) - \frac{dQ_{k-1}}{d\nu}(x)\right|\nu(dx)}.$$

If ν satisfies

$$\begin{aligned}&\int_{\mathcal{X}^*} h(y_k,x)\left|\frac{dP_{k-1}}{d\nu}(x)-\frac{dQ_{k-1}}{d\nu}(x)\right|\nu(dx)\\&\quad\le \int_{\mathcal{X}^*} h(y_k,x)\nu(dx)\int_{\mathcal{X}^*}\left|\frac{dP_{k-1}}{d\nu}(x)-\frac{dQ_{k-1}}{d\nu}(x)\right|\nu(dx),\end{aligned}$$

and

$$\int_{\mathcal{X}^*} h(y_k,x)\nu(dx)\le \int_{\mathcal{X}} h(y_k,x)P_{k-1}(dx)\vee\int_{\mathcal{X}} h(y_k,x)Q_{k-1}(dx),$$

then we have

$$\begin{aligned}\frac{d_{TV}(P_k,Q_k^*)}{d_{TV}(P_{k-1},Q_{k-1})}\le&\frac{1}{Z_k(P_{k-1})\vee Z_k(Q_{k-1})}\frac{\int_{\mathcal{X}^*} h(y_k,x)\left|\frac{dP_{k-1}}{d\nu}(x)-\frac{dQ_{k-1}}{d\nu}(x)\right|\nu(dx)}{\int_{\mathcal{X}^*}\left|\frac{dP_{k-1}}{d\nu}(x)-\frac{dQ_{k-1}}{d\nu}(x)\right|\nu(dx)}\\=&\frac{\int_{\mathcal{X}^*} h(y_k,x)\nu(dx)}{\int_{\mathcal{X}} h(y_k,x)P_{k-1}(dx)\vee\int_{\mathcal{X}} h(y_k,x)Q_{k-1}(dx)}\\\le&1.\end{aligned}$$

The statement of the theorem is proved. ■

Proposition 53 *In state estimation problems, given the data y_k, assume $P_{k-1}\in\bar{\mathcal{P}}_k(\mathcal{X})$ and $Q_{k-1}\in\bar{\mathcal{P}}_k(\mathcal{X})$. Suppose there exists a σ-finite measure ν defined on $(\mathcal{X},\mathcal{B}(\mathcal{X}))$ such that $P_{k-1}\ll\nu, Q_{k-1}\ll\nu$, and*

$$\int_{\mathcal{X}^*} g(x,y_k)\left|\frac{dP_{k-1}}{d\nu}(x)-\frac{dQ_{k-1}}{d\nu}(x)\right|\nu(dx) \tag{118}$$

$$\le \int_{\mathcal{X}^*} g(x,y_k)\nu(dx)\int_{\mathcal{X}^*}\left|\frac{dP_{k-1}}{d\nu}(x)-\frac{dQ_{k-1}}{d\nu}(x)\right|\nu(dx), \tag{119}$$

and

$$\int_{\mathcal{X}^*} g(x,y_k)\nu(dx)\le \int_{\mathcal{X}} g(x,y_k)P_{k-1}(dx)\vee\int_{\mathcal{X}} g(x,y_k)Q_{k-1}(dx), \tag{120}$$

where $g(x,y_k):=\int_{\mathcal{X}} h_k(y_k,x_k)T_k(x_k,x)dx_k$, and

$$\mathcal{X}^*:=\begin{cases}\{x\in\mathcal{X}:\frac{dP_{k-1}}{d\nu}(x)\ge\frac{dQ_{k-1}}{d\nu}(x)\}, & \text{if } Z_k(P_{k-1})\ge Z_k(Q_{k-1}),\\ \{x\in\mathcal{X}:\frac{dP_{k-1}}{d\nu}(x)\le\frac{dQ_{k-1}}{d\nu}(x)\}, & \text{otherwise.}\end{cases}$$

Then the posteriors P_k and Q_k^* satisfy

$$d_{TV}(P_k,Q_k^*)\le d_{TV}(P_{k-1},Q_{k-1}).$$

Proof Let ν be a σ-finite measure defined on $(\mathcal{X},\mathcal{B}(\mathcal{X}))$ such that $P_{k-1}\ll\nu$, $Q_{k-1}\ll\nu$. Since $P_k=F_k(P_{k-1})$ and $Q_k^*=F_k(Q_{k-1})$, given Equations (52), (53), (54), and (55), and

assuming that ν satisfies the conditions in Equations (119) and (120), the conclusion of Proposition 53 follows by an argument analogous to that used in the proof of Proposition 52. ■

Proposition 54 *In parameter-state estimation problems, given the data y_k, suppose $P_{k-1} \in \bar{\mathcal{P}}_k(\mathcal{X} \times \mathcal{W})$ and $Q_{k-1} \in \bar{\mathcal{P}}_k(\mathcal{X} \times \mathcal{W})$. Assume the state transition model h_k, the observation model T_k, and priors P_{k-1} and Q_{k-1} satisfy*

$$\begin{aligned}\int_{\mathcal{S}^*} g(x,w,y_k)\left|\frac{dP_{k-1}}{d\lambda}(x,w) - \frac{dQ_{k-1}}{d\lambda}(x,w)\right| dxdw \\ \leq \int_{\mathcal{S}^*} g(x,w,y_k)dxdw \int_{\mathcal{S}^*}\left|\frac{dP_{k-1}}{d\lambda}(x,w) - \frac{dQ_{k-1}}{d\lambda}(x,w)\right| dxdw,\end{aligned} \tag{121}$$

and

$$\int_{\mathcal{S}^*} g(x,w,y_k)dxdw \leq \int_{\mathcal{X}\times\mathcal{W}} g(x,w,y_k)dP_{k-1}(x,w) \vee \int_{\mathcal{X}\times\mathcal{W}} g(x,w,y_k)dQ_{k-1}(x,w), \tag{122}$$

where λ is the Lebesgue measure defined on the space $(\mathcal{X}\times\mathcal{W}, \mathcal{B}(\mathcal{X}\times\mathcal{W}))$, and

$$g(x,w,y_k) := \int_{\mathcal{X}} h_k(y_k,x_k,w)T_k(x_k,x,w)dx_k,$$

and

$$\mathcal{S}^* := \begin{cases}\{(x,w)\in\mathcal{X}\times\mathcal{W} : \frac{dP_{k-1}}{d\lambda}(x,w) \geq \frac{dQ_{k-1}}{d\lambda}(x,w)\}, & \text{if } Z_k(P_{k-1}) \geq Z_k(Q_{k-1}),\\ \{(x,w)\in\mathcal{X}\times\mathcal{W} : \frac{dP_{k-1}}{d\lambda}(x,w) \leq \frac{dQ_{k-1}}{d\lambda}(x,w)\}, & \text{otherwise.}\end{cases}$$

Then the posteriors P_k and Q_k^* satisfy

$$d_{TV}(P_k, Q_k^*) \leq d_{TV}(P_{k-1}, Q_{k-1}).$$

Proof Since $P_k = F_k(P_{k-1})$ and $Q_k^* = F_k(Q_{k-1})$, given Equations (57), (58), (59), and (60), if Equations (121) and (122) hold, the conclusion of Proposition 54 follows by an argument analogous to that used in the proof of Proposition 52. ■

D.2 Proof of Theorem 20

D.2.1 Proof of Theorem 20 for inverse problems

The proof of Theorem 20 for inverse problems is as follows.

Proof First, we prove that the conclusion of Theorem 20 holds if Assumption ER.H.1 is satisfied. Let ν be a measure which satisfies the conditions in Assumption ER.H.1. By the

definition of Hellinger distance, we can write $d_H^2(P_k, Q_k^*)$ as

$$\begin{aligned} d_H^2(P_k, Q_k^*) =& \frac{1}{2}\int_{\mathcal{X}} \left(\sqrt{\frac{dP_k}{d\nu}(x)} - \sqrt{\frac{dQ_k^*}{d\nu}(x)}\right)^2 \nu(dx) \\ =& \frac{1}{2}\int_{\mathcal{X}} \frac{dP_k}{d\nu}(x)\nu(dx) + \frac{1}{2}\int_{\mathcal{X}} \frac{dQ_k^*}{d\nu}(x)\nu(dx) \\ & - \int_{\mathcal{X}} \sqrt{\frac{dP_k}{d\nu}(x)}\sqrt{\frac{dQ_k^*}{d\nu}(x)}\nu(dx). \end{aligned}$$

Note that we have

$$\int_{\mathcal{X}} \frac{dP_k}{d\nu}(x)\nu(dx) = 1, \qquad \int_{\mathcal{X}} \frac{dQ_k^*}{d\nu}(x)\nu(dx) = 1.$$

Therefore, $d_H^2(P_k, Q_k^*)$ can be re-written as

$$\begin{aligned} d_H^2(P_k, Q_k^*) &= 1 - \int_{\mathcal{X}} \sqrt{\frac{dP_k}{d\nu}(x)}\sqrt{\frac{dQ_k^*}{d\nu}(x)}\nu(dx) \\ &= 1 - \frac{1}{\sqrt{Z_k(P_{k-1})Z_k(Q_{k-1})}}\int_{\mathcal{X}} h(y_k, x)\sqrt{\frac{dP_{k-1}}{d\nu}(x)\frac{dQ_{k-1}}{d\nu}(x)}\nu(dx). \end{aligned}$$

Similarly, we can write $d_H^2(P_{k-1}, Q_{k-1})$ as

$$d_H^2(P_{k-1}, Q_{k-1}) = 1 - \int_{\mathcal{X}} \sqrt{\frac{dP_{k-1}}{d\nu}(x)\frac{dQ_{k-1}}{d\nu}(x)}\nu(dx).$$

Then the term $d_H^2(P_k, Q_k^*) - d_H^2(P_{k-1}, Q_{k-1})$ can be expressed as

$$\begin{aligned} d_H^2(P_k, Q_k^*) - d_H^2(P_{k-1}, Q_{k-1}) &= \int_{\mathcal{X}} \sqrt{\frac{dP_{k-1}}{d\nu}(x)\frac{dQ_{k-1}}{d\nu}(x)}\nu(dx) \\ & - \frac{1}{\sqrt{Z_k(P_{k-1})Z_k(Q_{k-1})}}\int_{\mathcal{X}} h(y_k, x)\sqrt{\frac{dP_{k-1}}{d\nu}(x)\frac{dQ_{k-1}}{d\nu}(x)}\nu(dx). \end{aligned}$$

As ν satisfies the conditions in Assumption ER.H.1, we have

$$\begin{aligned} &\int_{\mathcal{X}} \sqrt{\frac{dP_{k-1}}{d\nu}(x)\frac{dQ_{k-1}}{d\nu}(x)}\nu(dx) \\ &\quad - \frac{1}{\sqrt{Z_k(P_{k-1})Z_k(Q_{k-1})}}\int_{\mathcal{X}} h(y_k, x)\sqrt{\frac{dP_{k-1}}{d\nu}(x)\frac{dQ_{k-1}}{d\nu}(x)}\nu(dx) \\ &\quad \leq \int_{\mathcal{X}} \sqrt{\frac{dP_{k-1}}{d\nu}(x)\frac{dQ_{k-1}}{d\nu}(x)}\nu(dx) \\ &\qquad - \frac{\int_{\mathcal{X}} h(y_k, x)\nu(dx)}{\sqrt{Z_k(P_{k-1})Z_k(Q_{k-1})}}\int_{\mathcal{X}} \sqrt{\frac{dP_{k-1}}{d\nu}(x)\frac{dQ_{k-1}}{d\nu}(x)}\nu(dx) \\ &\quad = \int_{\mathcal{X}} \sqrt{\frac{dP_{k-1}}{d\nu}(x)\frac{dQ_{k-1}}{d\nu}(x)}\nu(dx)\left(1 - \frac{\int_{\mathcal{X}} h(y_k, x)\nu(dx)}{\sqrt{Z_k(P_{k-1})Z_k(Q_{k-1})}}\right) \leq 0. \end{aligned}$$

Next, we prove that if Assumption ER.H.2 is satisfied, we then have $d_H(P_k, Q_k^*) \leq d_H(P_{k-1}, Q_{k-1})$.

We can re-write $d_H^2(P_k, Q_k^*)$ as

$$\begin{aligned}
d_H^2(P_k, Q_k^*) &= 1 - \frac{1}{\sqrt{Z_k(P_{k-1})Z_k(Q_{k-1})}} \int_{\mathcal{X}} h(y_k, x) \sqrt{\frac{dP_{k-1}}{d\nu}(x) \frac{dQ_{k-1}}{d\nu}(x)} \nu(dx) \\
&= 1 + \frac{1}{2} \frac{1}{\sqrt{Z_k(P_{k-1})Z_k(Q_{k-1})}} \int_{\mathcal{X}} h(y_k, x) \left(\left(\sqrt{\frac{dP_{k-1}}{d\nu}(x)} - \sqrt{\frac{dQ_{k-1}}{d\nu}(x)} \right)^2 \right. \\
&\qquad \left. - \frac{dP_{k-1}}{d\nu}(x) - \frac{dQ_{k-1}}{d\nu}(x) \right) \nu(dx) \\
&= 1 - \frac{1}{2} \left(\sqrt{\frac{Z_k(P_{k-1})}{Z_k(Q_{k-1})}} + \sqrt{\frac{Z_k(Q_{k-1})}{Z_k(P_{k-1})}} \right) \\
&\qquad + \frac{1}{2\sqrt{Z_k(P_{k-1})Z_k(Q_{k-1})}} \int_{\mathcal{X}} h(y_k, x) \left(\sqrt{\frac{dP_{k-1}}{d\nu}(x)} - \sqrt{\frac{dQ_{k-1}}{d\nu}(x)} \right)^2 \nu(dx) \\
&\leq \frac{1}{2\sqrt{Z_k(P_{k-1})Z_k(Q_{k-1})}} \int_{\mathcal{X}} h(y_k, x) \left(\sqrt{\frac{dP_{k-1}}{d\nu}(x)} - \sqrt{\frac{dQ_{k-1}}{d\nu}(x)} \right)^2 \nu(dx).
\end{aligned}$$

Hence, the term $d_H^2(P_k, Q_k^*) - d_H^2(P_{k-1}, Q_{k-1})$ satisfies

$$\begin{aligned}
&d_H^2(P_k, Q_k^*) - d_H^2(P_{k-1}, Q_{k-1}) \\
&\quad \leq \frac{1}{2\sqrt{Z_k(P_{k-1})Z_k(Q_{k-1})}} \int_{\mathcal{X}} h(y_k, x) \left(\sqrt{\frac{dP_{k-1}}{d\nu}(x)} - \sqrt{\frac{dQ_{k-1}}{d\nu}(x)} \right)^2 \nu(dx) \\
&\qquad - \frac{1}{2} \int_{\mathcal{X}} \left(\sqrt{\frac{dP_{k-1}}{d\nu}(x)} - \sqrt{\frac{dQ_{k-1}}{d\nu}(x)} \right)^2 \nu(dx). \qquad (123)
\end{aligned}$$

When ν satisfies the conditions in Assumption ER.H.2, the right-hand side of Equation (123) satisfies

$$\begin{aligned}
&\frac{1}{2\sqrt{Z_k(P_{k-1})Z_k(Q_{k-1})}} \int_{\mathcal{X}} h(y_k, x) \left(\sqrt{\frac{dP_{k-1}}{d\nu}(x)} - \sqrt{\frac{dQ_{k-1}}{d\nu}(x)} \right)^2 \nu(dx) \\
&\quad - \frac{1}{2} \int_{\mathcal{X}} \left(\sqrt{\frac{dP_{k-1}}{d\nu}(x)} - \sqrt{\frac{dQ_{k-1}}{d\nu}(x)} \right)^2 \nu(dx) \\
&\qquad \leq \frac{1}{2} \int_{\mathcal{X}} \left(\sqrt{\frac{dP_{k-1}}{d\nu}(x)} - \sqrt{\frac{dQ_{k-1}}{d\nu}(x)} \right)^2 \nu(dx) \\
&\qquad\quad \cdot \left(\frac{\int_{\mathcal{X}} h(y_k, x) \nu(dx)}{\sqrt{\int_{\mathcal{X}} h(y_k, x) P_{k-1}(dx)} \sqrt{\int_{\mathcal{X}} h(y_k, x) Q_{k-1}(dx)}} - 1 \right) \\
&\qquad \leq 0.
\end{aligned}$$

The statement of Theorem 20 for inverse problems is proved. ■

D.2.2 Proof of Theorem 20 for state estimation problems

The proof of Theorem 20 for state estimation problems is as follows.

Proof First, we prove that if Assumption ER.H.1 is satisfied, we have $d_H(P_k, Q_k^*) \leq d_H(P_{k-1}, Q_{k-1})$. Let ν be a measure which satisfies the conditions in Assumption SE.ER.1. Let π and π' denote the Radon-Nikodym derivatives $\frac{dP_{k-1}}{d\nu}$ and $\frac{dQ_{k-1}}{d\nu}$, respectively. Then we can write $d_H^2(P_k, Q_k^*)$ as

$$\begin{aligned} d_H^2(P_k, Q_k^*) =& \frac{1}{2}\int_{\mathcal{X}} \left(\sqrt{\frac{h_k(y_k, x)\int_{\mathcal{X}} T_k(x, x_{k-1}) P_{k-1}(dx_{k-1})}{Z_k(P_{k-1})}} \right. \\ & \left. - \sqrt{\frac{h_k(y_k, x)\int_{\mathcal{X}} T_k(x, x_{k-1}) Q_{k-1}(dx_{k-1})}{Z_k(Q_{k-1})}} \right)^2 dx \\ =& 1 - \frac{1}{\sqrt{Z_k(P_{k-1}) Z_k(Q_{k-1})}} \int_{\mathcal{X}} h_k(y_k, x) \sqrt{\int_{\mathcal{X}} T_k(x, x_{k-1}) \pi(x_{k-1}) \nu(dx_{k-1})} \\ & \cdot \sqrt{\int_{\mathcal{X}} T_k(x, x_{k-1}) \pi'(x_{k-1}) \nu(dx_{k-1})} dx. \end{aligned} \tag{124}$$

Using Hölder's inequality, we obtain

$$\begin{aligned} & \sqrt{\int_{\mathcal{X}} T_k(x, x_{k-1}) \pi(x_{k-1}) \nu(dx_{k-1})} \sqrt{\int_{\mathcal{X}} T_k(x, x_{k-1}) \pi'(x_{k-1}) \nu(dx_{k-1})} \\ & \quad \geq \int_{\mathcal{X}} T_k(x, x_{k-1}) \sqrt{\pi(x_{k-1}) \pi'(x_{k-1})} \nu(dx_{k-1}) \end{aligned} \tag{125}$$

Substituting Equation (125) into Equation (124) yields

$$\begin{aligned} & d_H^2(P_k, Q_k^*) \\ & \quad \leq 1 - \frac{\int_{\mathcal{X}} h_k(y_k, x) \left(\int_{\mathcal{X}} T_k(x, x_{k-1}) \sqrt{\pi(x_{k-1})} \sqrt{\pi'(x_{k-1})} \nu(dx_{k-1}) \right) dx}{\sqrt{Z_k(P_{k-1}) Z_k(Q_{k-1})}}. \end{aligned}$$

By the definition of the Hellinger distance, we can write $d_H^2(P_{k-1}, Q_{k-1})$ as

$$\begin{aligned} d_H^2(P_{k-1}, Q_{k-1}) =& \frac{1}{2}\int_{\mathcal{X}} \left(\sqrt{\pi(x_{k-1})} - \sqrt{\pi'(x_{k-1})} \right)^2 \nu(dx_{k-1}) \\ =& 1 - \int_{\mathcal{X}} \sqrt{\pi(x_{k-1}) \pi'(x_{k-1})} \nu(dx_{k-1}). \end{aligned}$$

Thus the term $d_H^2(P_k, Q_k^*) - d_H^2(P_{k-1}, Q_{k-1})$ satisfies

$$\begin{aligned} &d_H^2(P_k, Q_k^*) - d_H^2(P_{k-1}, Q_{k-1}) \\ &\quad \le \int_{\mathcal{X}} \sqrt{\pi(x_{k-1})\pi'(x_{k-1})}\nu(dx_{k-1}) \\ &\qquad\qquad - \frac{\int_{\mathcal{X}} h_k(y_k, x)\left(\int_{\mathcal{X}} T_k(x, x_{k-1})\sqrt{\pi(x_{k-1})}\sqrt{\pi'(x_{k-1})}\nu(dx_{k-1})\right)dx}{\sqrt{Z_k(P_{k-1})Z_k(Q_{k-1})}}. \end{aligned}$$

By Tonelli's theorem, we have

$$\begin{aligned} &\int_{\mathcal{X}} h_k(y_k, x)\left(\int_{\mathcal{X}} T_k(x, x_{k-1})\sqrt{\pi(x_{k-1})}\sqrt{\pi'(x_{k-1})}\nu(dx_{k-1})\right)dx \\ &\quad = \int_{\mathcal{X}}\left(\int_{\mathcal{X}} h_k(y_k, x)T_k(x, x_{k-1})dx\right)\sqrt{\pi(x_{k-1})\pi'(x_{k-1})}\nu(dx_{k-1}). \end{aligned}$$

When ν satisfies the conditions in Assumption ER.H.1, an argument analogous to that used in the proof of Theorem 20 for inverse problems yields

$$d_H^2(P_k, Q_k^*) - d_H^2(P_{k-1}, Q_{k-1}) \le 0.$$

Next, we prove that under Assumption ER.H.2, we have $d_H(P_k, Q_k^*) \le d_H(P_{k-1}, Q_{k-1})$.

Following an argument similar to that is used in the proof for Theorem 20 for inverse problems, we can show that

$$\begin{aligned} &d_H^2(P_k, Q_k^*) \\ &\le \frac{1}{2\sqrt{Z_k(P_{k-1})Z_k(Q_{k-1})}}\int_{\mathcal{X}} h_k(y_k, x) \\ &\quad \cdot \left(\sqrt{\int_{\mathcal{X}} T_k(x, x_{k-1})\pi(x_{k-1})\nu(dx_{k-1})} - \sqrt{\int_{\mathcal{X}} T_k(x, x_{k-1})\pi'(x_{k-1})\nu(dx_{k-1})}\right)^2 dx \\ &\le \frac{\int_{\mathcal{X}} h_k(y_k, x)\left(\int_{\mathcal{X}} T_k(x, x_{k-1})\left(\sqrt{\pi(x_{k-1})} - \sqrt{\pi'(x_{k-1})}\right)^2\nu(dx_{k-1})\right)dx}{2\sqrt{Z_k(P_{k-1})Z_k(Q_{k-1})}}. \end{aligned} \tag{126}$$

The second inequality in Equation (126) is obtained according to Equation (125).

By Tonelli's theorem, we have

$$\begin{aligned} &\int_{\mathcal{X}} h_k(y_k, x)\left(\int_{\mathcal{X}} T_k(x, x_{k-1})\left(\sqrt{\pi(x_{k-1})} - \sqrt{\pi'(x_{k-1})}\right)^2\nu(dx_{k-1})\right)dx \\ &\quad = \int_{\mathcal{X}}\left(\int_{\mathcal{X}} h_k(y_k, x)T_k(x, x_{k-1})dx\right)\left(\sqrt{\pi(x_{k-1})} - \sqrt{\pi'(x_{k-1})}\right)^2\nu(dx_{k-1}). \end{aligned}$$

Then we can prove that $d_H(P_k, Q_k^*) \le d_H(P_{k-1}, Q_{k-1})$ holds under Assumption ER.H.2 by following an argument analogous to that used in the proof of Theorem 20 for inverse problems. ■

D.2.3 Proof of Theorem 20 for parameter-state estimation problems

The proof of Theorem 20 for parameter-state estimation problems is analogous to the proof of Theorem 20 for state estimation problems.

D.3 Proof of Theorem 22

Proof For the metric space $(\mathcal{X}, d_{\mathcal{X}})$, the 1-Wasserstein distance between the posteriors P_k and Q_k^* is defined as

$$W_1(P_k, Q_k^*) := \inf_{\gamma \in \Gamma(P_k, Q_k^*)} \int_{\mathcal{X}\times\mathcal{X}} d_{\mathcal{X}}(x_1, x_2)\gamma(dx_1, dx_2),$$

where $\Gamma(P_k, Q_k^*)$ denotes the set of all couplings of P_k and Q_k^*. Define a probability measure $\gamma_0 : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ as

$$\gamma_0(dx, dx') := P_k(dx)Q_k^*(dx').$$

Then γ_0 is a coupling of P_k and Q_k^*. It follows that

$$\begin{aligned} W_1(P_k, Q_k^*) &\leq \int_{\mathcal{X}\times\mathcal{X}} d_{\mathcal{X}}(x, x')P_k(dx)Q_k^*(dx') \\ &= \frac{1}{Z_k(P_{k-1})Z_k(Q_{k-1})} \int_{\mathcal{X}\times\mathcal{X}} d_{\mathcal{X}}(x, x')h(y_k, x)h(y_k, x')P_{k-1}(dx)Q_{k-1}(dx') \\ &= \frac{\mathbb{E}_{(X,X')\sim P_{k-1}\otimes Q_{k-1}}[d_{\mathcal{X}}(X, X')h(y_k, X)h(y_k, X')]}{E_{X\sim P_{k-1}}[h(y_k, X)]E_{X\sim Q_{k-1}}[h(y_k, X)]}. \end{aligned} \tag{127}$$

According to the Kantorovich-Rubinstein duality, the 1-Wasserstein distance between the two priors P_{k-1} and Q_{k-1} can be written as

$$W_1(P_{k-1}, Q_{k-1}) = \sup_{f:\mathcal{X}\to\mathbb{R}, \|f\|_{\text{Lip}}\leq 1} \left| \int_{\mathcal{X}} f(x)P_{k-1}(dx) - \int_{\mathcal{X}} f(x)Q_{k-1}(dx) \right|.$$

Define a function $g : \mathcal{X} \to \mathbb{R}$ as

$$g(x) := d_{\mathcal{X}}(x, x_0),$$

where $x_0 \in \mathcal{X}$ is an arbitrary point in space $\mathcal{X}$. We first show that the function g is Lipschitz continuous on metric space $(\mathcal{X}, d_{\mathcal{X}})$ and its best Lipschitz constant $\|g\|_{\text{Lip}}$ is no greater than 1.

For two arbitrary points $x, x' \in \mathcal{X}$, we have

$$|g(x) - g(x')| = |d_{\mathcal{X}}(x, x_0) - d_{\mathcal{X}}(x', x_0)|.$$

Since metric $d_{\mathcal{X}}$ is symmetric and satisfies triangle inequality, we have

$$d_{\mathcal{X}}(x, x_0) \leq d_{\mathcal{X}}(x, x') + d_{\mathcal{X}}(x', x_0),$$

and

$$d_{\mathcal{X}}(x', x_0) \leq d_{\mathcal{X}}(x', x) + d_{\mathcal{X}}(x, x_0) = d_{\mathcal{X}}(x, x') + d_{\mathcal{X}}(x, x_0).$$

Then $d_{\mathcal{X}}(x, x')$ satisfies

$$d_{\mathcal{X}}(x, x') \geq |d_{\mathcal{X}}(x, x_0) - d_{\mathcal{X}}(x', x_0)|.$$

It follows that

$$\|g\|_{\text{Lip}} = \sup_{x,x' \in \mathcal{X}, x \neq x'} \frac{|d_{\mathcal{X}}(x, x_0) - d_{\mathcal{X}}(x', x_0)|}{d_{\mathcal{X}}(x, x')} \leq 1.$$

Then we have

$$W_1(P_{k-1}, Q_{k-1}) \geq \left| \int_{\mathcal{X}} d_{\mathcal{X}}(x, x_0) P_{k-1}(dx) - \int_{\mathcal{X}} d_{\mathcal{X}}(x, x_0) Q_{k-1}(dx) \right|.$$

As $x_0 \in \mathcal{X}$ is arbitrary, $W_1(P_{k-1}, Q_{k-1})$ satisfies

$$\begin{aligned} W_1(P_{k-1}, Q_{k-1}) &\geq \sup_{x_0 \in \mathcal{X}} \left| \int_{\mathcal{X}} d_{\mathcal{X}}(x, x_0) P_{k-1}(dx) - \int_{\mathcal{X}} d_{\mathcal{X}}(x, x_0) Q_{k-1}(dx) \right| \\ &= \sup_{x_0 \in \mathcal{X}} \left| \mathbb{E}_{X \sim P_{k-1}}[d_{\mathcal{X}}(X, x_0)] - \mathbb{E}_{X \sim Q_{k-1}}[d_{\mathcal{X}}(X, x_0)] \right|. \end{aligned} \tag{128}$$

Given Equations (127) and (128), if the following inequality holds:

$$\begin{aligned} &\frac{\mathbb{E}_{(X,X') \sim P_{k-1} \otimes Q_{k-1}}[d_{\mathcal{X}}(X, X') h(y_k, X) h(y_k, X')]}{E_{X \sim P_{k-1}}[h(y_k, X)] E_{X \sim Q_{k-1}}[h(y_k, X)]} \\ &\leq \sup_{x_0 \in \mathcal{X}} \left| \mathbb{E}_{X \sim P_{k-1}}[d_{\mathcal{X}}(X, x_0)] - \mathbb{E}_{X \sim Q_{k-1}}[d_{\mathcal{X}}(X, x_0)] \right|, \end{aligned}$$

then we have

$$W_1(P_k, Q_k^*) \leq W_1(P_{k-1}, Q_{k-1}).$$

■

D.4 Proof of Theorem 23

Proof First, we prove that Theorem 23 holds for state estimation problems.

For state estimation problems, define a probability measure γ_0 as

$$\gamma_0(dx, dx') = P_k(dx) Q_k^*(dx').$$

Then γ_0 is a coupling of P_k and Q_k^*. According to the definition of 1-Wasserstein, we have

$$W_1(P_k, Q_k^*) \leq \int_{\mathcal{X} \times \mathcal{X}} d_{\mathcal{X}}(x, x') P_k(dx) Q_k^*(dx').$$

By the definition of F_k for state estimation problems, $P_k(dx)$ can be written as

$$P_k(dx) = \frac{h_k(y_k, x) P_k^-(dx)}{Z_k(P_{k-1})}.$$

Similarly, we can express $Q_k^*(dx')$ as

$$Q_k^*(dx') = \frac{h_k(y_k, x')Q_k^-(dx')}{Z_k(Q_{k-1})}.$$

It follows that

$$\begin{aligned} W_1(P_k, Q_k^*) \leq & \frac{1}{Z_k(P_{k-1})Z_k(Q_{k-1})} \int_{\mathcal{X}\times\mathcal{X}} h_k(y_k, x)h_k(y_k, x')P_k^-(dx)Q_k^-(dx') \\ = & \frac{\mathbb{E}_{(X,X')\sim P_k^-\otimes Q_k^-}[d_{\mathcal{X}}(X, X')h_k(y_k, X)h_k(y_k, X')]}{\mathbb{E}_{X\sim P_k^-}[h_k(y_k, X)]\,\mathbb{E}_{X\sim Q_k^-}[h_k(y_k, X)]}. \end{aligned} \tag{129}$$

According to the Kantorovich-Rubinstein duality, the 1-Wasserstein distance between the two priors P_{k-1} and Q_{k-1} can be written as

$$W_1(P_{k-1}, Q_{k-1}) = \sup_{f:\mathcal{X}\to\mathbb{R}, \|f\|_{\text{Lip}}\leq 1} \left| \int_{\mathcal{X}} f(x)p_{k-1}(x)dx - \int_{\mathcal{X}} f(x)q_{k-1}(x)dx \right|.$$

Following an arguments similar to that used in the proof of Theorem 22, we can show that

$$W_1(P_{k-1}, Q_{k-1}) \geq \sup_{\hat{x}\in\mathcal{X}} \left| \mathbb{E}_{X\sim P_{k-1}}[d_{\mathcal{X}}(X, \hat{x})] - \mathbb{E}_{X\sim Q_{k-1}}[d_{\mathcal{X}}(X, \hat{x})] \right|. \tag{130}$$

Under the conditions given in Theorem 23, we have

$$\begin{aligned} & \frac{\mathbb{E}_{(X,X')\sim P_k^-\otimes Q_k^-}[d_{\mathcal{X}}(X, X')h_k(y_k, X)h_k(y_k, X')]}{\mathbb{E}_{X\sim P_k^-}[h_k(y_k, X)]\,\mathbb{E}_{X\sim Q_k^-}[h_k(y_k, X)]} \\ & \leq \sup_{\hat{x}\in\mathcal{X}} \left| \mathbb{E}_{X\sim P_{k-1}}[d_{\mathcal{X}}(X, \hat{x})] - \mathbb{E}_{X\sim Q_{k-1}}[d_{\mathcal{X}}(X, \hat{x})] \right|. \end{aligned}$$

Given Equations (129) and (130), it follows that

$$W_1(P_k, Q_k^*) \leq W_1(P_{k-1}, Q_{k-1}).$$

Analogously, Theorem 23 can be shown to hold for parameter-state estimation problems. ■

Appendix E. Proofs of Corollary 25 and Corollary 26

E.1 Proof of Corollary 25

Proof Given the underlying system, for any integer $k \geq 1$, we have

$$h_k(y_k, x, w) = p_{\mathcal{N}}(y_k \mid \Phi_k(x, w), \Gamma) \leq (2\pi)^{-\frac{r}{2}} \det(\Gamma)^{-\frac{1}{2}}.$$

It follows that

$$\begin{aligned} \tilde{C}_{Th}(y_k; k) = & \sup_{x_{k-1}\in\mathcal{X}, w\in\mathcal{W}} \int_{\mathcal{X}} h_k(y_k, x, w)T_k(x, x_{k-1}, w)dx \\ & \leq (2\pi)^{-\frac{r}{2}} \det(\Gamma)^{-\frac{1}{2}} \sup_{x_{k-1}\in\mathcal{X}, w\in\mathcal{W}} \int_{\mathcal{X}} T_k(x, x_{k-1}, w)dx \\ & = (2\pi)^{-\frac{r}{2}} \det(\Gamma)^{-\frac{1}{2}} < \infty. \end{aligned} \tag{131}$$

Hence Assumption PS.1 is satisfied for all $k \geq 1$ and Theorem 16 can be applied to this problem under the total variation and Hellinger distance.

Next, we derive an upper bound of the approximation error $d(Q_k, Q_k^*)$ for any integer $k \geq 1$.

Let q_k^* denote the density of Q_k^* with respect to the Lebesgue measure. The Kullback-Leibler (KL) divergence between Q_k and Q_k^* is given by

$$\begin{aligned} KL(Q_k \| Q_k^*) &= \mathbb{E}_{q_k(x_k,w)}\left[\log \frac{q_k(x_k,w)}{q_k^*(x_k,w)}\right] \\ &= \mathbb{E}_{q_k(x_k,w)}\left[\log\left(\int_{\mathcal{X}\times\mathcal{W}} p_{\mathcal{N}}(y_k \mid \Phi_k(x_k,w),\Gamma) q_k^{*-}(x_k,w) dx_k dw\right)\right] \\ &\quad - \mathbb{E}_{q_k(x_k,w)}\left[\log\left(p_{\mathcal{N}}(y_k \mid \Phi_k(x_k,w),\Gamma) q_k^{*-}(x_k,w)\right)\right] + \mathbb{E}_{q_k(x_k,w)}[\log q_k(x_k,w)] \\ &= \log\left(\int_{\mathcal{X}\times\mathcal{W}} p_{\mathcal{N}}(y_k \mid \Phi_k(x_k,w),\Gamma) q_k^{*-}(x_k,w) dx_k dw\right) - \mathcal{L}_k(Q_k). \end{aligned} \tag{132}$$

The first term on the right-hand side of Equation (132) satisfies

$$\begin{aligned} &\log(\int_{\mathcal{X}\times\mathcal{W}} p_{\mathcal{N}}(y_k \mid \Phi_k(x_k,w),\Gamma) q_k^{*-}(x_k,w) dx_k dw \\ &\quad \leq \log\left((2\pi)^{-\frac{r}{2}} \det(\Gamma)^{-\frac{1}{2}} \int_{\mathcal{X}\times\mathcal{W}} q_k^{*-}(x_k,w) dx_k dw\right) \\ &\quad = -\frac{r}{2}\log(2\pi) - \frac{1}{2}\log(\det(\Gamma)). \end{aligned} \tag{133}$$

Substituting Equation (133) and $\mathcal{L}_k(Q_k) \geq \epsilon_k$ into Equation (132) yields:

$$KL(Q_k \| Q_k^*) \leq -\frac{r}{2}\log(2\pi) - \frac{1}{2}\log(\det(\Gamma)) - \epsilon_k.$$

Therefore, we have

$$d_{TV}(Q_k^*, Q_k) \leq \frac{1}{\sqrt{2}}\sqrt{KL(Q_k \| Q_k^*)} \leq \frac{1}{\sqrt{2}}\sqrt{-\frac{r}{2}\log(2\pi) - \frac{1}{2}\log(\det(\Gamma)) - \epsilon_k}, \tag{134}$$

and

$$d_H(Q_k^*, Q_k) \leq \frac{1}{\sqrt{2}}\sqrt{KL(Q_k \| Q_k^*)} \leq \frac{1}{\sqrt{2}}\sqrt{-\frac{r}{2}\log(2\pi) - \frac{1}{2}\log(\det(\Gamma)) - \epsilon_k}. \tag{135}$$

Combining Theorem 16 and Equations (131), (134), and (135) completes the proof of Corollary 25 under the total variation and Hellinger distances.

According to Gibbs and Edward (2002), if the metric space $(\mathcal{X}\times\mathcal{W}, d_{\mathcal{X}\times\mathcal{W}})$ satisfies

$$\sup_{(x,w),(x',w')\in\mathcal{X}\times\mathcal{W}} d_{\mathcal{X}\times\mathcal{W}}((x,w),(x',w')) = D < \infty,$$

then we have

$$W_1(P_k, Q_k) \leq D d_{TV}(P_k, Q_k). \tag{136}$$

The statement of Corollary 25 for the 1-Wasserstein distance is directly proved by combining the result of Corollary 25 for the total variation distance and Equation (136). ■

E.2 Proof of Corollary 26

Proof For any time step $k \geq 1$, we have

$$h_k(y_k, x) = p_{\mathcal{N}}(y_k \mid \Phi_k(x), \Gamma) \leq (2\pi)^{-\frac{r}{2}} \det(\Gamma)^{-\frac{1}{2}}.$$

Thus the underlying system with the true system parameter $\bar{w}$ satisfies

$$\begin{aligned} C_{Th}(y_k; k) &= \sup_{x_{k-1} \in \mathcal{X}} \int_{\mathcal{X}} h_k(y_k, x) T_k(x, x_{k-1}, \bar{w}) dx \\ &\leq (2\pi)^{-\frac{r}{2}} \det(\Gamma)^{-\frac{1}{2}} \sup_{x_{k-1} \in \mathcal{X}} \int_{\mathcal{X}} T_k(x, x_{k-1}, \bar{w}) dx \\ &= (2\pi)^{-\frac{r}{2}} \det(\Gamma)^{-\frac{1}{2}} < \infty. \end{aligned} \tag{137}$$

Therefore, the underlying system with true parameter $\bar{w}$ satisfies Assumption SE.2. Next, we derive an upper bound for the approximation error (Q_k^*, Q_k). Here, Q_k^* is the exact posterior obtained by propagating the prior Q_{k-1} under the system with the true parameter $\bar{w}$. Let $Q_{k,\hat{w}_k}^*$ be the exact posterior obtained using the same prior Q_{k-1} but under system with the estimated parameter $\hat{w}_k$. By the triangle inequality satisfied by the metric d, we have

$$d(Q_k^*, Q_k) \leq d(Q_k^*, Q_{k,\hat{w}_k}^*) + d(Q_{k,\hat{w}_k}^*, Q_k). \tag{138}$$

First, we derive an upper bound of $d(Q_k^*, Q_{k,\hat{w}_k}^*)$. Note that $Z_k(Q_{k-1})$ and $Z_{k,\hat{w}_k}(Q_{k-1})$ are corresponding evidences of Q_k^* and $Q_{k,\hat{w}_k}^*$, respectively. Then we can write $d_{TV}(Q_k^*, Q_{k,\hat{w}_k}^*)$ as:

$$\begin{aligned} d_{TV}(Q_k^*, Q_{k,\hat{w}_k}^*) =& \frac{1}{2} \int_{\mathcal{X}} \left| \frac{h_k(y_k, x)}{Z_{k,\hat{w}_k}(Q_{k-1})} \int_{\mathcal{X}} T_k(x, x_{k-1}, \hat{w}_k) q_{k-1}(x_{k-1}) dx_{k-1} \right. \\ & \left. - \frac{h_k(y_k, x)}{Z_k(Q_{k-1})} \int_{\mathcal{X}} T_k(x, x_{k-1}, \bar{w}) q_{k-1}(x_{k-1}) dx_{k-1} \right| dx \\ \leq & I_1 + I_2, \end{aligned} \tag{139}$$

where

$$\begin{aligned} I_1 = \frac{1}{2 Z_{k,\hat{w}_k}(Q_{k-1})} \int_{\mathcal{X}} h_k(y_k, x) \left| \int_{\mathcal{X}} T_k(x, x_{k-1}, \hat{w}_k) q_{k-1}(x_{k-1}) dx_{k-1} \right. \\ \left. - \int_{\mathcal{X}} T_k(x, x_{k-1}, \bar{w}) q_{k-1}(x_{k-1}) dx_{k-1} \right| dx, \end{aligned}$$

and

$$\begin{aligned} I_2 = \frac{1}{2} \left| \frac{1}{Z_{k,\hat{w}_k}(Q_{k-1})} - \frac{1}{Z_k(Q_{k-1})} \right| \\ \cdot \int_{\mathcal{X}} h_k(y_k, x) \left(\int_{\mathcal{X}} T_k(x, x_{k-1}, \bar{w}) q_{k-1}(x_{k-1}) dx_{k-1} \right) dx. \end{aligned}$$

First, we derive an upper bound for I_1. By Assumption VI.1, I_1 satisfies

$$\begin{aligned}
I_1 =& \frac{1}{2Z_{k,\hat{w}_k}(Q_{k-1})} \int_{\mathcal{X}} h_k(y_k, x) \\
&\cdot \left| \int_{\mathcal{X}} q_{k-1}(x_{k-1}) \left(T_k(x, x_{k-1}, \hat{w}_k) - T_k(x, x_{k-1}, \bar{w})\right) dx_{k-1} \right| dx \\
\leq& \frac{1}{2Z_{k,\hat{w}_k}(Q_{k-1})} \int_{\mathcal{X}} h_k(y_k, x) \\
&\cdot \left(\int_{\mathcal{X}} q_{k-1}(x_{k-1}) \left| T_k(x, x_{k-1}, \hat{w}_k) - T_k(x, x_{k-1}, \bar{w}) \right| dx_{k-1} \right) dx \\
\leq& \frac{\|\hat{w}_k - \bar{w}\|_m}{2Z_{k,\hat{w}_k}(Q_{k-1})} \int_{\mathcal{X}} h_k(y_k, x) g_{\text{Lip}}(x; k) \left(\int_{\mathcal{X}} q_{k-1}(x_{k-1}) dx_{k-1} \right) dx \\
=& \frac{\|\hat{w}_k - \bar{w}\|_m}{2Z_{k,\hat{w}_k}(Q_{k-1})} \int_{\mathcal{X}} h_k(y_k, x) g_{\text{Lip}}(x; k) dx = \frac{\tilde{C}_{VI}(y_k; k) \|\hat{w}_k - \bar{w}\|_m}{2Z_{k,\hat{w}_k}(Q_{k-1})}.
\end{aligned} \tag{140}$$

We then derive an upper bound for I_2. Note that

$$\int_{\mathcal{X}} h_k(y_k, x) \left(\int_{\mathcal{X}} T_k(x, x_{k-1}, \bar{w}) q_{k-1}(x_{k-1}) dx_{k-1} \right) dx = Z_k(Q_{k-1}).$$

Then I_2 satisfies

$$\begin{aligned}
I_2 =& \frac{\left| Z_{k,\hat{w}_k}(Q_{k-1}) - Z_k(Q_{k-1}) \right|}{2Z_{k,\hat{w}_k}(Q_{k-1})} \\
=& \frac{1}{2Z_{k,\hat{w}_k}(Q_{k-1})} \\
&\cdot \left| \int_{\mathcal{X}} h_k(y_k, x) \left(\int_{\mathcal{X}} \left(T_k(x, x_{k-1}, \hat{w}_k) - T_k(x, x_{k-1}, \bar{w})\right) q_{k-1}(x_{k-1}) dx_{k-1} \right) dx \right| \\
\leq& \frac{1}{2Z_{k,\hat{w}_k}(Q_{k-1})} \\
&\cdot \int_{\mathcal{X}} h_k(y_k, x) \left(\int_{\mathcal{X}} \left| T_k(x, x_{k-1}, \hat{w}_k) - T_k(x, x_{k-1}, \bar{w}) \right| q_{k-1}(x_{k-1}) dx_{k-1} \right) dx \\
\leq& \frac{\tilde{C}_{VI}(y_k; k) \|\hat{w}_k - \bar{w}\|_m}{2Z_{k,\hat{w}_k}(Q_{k-1})}.
\end{aligned} \tag{141}$$

Substituting Equations (140) and (141) into Equation (139) yields:

$$d_{TV}(Q_k^*, Q_{k,\hat{w}_k}^*) \leq \frac{\tilde{C}_{VI}(y_k; k) \|\hat{w}_k - \bar{w}\|_m}{Z_{k,\hat{w}_k}(Q_{k-1})}. \tag{142}$$

Subsequently, we derive an upper bound for $d_H(Q_k^*, Q_{k,\hat{w}_k}^*)$. Let q_k^* and $q_{k,\hat{w}_k}^*$ denote the Lebesgue densities of Q_k^* and $Q_{k,\hat{w}_k}^*$, respectively. Let $\tilde{Q}_k^*$ and $\tilde{Q}_{k,\hat{w}_k}^*$ be two measures with Lebesgue densities $\tilde{q}_k^*(x) = Z_k(Q_{k-1}) q_k^*(x)$ and $\tilde{q}_{k,\hat{w}_k}^* = Z_{k,\hat{w}_k}(Q_{k-1}) q_{k,\hat{w}_k}^*(x)$. Then

the distance $\tilde{d}_H(\tilde{Q}_k^*, \tilde{Q}_{k,\hat{w}_k}^*)$ satisfies

$$\begin{aligned}\tilde{d}_H^2(\tilde{Q}_k^*, \tilde{Q}_{k,\hat{w}_k}^*) &= \frac{1}{2}\int_{\mathcal{X}} \left(\sqrt{h_k(y_k,x)\int_{\mathcal{X}} T_k(x, x_{k-1}, \hat{w}_k) q_{k-1}(x_{k-1}) dx_{k-1}} \right. \\ &\qquad \left. - \sqrt{h_k(y_k,x)\int_{\mathcal{X}} T_k(x, x_{k-1}, \bar{w}) q_{k-1}(x_{k-1}) dx_{k-1}} \right)^2 dx \\ &= \frac{1}{2}\int_{\mathcal{X}} h_k(y_k,x) \left(\sqrt{\int_{\mathcal{X}} T_k(x, x_{k-1}, \hat{w}_k) q_{k-1}(x_{k-1}) dx_{k-1}} \right. \\ &\qquad \left. - \sqrt{\int_{\mathcal{X}} T_k(x, x_{k-1}, \bar{w}) q_{k-1}(x_{k-1}) dx_{k-1}} \right)^2 dx.\end{aligned}$$

As we have

$$\begin{aligned}&\left(\sqrt{\int_{\mathcal{X}} T_k(x, x_{k-1}, \hat{w}_k) q_{k-1}(x_{k-1}) dx_{k-1}} - \sqrt{\int_{\mathcal{X}} T_k(x, x_{k-1}, \bar{w}) q_{k-1}(x_{k-1}) dx_{k-1}} \right)^2 \\ &= \int_{\mathcal{X}} T_k(x, x_{k-1}, \hat{w}_k) q_{k-1}(x_{k-1}) dx_{k-1} + \int_{\mathcal{X}} T_k(x, x_{k-1}, \bar{w}) q_{k-1}(x_{k-1}) dx_{k-1} \\ &\qquad - 2\sqrt{\int_{\mathcal{X}} T_k(x, x_{k-1}, \hat{w}_k) q_{k-1}(x_{k-1}) dx_{k-1}} \sqrt{\int_{\mathcal{X}} T_k(x, x_{k-1}, \bar{w}) q_{k-1}(x_{k-1}) dx_{k-1}} \\ &\leq \int_{\mathcal{X}} T_k(x, x_{k-1}, \hat{w}_k) q_{k-1}(x_{k-1}) dx_{k-1} + \int_{\mathcal{X}} T_k(x, x_{k-1}, \bar{w}) q_{k-1}(x_{k-1}) dx_{k-1} \\ &\qquad - 2\int_{\mathcal{X}} \sqrt{T_k(x, x_{k-1}, \hat{w}_k) T_k(x, x_{k-1}, \bar{w})} q_{k-1}(x_{k-1}) dx_{k-1} \\ &= \int_{\mathcal{X}} q_{k-1}(x_{k-1}) \left(\sqrt{T_k(x, x_{k-1}, \hat{w}_k)} - \sqrt{T_k(x, x_{k-1}, \bar{w})} \right)^2 dx_{k-1},\end{aligned}$$

where the first inequality is obtained according to Hölder's inequality, $\tilde{d}_H(\tilde{Q}_k^*, \tilde{Q}_{k,\hat{w}_k}^*)$ satisfies

$$\begin{aligned}&\tilde{d}_H^2(\tilde{Q}_k^*, \tilde{Q}_{k,\hat{w}_k}^*) \\ &\leq \frac{1}{2}\int_{\mathcal{X}} h_k(y_k,x) \left(\int_{\mathcal{X}} q_{k-1}(x_{k-1}) \left(\sqrt{T_k(x, x_{k-1}, \hat{w}_k)} - \sqrt{T_k(x, x_{k-1}, \bar{w})} \right)^2 dx_{k-1} \right) dx.\end{aligned}$$

The term $\left(\sqrt{T_k(x, x_{k-1}, \hat{w}_k)} - \sqrt{T_k(x, x_{k-1}, \bar{w})} \right)^2$ satisfies

$$\begin{aligned}&\left(\sqrt{T_k(x, x_{k-1}, \hat{w}_k)} - \sqrt{T_k(x, x_{k-1}, \bar{w})} \right)^2 \\ &\leq \left| \sqrt{T_k(x, x_{k-1}, \hat{w}_k)} - \sqrt{T_k(x, x_{k-1}, \bar{w})} \right| \cdot \left| \sqrt{T_k(x, x_{k-1}, \hat{w}_k)} + \sqrt{T_k(x, x_{k-1}, \bar{w})} \right| \\ &= \left| T_k(x, x_{k-1}, \hat{w}_k) - T_k(x, x_{k-1}, \bar{w}) \right| \\ &\leq g_{\mathrm{Lip}}(x; k) \|\hat{w}_k - \bar{w}\|_m.\end{aligned}$$

It follows that

$$\begin{aligned}\tilde{d}_H^2(\tilde{Q}_k^*, \tilde{Q}_{k,\hat{w}_k}^*) &\leq \frac{1}{2}\|\hat{w}_k - \bar{w}\|_m \int_{\mathcal{X}} h_k(y_k, x) g_{\text{Lip}}(x; k) dx \\ &= \frac{1}{2}\tilde{C}_{VI}(y_k; k)\|\hat{w}_k - \bar{w}\|_m.\end{aligned}$$

Therefore, by Lemma 43, we have

$$\begin{aligned}d_H(Q_k^*, Q_{k,\hat{w}_k}^*) &\leq \frac{2}{\sqrt{Z_{k,\hat{w}_k}(Q_{k-1})}} \tilde{d}_H(\tilde{Q}_k^*, \tilde{Q}_{k,\hat{w}_k}^*) \\ &= \sqrt{\frac{2\tilde{C}_{VI}(y_k; k)\|\hat{w}_k - \bar{w}\|_m}{Z_{k,\hat{w}_k}(Q_{k-1})}}. \end{aligned} \tag{143}$$

We next derive an upper bound of $d(Q_{k,\hat{w}_k}^*, Q_k)$. The KL divergence between Q_k and $Q_{k,\hat{w}_k}^*$ satisfies:

$$\begin{aligned}KL(Q_k \| Q_{k,\hat{w}_k}^*) &= \mathbb{E}_{q_k(x_k)}\left[\log \frac{q_k(x_k)}{q_{k,\hat{w}_k}^*(x_k)}\right] \\ &= \mathbb{E}_{q_k(x_k)}\left[\log\left(\int_{\mathcal{X}} p_{\mathcal{N}}(y_k \mid \Phi_k(x_k), \Gamma) q_{k,\hat{w}_k}^{*-}(x_k) dx_k\right)\right] \\ &\quad - \mathbb{E}_{q_k(x_k)}\left[\log\left(p_{\mathcal{N}}(y_k \mid \Phi_k(x_k), \Gamma) q_{k,\hat{w}_k}^{*-}(x_k)\right)\right] + \mathbb{E}_{q_k(x_k)}[\log q_k(x_k)] \\ &= \log\left(\int_{\mathcal{X}} p_{\mathcal{N}}(y_k \mid \Phi_k(x_k), \Gamma) q_{k,\hat{w}_k}^{*-}(x_k) dx_k\right) - \mathcal{L}_k(Q_k, \hat{w}_k). \end{aligned} \tag{144}$$

The first term on the right-hand side of Equation (144) satisfies

$$\log\left(\int_{\mathcal{X}} p_{\mathcal{N}}(y_k \mid \Phi_k(x_k), \Gamma) q_{k,\hat{w}_k}^{*-}(x_k) dx_k\right) \leq -\frac{r}{2}\log(2\pi) - \frac{1}{2}\log(\det(\Gamma)).$$

As we also have $\mathcal{L}_k(Q_k, \hat{w}_k) \leq \epsilon_k$, the KL divergence $KL(Q_k \| Q_{k,\hat{w}_k}^*)$ satisfies

$$KL(Q_k \| Q_{k,\hat{w}_k}^*) \leq -\frac{r}{2}\log(2\pi) - \frac{1}{2}\log(\det(\Gamma)) - \epsilon_k.$$

It follows that

$$d_{TV}(Q_k, Q_{k,\hat{w}_k}^*) \leq \frac{1}{\sqrt{2}}\sqrt{KL(Q_k \| Q_{k,\hat{w}_k}^*)} \leq \frac{1}{\sqrt{2}}\sqrt{-\frac{r}{2}\log(2\pi) - \frac{1}{2}\log(\det(\Gamma)) - \epsilon_k}, \tag{145}$$

and

$$d_H(Q_k, Q_{k,\hat{w}_k}^*) \leq \frac{1}{\sqrt{2}}\sqrt{KL(Q_k \| Q_{k,\hat{w}_k}^*)} \leq \frac{1}{\sqrt{2}}\sqrt{-\frac{r}{2}\log(2\pi) - \frac{1}{2}\log(\det(\Gamma)) - \epsilon_k}. \tag{146}$$

Combining Equation (142) with Equation (145) gives

$$d_{TV}(Q_k^*, Q_k) \leq \frac{\tilde{C}_{VI}(y_k; k)\|\hat{w}_k - \bar{w}\|_m}{Z_{k,\hat{w}_k}(Q_{k-1})} + \frac{1}{\sqrt{2}}\sqrt{-\frac{r}{2}\log(2\pi) - \frac{1}{2}\log(\det(\Gamma)) - \epsilon_k}. \tag{147}$$

Similarly, the bound of $d_H(Q_k^*, Q_k)$ can be obtained by combining Equation (143) and Equation (146):

$$d_H(Q_k^*, Q_k) \leq \sqrt{\frac{2\tilde{C}_{VI}(y_k;k)\|\hat{w}_k - \bar{w}\|_m}{Z_{k,\hat{w}_k}(Q_{k-1})}} + \frac{1}{\sqrt{2}}\sqrt{-\frac{r}{2}\log(2\pi) - \frac{1}{2}\log(\det(\Gamma)) - \epsilon_k}. \quad (148)$$

Combining Theorem 16 with Equations (137), (147), and (148) completes the proof the Corollary 26 for the total variation and Hellinger distances.

According to Gibbs and Edward (2002), if the metric space $(\mathcal{X}, d_{\mathcal{X}})$ satisfies

$$\sup_{x,x' \in \mathcal{X}} d_{\mathcal{X}}(x, x') = D < \infty,$$

then we have

$$W_1(P_k, Q_k) \leq D d_{TV}(P_k, Q_k). \quad (149)$$

The conclusion of Corollary 26 for the 1-Wasserstein distance is proved by combining the results of Corollary 26 for the total variation distance and inequality (149).

■

References

Optimal transport and Wasserstein distance. `https://www.stat.cmu.edu/~larry/=sml/Opt.pdf`.

Manisha Aggarwal and Stefan Cobzaș. On some Lipschitz-type functions. *Journal of Mathematical Analysis and Applications*, 517(2):126631, 2023. doi: 10.1016/j.jmaa.2022.126631.

Mohammad Al-Jarrah, Bamdad Hosseini, and Amirhossein Taghvaei. Data-driven approximation of stationary nonlinear filters with optimal transport maps. In *IEEE 63rd Conference on Decision and Control (CDC)*, pages 2727–2733, 2024.

Mohammad Al-Jarrah, Bamdad Hosseini, and Amirhossein Taghvaei. Fast filtering of non-Gaussian models using amortized optimal transport maps. *arXiv:2503.12633*, 2025.

Sanjib Basu. Stability and infinitesimal robustness of posterior distribution and posterior quantities. *Statistics & Probability Letters*, 71:151–162, 1998. doi: 10.1016/S0378-3758(98)00090-1.

Sanjib Basu. Uniform stability of posteriors. *Statistics & Probability Letters*, 46:53–58, 2000. doi: 10.1016/S0167-7152(99)00086-3.

James O. Berger and Elías Moreno. An overview of robust Bayesian analysis. *Test*, pages 5–124, 1994. doi: 10.1007/BF02562676.

S Bose. Bayesian robustness with more than one class of contamination. *J. Statist. Planning and Inference*, 40:177–187, 1994. doi: 10.1016/0378-3758(94)90119-8.

David R. Burt, Carl Edward Rasmussen, and Mark van der Wilk. Convergence of sparse variational inference in gaussian processes regression. *Journal of Machine Learning Research*, 21(131):1–63, 2020. URL `http://jmlr.org/papers/v21/19-1015.html`.

Andrew Campbell, Yuyang Shi, Tom Rainforth, and Arnaud Doucet. Online variational filtering and parameter learning. In *35th Conference on Neural Information Processing Systems*, 2021.

J.A. Carrillo, F. Hoffmann, A. M. Stuart, and U. Vaes. Statistical accuracy of approximate filtering methods. *arXiv:2402.01593*, 2024.

Marta Catalano, Antonio Lijoi, and Igor Prünster. Measuring dependence in the Wasserstein distance for Bayesian nonparametric models. *Annals of Statistics*, (5):2916–2947, 2021. doi: 10.1214/21-AOS2065.

Mathis Chagneux, Élisabeth Gassiat, Pierre Gloaguen, and Sylvain Le Corff. Additive smoothing error in backward variational inference for general state-space models. *Journal of Machine Learning Research*, 25(28):1–33, 2024. URL `https://www.jmlr.org/papers/volume25/22-1392/22-1392.pdf`.

Dan Crisan and Arnaud Doucet. A survey of convergence results on particle filtering. *IEEE Transactions on Signal Processing*, 50(3):736–746, 2002. doi: 10.1109/78.984773.

Dan Crisan, Alberto López-Yela, and Joaquin Miguez. Stable approximation schemes for optimal filters. *SIAM/ASA Journal on Uncertainty Quantification*, 8(1):483–509, 2020. doi: 10.1137/19M1255410.

Tiangang Cui and Sergey Dolgov. Deep composition of tensor-trains using squared inverse Rosenblatt transport. *Foundations of Computational Mathematics*, 22(6):1863–1922, 2022. doi: 10.1007/s10208-021-09537-5.

M Dowling, Y. Zhao, and II. M Park. real-time variational method for learning neural trajectory and its dynamics. In *11th International Conference on Learning Representations*, 2023.

Víctor Elvira, Joaquín Miguez, and Petar Djurić. On the performance of particle filters with adaptive number of particles. *Statistics and Computing*, 31(81), 2021. doi: 10.1007/s11222-021-10056-0.

Nando Freitas and Gordon Neil. *Sequential Monte Carlo Methods in Practice*, volume 1. New York: springer, 2001.

Alfredo Garbuno-Inigo, Tapio Helin, Franca Hoffmann, and Bamdad Hosseini. Bayesian posterior perturbation analysis with integral probability metrics. *arXiv:2303.01512*, 2023.

Alison L. Gibbs and Su Francis Edward. On choosing and bounding probability metrics. *Internation Statistical Reivew*, 70(3):419–435, 2002. doi: 10.2307/1403865.

J. Grashorn, M. Broggi, L. Chamoin, and M. Beer. Transport map coupling filter for state-parameter estimation. In *Advances in Reliability, Safety and Security ESREL*, 2024.

P. Gustafson and Wasserman L. Local sensitivity diagnostics for Bayesian inference. *Ann. Stats.*, (6):2153–2167, 1995. doi: 10.1214/aos/1034713652.

K Ito and K Xiong. Gaussian filters for nonlinear filtering problems. *IEEE Transactions on Automatic Control*, 45(5):910–927, 2000. doi: 10.1109/9.855552.

Anya Katsevich and Philippe Rigollet. On the approximation accuracy of gaussian variational inference. *Annals of Statistics*, (4):1384–1409, 2024. doi: 10.1214/24-AOS2393.

K. Law, A. Stuart, and K. Zygalakis. *Data Assimilation: A Mathematical Introduction*, volume 62. Springer, Cham, 2015.

F. Le Gland, V. Monbet, and V.D. Tran. *Large Sample Asymptotics for the Ensemble Kalman Filter*. Diss. INRIA, 2009.

François Le Gland and Nadia Oudjane. Stability and uniform approximation of nonlinear filters using the Hilbert metric and application to particle filters. *Annals of Applied Probability*, 14(1):144–187, 2004. doi: 10.1214/aoap/1075828050.

E E Leamer. Sets of posterior means with bounded variance prior. *Econometrica*, 50: 725–736, 1982. doi: 10.2307/1912610.

Julian Neri, Roland Badeau, and Philippe Depalle. Probabilistic filter and smoother for variational inference of Bayesian linear dynamical system. In *2020 IEEE International Conference on Acoustics, Speech and Signal Processing*, 2020.

Ning Ning and Edward L. Ionides. Iterated block particle filter for high-dimensional parameter learning: beating the curse of dimensionality. *Journal of Machine Learning Research*, 24(82):1–76, 2023.

Patrick Rebesschini and Ramon van Handel. Can local particle filters beat the curse of dimensionality? *Annals of Applied Probability*, 25(5):2809–2866, 2015. doi: 10.1214/14-AAP1061.

Judith Rousseau and Catia Scricciolo. Wasserstein convergence in Bayesian and frequentist deconvolution models. *Annals of Statistics*, (4):1691–1715, 2024. doi: 10.1214/24-AOS2413.

F Ruggeri. Posterior ranges of functions of parameters under priors with specified quantiles. *Comm. Statist. Theory and methods*, 19:127–144, 1990. doi: 10.1080/03610929008830192.

D. Sanz-Alonso, A. Stuart, and A. Taeb. *Inverse Problems and Data Assimilation*, volume 107. Cambridge University Press, 2023.

T. Sedighi, A. Hosseinian-Far, and A. Daneshkhah. Measuring local sensitivity in Bayesian inference using a new class of metrics. *Communications in Statistics - Theory and Methods*, (11):3581–3597, 2021. doi: 10.1080/03610926.2021.1977956.

S Sivaganesan. Bounds on posterior expectation for density bounded classes with constant bandwith. *J. Statist. Planning and Inference*, 40:331–343, 1994. doi: 10.1016/0378-3758(94)90129-5.

Björn Sprungk. On the local lipschitz stability of Bayesian inverse problems. *Inverse Problems*, 36(5):055015, 2020. doi: 10.1088/1361-6420/ab6f43.

Simo Säkkä. *Bayesian Filtering and Smoothing*. Cambridge University Press, 2013.

C Vallani. *Optimal Transport: Old and New*. Grundlehren der mathematischen Wissenschaften. Berlin, Germany: Springer, 2009.

Liliang Wang and Alex Gorodetsky. Factorization-based online variational inference for state-parameter estimation of partially observable nonlinear dynamical systems. In *AIAA Scitech 2025 Forum*, 2025. doi: 10.2514/6.2025-1960.

Fengshuo Zhang and Chao Gao. Convergence rates of variational posterior distributions. *Annals of Statistics*, (4):2180–2207, 2020. doi: 10.1214/19-AOS1883.

Yiran Zhao and Tiangang Cui. Tensor-train methods for sequential state and parameter learning in state-space models. *Journal of Machine Learning Research*, 25(244):1–51, 2024.