SIAM J. CONTROL OPTIM.
Vol. 58, No. 2, pp. 1144–1182
© 2020 Society for Industrial and Applied Mathematics

Downloaded 05/17/20 to 130.15.241.167. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

ROBUSTNESS TO INCORRECT SYSTEM MODELS IN STOCHASTIC CONTROL*

ALI D. KARA† AND SERDAR YÜKSEL†

Abstract. In stochastic control applications, typically only an ideal model (controlled transition kernel) is assumed and the control design is based on the given model, raising the problem of performance loss due to the mismatch between the assumed model and the actual model. Toward this end, we study continuity properties of discrete-time stochastic control problems with respect to system models (i.e., controlled transition kernels) and robustness of optimal control policies designed for incorrect models applied to the true system. We study both fully observed and partially observed setups under an infinite horizon discounted expected cost criterion. We show that continuity can be established under total variation convergence of the transition kernels under mild assumptions and with further restrictions on the dynamics and observation model under weak and setwise convergence of the transition kernels. Using these continuity properties, we establish convergence results and error bounds due to mismatch that occurs by the application of a control policy which is designed for an incorrectly estimated system model to a true model, thus establishing positive and negative results on robustness. Compared to the existing literature, we obtain strictly refined robustness results that are applicable even when the incorrect models can be investigated under weak convergence and setwise convergence criteria (with respect to a true model), in addition to the total variation criteria. These entail positive implications on empirical learning in (data-driven) stochastic control since often system models are learned through empirical training data where typically a weak convergence criterion applies but stronger convergence criteria do not.

Key words. stochastic control, robustness, incorrect models, partially observed models, learning

AMS subject classifications. 93E20, 93E03, 93E11, 62G35

DOI. 10.1137/18M1208058

1. Introduction.

1.1. Preliminaries. In this paper, we study continuity properties of stochastic control problems with respect to transition kernels and applications of these to robustness of optimal control policies applied to systems with incomplete or incorrect probabilistic models.

We start with the probabilistic setup of the problem. Let $\mathbb{X} \subset \mathbb{R}^m$ denote a Borel set which is the state space of a partially observed controlled Markov process. Here and throughout the paper $\mathbb{Z}_+$ denotes the set of nonnegative integers and $\mathbb{N}$ denotes the set of positive integers. Let $\mathbb{Y} \subset \mathbb{R}^n$ be a Borel set denoting the observation space of the model, and let the state be observed through an observation channel Q. The observation channel, Q, is defined as a stochastic kernel (regular conditional probability) from $\mathbb{X}$ to $\mathbb{Y}$, such that $Q(\cdot|x)$ is a probability measure on the (Borel) σ-algebra $\mathcal{B}(\mathbb{Y})$ of $\mathbb{Y}$ for every $x \in \mathbb{X}$, and $Q(A|\cdot) : \mathbb{X} \to [0,1]$ is a Borel measurable function for every $A \in \mathcal{B}(\mathbb{Y})$. A decision maker (DM) is located at the output of the channel Q, and hence it only sees the observations $\{Y_t, t \in \mathbb{Z}_+\}$ and chooses its actions from $\mathbb{U}$, the action space which is a Borel subset of some Euclidean space.

*Received by the editors August 17, 2018; accepted for publication (in revised form) January 13, 2020; published electronically April 27, 2020. Some of the results in this paper were presented at the 2018 IEEE Conference on Decision and Control.
https://doi.org/10.1137/18M1208058
Funding: This research was supported in part by the Natural Sciences and Engineering Research Council (NSERC) of Canada.
†Mathematics and Statistics, Queen's University, Kingston, Ontario, K7L 3N6 (16adk@queensu.ca, yuksel@mast.queensu.ca).

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 05/17/20 to 130.15.241.167. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

An *admissible policy* γ is a sequence of control functions $\{\gamma_t,\, t \in \mathbb{Z}_+\}$ such that γ_t is measurable with respect to the σ-algebra generated by the information variables

$$I_t = \{Y_{[0,t]}, U_{[0,t-1]}\}, \quad t \in \mathbb{N}, \qquad I_0 = \{Y_0\},$$

where

$$U_t = \gamma_t(I_t), \quad t \in \mathbb{Z}_+, \tag{1.1}$$

are the $\mathbb{U}$-valued control actions and

$$Y_{[0,t]} = \{Y_s,\, 0 \le s \le t\}, \quad U_{[0,t-1]} = \{U_s,\, 0 \le s \le t-1\}.$$

We define Γ to be the set of all such admissible policies.

The update rules of the system are determined by (1.1) and the following relationships:

$$\Pr\big((X_0, Y_0) \in B\big) = \int_B P(dx_0)Q(dy_0|x_0), \quad B \in \mathcal{B}(\mathbb{X} \times \mathbb{Y}),$$

where P is the (prior) distribution of the initial state X_0, and

$$\begin{aligned}&\Pr\left((X_t, Y_t) \in B \,\middle|\, (X, Y, U)_{[0,t-1]} = (x, y, u)_{[0,t-1]}\right)\\ &= \int_B \mathcal{T}(dx_t|x_{t-1}, u_{t-1})Q(dy_t|x_t), B \in \mathcal{B}(\mathbb{X} \times \mathbb{Y}), t \in \mathbb{N},\end{aligned}$$

where $\mathcal{T}$ is the transition kernel of the model which is a stochastic kernel from $\mathbb{X} \times \mathbb{U}$ to $\mathbb{X}$.

Using stochastic realization results (see Lemma 1.2 in [21] or Lemma 3.1 of [11]), the process defined above can be represented in functional form as follows:

$$X_{t+1} = f(X_t, U_t, W_t), \qquad Y_t = g(X_t, V_t) \tag{1.2}$$

for some measurable functions f, g with $\{W_t\}$ being an independent and identically distributed (i.i.d.) system noise process and $\{V_t\}$ an i.i.d. disturbance process, which are independent of X_0 and each other. Here, the first equation represents the transition kernel $\mathcal{T}$ as it gives the relation of the most recent state and action variables to the upcoming state. From this representation it can be seen that the probabilistic nature of the kernel is determined by the function f and the probability model of the noise W_t. The second equation represents the measurement channel Q, as it describes the relation between the state and observation variables. We let the objective of the agent (decision maker) be the minimization of the infinite horizon discounted cost,

$$J_\beta(P, \mathcal{T}, \gamma) = E_P^{\mathcal{T},\gamma}\left[\sum_{t=0}^{\infty} \beta^t c(X_t, U_t)\right]$$

for some discount factor $\beta \in (0, 1)$, over the set of admissible policies $\gamma \in \Gamma$, where $c : \mathbb{X} \times \mathbb{U} \to \mathbb{R}$ is a Borel-measurable stagewise cost function and $E_P^{\mathcal{T},\gamma}$ denotes the expectation with initial state probability measure P and transition kernel $\mathcal{T}$ under policy γ. Note that $P \in \mathcal{P}(\mathbb{X})$, where we let $\mathcal{P}(\mathbb{X})$ denote the set of probability measures on $\mathbb{X}$.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 05/17/20 to 130.15.241.167. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

We define the optimal cost for the discounted infinite horizon setup as a function of the priors and the transition kernels as

$$J^*_\beta(P,\mathcal{T}) = \inf_{\gamma\in\Gamma} J_\beta(P,\mathcal{T},\gamma).$$

The focus of the paper will be to address the following problems.

Problem P1: Continuity of $J^*_\beta(P,\mathcal{T})$ under the convergence of the transition kernels. Let $\{\mathcal{T}_n, n\in\mathbb{N}\}$ be a sequence of transition kernels which converge in some sense to another transition kernel $\mathcal{T}$. Does that imply that

$$J^*_\beta(P,\mathcal{T}_n) \to J^*_\beta(P,\mathcal{T})?$$

Problem P2: Robustness to incorrect models. A problem of major practical importance is robustness of an optimal controller to modeling errors. Suppose that an optimal policy is constructed according to a model which is incorrect: how does the application of the control to the true model affect the system performance and does the error decrease to zero as the models become closer to each other? In particular, suppose that γ^*_n is an optimal policy designed for $\mathcal{T}_n$, an incorrect model for a true model $\mathcal{T}$. Is it the case that if $\mathcal{T}_n \to \mathcal{T}$, then $J_\beta(P,\mathcal{T},\gamma^*_n) \to J^*_\beta(P,\mathcal{T})$?

Problem P3: Empirical consistency of learned probabilistic models and data-driven stochastic control. Let $\mathcal{T}(\cdot|x,u)$ be a transition kernel given previous state and action variables $x\in\mathbb{X}, u\in\mathbb{U}$, which is unknown to the decision maker. Suppose the DM builds a model for the transition kernels, $\mathcal{T}_n(\cdot|x,u)$, for all possible $x\in\mathbb{X}, u\in\mathbb{U}$ by collecting training data (e.g., from the evolving system). Do we have that the cost calculated under $\mathcal{T}_n$ converges to the true cost (i.e., do we have that the cost obtained from applying the optimal policy for the empirical model converges to the true cost as the training length increases)?

1.2. Literature review. Robustness is a desired property for the optimal control of stochastic or deterministic systems when a given model does not reflect the actual system perfectly, as is usually the case in practice.

A common approach in the literature has been to design controllers that work sufficiently well for all possible uncertain systems under some structured constraints, such as H_∞ norm bounded perturbations (see [3, 58]). The design for robust controllers has often been developed through a game theoretic formulation where the minimizer is the controller and the maximizer is the uncertainty. The connections of this formulation to risk sensitive control were established in [30, 15]. Using Legendre-type transforms, relative entropy constraints came in to the literature to probabilistically model the uncertainties (see, e.g., [41, equation (4)] or [15, equations (2)–(3)]. Here, one selects a nominal system which satisfies a relative entropy bound between the actual measure and the nominal measure and solves a risk sensitive optimal control problem, and this solution value provides an upper bound for the original system performance. As such, a common approach in robust stochastic control has been to consider all models which satisfy certain bounds in terms of relative entropy pseudodistance (or Kullback–Leibler divergence); see, e.g., [15, 41, 40, 10] among others.

Other metrics or criteria, different from the relative entropy pseudodistance, have also been used to quantify the uncertainty in the system models. Reference [51] has studied a min-max formulation for robust control where the one-stage transition kernel belongs to a ball under the total variation metric for each state action pair. For distributionally robust stochastic optimization problems, it is assumed that the

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 05/17/20 to 130.15.241.167. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

underlying probability measure of the system lies within an ambiguity set and a worst case single-stage optimization is made considering the probability measures in the ambiguity set. To construct ambiguity sets, [9, 18] use the Wasserstein metric (see section 1.3), [17] uses the Prokhorov metric which metrizes the weak topology, [50] uses the total variation distance, and [34] works with relative entropy. [55, 29, 37] have studied robust dynamic programming approaches through a min-max formulation for fully observed finite state-action space models with uncertain transition probabilities. Further related work with model uncertainty includes [38, 4, 56], with some further work in the economics literature [26, 22]. In [33], an optimal filtering problem for a control-free system with uncertainties in transition kernels and measurement channels is considered.

Further related studies include [12], which studies the optimal control of systems with unknown dynamics for a linear quadratic regulator setup and proposes an algorithm to learn the system from observed data with quantitative convergence bounds. [52] considers stochastic uncertainties while [47] considers deterministic structured uncertainties in robust control; some connections between these and our paper can be seen in the examples presented in section 1.3.2.

For fully observed models, [35, Theorem 5.1] establishes continuity results for approximate models and gives a set convergence result for sets of optimal control actions, but this set convergence result is inconclusive for robustness without further assumptions on the true system model. Reference [36] is another related work which studies continuity of the value function for fully observed models under a general metric defined as the integral probability metric, which captures both the total variation metric or the Kantorovich metric with different setups (which is not weaker than the metrics leading to weak convergence). We also note that approximation methods for stochastic control problems with standard Borel spaces through quantization, which lead to finite models, can be viewed as approximations of transition kernels, but this interpretation requires caution: indeed, [46, 1, 2], among many others, study approximation methods for Markov decision processes (MDPs) where the convergence of approximate models is satisfied in a particularly constructed fashion. In particular, [46] presents a construction for the approximate models through quantizing the actual model with continuous spaces (leading to a finite space model), which allows for continuity and robustness results with only a weak continuity assumption on the true transition kernel which, in turn, leads to the weak convergence of the approximate models. A detailed analysis of approximation methods for continuous state and action spaces can be found in [44] for both fully observed and partially observed models. However, these positive results on the weak convergence of approximate kernels to the true one in such studies do not directly apply to the robustness of an arbitrary sequence of models which converges weakly to a true model, as our counterexamples in this paper will demonstrate.

Related work also includes our recent studies [57, 32]. [57] considers various topologies on the sets of observation channels and quantizers in partially observed stochastic control and provides some supporting results. [32] presents robustness and continuity properties for stochastic control problems with respect to the prior measures with fixed transition kernels; that is, the robustness only to initial priors is studied in [32]. Different from these studies, here we study continuity in and robustness to incorrect transition kernels, which requires significantly different analytical tools due to the dynamic nature of the problems.

In section 1.4 we will present the contributions, which will also make the comparison with the reviewed literature more explicit.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 05/17/20 to 130.15.241.167. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

1.3. Some examples and convergence criteria for transition kernels.

1.3.1. Convergence criteria for transition kernels. Before presenting convergence criteria for controlled transition kernels, we first review the convergence of probability measures. Three important notions of convergences for sets of probability measures to be studied in the paper are weak convergence, setwise convergence, and convergence under total variation. For $N \in \mathbb{N}$, a sequence $\{\mu_n, n \in \mathbb{N}\}$ in $\mathcal{P}(\mathbb{R}^N)$ is said to converge to $\mu \in \mathcal{P}(\mathbb{R}^N)$ *weakly* if

$$\int_{\mathbb{R}^N} c(x)\mu_n(\mathrm{d}x) \to \int_{\mathbb{R}^N} c(x)\mu(\mathrm{d}x)$$

for every continuous and bounded $c : \mathbb{R}^N \to \mathbb{R}$. $\{\mu_n\}$ is said to converge *setwise* to $\mu \in \mathcal{P}(\mathbb{R}^N)$ if the above holds for all measurable and bounded $c : \mathbb{R}^N \to \mathbb{R}$. For probability measures $\mu, \nu \in \mathcal{P}(\mathbb{R}^N)$, the *total variation* metric is given by

$$\|\mu - \nu\|_{TV} = 2 \sup_{B \in \mathcal{B}(\mathbb{R}^N)} |\mu(B) - \nu(B)| = \sup_{f : \|f\|_\infty \leq 1} |\int f(x)\mu(\mathrm{d}x) - \int f(x)\nu(\mathrm{d}x)|,$$

where the supremum is taken over all measurable real f such that

$$\|f\|_\infty = \sup_{x \in \mathbb{R}^N} |f(x)| \leq 1.$$

A sequence $\{\mu_n\}$ is said to converge in total variation to $\mu \in \mathcal{P}(\mathbb{R}^N)$ if $\|\mu_n - \mu\|_{TV} \to 0$. Total variation defines a stringent metric for convergence; for example, a sequence of discrete probability measures does not converge in total variation to a probability measure which admits a density function. Setwise convergence, though, induces a topology on the space of probability measures which is not metrizable [20, p. 59]. However, the space of probability measures on a complete, separable, metric (Polish) space endowed with the topology of weak convergence is itself complete, separable, and metric [39]. We also note here that relative entropy convergence, through Pinsker's inequality [23, Lemma 5.2.8], is stronger than even total variation convergence, which has also been studied in robust stochastic control as reviewed earlier. Another metric for probability measures is the Wasserstein distance: For compact spaces, the Wasserstein distance of order 1, denoted by W_1, metrizes the weak topology (see [53, Theorem 6.9]). For noncompact spaces convergence in the W_1 metric implies weak convergence (in particular this metric bounds from above the bounded-Lipschitz metric [53, p. 109]). Considering these relations, our results in this paper can be directly generalized to the relative entropy distance or the Wasserstein distance. Building on the above, we introduce the following convergence notions for (controlled) transition kernels.

Definition 1. *For a sequence of transition kernels $\{\mathcal{T}_n, n \in \mathbb{N}\}$, we say that*
(i) *$\mathcal{T}_n \to \mathcal{T}$ weakly if $\mathcal{T}_n(\cdot|x,u) \to \mathcal{T}(\cdot|x,u)$ weakly for all $x \in \mathbb{X}$ and $u \in \mathbb{U}$,*
(ii) *$\mathcal{T}_n \to \mathcal{T}$ setwise if $\mathcal{T}_n(\cdot|x,u) \to \mathcal{T}(\cdot|x,u)$ setwise for all $x \in \mathbb{X}$ and $u \in \mathbb{U}$,*
(iii) *$\mathcal{T}_n \to \mathcal{T}$ under the total variation distance if $\mathcal{T}_n(\cdot|x,u) \to \mathcal{T}(\cdot|x,u)$ under total variation for all $x \in \mathbb{X}$ and $u \in \mathbb{U}$.*

1.3.2. Examples. Let a controlled model be given as

$$x_{t+1} = F(x_t, u_t, w_t),$$

where $\{w_t\}$ is an i.i.d. noise process. The uncertainty on the transition kernel for such a system may arise from lack of information on F or the i.i.d. noise process w_t or both:

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 05/17/20 to 130.15.241.167. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

(i) Let $\{F_n\}$ denote an approximating sequence for F, so that $F_n(x,u,w) \to F(x,u,w)$ pointwise. Assume that the probability measure of the noise is known. Then, corresponding kernels $\mathcal{T}_n$ converge weakly to $\mathcal{T}$: If we denote the probability measure of w with μ, for any $g \in C_b(\mathbb{X})$ and for any $(x_0,u_0) \in \mathbb{X} \times \mathbb{U}$ using the dominated convergence theorem we have

$$\lim_{n\to\infty} \int g(x_1)\mathcal{T}_n(dx_1|x_0,u_0) = \lim_{n\to\infty} \int g(F_n(x_0,u_0,w))\mu(dw)$$
$$= \int g(F(x_0,u_0,w))\mu(dw) = \int g(x_1)\mathcal{T}(dx_1|x_0,u_0).$$

(ii) Much of the robust control literature deals with deterministic systems where the nominal model is a deterministic perturbation of the actual model (see, e.g., [47, 12]). The considered model is in the following form: $\tilde{F}(x_t,u_t) = F(x_t,u_t) + \Delta F(x_t,u_t)$, where F represents the nominal model and ΔF is the model uncertainty satisfying some norm bounds. For such deterministic systems, pointwise convergence of $\tilde{F}$ to the nominal model F, i.e., $\Delta F(x_t,u_t) \to 0$, can be viewed as weak convergence for deterministic systems by the discussion in (i). It is evident, however, that total variation convergence would be too strong for such a convergence criterion, since $\delta_{\tilde{F}(x_t,u_t)} \to \delta_{F(x_t,u_t)}$ weakly but $\|\delta_{\tilde{F}(x_t,u_t)} - \delta_{F(x_t,u_t)}\|_{TV} = 2$ for all $\Delta F(x_t,u_t) \neq 0$ where δ denotes the Dirac measure.

(iii) Let $F(x_t,u_t,w_t) = f(x_t,u_t) + w_t$ be such that the function f is known and $w_t \sim \mu$ is not known correctly and an incorrect model μ_n is assumed. If $\mu_n \to \mu$ weakly, setwise, or in total variation, then the corresponding transition kernels $\mathcal{T}_n$ converge in the same sense to $\mathcal{T}$. Observe the following:

$$\int g(x_1)\mathcal{T}_n(dx_1|x_0,u_0) - \int g(x_1)\mathcal{T}(dx_1|x_0,u_0)$$
$$= \int g(w_0 + f(x_0,u_0))\mu_n(dw_0) - \int g(w_0 + f(x_0,u_0))\mu(dw_0). \tag{1.3}$$

(a) Suppose $\mu_n \to \mu$ weakly. If g is a continuous and bounded function, then $g(\cdot + f(x_0,u_0))$ is a continuous and bounded function for all $(x_0,u_0) \in \mathbb{X} \times \mathbb{U}$. Thus, (1.3) goes to 0. Note that f does not need to be continuous. (b) Suppose $\mu_n \to \mu$ setwise. If g is a measurable and bounded function, then $g(\cdot + f(x_0,u_0))$ measurable and bounded for all $(x_0,u_0) \in \mathbb{X} \times \mathbb{U}$. Thus, (1.3) goes to 0. (c) Finally, assume $\mu_n \to \mu$ in total variation. If g is bounded, (1.3) converges to 0, as in item (b). As a special case, assume that μ_n and μ admit densities h_n and h, respectively; then the pointwise convergence of h_n to h implies the convergence of μ_n to μ in total variation by Scheffé's theorem.

(iv) Suppose now neither F nor the probability model of w_t is known perfectly. It is assumed that w_t admits a measure μ_n and $\mu_n \to \mu$ weakly. For the function F we again have an approximating sequence $\{F_n\}$. If $F_n(x,u,w_n) \to F(x,u,w)$ for all $(x,u) \in \mathbb{X} \times \mathbb{U}$ and for any $w_n \to w$, then the transition kernel $\mathcal{T}_n$ corresponding to the model F_n converges weakly to the one of F, $\mathcal{T}$: For any $g \in C_b(\mathbb{X})$,

$$\lim_{n\to\infty} \int g(x_1)\mathcal{T}_n(dx_1|x_0,u_0) = \lim_{n\to\infty} \int g(F_n(x_0,u_0,w))\mu_n(dw)$$
$$= \int g(F(x_0,u_0,w))\mu(dw) = \int g(x_1)\mathcal{T}(dx_1|x_0,u_0).$$

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 05/17/20 to 130.15.241.167. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

In the analysis above we used a generalized dominated convergence result, Lemma A.1, to be presented later building on [35, Theorem 3.5] and [49, Theorem 3.5].

(v) Let again $\{F_n\}$ denote an approximating sequence for F and suppose now $F_{x_0,u_0,n}(\cdot) := F_n(x_0,u_0,\cdot) : \mathbb{W} \to \mathbb{X}$ is invertible for all $x_0, u_0 \in \mathbb{X} \times \mathbb{U}$ and $F^{-1}_{(x_0,u_0),n}(\cdot)$ denotes the inverse for fixed (x_0,u_0). It is assumed that $F^{-1}_{(x_0,u_0),n}(x_1) \to F^{-1}_{x_0,u_0}(x_1)$ pointwise for all (x_0,u_0). Suppose further that the noise process w_t admits a continuous density $f_W(w)$. The Jacobian matrix, $\frac{\partial x_1}{\partial w}$, is the matrix whose components are the partial derivatives of x_1, i.e., with $x_1 \in \mathbb{X} \subset \mathbb{R}^m$ and $w \in \mathbb{W} \subset \mathbb{R}^m$, it is an $m \times m$ matrix with components $\frac{\partial (x_1)_i}{\partial w_j}$, $1 \le i,j \le m$. If the Jacobian matrix of derivatives $\frac{\partial x_1}{\partial w}(w)$ is continuous in w and nonsingular for all w, then by the inverse function theorem of vector calculus (see [25, section 1.11]), we have that the density of the state variables can be written as

$$f_{X_1,n,(x_0,u_0)}(x_1) = f_W(F^{-1}_{x_0,u_0,n}(x_1))\left|\frac{\partial x_1}{\partial w}(F^{-1}_{x_0,u_0,n}(x_1))\right|^{-1},$$

$$f_{X_1,(x_0,u_0)}(x_1) = f_W(F^{-1}_{x_0,u_0}(x_1))\left|\frac{\partial x_1}{\partial w}(F^{-1}_{x_0,u_0}(x_1))\right|^{-1}.$$

With the above, $f_{X_1,n,(x_0,u_0)}(x_1) \to f_{X_1,(x_0,u_0)}(x_1)$ pointwise for all fixed (x_0,u_0). Therefore, by Scheffé's theorem, the corresponding kernels $\mathcal{T}_n(\cdot|x_0,u_0) \to \mathcal{T}(\cdot|x_0,u_0)$ in total variation for all (x_0,u_0).

(vi) These studies will be used and analyzed in detail in section 5, where data-driven stochastic control problems will be considered where estimated models are obtained through empirical measurements of the state action variables.

1.4. Summary of results and contributions. We now introduce the main assumptions that will be occasionally used for our technical results in the paper.

Assumption 1.1.

(a) *The sequence of transition kernels $\mathcal{T}_n$ satisfies the following: $\{\mathcal{T}_n(\cdot|x_n,u_n), n \in \mathbb{N}\}$ converges weakly to $\mathcal{T}(\cdot|x,u)$ for any sequence $\{x_n,u_n\} \subset \mathbb{X} \times \mathbb{U}$ and $x,u \in \mathbb{X} \times \mathbb{U}$ such that $(x_n,u_n) \to (x,u)$.*

(b) *The stochastic kernels $\mathcal{T}(\cdot|x,u)$ and $\{\mathcal{T}_n(\cdot|x,u)\}_n$ are weakly continuous in (x,u).*

(c) *The stagewise cost function $c(x,u)$ is nonnegative, bounded, and continuous on $\mathbb{X} \times \mathbb{U}$.*

(d) *$\mathbb{U}$ is compact.*

Assumption 1.2. *The observation channel $Q(\cdot|x)$ is continuous in total variation, i.e., if $x_k \to x$, then $Q(\cdot\,|x_k) \to Q(\cdot\,|x)$ in total variation (only for partially observed models).*

Assumption 1.3.

(a) *The sequence of transition kernels $\mathcal{T}_n$ satisfies the following: $\{\mathcal{T}_n(\cdot|x,u_n), n \in \mathbb{N}\}$ converges setwise to $\mathcal{T}(\cdot|x,u)$ for any sequence $\{u_n\} \subset \mathbb{U}$ and $x,u \in \mathbb{X} \times \mathbb{U}$ such that $u_n \to u$.*

(b) *The stochastic kernels $\mathcal{T}(\cdot|x,u)$ and $\{\mathcal{T}_n(\cdot|x,u)\}_n$ are setwise continuous in u.*

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 05/17/20 to 130.15.241.167. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

(c) *The stagewise cost function* $c(x,u)$ *is nonnegative, bounded, and continuous on* $\mathbb{U}$.
(d) $\mathbb{U}$ *is compact.*

Assumption 1.4.
(a) *The sequence of transition kernels* $\mathcal{T}_n$ *satisfies the following:* $\|\mathcal{T}_n(\cdot|x,u_n) - \mathcal{T}(\cdot|x,u)\|_{TV} \to 0$ *for any sequence* $\{u_n\} \subset \mathbb{U}$ *and* $x,u \in \mathbb{X} \times \mathbb{U}$ *such that* $u_n \to u$.
(b) *The stochastic kernels* $\mathcal{T}(\cdot|x,u)$ *and* $\{\mathcal{T}_n(\cdot|x,u)\}_n$ *are continuous in total variation in* u.
(c) *The stagewise cost function* $c(x,u)$ *is nonnegative, bounded, and continuous on* $\mathbb{U}$.
(d) $\mathbb{U}$ *is compact.*

In sections 2 and 3 we study continuity (Problem P1) and robustness (Problem P2) for partially observed models. In particular we show the following:

- Continuity and robustness do not hold in general under weak convergence of kernels (Theorem 2.1).
- Under Assumptions 1.1 and 1.2, continuity and robustness hold (Theorems 2.3, 3.3).
- Continuity and robustness do not hold in general under setwise convergence of the kernels (Theorem 2.4).
- Continuity and robustness do not hold in general under total variation convergence of the kernels (Example 4.1).
- Under Assumption 1.4, continuity and robustness hold (Theorems 2.5, 3.2).

In section 4, we study continuity (Problem P1) and robustness (Problem P2) for fully observed models. In particular we show the following:

- Continuity and robustness do not hold in general under weak convergence of kernels (Theorem 4.1, Example 4.1).
- Under Assumption 1.1, continuity holds (Theorem 4.2), and under Assumption 1.1, robustness holds if the optimal policies for every initial point are identical (Theorem 4.4).
- Continuity and robustness do not hold in general under setwise convergence of the kernels (Theorems 4.5, 4.7).
- Under Assumption 1.3, continuity holds (Theorem 4.6), and under Assumption 1.3, robustness holds if the optimal policies for every initial point are identical (Theorem 4.8).
- Continuity and robustness do not hold in general under total variation convergence of the kernels (Example 4.1).
- Under Assumption 1.4, continuity and robustness hold (subsection 4.5).

Compared to the existing literature reviewed earlier, the above results use strictly more relaxed and refined convergence criteria to study robustness. In section 5, these results will be applied to arrive at positive and negative implications on empirical consistency and data-driven learning in stochastic control since often system models are learned through empirical training data where typically weak convergence criterion applies (in an almost sure sense) but stronger convergence criteria do not.

2. Continuity of optimal cost with respect to convergence of transition kernels (partially observed case). In this section, we will study continuity of the optimal discounted cost under the convergence of transition kernels for partially observed models.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 05/17/20 to 130.15.241.167. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

2.1. Weak convergence.

2.1.1. Absence of continuity under weak convergence. The following shows that the optimal cost may not be continuous under weak convergence of transition kernels.

THEOREM 2.1. *Let $\mathcal{T}_n \to \mathcal{T}$ weakly, then it is not necessarily true that $J^*_\beta(P, \mathcal{T}_n) \to J^*_\beta(P, \mathcal{T})$ even when the prior distributions are the same, the measurement channel Q is continuous in total variation, and $c(x, u)$ is continuous and bounded on $\mathbb{X} \times \mathbb{U}$.*

Proof. We prove the result with a counterexample. Letting $\mathbb{X} = \mathbb{U} = \mathbb{Y} = [-1, 1]$ and $c(x, u) = (x - u)^2$, the observation channel is chosen to be uniformly distributed over [-1,1], $Q \sim U([-1, 1])$, the initial distributions of the state variable are chosen to be same as $P \sim \delta_1$, where $\delta_x(A) := \mathbb{1}_{\{x \in A\}}$ for Borel A, and the transition kernels are

$$\begin{aligned}\mathcal{T}(\cdot|x, u) &= \delta_{-1}(x)\left[\frac{1}{2}\delta_1(\cdot) + \frac{1}{2}\delta_{-1}(\cdot)\right] + \delta_1(x)\left[\frac{1}{2}\delta_1(\cdot) + \frac{1}{2}\delta_{-1}(\cdot)\right] \\ &\quad + (1 - \delta_{-1}(x))(1 - \delta_1(x))\delta_0(\cdot), \\ \mathcal{T}_n(\cdot|x, u) &= \delta_{-1}(x)\left[\frac{1}{2}\delta_{(1-1/n)}(\cdot) + \frac{1}{2}\delta_{(-1+1/n)}(\cdot)\right] + \delta_1(x)\left[\frac{1}{2}\delta_{(1-1/n)}(\cdot) + \frac{1}{2}\delta_{(-1+1/n)}(\cdot)\right] \\ &\quad + (1 - \delta_{-1}(x))(1 - \delta_1(x))\delta_0(\cdot).\end{aligned}$$

In other words, for $\mathcal{T}$

$$\begin{aligned}&\text{if } x_t \in \{1, -1\}, \text{ then } \quad x_{t+1} = 1 \text{ or } -1 \\ &\text{else } x_{t+1} = 0,\end{aligned}$$

for $\mathcal{T}_n$

$$\begin{aligned}&\text{if } x_t \in \{1, -1\}, \text{ then } \quad x_{t+1} = 1 - \frac{1}{n} \text{ or } -1 + \frac{1}{n} \\ &\text{else } x_{t+1} = 0\end{aligned}$$

independent of the control, where the events noted above with *or* are equally likely. It can be seen that $\mathcal{T}_n \to \mathcal{T}$ weakly according to Definition 1(i). Since the cost function is mean square error, control does not affect the dynamics and the channel is noninformative, and the optimal policy is

$$\gamma^*_k(y_{[0,k]}) = E[X_k] = \begin{cases} 0 & \text{if } k > 0, \\ 1 & \text{if } k = 0. \end{cases}$$

Note that the cost function is continuous, and the measurement channel is continuous in total variation. The optimal discounted costs can be found as

$$\begin{aligned}J^*_\beta(P, \mathcal{T}) &= \sum_{k=1}^{\infty} E^{\mathcal{T}}_P[\beta^k X_k^2] = \sum_{k=1}^{\infty} \beta^k = \frac{\beta}{1 - \beta}, \\ J^*_\beta(P, \mathcal{T}_n) &= \sum_{k=1}^{\infty} E^{\mathcal{T}_n}_P[\beta^k X_k^2] = \beta\left[\frac{1}{2}\left(1 - \frac{1}{n}\right)^2 + \frac{1}{2}\left(-1 + \frac{1}{n}\right)^2\right].\end{aligned}$$

Then we have $J^*_\beta(P, \mathcal{T}_n) \to \beta \neq \frac{\beta}{1-\beta}$. ∎

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 05/17/20 to 130.15.241.167. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

2.2. A sufficient condition for continuity under weak convergence. In the following, we will establish and utilize some regularity properties for the optimal cost with respect to the convergence of transition kernels.

ASSUMPTION 2.1.
(a) *The stochastic kernel* $\mathcal{T}(\cdot|x,u)$ *is weakly continuous in* (x,u).
(b) *The observation channel* $Q(\cdot|x)$ *is continuous in total variation, i.e., if* $x_k \to x$, *then* $Q(\,\cdot\,|x_k) \to Q(\,\cdot\,|x)$ *in total variation.*
(c) *The stagewise cost function* $c(x,u)$ *is nonnegative, bounded, and continuous on* $\mathbb{X} \times \mathbb{U}$
(d) $\mathbb{U}$ *is compact.*

It is a standard result that any partially observed Markov decision process can be reduced to a (completely observable) MDP, whose states are the posterior state distributions or *beliefs* of the observer; that is, the state at time t is $Z_t(\,\cdot\,) := \mathsf{Pr}\{X_t \in \cdot\,|Y_0,\ldots,Y_t,U_0,\ldots,U_{t-1}\} \in \mathcal{P}(\mathbb{X})$. We call this equivalent MDP the belief-MDP, The belief-MDP has state space $\mathbb{Z} = \mathcal{P}(\mathbb{X})$ and action space $\mathbb{U}$. Under the topology of weak convergence, since $\mathbb{X}$ is a Borel space, $\mathbb{Z}$ is metrizable with the Prokhorov metric which makes $\mathbb{Z}$ into a Borel space [39]. The transition probability η of the belief-MDP can be constructed through nonlinear filtering equations [45, pp. 334–335]. The one-stage cost function c of the belief-MDP is given by $\tilde{c}(z,u) := \int_{\mathbb{X}} c(x,u)z(dx)$. By [5, Proposition 7.30], the one stage cost function $\tilde{c}$ of the belief-MDP is continuous and bounded, that is, in $C_b(\mathbb{Z} \times \mathbb{U})$, under Assumption 2.1(c). By [19, Theorem 3.7, Example 4.1] (see also [31]), under Assumption 2.1, the stochastic kernel η for belief-MDP is weakly continuous in (z,u). For an MDP with weakly continuous transition probabilities and compact action spaces, it follows that an optimal control policy exists: This follows because the *discounted cost optimality operator* $T : C_b(\mathbb{Z}) \to C_b(\mathbb{Z})$ (see, e.g., [28, Chapter 8.5]),

$$(T(f))(z) = \min_u(\tilde{c}(z,u) + \beta E[f(z_1)|z_0 = z, u_0 = u]), \tag{2.1}$$

is a contraction from $C_b(\mathbb{Z})$ to itself under the supremum norm. As a result, there exists a fixed point, the value function, and an optimal control policy exists. In view of this existence result, in the following we will consider optimal policies. We note though that for the results which do not use the assumption, one may use ϵ-optimal policies without affecting the results.

THEOREM 2.2. *Under Assumptions* 1.1 *and* 1.2,

$$\sup_{\gamma\in\Gamma} |J_\beta(P,\mathcal{T}_n,\gamma) - J_\beta(P,\mathcal{T},\gamma)| \to 0.$$

Proof.

$$\begin{aligned}
&\sup_{\gamma\in\Gamma} |J_\beta(P,\mathcal{T}_n,\gamma) - J_\beta(P,\mathcal{T},\gamma)| \\
&\quad = \sup_{\gamma\in\Gamma} \left| \sum_{t=0}^{\infty} \beta^t \left(E_P^{\mathcal{T}}\Big[c\big(X_t,\gamma(Y_{[0,t]})\big)\Big] - E_P^{\mathcal{T}_n}\Big[c\big(X_t,\gamma(Y_{[0,t]})\big)\Big] \right) \right| \\
&\quad \le \sum_{t=0}^{\infty} \beta^t \sup_{\gamma\in\Gamma} \left| E_P^{\mathcal{T}}\Big[c\big(X_t,\gamma(Y_{[0,t]})\big)\Big] - E_P^{\mathcal{T}_n}\Big[c\big(X_t,\gamma(Y_{[0,t]})\big)\Big] \right|.
\end{aligned}$$

Recall that an *admissible policy* γ is a sequence of control functions $\{\gamma_t,\, t \in \mathbb{Z}_+\}$. At the last step above, we make a slight abuse of notation; the sup at the first step is

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

over all sequence of control functions $\{\gamma_t,\, t \in \mathbb{Z}_+\}$ whereas the sup at the last step is over all sequence of control functions $\{\gamma_{t'},\, t' \leq t\}$, but we will use the same notation, γ, in the rest of the proof. In Appendix A.1, we show the following for any $t \geq 0$:

$$\sup_{\gamma\in\Gamma}\left|E_P^{\mathcal{T}}\left[c\big(X_t,\gamma(Y_{[0,t]})\big)\right] - E_P^{\mathcal{T}_n}\left[c\big(X_t,\gamma(Y_{[0,t]})\big)\right]\right| \to 0. \tag{2.2}$$

For any $\epsilon > 0$, we choose a $K < \infty$ such that $\sum_{t=K+1}^{\infty} \beta^k 2\|c\|_\infty \leq \epsilon/2$. For the chosen K, we choose an $N < \infty$ such that

$$\sup_{\gamma\in\Gamma}\left|E_P^{\mathcal{T}}\left[c\big(X_t,\gamma(Y_{[0,t]})\big)\right] - E_P^{\mathcal{T}_n}\left[c\big(X_t,\gamma(Y_{[0,t]})\big)\right]\right| \leq \epsilon/2K$$

for all time stages $t \leq K$ and for all $n > N$. Thus, we have that $\sup_{\gamma\in\Gamma}\left|J_\beta(P,\mathcal{T}_n,\gamma) - J_\beta(P,\mathcal{T},\gamma)\right. \to 0$ as $n \to \infty$. ∎

Now we give the main result of this section.

THEOREM 2.3. *Suppose the conditions of Theorem* 2.2 *hold. Then*

$$\lim_{n\to\infty} |J_\beta^*(P,\mathcal{T}_n) - J_\beta^*(P,\mathcal{T})| = 0.$$

Proof. We start with the following bound:

$$\begin{aligned} &|J_\beta^*(P,\mathcal{T}_n) - J_\beta^*(P,\mathcal{T})| \\ &\qquad \leq \max\Big(J_\beta(P,\mathcal{T}_n,\gamma^*) - J_\beta(P,\mathcal{T},\gamma^*), J_\beta(P,\mathcal{T},\gamma_n^*) - J_\beta(P,\mathcal{T}_n,\gamma_n^*)\Big), \end{aligned} \tag{2.3}$$

where γ^* and γ_n^* are the optimal policies, respectively, for $\mathcal{T}$ and $\mathcal{T}_n$. Both terms go to 0 by Theorem 2.2. ∎

2.3. Absence of continuity under setwise convergence. We now show that continuity of optimal costs may fail under the setwise convergence of transition kernels. Theorem 4.5 in the next section establishes this result for fully observed models. As we note later, a fully observed system can be viewed as a partially observed system with the measurement being the state itself through (4.1), and therefore, in view of space constraints, a separate proof will not be provided for the following result.

THEOREM 2.4. *Let $\mathcal{T}_n \to \mathcal{T}$ setwise. Then, it is not true in general that $J_\beta^*(P,\mathcal{T}_n) \to J_\beta^*(P,\mathcal{T})$, even when $\mathbb{X}, \mathbb{Y}$, and $\mathbb{U}$ are compact and $c(x,u)$ is continuous and bounded in $\mathbb{X}\times\mathbb{U}$.*

2.4. Continuity under total variation. We have the following results.

THEOREM 2.5. *Under Assumption* 1.4

$$J_\beta^*(P,\mathcal{T}_n) \to J_\beta^*(P,\mathcal{T}).$$

Proof. We start with the following bound:

$$|J_\beta^*(\mathcal{T}_n) - J_\beta^*(\mathcal{T})| \leq \max\Big(J_\beta(\mathcal{T}_n,\gamma^*) - J_\beta(\mathcal{T},\gamma^*), J_\beta(\mathcal{T},\gamma_n^*) - J_\beta(\mathcal{T}_n,\gamma_n^*)\Big),$$

where γ^* and γ_n^* are the optimal policies, respectively, for $\mathcal{T}$ and $\mathcal{T}_n$.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 05/17/20 to 130.15.241.167. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Downloaded 05/17/20 to 130.15.241.167. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

We now study the following:

$$\begin{aligned}
&\sup_{\gamma\in\Gamma}|J_\beta(P,\mathcal{T}_n,\gamma)-J_\beta(P,\mathcal{T},\gamma)|\\
&\quad=\sup_{\gamma\in\Gamma}\left|\sum_{t=0}^{\infty}\beta^t\left(E_P^{\mathcal{T}}\Big[c\big(X_t,\gamma(Y_{[0,t]})\big)\Big]-E_P^{\mathcal{T}_n}\Big[c\big(X_t,\gamma(Y_{[0,t]})\big)\Big]\right)\right|\\
&\quad\le\sum_{t=0}^{\infty}\beta^t\sup_{\gamma\in\Gamma}\left|E_P^{\mathcal{T}}\Big[c\big(X_t,\gamma(Y_{[0,t]})\big)\Big]-E_P^{\mathcal{T}_n}\Big[c\big(X_t,\gamma(Y_{[0,t]})\big)\Big]\right|.
\end{aligned}$$

In Appendix A.2 we show that for all $t<\infty$

$$\sup_{\gamma\in\Gamma}\left|E_P^{\mathcal{T}}\Big[c\big(X_t,\gamma(Y_{[0,t]})\big)\Big]-E_P^{\mathcal{T}_n}\Big[c\big(X_t,\gamma(Y_{[0,t]})\big)\Big]\right|\to 0. \tag{2.4}$$

For any $\epsilon>0$, we choose a $K<\infty$ such that $\sum_{t=K+1}^{\infty}\beta^t 2\|c\|_\infty\le\epsilon/2$. For the chosen K, we choose an $N<\infty$ such that

$$\sup_{\gamma\in\Gamma}\left|E_P^{\mathcal{T}}\Big[c\big(X_t,\gamma(Y_{[0,t]})\big)\Big]-E_P^{\mathcal{T}_n}\Big[c\big(X_t,\gamma(Y_{[0,t]})\big)\Big]\right|\le\epsilon/2K$$

for all time stages $t\le K$ and for all $n>N$. Therefore, we have that for any given $\epsilon>0$, for $n>N$

$$\sup_{\gamma\in\Gamma}\big|J_\beta(\mathcal{T}_n,\gamma)-J_\beta(\mathcal{T},\gamma)\big|<\epsilon. \tag{2.5}$$

Thus, the result follows. ∎

We now present a result on the rate of convergence. For stochastic control problems, *strategic measures* are defined [48] as the set of probability measures induced on the product spaces of the state and action pairs by admissible control policies: Given an initial distribution on the state, and a policy, one can uniquely define a probability measure on the infinite product space consistent with finite dimensional distributions, by the Ionescu–Tulcea theorem [27, Proposition C.10]. Now, define a strategic measure under a policy $\gamma^n=\{\gamma_0^n,\gamma_1^n,\ldots,\gamma_k^n,\ldots\}$ as a probability measure defined on $\mathcal{B}(\mathbb{X}\times\mathbb{Y}\times\mathbb{U})^{\mathbb{Z}_+}$ by

$$\begin{aligned}
&P_{\mathcal{T}}^{\gamma^n}(\mathrm{d}(x_0,y_0,u_0),\mathrm{d}(x_1,y_1,u_1),\ldots)\\
&\quad=P(\mathrm{d}x_0)Q(\mathrm{d}y_0|x_0)1_{\{\gamma^n(y_0)\in\mathrm{d}u_0\}}\mathcal{T}(\mathrm{d}x_1|x_0,u_0)Q(\mathrm{d}y_1|x_1)1_{\{\gamma^n(y_0,y_1)\in\mathrm{d}u_1\}}\cdots.
\end{aligned}$$

Next, with uniformity in the total variation convergence, Theorem 2.5 is enhanced.

THEOREM 2.6. *If the cost function c is bounded,*

$$|J_\beta^*(P,\mathcal{T}_n)-J_\beta^*(P,\mathcal{T})|\le\|c\|_\infty\frac{\beta}{(\beta-1)^2}\sup_{x\in\mathbb{X},u\in\mathbb{U}}\|\mathcal{T}_n(.|x,u)-\mathcal{T}(.|x,u)\|_{TV}.$$

Proof. We start with the following bound as before:

$$|J_\beta^*(\mathcal{T}_n)-J_\beta^*(\mathcal{T})|\le\max\Big(J_\beta(\mathcal{T}_n,\gamma^*)-J_\beta(\mathcal{T},\gamma^*),J_\beta(\mathcal{T},\gamma_n^*)-J_\beta(\mathcal{T}_n,\gamma_n^*)\Big),$$

where γ^* and γ_n^* are the optimal policies, respectively, for $\mathcal{T}$ and $\mathcal{T}_n$.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 05/17/20 to 130.15.241.167. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Then, with $P^{\gamma}_{\mathcal{T}_n}$ and $P^{\gamma}_{\mathcal{T}}$ denoting the strategic measures for two chains with a policy γ and kernels $\mathcal{T}_n$ and $\mathcal{T}$, we have

$$\begin{aligned}
&|J_\beta(\mathcal{T}_n,\gamma) - J_\beta(\mathcal{T},\gamma)| \\
&\leq \sum_k \beta^k |\int c(x_k,\gamma(y_{[0,k]}))P^{\gamma}_{\mathcal{T}_n}(dx_k,dy_{[0,k]}) - \int c(x_k,\gamma(y_{[0,k]}))P^{\gamma}_{\mathcal{T}}(dx_k,dy_{[0,k]})| \\
&\leq \sum_k \beta^k \|c\|_\infty \|P^{\gamma}_{\mathcal{T}_n}(d(x,y,u)_{[0,k]}) - P^{\gamma}_{\mathcal{T}}(d(x,y,u)_{[0,k]})\|_{TV}.
\end{aligned}$$

In Appendix A.3 we establish the following relation:

$$\|P^{\gamma}_{\mathcal{T}_n}(d(x,y,u)_{[0,k]}) - P^{\gamma}_{\mathcal{T}}(d(x,y,u)_{[0,k]})\|_{TV} \leq k \sup_{x\in\mathbb{X},u\in\mathbb{U}} \|\mathcal{T}_n(.|x,u) - \mathcal{T}(.|x,u)\|_{TV}. \tag{2.6}$$

Using this bound, we will have

$$\begin{aligned}
\|J^*_\beta(\mathcal{T}_n) - J^*_\beta(\mathcal{T})| &\leq \sum_k \beta^k \|c\|_\infty k \sup_{x\in\mathbb{X},u\in\mathbb{U}} \|\mathcal{T}(.|x,u) - \mathcal{T}_n(.|x,u)\|_{TV} \\
&= \|c\|_\infty \frac{\beta}{(\beta-1)^2} \sup_{x\in\mathbb{X},u\in\mathbb{U}} \|\mathcal{T}(.|x,u) - \mathcal{T}_n(.|x,u)\|_{TV}. \qquad \square
\end{aligned}$$

3. Robustness to incorrect transition kernels (partially observed case).

Here, we consider the robustness problem P2: Suppose we design an optimal policy, γ^*_n, for a transition kernel, $\mathcal{T}_n$, assuming it is the correct model and apply the policy to the true model whose transition kernel is $\mathcal{T}$. We study the robustness of the suboptimal policy γ^*_n.

3.1. Total variation.

THEOREM 3.1. *Suppose the stagewise cost function $c(x,u)$ is bounded in $\mathbb{X}\times\mathbb{U}$, then*

$$|J_\beta(P,\mathcal{T},\gamma^*_n) - J^*_\beta(P,\mathcal{T})| \leq 2\|c\|_\infty \frac{\beta}{(\beta-1)^2} \sup_{x\in\mathbb{X},u\in\mathbb{U}} \|\mathcal{T}(.|x,u) - \mathcal{T}_n(.|x,u)\|_{TV}$$

for a fixed prior distribution $P\in\mathcal{P}(\mathbb{X})$, where γ^*_n is the optimal policy designed for the transition kernel $\mathcal{T}_n$.

Proof. We begin with the following:

$$|J_\beta(\mathcal{T},\gamma^*_n) - J^*_\beta(\mathcal{T})| \leq |J_\beta(\mathcal{T},\gamma^*_n) - J_\beta(\mathcal{T}_n,\gamma^*_n)| + |J_\beta(\mathcal{T}_n,\gamma^*_n) - J_\beta(\mathcal{T},\gamma^*)|.$$

The second term is bounded using Theorem 2.6. For the first term, we use the proof of Theorem 2.6, where we showed that for any $\gamma\in\Gamma$

$$|J_\beta(\mathcal{T},\gamma) - J_\beta(\mathcal{T}_n,\gamma)| \leq \|c\|_\infty \frac{\beta}{(\beta-1)^2} \sup_{x\in\mathbb{X},u\in\mathbb{U}} \|\mathcal{T}(.|x,u) - \mathcal{T}_n(.|x,u)\|_{TV}.$$

Thus, the result follows. $\square$

The next theorem gives an asymptotic robustness result.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 05/17/20 to 130.15.241.167. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

THEOREM 3.2. *Under Assumption* 1.4

$$|J_\beta(P,\mathcal{T},\gamma_n^*) - J_\beta^*(P,\mathcal{T})| \to 0,$$

where γ_n^* *is the optimal policy designed for the kernel* $\mathcal{T}_n$.

Proof. We write the following:

$$|J_\beta(P,\mathcal{T},\gamma_n^*) - J_\beta^*(P,\mathcal{T})| \leq |J_\beta(P,\mathcal{T},\gamma_n^*) - J_\beta^*(P,\mathcal{T}_n)| + |J_\beta^*(P,\mathcal{T}_n) - J_\beta^*(P,\mathcal{T})|.$$

The second term goes to 0 by Theorem 2.5 and the first term goes to 0 using (2.5) again from the proof of Theorem 2.5. ∎

3.2. Setwise convergence. Theorem 4.7 in the next section establishes the lack of robustness under the setwise convergence of kernels. As we note later, a fully observed system can be viewed as a partially observed system with the measurement being the state itself (see (4.1)).

3.3. Weak convergence.

THEOREM 3.3. *Under Assumptions* 1.1 *and* 1.2, $|J_\beta(\mathcal{T},\gamma_n^*) - J_\beta^*(\mathcal{T})| \to 0$, *where* γ_n^* *is the optimal policy designed for the transition kernel* $\mathcal{T}_n$.

Proof. We write

$$|J_\beta(\mathcal{T},\gamma_n^*) - J_\beta^*(\mathcal{T})| \leq |J_\beta(\mathcal{T},\gamma_n^*) - J_\beta(\mathcal{T}_n,\gamma_n^*)| + |J_\beta(\mathcal{T}_n,\gamma_n^*) - J_\beta(\mathcal{T},\gamma^*)|.$$

The first term goes to 0 by Theorem 2.2. For the second term we use Theorem 2.3. ∎

Remark 3.1. In this paper we study the case where the channel is known to the controller; that is, the true channel model Q is available to the controller. For the case where this is no longer true, the following analysis can be made. If the transition kernel $\mathcal{T}$ and the channel Q are not known, the controller would have an approximating sequence $\mathcal{T}_nQ_n(x_{t+1},y_{t+1} \in \cdot \times \cdot|x_t,u_t) \in \mathcal{P}(\mathbb{X}\times\mathbb{Y})$ for the true joint measure $\mathcal{T}Q(x_{t+1},y_{t+1} \in \cdot \times \cdot|x_t,u_t) \in \mathcal{P}(\mathbb{X}\times\mathbb{Y})$ for all (x_t,u_t). The question then becomes analyzing the convergence of $\mathcal{T}_nQ_n \to \mathcal{T}Q$. Due to space constraints, we do not present explicit results on this problem; however, we note that in [57], a similar joint convergence is studied for convergence of measurement channels and fixed model/prior distributions. The reader can refer to [57, Lemma 2.2] for an analysis on the convergence of $PQ_n \to PQ$ for a single stage problem, and for a multistage problem, the following can be considered: With

$$|J_\beta^*(\mathcal{T}_nQ_n) - J_\beta^*(\mathcal{T}Q)| \leq |J_\beta^*(\mathcal{T}_nQ_n) - J_\beta^*(\mathcal{T}Q_n)| + |J_\beta^*(\mathcal{T}Q_n) - J_\beta^*(\mathcal{T}Q)|,$$

[57, Theorem 6.2] presents sufficient conditions to guarantee the convergence of the second term above. For the first term, the total variation convergence results in this paper provide an analysis on the uniform convergence over a class of channels, thus establishing positive results on continuity under the joint convergence of both transition kernels and measurement channels.

4. Continuity and robustness in the fully observed case. In this section, we consider the fully observed case where the controller has direct access to the state variables. We present the results for this case separately, since here we cannot utilize the regularity properties of measurement channels, which allows for stronger continuity and robustness results. Similar to the discussions related to (2.1), that

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 05/17/20 to 130.15.241.167. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

is, as the operator defined in (2.1) is a contraction and as it admits a fixed point (value function), under measurable selection conditions due to weak or strong (setwise) continuity of transition kernels [27, section 3.3], for infinite horizon discounted cost problems, optimal policies can be selected from those which are stationary and deterministic. Therefore we will restrict the policies to be stationary and deterministic so that $U_t = \gamma(X_t)$ for some measurable function γ. Notice also that fully observed models can be viewed as partially observed with the measurement channel thought to be

$$Q(\cdot|x) = \delta_x(\cdot), \tag{4.1}$$

which is only weakly continuous, and thus it does not satisfy Assumption 1.2.

4.1. Weak convergence.

4.1.1. Absence of continuity under weak convergence.

We start with a negative result.

THEOREM 4.1. *For $\mathcal{T}_n \to \mathcal{T}$ weakly, it is not necessarily true that $J^*_\beta(\mathcal{T}_n) \to J^*_\beta(\mathcal{T})$ even when the prior distributions are the same and $c(x,u)$ is continuous and bounded in $\mathbb{X} \times \mathbb{U}$.*

Proof. We prove the result with a counterexample, similar to the model used in the proof of Theorem 2.1. Letting $\mathbb{X} = [-1,1]$, $\mathbb{U} = \{-1,1\}$, and $c(x,u) = (x-u)^2$, the initial distributions are given by $P \sim \delta_1$, that is, $X_0 = 1$, and the transition kernels are

$$\begin{aligned}\mathcal{T}(\cdot|x,u) = \delta_{-1}(x)\left[\frac{1}{2}\delta_1(\cdot) + \frac{1}{2}\delta_{-1}(\cdot)\right] + \delta_1(x)\left[\frac{1}{2}\delta_1(\cdot) + \frac{1}{2}\delta_{-1}(\cdot)\right]\\ + (1-\delta_{-1}(x))(1-\delta_1(x))\delta_0(\cdot),\\ \mathcal{T}_n(\cdot|x,u) = \delta_{-1}(x)\left[\frac{1}{2}\delta_{(1-1/n)}(\cdot) + \frac{1}{2}\delta_{(-1+1/n)}(\cdot)\right] + \delta_1(x)\left[\frac{1}{2}\delta_{(1-1/n)}(\cdot) + \frac{1}{2}\delta_{(-1+1/n)}(\cdot)\right]\\ + (1-\delta_{-1}(x))(1-\delta_1(x))\delta_0(\cdot).\end{aligned}$$

It can be seen that $\mathcal{T}_n \to \mathcal{T}$ weakly according to Definition 1(i). Under this setup we can calculate the optimal costs as follows:

$$J^*_\beta(\mathcal{T}_n) = \frac{1}{n^2} + \sum_{k=2}^{\infty} \beta^k = \frac{1}{n^2} + \frac{\beta^2}{1-\beta},$$

and $J^*_\beta(\mathcal{T}) = 0$. Thus, continuity does not hold. ∎

We now present another counterexample emphasizing the importance of continuous convergence in the actions. The following counterexample shows that without the continuous convergence and regularity assumptions on the kernel $\mathcal{T}$, continuity fails even when $\mathcal{T}_n(\cdot|x,u) \to \mathcal{T}(\cdot|x,u)$ pointwise (for x,u) in total variation (also setwise and weakly) and even when the cost function $c(x,u)$ is continuous and bounded. Notice that this example also holds for setwise and weak convergence.

EXAMPLE 4.1. *Assume that the kernels are given by*

$$\mathcal{T}_n(\cdot|x,u) \sim U([u^n, 1+u^n]),$$
$$\mathcal{T}(\cdot|x,u) \sim \begin{cases} U([0,1]) & \text{if } u \neq 1,\\ U([1,2]) & \text{if } u = 1,\end{cases}$$

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 05/17/20 to 130.15.241.167. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

where $\mathbb{U} = [0,1]$ *and* $\mathbb{X} = \mathbb{R}$. *We note first that* $\mathcal{T}_n(\cdot|x,u) \to \mathcal{T}(\cdot|x,u)$ *in total variation for every fixed* x *and* u.

The cost function is in the following form:

$$c(x,u) = \begin{cases} 2 & \text{if } x \le \frac{1}{e}, \\ 2 - \frac{x - \frac{1}{e}}{0.1} & \text{if } \frac{1}{e} < x \le 0.1 + \frac{1}{e}, \\ 1 & \text{if } 0.1 + \frac{1}{e} < x \le 1 + \frac{1}{e} - 0.1, \\ 2 - \frac{1 + \frac{1}{e} - x}{0.1} & \text{if } 1 + \frac{1}{e} - 0.1 < x \le 1 + \frac{1}{e}, \\ 2 & \text{if } 1 + \frac{1}{e} < x. \end{cases}$$

Notice that $c(x,u)$ *is a continuous function.*

With this setup, $\gamma^*(x) = 0$ *is an optimal policy for* $\mathcal{T}$ *since on the* $[0,1]$ *interval the induced cost is less than the cost induced on the* $[1,2]$ *interval. The cost under this policy is*

$$J^*_\beta(\mathcal{T}) = \sum_{t=0}^{\infty} \beta^t \left(2 \times \frac{1}{e} + \frac{0.3}{2} + 0.9 - \frac{1}{e}\right) = \frac{1}{1-\beta}\left(1.05 + \frac{1}{e}\right).$$

For $\mathcal{T}_n$, $\gamma^*_n(x) = e^{-\frac{1}{n}}$ *is an optimal policy for every* n *as* $e^{-\frac{1}{n} \times n} = \frac{1}{e}$ *and thus the state is distributed between* $\frac{1}{e} < x \le 1 + \frac{1}{e}$ *in which interval the cost is the least. Hence, we can write*

$$\lim_{n\to\infty} J_\beta(\mathcal{T}_n, \gamma^*_n) = \sum_{t=0}^{\infty} \beta^t \left(0.3 + 1 - 0.2\right) = \frac{1.1}{1-\beta} \neq \frac{1}{1-\beta}\left(1.05 + \frac{1}{e}\right) = J^*_\beta(\mathcal{T}). \quad \square$$

4.1.2. A sufficient condition for continuity under weak convergence. We will now establish that if the kernels and the model components have some further regularity, continuity does hold.

The assumptions of the following result are the same as the assumptions for the partially observed case (Theorem 2.3) except for the assumption on the measurement channel Q.

THEOREM 4.2. *Under Assumption* 1.1, $J_\beta(\mathcal{T}_n, \gamma^*_n) \to J_\beta(\mathcal{T}, \gamma^*)$ *for any initial state* x_0, *as* $n \to \infty$.

Proof. We build on the proof of [43, Proposition 3.10]. We will use the successive approximations for an inductive argument.

Recall *discounted cost optimality operator* $T : C_b(\mathbb{Z}) \to C_b(\mathbb{Z})$ from (2.1)

$$(T(v))(x) = \inf_u (c(x,u) + \beta E[v(x_1)|x_0 = x, u_0 = u]),$$

which is a contraction from $C_b(\mathbb{X})$ to itself under the supremum norm and has a fixed point, the value function.

For the kernel $\mathcal{T}$, we will denote the approximation functions by

$$v^k(x) = T(v^{k-1})(x)$$

and for the kernel $\mathcal{T}_n$ we will use $v^k_n(x)$ to denote the approximation functions; notice that the operator T also depends on n for the model $\mathcal{T}_n$, but we will continue using it as T in what follows.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 05/17/20 to 130.15.241.167. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

We wish to show that the approximation functions for $\mathcal{T}_n$ continuously converge to the ones for $\mathcal{T}$. Then, for the first step of the induction we have

$$v^1(x) = c(x, u^*) \quad v_n^1(x_n) = c(x_n, u_n^*),$$

and thus we can write

$$|v^1(x) - v_n^1(x_n)| \leq \sup_{u \in \mathbb{U}} \big|c(x, u) - c(x_n, u)\big|$$

since $c \in C_b(\mathbb{X} \times \mathbb{U})$ and the action space, $\mathbb{U}$, is compact, and the first step of the induction holds, i.e., $\lim_{n\to\infty} |v^1(x) - v_n^1(x_n)| = 0$.

For the kth step we have

$$v^k(x) = T(v^{k-1})(x) = \inf_u \big[c(x, u) + \beta \int_{\mathbb{X}} v^{k-1}(x^1)\mathcal{T}(dx^1|x, u)\big],$$
$$v_n^k(x_n) = T(v_n^{k-1})(x_n) = \inf_u \big[c(x_n, u) + \beta \int_{\mathbb{X}} v_n^{k-1}(x^1)\mathcal{T}_n(dx^1|x_n, u)\big].$$

Note that the assumptions of the theorem satisfy the measurable selection criteria and hence we can choose minimizing selectors [27, section 3.3]. If we denote the selectors by u^* and u_n^*, we can write

$$\begin{aligned}
&|v^k(x) - v_n^k(x_n)| \\
&\leq \max\Bigg(\Bigg[|c(x, u^*) - c(x_n, u^*)| + \beta\Big|\int_{\mathbb{X}} v^{k-1}(x^1)\mathcal{T}(dx^1|x, u^*) \\
&\qquad\qquad - \int_{\mathbb{X}} v_n^{k-1}(x^1)\mathcal{T}_n(dx^1|x_n, u^*)\Big|\Bigg], \\
&\Bigg[|c(x, u_n^*) - c(x_n, u_n^*)| + \beta\Big|\int_{\mathbb{X}} v^{k-1}(x^1)\mathcal{T}(dx^1|x, u_n^*) - \int_{\mathbb{X}} v_n^{k-1}(x^1)\mathcal{T}_n(dx^1|x_n, u_n^*)\Big|\Bigg]\Bigg).
\end{aligned}$$

Hence, we can write

(4.2)

$$\begin{aligned}
&|v^k(x) - v_n^k(x_n)| \\
&\leq \sup_{u \in \mathbb{U}} \Bigg[|c(x, u) - c(x_n, u)| + \beta\Big|\int_{\mathbb{X}} v^{k-1}(x^1)\mathcal{T}(dx^1|x, u) - \int_{\mathbb{X}} v_n^{k-1}(x^1)\mathcal{T}_n(dx^1|x_n, u)\Big|\Bigg],
\end{aligned}$$

and the first term goes to 0 as $c(x, u)$ is continuous in x uniformly over all $u \in \mathbb{U}$. For the second term we write

$$\begin{aligned}
&\sup_{u \in \mathbb{U}} \Big|\int_{\mathbb{X}} v^{k-1}(x^1)\mathcal{T}(dx^1|x, u) - \int_{\mathbb{X}} v_n^{k-1}(x^1)\mathcal{T}_n(dx^1|x_n, u)\Big| \\
&\leq \sup_{u \in \mathbb{U}} \Big|\int_{\mathbb{X}} \big(v^{k-1}(x^1) - v_n^{k-1}(x^1)\big)\mathcal{T}_n(dx^1|x_n, u)\Big| \\
&\quad + \sup_{u \in \mathbb{U}} \Big|\int_{\mathbb{X}} v^{k-1}(x^1)\mathcal{T}(dx^1|x, u) - \int_{\mathbb{X}} v^{k-1}(x^1)\mathcal{T}_n(dx^1|x_n, u)\Big|,
\end{aligned}$$

where, here for the first term, by the induction argument for any $x_n^1 \to x^1$, $\big|v^{k-1}(x^1) - v_n^{k-1}(x_n^1)\big| \to 0$ (i.e., we have continuous convergence). We also have that $\mathcal{T}_n(\cdot|x_n, u) \to$

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 05/17/20 to 130.15.241.167. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

$\mathcal{T}(\cdot|x,u)$ weakly uniformly over $u \in \mathbb{U}$ as $\mathbb{U}$ is compact. Therefore, using Lemma A.1 the first term goes to 0. For the second term we again use that $\mathcal{T}_n(\cdot|x_n,u)$ converges weakly to $\mathcal{T}(\cdot|x,u)$ uniformly over $u \in \mathbb{U}$. With an almost identical induction argument it can also be shown that $v^{k-1}(x^1)$ is continuous in x^1, and thus the second term also goes to 0.

So far, we have showed that for any $k \in \mathbb{N}$, $\lim_{n\to\infty} \left|v_n^k(x_n) - v^k(x)\right| = 0$ for any $x_n \to x$, in particular it is also true that $\lim_{n\to\infty} \left|v_n^k(x) - v^k(x)\right| = 0$ for any x.

As we have stated earlier, it can be shown that the approximation operator T is a contractive operator under supremum norm with modulus β and it converges to a fixed point which is the value function. Thus, we have

$$\left|J_\beta(\mathcal{T},\gamma^*) - v^k(x)\right| \le \|c\|_\infty \frac{\beta^k}{1-\beta}, \quad \left|J_\beta^*(\mathcal{T}_n,\gamma_n^*) - v_n^k(x)\right| \le \|c\|_\infty \frac{\beta^k}{1-\beta}. \tag{4.3}$$

Combining the results,

$$\begin{aligned}&|J_\beta(\mathcal{T}_n,\gamma_n^*) - |J_\beta(\mathcal{T},\gamma^*)| \\ &\qquad \le |J_\beta(\mathcal{T}_n,\gamma_n^*) - v_n^k(x)| + |v_n^k(x) - v^k(x)| + |J_\beta(\mathcal{T},\gamma^*) - v^k(x)|.\end{aligned}$$

Note that the first and the last term can be made arbitrarily small since (4.3) holds for all $k \in \mathbb{N}$; the second term goes to 0 with an inductive argument for all $k \in \mathbb{N}$. ∎

4.1.3. Absence of robustness under weak convergence. The following result shows that the conditions that satisfy the continuity are not sufficient for robustness in the fully observed models.

THEOREM 4.3. *Supposing $\mathcal{T}_n(\cdot|x_n,u_n) \to \mathcal{T}(\cdot|x,u)$ weakly for every $x \in \mathbb{X}$ and $u \in \mathbb{U}$ and $(x_n,u_n) \to (x,u)$, then it is not true in general that $J_\beta(\mathcal{T},\gamma_n^*) \to J_\beta(\mathcal{T},\gamma^*)$, even when $\mathbb{X}$ and $\mathbb{U}$ are compact and $c(x,u)$ is continuous and bounded in $\mathbb{X}\times\mathbb{U}$.*

Proof. We prove the result with a counterexample. Take $\mathbb{X} = [0,2]$ and $\mathbb{U} = \{0,1,2\}$.

Suppose the kernels are given in the following form for $n \ge 1$:

$$\begin{aligned}\mathcal{T}_n(\cdot|x,u) =& \delta_{1+1/n}(\cdot)\mathbb{1}_{\{x\ge 1+1/n, u=1\}} + \delta_{1-1/n}(\cdot)\mathbb{1}_{\{x\ge 1+1/n, u=0\}} + \delta_1(\cdot)\mathbb{1}_{\{x\ge 1+1/n, u=2\}} \\ &+ \delta_{1-1/n}\mathbb{1}_{\{x\le 1-1/n, u=1\}} + \delta_{1+1/n}\mathbb{1}_{\{x\le 1-1/n, u=0\}} + \delta_1\mathbb{1}_{\{x\le 1-1/n, u=2\}} \\ &+ \delta_1\mathbb{1}_{\{1-1/n<x<1+1/n\}} \\ \mathcal{T}(\cdot|x,u) =& \delta_1(\cdot).\end{aligned}$$

The cost function is given by

$$c(x,u) = \begin{cases}(x-1)\mathbb{1}_{x\ge 1} + 0\mathbb{1}_{x<1} & \text{if } u = 0,1, \\ 3 & \text{if } u = 2.\end{cases}$$

With this setup, an optimal policy for $\mathcal{T}_n$ when the initial state is $x = 0$ is given by

$$\gamma_n^*(x) = \begin{cases}1 & \text{if } x \le 1 - 1/n, \\ 0 & \text{if } x \ge 1 + 1/n, \\ 2 & \text{otherwise.}\end{cases}$$

When the initial state is 0, the cost under this policy is $J_\beta(\mathcal{T}_n,\gamma_n^*) = 0$; therefore the policy γ_n^* is indeed optimal for $\mathcal{T}_n$. An optimal policy for $\mathcal{T}$ is given by $\gamma(x) = 1, \forall x$.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 05/17/20 to 130.15.241.167. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Thus, the discounted cost values can be calculated as

$$J_\beta(\mathcal{T},\gamma_n^*) = \sum_{t=0}^{\infty} \beta^t E[c(X_t,\gamma_n^*(X_t))] = \sum_{t=0}^{\infty} \beta^t c(1,\gamma_n^*(1)) = \sum_{t=0}^{\infty} \beta^t 3 = \frac{3}{1-\beta},$$
$$J_\beta(\mathcal{T},\gamma^*) = 0. \qquad \square$$

4.2. A sufficient condition for robustness under weak convergence. We now present a result that establishes robustness if the optimal policies for every initial point are identical. That is, for every n, γ_n^* is optimal for every $x_0 \in \mathbb{X}$ (under the model $\mathcal{T}_n$). Notice that in the counterexample used for Theorem 4.3, γ_n^* is not optimal if the initial point is between $1-1/n$ and $1+1/n$. A sufficient condition for this property is that γ_n^* solves the discounted cost optimality equation (DCOE) for every initial point.

A policy $\gamma^* \in \Gamma$ solves the discounted cost optimality equation and is optimal if it satisfies

$$J_\beta^*(\mathcal{T},x) = c(x,\gamma^*(x)) + \beta \int J_\beta^*(\mathcal{T},x_1)\mathcal{T}(dx_1|x,\gamma^*(x)).$$

Thus, a policy is optimal for every initial point if it satisfies the DCOE for all initial points $x \in \mathbb{X}$.

THEOREM 4.4. *Under Assumption* 1.1, $J_\beta(\mathcal{T},\gamma_n^*) \to J_\beta(\mathcal{T},\gamma^*)$ *for any initial point* x_0 *if* γ_n^* *is optimal for any initial point for the kernel* $\mathcal{T}_n$.

Proof. Define the following operator for γ_n^*, an optimal policy for $\mathcal{T}_n$,

$$(T_n(v))(x_0) = c(x_0,\gamma_n^*(x_0)) + \beta \int v(x_1)\mathcal{T}(dx_1|x_0,\gamma_n^*(x_0)), \tag{4.4}$$

which is a contraction from $C_b(\mathbb{X})$ to itself under the supremum norm with modulus β and has a unique fixed point. One can show that the fixed point is $J_\beta(\mathcal{T},\gamma_n^*)$.

In Appendix A.5 we show that

$$T_n^k(J_\beta^*(\mathcal{T}))(x_n) \to J_\beta^*(\mathcal{T},x)) \tag{4.5}$$

for any fixed $k < \infty$ as $n \to \infty$ for some $x_n \to x$, where T_n^k denotes the operator T_n applied k consecutive times.

Our next claim is that $T_n^k(J_\beta^*(\mathcal{T}))(x) \to J_\beta(\mathcal{T},\gamma_n^*,x)$ as $k \to \infty$. This is true as T_n is a contraction with modulus β and $J_\beta(\mathcal{T},\gamma_n^*)$ is its unique fixed point. Thus $T_n^k(J_\beta^*(\mathcal{T}))(x) \to J_\beta(\mathcal{T},\gamma_n^*,x)$ as $k \to \infty$ and this convergence is uniform over n as the contraction rate β does not depend on n.

Now, we write

$$J_\beta(\mathcal{T},\gamma_n^*) - J_\beta^*(\mathcal{T}) \le |J_\beta(\mathcal{T},\gamma_n^*) - T_n^k(J_\beta^*(\mathcal{T}))| + |T_n^k(J_\beta^*(\mathcal{T})) - J_\beta^*(\mathcal{T})|.$$

We can make the first term arbitrarily small by choosing k large enough uniformly over n. For the chosen k, the second term goes to 0 as $n \to \infty$ by (4.5). $\square$

Some remarks are in order.

Remark 4.1. For the partially observed case, the proof approach we use makes use of policy exchange (e.g., (2.3)) and for this approach the total variation continuity of channel $Q(\cdot|x)$ is a key step to deal with the uniform convergence over policies. As

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 05/17/20 to 130.15.241.167. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

we stated before, the channel for fully observed models can be considered in the form of (4.1), which is only weakly continuous and not continuous in total variation. Thus, it may lead to negative results as in Theorem 4.3. However, for the fully observed models we can reach continuity and robustness (Theorems 4.2, 4.4) using the value iteration approach. With this approach, instead of exchanging policies and analyzing uniform convergence over all policies, we can exchange control actions (e.g., (4.2)) and analyze uniform convergence over the action space $\mathbb{U}$ by using the discounted optimality operator (2.1). Hence, we are only able to show convergence over optimal policies for the fully observed case, i.e., $J_\beta(\mathcal{T}_n, \gamma_n^*) \to J_\beta(\mathcal{T}, \gamma^*)$ or $J_\beta(\mathcal{T}, \gamma_n^*) \to J_\beta(\mathcal{T}, \gamma^*)$, where γ_n^* and γ^* are optimal policies, whereas, for partially observed models, the regularity of the channel allows us to show convergence over any sequence of policies, i.e., $\sup_{\gamma\in\Gamma} |J_\beta(\mathcal{T}_n, Q, \gamma) - J_\beta(\mathcal{T}, Q, \gamma)| \to 0$.

Remark 4.2. As we have discussed in subsection 2.2, a partially observed model can be reduced to a fully observed process where the state process (beliefs) becomes probability measure valued. Consider the partially observed models with transition kernels $\mathcal{T}_n$ and $\mathcal{T}$ (with a channel Q) and their corresponding fully observed transition kernels η_n and η: following the discussions and techniques in [16] and [31], one can show that η_n and η satisfy the conditions of Theorems 4.4 and 4.2, that is, $\eta_n(\cdot|z_n, u_n) \to \eta(\cdot|z, u)$ for any $(z_n, u_n) \to (z, u)$ under the following set of assumptions:

- $\mathcal{T}_n(\cdot|x_n, u_n) \to \mathcal{T}(\cdot|x, u)$ for any $(x_n, u_n) \to (x, u)$,
- $Q(\cdot|x)$ is continuous on total variation in x.

We remark that these conditions also agree with the conditions presented for the continuity and robustness of the partially observed models (Theorems 2.3 and 3.3).

Remark 4.3. It can be shown that if we restrict the set of policies to an equicontinuous family of functions, robustness can also be achieved: Under the conditions of Theorem 4.2, in this case, $|J_\beta(\mathcal{T}, \gamma_n^*) - J_\beta^*(\mathcal{T})| \to 0$. A short proof for this result can be found in Appendix A.4.

4.3. Setwise convergence.

4.3.1. Absence of continuity under setwise convergence. We give a negative result similar to Theorem 2.4.

THEOREM 4.5. *Letting $\mathcal{T}_n \to \mathcal{T}$ setwise, then it is not necessarily true that $J_\beta^*(\mathcal{T}_n) \to J_\beta^*(\mathcal{T})$ even when $c(x, u)$ is continuous and bounded in $\mathbb{X} \times \mathbb{U}$.*

Proof. See Example 4.1. ∎

4.3.2. A sufficient condition for continuity under setwise convergence.

THEOREM 4.6. *Under Assumption* 1.3 *$J_\beta(\mathcal{T}_n, \gamma_n^*) \to J_\beta(\mathcal{T}, \gamma^*)$ for any initial state x_0, as $n \to \infty$.*

Proof. We use the same value iteration technique that we used to prove Theorem 4.2.

We wish to show that the approximation functions for $\mathcal{T}_n$ converge pointwise to the ones for $\mathcal{T}$. Then, for the first step of the induction we have

$$v^1(x) = \inf_u c(x, u), \quad v_n^1(x) = \inf_u c(x, u),$$

and thus we can write

$$|v^1(x) - v_n^1(x)| = 0.$$

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 05/17/20 to 130.15.241.167. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

For step k we have

$$v^k(x) = \inf_u \left[c(x,u) + \int_{\mathbb{X}} v^{k-1}(x^1)\mathcal{T}(dx^1|x,u) \right],$$
$$v_n^k(x) = \inf_u \left[c(x,u) + \int_{\mathbb{X}} v_n^{k-1}(x^1)\mathcal{T}_n(dx^1|x,u) \right].$$

Note that the assumptions of the theorem satisfy the measurable selection criteria and hence we can choose minimizing selectors [27, section 3.3]. If we denote the selectors by u^* and u_n^*, we can write

$$\begin{aligned}
&|v^k(x) - v_n^k(x)| \\
&\leq \max\Bigg(\left[|c(x,u^*) - c(x,u^*)| + |\int_{\mathbb{X}} v^{k-1}(x^1)\mathcal{T}(dx^1|x,u^*) - \int_{\mathbb{X}} v_n^{k-1}(x^1)\mathcal{T}_n(dx^1|x,u^*)| \right], \\
&\quad \left[|c(x,u_n^*) - c(x,u_n^*)| + |\int_{\mathbb{X}} v^{k-1}(x^1)\mathcal{T}(dx^1|x,u_n^*) - \int_{\mathbb{X}} v_n^{k-1}(x^1)\mathcal{T}_n(dx^1|x,u_n^*)| \right] \Bigg).
\end{aligned}$$

For the first term we use [42, Theorem 20], since $\mathcal{T}_n(\cdot|x,u^*) \to \mathcal{T}(\cdot|x,u^*)$ setwise and $v_n^{k-1} \to v^{k-1}$ pointwise.

For the second term, we use a contradiction argument. Assume that there exists an $\epsilon > 0$ and some subsequence (identified with n_k) such that

$$|\int_{\mathbb{X}} v^{k-1}(x^1)\mathcal{T}(dx^1|x,u_{n_k}^*) - \int_{\mathbb{X}} v_n^{k-1}(x^1)\mathcal{T}_{n_k}(dx^1|x,u_{n_k}^*)| > \epsilon. \tag{4.6}$$

Now, take a further subsequence $u_{n_k'}^*$ of this sequence, which converges to some u whose existence follows from the compactness of $\mathbb{U}$. Notice that along this subsequence $\mathcal{T}(\cdot|x,u_{n_k'}^*) \to \mathcal{T}(\cdot|x,u)$ and $\mathcal{T}_{n_k'}(\cdot|x,u_{n_k'}^*) \to \mathcal{T}(\cdot|x,u)$. Thus, using the induction step and [42, Theorem 20] the above term converges to 0 along the subsequence indexed by n_k', which contradicts (4.6). The rest of the proof follows from the arguments in Theorem 5.2. ∎

4.3.3. Absence of robustness under setwise convergence. Now, we give a result showing that even if the continuity holds under the setwise convergence of the kernels, the robustness may not be satisfied.

THEOREM 4.7. *Supposing $\mathcal{T}_n(\cdot|x_n,u_n) \to \mathcal{T}(\cdot|x,u)$ setwise for every $x \in \mathbb{X}$ and $u \in \mathbb{U}$ and $(x_n,u_n) \to (x,u)$, then it is not true in general that $J_\beta(\mathcal{T},\gamma_n^*) \to J_\beta(\mathcal{T},\gamma^*)$, even when $\mathbb{X}$ and $\mathbb{U}$ are compact and $c(x,u)$ is continuous and bounded in $\mathbb{X} \times \mathbb{U}$.*

Proof. We prove the result with a counterexample. Define

$$L_{n,k} = \left[\frac{2k-2}{2n}, \frac{2k-1}{2n} \right), \; R_{n,k} = \left[\frac{2k-1}{2n}, \frac{k}{n} \right).$$

Let $L = \{y \in \cup_{k=1}^n L_{n,k}\}$ and $R = \{y \in \cup_{k=1}^n R_{n,k}\}$. Next, define the square-wave function by

$$h_n(t) = 1_{\{t \in L\}} - 1_{\{t \in R\}}.$$

Define two sequences of probability density functions as

$$f_n(t) = (1 + h_n(t))1_{\{t \in [0,1]\}} \quad , \quad g_n(t) = (1 - h_n(t))1_{\{t \in [0,1]\}}.$$

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 05/17/20 to 130.15.241.167. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Consider the kernels given in the following form for $n \geq 1$:

$$\begin{aligned}\mathcal{T}_n(\cdot|x,u) \sim & \mathbb{1}_{\{x\in L,u=1\}} f_n(\cdot) + \mathbb{1}_{\{x\in L,u=0\}} g_n(\cdot) \\ & + \mathbb{1}_{\{x\in R,u=0\}} f_n(\cdot) + \mathbb{1}_{\{x\in R,u=1\}} g_n(\cdot) \\ \mathcal{T}(\cdot|x,u) \sim & U([0,1]).\end{aligned}$$

By the proof of the Riemann–Lebesgue lemma [54, Theorem 12.21]

$$\lim_{n\to\infty} \int_0^1 h_n(t) g(t) \mathrm{d}t = 0 \ \forall g \in L_1([0,1],\mathbb{R}),$$

and therefore

$$\lim_{n\to\infty} \int_0^1 f_n(t) g(t) \mathrm{d}t = \int_0^1 g(t) \ \forall g \in L_1([0,1],\mathbb{R}).$$

As a result, $\mathcal{T}_n(\cdot|x,u) \to \mathcal{T}(\cdot|x,u)$ setwise for every $x \in \mathbb{X}$ and $u \in \mathbb{U}$.

The cost function is given by

$$c(x,u) = \begin{cases} 2 & \text{if } u = 0, \\ x & \text{if } u = 1. \end{cases}$$

Notice that if the system starts anywhere on L, it does not matter how we define γ_n^* for $x \in R$ as the state always stays at L. Thus, with this setup, it can be seen that an optimal policy for $\mathcal{T}_n$ is given by

$$\gamma_n^*(x) = \begin{cases} 1 & \text{if } x \in L, \\ 0 & \text{if } x \in R \end{cases}$$

when the initial state is $x = 0$, which belongs to L for any $n \geq 1$. The optimal policy for $\mathcal{T}$ is given by $\gamma^*(x) = 1$. The discounted cost values can be calculated as follows:

$$J_\beta(\mathcal{T},\gamma^*) = \sum_{t=0}^{\infty} \beta^t E_{\mathcal{T}}[c(X,1)] = \sum_{t=0}^{\infty} \beta^t \int_0^1 x dx = \frac{1}{2(1-\beta)}.$$

Building on the calculations in [32], the cost under the policy γ_n^* is calculated as

$$\begin{aligned}J_\beta(\mathcal{T},\gamma_n^*) &= \sum_{t=0}^{\infty} \beta^t E_{\mathcal{T}}[c(X,\gamma_n^*(X))] = \frac{1}{1-\beta}\left(\int_L c(x,1)dx + \int_R c(x,0)dx\right) \\ &= \frac{1}{1-\beta}\left(\int_L x dx + \int_R 2dx\right) = \frac{1}{1-\beta}\left(\frac{1}{4} - \frac{1}{8n} + 1\right) \to \frac{5}{4(1-\beta)},\end{aligned}$$

which completes the proof. ∎

4.4. A sufficient condition for robustness under setwise convergence. We now present a similar result to Theorem 4.4, that is, we show that under the conditions of Theorem 4.6, if further for every n, γ_n^* is optimal for every $x_0 \in \mathbb{X}$ (under the model $\mathcal{T}_n$), then robustness holds under setwise convergence.

THEOREM 4.8. *Supposing Assumption* 1.3 *holds, if further we have that for every* n, γ_n^* *is optimal for every* $x_0 \in \mathbb{X}$ *(under the model* $\mathcal{T}_n$*), then* $J_\beta(\mathcal{T},\gamma_n^*) \to J_\beta(\mathcal{T},\gamma^*)$.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 05/17/20 to 130.15.241.167. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Proof. We use the same proof technique as we used for Theorem 4.4. Define the following operator for γ_n^*, an optimal policy for $\mathcal{T}_n$ as in the proof of Theorem 4.4,

$$(T_n(v))(x_0) = c(x_0, \gamma_n^*(x_0)) + \beta \int v(x_1)\mathcal{T}(dx_1|x_0, \gamma_n^*(x_0)),$$

which is a contraction from $C_b(\mathbb{X})$ to itself under the supremum norm with modulus β and has a unique fixed point which is $J_\beta(\mathcal{T}, \gamma_n^*)$. Hence, $T_n^k(J_\beta^*(\mathcal{T}))(x) \to J_\beta(\mathcal{T}, \gamma_n^*, x)$ as $k \to \infty$ uniformly over n, where T_n^k is the operator T_n applied k consecutive times.

Using the properties of setwise convergence we show in section A.6 that

$$\lim_{n\to\infty} T_n^k(J_\beta^*(\mathcal{T}))(x) = J_\beta^*(\mathcal{T}, x) \quad \forall k < \infty. \tag{4.7}$$

Then, we write

$$J_\beta(\mathcal{T}, \gamma_n^*) - J_\beta^*(\mathcal{T}) \leq |J_\beta(\mathcal{T}, \gamma_n^*) - T_n^k(J_\beta^*(\mathcal{T}))| + |T_n^k(J_\beta^*(\mathcal{T})) - J_\beta^*(\mathcal{T})| \to 0.$$

We can make the second term arbitrarily small by choosing k large enough uniformly over all n since $T_n^k(J_\beta^*(\mathcal{T}))(x) \to J_\beta(\mathcal{T}, \gamma_n^*, x)$ as $k \to \infty$ uniformly over n. For the fixed k, the first term can be made arbitrarily small by choosing n large enough using (4.7). ∎

4.5. Total variation. The continuity result in Theorem 2.5 and the robustness result in Theorem 3.1 apply to this case since the fully observed model may be viewed as a partially observed model with the measurement channel Q given in (4.1).

Remark 4.4. We note that if the action and state spaces are finite, then total variation convergence and weak convergence coincide and thus Theorems 2.5, 2.6, and 3.1 from the partially observed case directly apply to this case considering the channel as a perfect channel. Thus, the only assumptions needed to establish continuity and robustness are

- $\mathcal{T}_n(\cdot|x, u_n) \to \mathcal{T}(\cdot|x, u)$ in total variation for all $x \in \mathbb{X}$,
- $\mathcal{T}(\cdot|x, u)$ is continuous in total variation in u for every given $x \in \mathbb{X}$,
- $\mathbb{U}$ is compact.

Since the spaces are finite, this set of assumptions reduces to

- $\mathcal{T}_n(\cdot|x, u) \to \mathcal{T}(\cdot|x, u)$ in total variation for all $x \in \mathbb{X}$ and $u \in \mathbb{U}$.

Remark 4.5. We note that all of the results we present in this paper also apply to finite horizon problems. If we define a finite horizon problem by

$$J(P, \mathcal{T}, \gamma) = \sum_{t=0}^{T} E_{P,Q}^{\mathcal{T}}[c(X_t, U_t)],$$

the continuity and robustness properties hold under the same conditions we have presented for the infinite horizon discounted problem.

5. Implications for data-driven learning methods in stochastic control. In practice, one might try to learn the kernel of a controlled Markov chain from empirical data; see, e.g., [6, 55, 24] for some related literature in the control-free and controlled contexts.

Let us briefly discuss the case where a random variable is repeatedly observed, but its probability measure is not known a priori. Let $\{(X_i), i \in \mathbb{N}\}$ be an $\mathbb{X}$-valued

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 05/17/20 to 130.15.241.167. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

i.i.d. random variable sequence generated according to some distribution μ. Defining for every (fixed) Borel $B \subset \mathbb{X}$, and $n \in \mathbb{N}$, the empirical occupation measures

$$\mu_n(B) = \frac{1}{n}\sum_{i=1}^{n} 1_{\{X_i \in B\}},$$

and one has $\mu_n(B) \to \mu(B)$ almost surely by the strong law of large numbers. Also, $\mu_n \to \mu$ weakly with probability one ([14, Theorem 11.4.1]). However, μ_n cannot converge to μ in total variation or setwise, in general. On the other hand, if we know that μ admits a density, we can find estimators to estimate μ under total variation [13, Chapter 3]. For a more detailed discussion on convergence of empirical occupation measures see [32, pp. 1950–1951]. In the previous sections, we established robustness results under the convergence of transition kernels in the topology of weak convergence and total variation. We build on these observations next.

5.1. Application of robustness results to data-driven learning.

COROLLARY 5.1 (to Theorems 2.6 and 3.1). *Suppose we are given the following dynamics for finite state space, $\mathbb{X}$, and finite action space, $\mathbb{U}$,*

$$x_{t+1} = f(x_t, u_t, w_t), \qquad y_t = g(x_t, v_t),$$

where $\{w_t\}$ and $\{v_t\}$ are i.i.d. noise processes and the noise models are unknown. Suppose that there is an initial training period so that under some policy, every x, u pair is visited infinitely often if training were to continue indefinitely, but that the training ends at some finite time. Let us assume that, through this training, we empirically learn the transition dynamics such that for every (fixed) Borel $B \subset \mathbb{X}$, for every $x \in \mathbb{X}$, $u \in \mathbb{U}$, and $n \in \mathbb{N}$, the empirical occupation measures are

$$\mathcal{T}_n(B|x_0 = x, u_0 = u) = \frac{\sum_{i=1}^{n} 1_{\{X_i \in B, X_{i-1}=x, U_{i-1}=u\}}}{\sum_{i=1}^{n} 1_{\{X_{i-1}=x, U_{i-1}=u\}}}.$$

Then we have that $J^*_\beta(\mathcal{T}_n) \to J^*_\beta(\mathcal{T})$ and $J_\beta(\mathcal{T}, \gamma^*_n) \to J^*_\beta(\mathcal{T})$, where γ^*_n is the optimal policy designed for $\mathcal{T}_n$. Since the channel model g has no restrictions, this result also applies to the fully observed model setup by taking $g(x_t, v_t) = x_t$.

Proof. We have that $\mathcal{T}_n(\cdot|x, u) \to \mathcal{T}(\cdot|x, u)$ weakly for every $x \in \mathbb{X}$, $u \in \mathbb{U}$ almost surely by the law of large numbers. Since the spaces are finite, we also have $\mathcal{T}_n(\cdot|x, u) \to \mathcal{T}(\cdot|x, u)$ under total variation. By Theorems 2.6 and 3.1, the results follow. ∎

The following holds for more general spaces.

COROLLARY 5.2 (to Theorems 3.3, 2.3, 4.2, and 4.4). *Suppose we are given the following dynamics with state space $\mathbb{X}$ and action space $\mathbb{U}$,*

$$x_{t+1} = f(x_t, u_t, w_t), \qquad y_t = g(x_t, v_t),$$

where $\{w_t\}$ and $\{v_t\}$ are i.i.d. noise processes and the noise models are unknown. Suppose that $f(x, u, \cdot) : \mathbb{W} \to \mathbb{X}$ is invertible for all fixed (x, u) and $f(x, u, w)$ is continuous and bounded on $\mathbb{X} \times \mathbb{U} \times \mathbb{W}$. We construct the empirical measures for the noise process w_t such that for every (fixed) Borel $B \subset \mathbb{W}$, and for every $n \in \mathbb{N}$, the empirical occupation measures are

$$\mu_n(B) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{f^{-1}_{x_{i-1}, u_{i-1}}(x_i) \in B\}}, \tag{5.1}$$

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 05/17/20 to 130.15.241.167. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

where $f^{-1}_{x_{i-1},u_{i-1}}(x_i)$ *denotes the inverse of* $f(x_{i-1}, u_{i-1}, w) : \mathbb{W} \to \mathbb{X}$ *for given* (x_{i-1}, u_{i-1}). *Using the noise measurements, we construct the empirical transition kernel estimates for any* (x_0, u_0) *and Borel* B *as*

$$\mathcal{T}_n(B|x_0, u_0) = \mu_n(f^{-1}_{x_0,u_0}(B)).$$

(i) *If the measurement channel (represented by the function* g*) is continuous in total variation, then* $J^*_\beta(\mathcal{T}_n) \to J^*_\beta(\mathcal{T})$ *and* $J_\beta(\mathcal{T}, \gamma^*_n) \to J^*_\beta(\mathcal{T})$, *where* γ^*_n *is the optimal policy designed for* $\mathcal{T}_n$ *for all initial points.*

(ii) *If the measurement channel is in the form* $g(x_t, v_t) = x_t$ *(i.e., fully observed), then* $J^*_\beta(\mathcal{T}_n) \to J^*_\beta(\mathcal{T})$, *and if further for every* n, γ^*_n *is optimal for every* $x_0 \in \mathbb{X}$ *(under the model* $\mathcal{T}_n$*), then* $J_\beta(\mathcal{T}, \gamma^*_n) \to J^*_\beta(\mathcal{T})$.

Proof. We have $\mu_n \to \mu$ weakly with probability one where μ is the model. We claim that the transition kernels are such that $\mathcal{T}_n(\cdot|x_n, u_n) \to \mathcal{T}(\cdot|x, u)$ weakly for any $(x_n, u_n) \to (x, u)$. To see that observe the following for $h \in C_b(\mathbb{X})$:

$$\begin{aligned}&\int h(x_1)\mathcal{T}_n(dx_1|x_n, u_n) - \int h(x_1)\mathcal{T}(dx_1|x, u)\\ &= \int h(f(x_n, u_n, w))\mu_n(dw) - \int h(f(x, u, w))\mu(dw) \to 0,\end{aligned}$$

where μ_n is the empirical measure for w_t and μ is the true measure again. For the last step, we used that $\mu_n \to \mu$ weakly and $h(f(x_n, u_n, w))$ continuously converge to $h(f(x, u, w))$, i.e., $h(f(x_n, u_n, w_n)) \to h(f(x, u, w)$ for some $w_n \to w$ since f and g are continuous functions. Similarly, it can be also shown that $\mathcal{T}_n(\cdot|x, u)$ and $\mathcal{T}(\cdot|x, u)$ are weakly continuous on (x, u). Thus, for the case where the channel is continuous in total variation by Theorems 3.3 and 2.3 if $c(x, u)$ is bounded and $\mathbb{U}$ is compact, the result follows.

For the fully observed case, $J^*_\beta(\mathcal{T}_n) \to J^*_\beta(\mathcal{T})$ by Theorem 4.2 and $J_\beta(\mathcal{T}, \gamma^*_n) \to J^*_\beta(\mathcal{T})$ by Theorem 4.4. ∎

Remark 5.1. We note here that the moment estimation method can also lead to consistency. Suppose that the distribution of W is determined by its moments, such that estimate models W_n have moments of all orders and $\lim_n = E[W^r_n] = E[W^r]$ for all $r \in \mathbb{Z}_+$. Then, we have that [8, Theorem 30.2] $W_n \to W$ weakly and thus $\mathcal{T}_n(\cdot|x_n, u_n) \to \mathcal{T}(\cdot|x, u)$ weakly for any $(x_n, u_n) \to (x, u)$ under the assumptions of above corollary. Hence, we reach continuity and robustness using the same arguments as in the previous result (Corollary 5.2).

Now, we give a similar result with the assumption that the noise process of the dynamics admits a continuous probability density function.

COROLLARY 5.3 (to Theorems 2.6 and 3.1). *Suppose we are given the following dynamics for real vector state space* $\mathbb{X}$ *and action space* $\mathbb{U}$*:*

$$x_{t+1} = f(x_t, u_t, w_t), \qquad y_t = g(x_t, v_t),$$

where $\{w_t\}$ *and* $\{v_t\}$ *are i.i.d. noise processes and the noise models are unknown, but it is known that the noise* w_t *admits a continuous probability density function. Suppose that* $f(x, u, \cdot) : \mathbb{W} \to \mathbb{X}$ *is invertible for all* (x, u). *We collect i.i.d. samples of* $\{w_t\}$ *as in* (5.1) *and use them to construct an estimator,* $\tilde{\mu}_n$, *as described in* [13], *which consistently estimates* μ *in total variation. Using these empirical estimates, we construct the empirical transition kernel estimates for any* (x_0, u_0) *and Borel* B *as*

$$\mathcal{T}_n(B|x_0, u_0) = \tilde{\mu}_n(f^{-1}_{x_0,u_0}(B)).$$

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 05/17/20 to 130.15.241.167. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Then independent of the channel, $J^*_\beta(\mathcal{T}_n) \to J^*_\beta(\mathcal{T})$ and $J_\beta(\mathcal{T}, \gamma^*_n) \to J^*_\beta(\mathcal{T})$, where γ^*_n is the optimal policy designed for $\mathcal{T}_n$. Since the channel model g has no restrictions, this result also applies to the fully observed model setup by taking $g(x_t, v_t) = x_t$.

Proof. By [13] we can estimate μ in total variation so that almost surely $\lim_{n\to\infty} \|\tilde{\mu}_n - \mu\|_{TV} = 0$. We claim that the convergence of $\tilde{\mu}_n$ to μ under the total variation metric implies the convergence of $\mathcal{T}_n$ to $\mathcal{T}$ in total variation uniformly over all $x \in \mathbb{X}$ and $u \in \mathbb{U}$, i.e., $\lim_{n\to\infty} \sup_{x,u} \|\mathcal{T}_n(\cdot|x,u) - \mathcal{T}(\cdot|x,u)\|_{TV} = 0$. Observe the following:

$$\begin{aligned}
&\sup_{x,u} \|\mathcal{T}_n(\cdot|x,u) - \mathcal{T}(\cdot|x,u)\|_{TV} \\
&= \sup_{x,u} \sup_{||h||_\infty \le 1} \Big| \int h(x_1)\mathcal{T}_n(dx_1|x,u) - \int h(x_1)\mathcal{T}(dx_1|x,u) \Big| \\
&= \sup_{x,u} \sup_{||h||_\infty \le 1} \Big| \int h(f(x,u,w))\tilde{\mu}_n(dw) - \int h(f(x,u,w))\mu(dw) \Big| \le \|\tilde{\mu}_n - \mu\|_{TV} \to 0.
\end{aligned}$$

Thus, by Theorems 2.6 and 3.1, the result follows. ∎

The following example presents some system and channel models which satisfy the requirements of the above corollaries.

EXAMPLE 5.1. *Let $\mathbb{X}, \mathbb{Y}, \mathbb{U}$ be real vector spaces with*

$$x_{t+1} = f(x_t, u_t) + w_t, \qquad y_t = h(x_t, v_t)$$

for unknown i.i.d. noise processes $\{w_t\}$ and $\{v_t\}$.

(i) *Suppose the channel is in the following form: $y_t = h(x_t, v_t) = x_t + v_t$, where v_t admits a density (e.g., Gaussian density). It can be shown by an application of Scheffé's theorem that the channels in this form are continuous in total variation. If further $f(x_t, u_t)$ is continuous and bounded, then the requirements of Corollary 5.2 hold for partially observed models.*

(ii) *If the channel is in the form $x_t = h(x_t, v_t)$, then the system is fully observed. If further $f(x_t, u_t)$ is continuous and bounded, then the requirements of Corollary 5.2 hold for fully observed models.*

(iii) *Supposing the function $f(x_t, u_t)$ is known, if the noise process w_t admits a continuous density, then one can estimate the noise model in total variation in a consistent way (see [13]). Hence, the conditions of Corollary 5.3 hold independent of the channel model.* ∎

6. Conclusion. We studied regularity properties of optimal stochastic control on the space of transition kernels and applications to robustness of optimal control policies designed for an incorrect model applied to an actual system.

Appendix A. Technical results.

A.1. Proof of (2.2) in Theorem 2.2. Before the proof we give a key lemma. The lemma we present generalizes the following result from [35, Theorem 3.5] and [49, Theorem 3.5].

LEMMA A.1. *Suppose $\{\mu_n\}_n \subset \mathcal{P}(\mathbb{X})$, where $\mathbb{X}$ is metric space, converges weakly to some $\mu \in \mathcal{P}(\mathbb{X})$. For a bounded real valued sequence of functions $\{f_n\}_n$ such that $\|f_n\|_\infty < C$ for all $n > 0$ with $C < \infty$, if $\lim_{n\to\infty} f_n(x_n) = f(x)$ for all $x_n \to x$, i.e., f_n continuously converges to f, then $\lim_{n\to\infty} \int_{\mathbb{X}} f_n(x)\mu_n(dx) = \int_{\mathbb{X}} f(x)\mu(dx)$.*

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 05/17/20 to 130.15.241.167. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

LEMMA A.2. *Suppose we have a uniformly bounded family of functions* $\{f_n^\gamma : \mathbb{X} \to \mathbb{R}, \gamma \in \Gamma, n > 0\}$ *such that* $\|f_n^\gamma\|_\infty < C$ *for all* $\gamma \in \Gamma$ *and for all* $n > 0$ *for some* $C < \infty$.

Further suppose we have another uniformly bounded family of functions $\{f^\gamma : \mathbb{X} \to \mathbb{R}, \gamma \in \Gamma\}$ *such that* $\|f^\gamma\|_\infty < C$ *for all* $\gamma \in \Gamma$ *for some* $C < \infty$ *under the following assumptions:*

(i) *For any* $x_n \to x$

$$\sup_{\gamma\in\Gamma} \left| f_n^\gamma(x_n) - f^\gamma(x) \right| \to 0, \tag{A.1}$$

$$\sup_{\gamma\in\Gamma} \left| f^\gamma(x_n) - f^\gamma(x) \right| \to 0. \tag{A.2}$$

(ii) $\mathcal{T}_n(\cdot|x_n, u_n)$ *converges weakly to* $\mathcal{T}(\cdot|x, u)$ *for any* $(x_n, u_n) \to (x, u)$.

(iii) $\mathcal{T}(\cdot|x, u)$ *is weakly continuous in* (x, u).

(iv) $\mathbb{U}$ *is compact.*

Then for fixed observation realizations, $y_{[0,t]} := \{y_0, \dots, y_t\}$ *and for some* $x_t^n \to x_t$

$$\sup_{\gamma\in\Gamma} \left| \int \mathcal{T}_n(dx_{t+1}|x_t^n, \gamma(y_{[0,t]})) f_n^\gamma(x_{t+1}) - \int \mathcal{T}(dx_{t+1}|x_t, \gamma(y_{[0,t]})) f^\gamma(x_{t+1}) \right| \to 0. \tag{A.3}$$

Proof. Using (A.2), we see that $\{f^\gamma\}$ is a equicontinuous family of functions. Thus, by the Arzela–Ascoli theorem, for any given compact set $K \subset \mathbb{X}$ and $\epsilon > 0$ there is a finite set of continuous functions $\mathbb{F} := \{f^1, \dots, f^N\}$ so that for any γ there is $f^i \in \mathbb{F}$ with

$$\sup_{x\in K} |f^\gamma(x) - f^i(x)| \le \epsilon.$$

Now, we claim that for the same $\epsilon > 0$, the same $f^i \in \mathbb{F}$, and the chosen compact set K, we can make $\sup_{x\in K} |f_n^\gamma(x) - f^i(x)| \le 3\epsilon/2$ for large enough n and for any $\gamma \in \Gamma$. To see this, observe the following:

$$\sup_{x\in K} |f_n^\gamma(x) - f^i(x)| \le \sup_{x\in K} |f_n^\gamma(x) - f^\gamma(x)| + \sup_{x\in K} |f^\gamma(x) - f^i(x)|;$$

the second term is less than ϵ with the argument we made in the first paragraph of the proof. The first term can also be made arbitrarily small since $f_n^\gamma \to f^\gamma$ uniformly on compact sets by (A.1). Now we wish to show that we can find a compact subset of $\mathbb{X}$ such that all probability measures (kernels) in the term (A.3) put their measure mainly on this compact set. Consider the set of measures $S := \cup_{\gamma\in\Gamma} S_\gamma$, where

$$S_\gamma = \{\mathcal{T}_n(\cdot|x_t^n, \gamma(y_{[0,t]})) : \mathcal{T}_n(\cdot|x_t^n, \gamma(y_{[0,t]})) \to \mathcal{T}(\cdot|x_t, \gamma(y_{[0,k]}))\}.$$

Here, notice that the set S depends on the sequence $\{x_t^n\}$ and the observation realizations $y_{[0,t]}$. To cover all the kernels in (A.3) we take $\{x_t^n\}$ and $y_{[0,t]}$ as they are given in the statement of the lemma.

For a sequence from the set S, say, $\mathcal{T}_{n_m}(\cdot|x_t^{n_m}, \gamma_m(y_{[0,t]}))$, since $\mathbb{U}$ is a compact set and the observations are fixed, there exists a subsequence in which $\gamma_{m_r}(y_{[0,t]}) \to u^*$ for some $u^* \in \mathbb{U}$. Now we focus on this subsequence $\mathcal{T}_{n_{m_r}}(\cdot|x_t^{n_{m_r}}, \gamma_{m_r}(y_{[0,t]}))$. By the assumption (ii) in the lemma statement $\{\mathcal{T}_n\}_n$ also satisfies the following: for any $(x_n, u_n) \to (x, u)$, $\mathcal{T}_n(\cdot|x_n, u_n) \to \mathcal{T}(\cdot|x, u)$. Thus,

$$\mathcal{T}_{n_{m_r}}(\cdot|x_t^{n_{m_r}}, \gamma_{m_r}(y_{[0,t]})) \to \mathcal{T}(\cdot|x_t, u^*).$$

Hence any sequence in the set S has a convergent subsequence, and thus S is a precompact family. Therefore, it is a tight family of functions by the Prokhorov

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 05/17/20 to 130.15.241.167. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

theorem [7, Theorem 5.2] (see also [14, Theorem 11.5.3]). Hence, for any $\epsilon > 0$, there exists a compact set K_ϵ such that for all n and uniformly for all $\gamma \in \Gamma$,

$$\int_{K_\epsilon} \mathcal{T}_n(dx_1|x_t^n, \gamma(y_{[0,t]})) \geq 1 - \epsilon.$$

Now, we fix an $\epsilon > 0$ and choose a compact set K_ϵ according to the above discussion such that all $\mathcal{T}_n$ put almost all their measure (more than $1-\epsilon$) on K_ϵ. We also fix a finite family of continuous functions $\mathbb{F} := \{f^1, \dots, f^N\}$ such that for any γ, we can find an $f^i \in \mathbb{F}$ with $\sup_{x_t \in K_\epsilon} |f^\gamma(x_t) - f^i(x_t)| \leq \epsilon$. Moreover, we choose a large N such that $\sup_{x \in K_\epsilon} |f_n^\gamma(x) - f^i(x)| \leq 3\epsilon/2$ for all $n \geq N$.

With this setup, we go back to the main statement:

$$\begin{aligned}
&\sup_{\gamma\in\Gamma}\Big|\int \mathcal{T}_n(dx_{t+1}|x_t^n, \gamma(y_{[0,t]}))f_n^\gamma(x_{t+1}) - \int \mathcal{T}(dx_{t+1}|x_t, \gamma(y_{[0,t]}))f^\gamma(x_{t+1})\Big| \\
&\leq \sup_{\gamma\in\Gamma}\Big|\int_{\mathbb{X}\setminus K_\epsilon} \mathcal{T}_n(dx_{t+1}|x_t^n, \gamma(y_{[0,t]}))f_n^\gamma(x_{t+1}) - \int_{\mathbb{X}\setminus K_\epsilon} \mathcal{T}(dx_{t+1}|x_t, \gamma(y_{[0,t]}))f^\gamma(x_{t+1})\Big| \\
&\qquad + \sup_{\gamma\in\Gamma}\Big|\int_{K_\epsilon} \mathcal{T}_n(dx_{t+1}|x_t^n, \gamma(y_{[0,t]}))f_n^\gamma(x_{t+1}) \\
&\qquad - \int_{K_\epsilon} \mathcal{T}(dx_{t+1}|x_t, \gamma(y_{[0,t]}))f^\gamma(x_{t+1})\Big| \\
&\leq 2\epsilon C + \sup_{\gamma\in\Gamma}\Big|\int_{K_\epsilon} \mathcal{T}_n(dx_{t+1}|x_t^n, \gamma(y_{[0,t]}))\big(f_n^\gamma(x_{t+1}) - f^i(x_{t+1})\big) \\
&\qquad\qquad + \int_{K_\epsilon} \mathcal{T}_n(dx_{t+1}|x_t^n, \gamma(y_{[0,t]}))f^i(x_{t+1}) \\
&\qquad\qquad\qquad - \int_{K_\epsilon} \mathcal{T}(dx_{t+1}|x_t, \gamma(y_{[0,t]}))f^i(x_{t+1}) \\
&\qquad\qquad\qquad + \int_{K_\epsilon} \mathcal{T}(dx_{t+1}|x_t, \gamma(y_{[0,t]}))\big(f^i(x_{t+1}) - f^\gamma(x_{t+1})\big)\Big| \\
&\leq 2\epsilon C + \sup_{\gamma\in\Gamma}\Big|\int_{K_\epsilon} \mathcal{T}_n(dx_{t+1}|x_t^n, \gamma(y_{[0,t]}))f^i(x_{t+1}) \\
&\qquad\qquad\qquad - \int_{K_\epsilon} \mathcal{T}(dx_{t+1}|x_t, \gamma(y_{[0,t]}))f^i(x_{t+1})\Big| + 5\epsilon/2 \leq 4\epsilon C + 7\epsilon/2,
\end{aligned}$$

where C is the uniform bound of f_n^γ and $f^i(x_{t+1})$ is chosen according to the discussion above such that f^i is ϵ close to $f^\gamma(x_{t+1})$ and the same f^i is $3\epsilon/2$ close to $f_n^\gamma(x_{t+1})$.

At the last step, we used the fact that $\mathcal{T}_n(dx_{t+1}|x_t^n, \gamma(y_{[0,t]}))$ converges weakly to $\mathcal{T}(dx_{t+1}|x_t, \gamma(y_{[0,t]}))$ uniformly on $\mathbb{U}$. Thus,

$$\begin{aligned}
&\sup_{\gamma\in\Gamma}\Big|\int_{K_\epsilon} \mathcal{T}_n(dx_{t+1}|x_t^n, \gamma(y_{[0,t]}))f^i(x_{t+1}) - \int_{K_\epsilon} \mathcal{T}(dx_{t+1}|x_t, \gamma(y_{[0,t]}))f^i(x_{t+1})\Big| \\
&\leq \sup_{\gamma\in\Gamma}\Big|\int_{\mathbb{X}-K_\epsilon} \mathcal{T}_n(dx_{t+1}|x_t^n, \gamma(y_{[0,t]}))f^i(x_{t+1}) - \int_{\mathbb{X}-K_\epsilon} \mathcal{T}(dx_{t+1}|x_t, \gamma(y_{[0,t]}))f^i(x_{t+1})\Big| \\
&\qquad + \sup_{\gamma\in\Gamma}\Big|\int_{\mathbb{X}} \mathcal{T}_n(dx_{t+1}|x_t^n, \gamma(y_{[0,t]}))f^i(x_{t+1}) \\
&\qquad - \int_{\mathbb{X}} \mathcal{T}(dx_{t+1}|x_t, \gamma(y_{[0,t]}))f^i(x_{t+1})\Big| \leq 2\epsilon C + \epsilon
\end{aligned}$$

for large enough n. As ϵ is arbitrary, the result follows. ∎

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

With this lemma, we go back to (2.2). For easiness of notation we will first study the case where $t = 3$.

$$\begin{aligned}
&\sup_{\gamma\in\Gamma}\Big|E_P^{\mathcal{T}}\Big[c\big(X_3,\gamma(Y_{[0,3]})\big)\Big] - E_P^{\mathcal{T}_n}\Big[c\big(X_3,\gamma(Y_{[0,3]})\big)\Big]\Big| \\
&= \sup_{\gamma\in\Gamma}\Big|\int P(dx_0)Q(dy_0|x_0)\mathcal{T}(dx_1|x_0,\gamma(y_0))Q(dy_1|x_1)\mathcal{T}(dx_2|x_1,\gamma(y_{[0,1]})) \\
&\qquad\qquad \times Q(dy_2|x_2)\mathcal{T}(dx_3|x_2,\gamma(y_{[0,2]}))Q(dy_3|x_3)c(x_3,\gamma(y_{[0,3]})) \\
&\qquad - \int P(dx_0)Q(dy_0|x_0)\mathcal{T}_n(dx_1|x_0,\gamma(y_0))Q(dy_1|x_1)\mathcal{T}_n(dx_2|x_1,\gamma(y_{[0,1]})) \\
&\qquad\qquad \times Q(dy_2|x_2)\mathcal{T}_n(dx_3|x_2,\gamma(y_{[0,2]}))Q(dy_3|x_3)c(x_3,\gamma(y_{[0,3]}))\Big|.
\end{aligned}$$

Using the dominated convergence theorem, it suffices to show that

$$\begin{aligned}
&\sup_{\gamma\in\Gamma}\Big|\int \mathcal{T}_n(dx_1|x_0,\gamma(y_0))Q(dy_1|x_1)\mathcal{T}_n(dx_2|x_1,\gamma(y_{[0,1]})) \\
&\qquad\qquad \times Q(dy_2|x_2)\mathcal{T}_n(dx_3|x_2,\gamma(y_{[0,2]}))Q(dy_3|x_3)c(x_3,\gamma(y_{[0,3]})) \\
&\qquad - \int \mathcal{T}(dx_1|x_0,\gamma(y_0))Q(dy_1|x_1)\mathcal{T}(dx_2|x_1,\gamma(y_{[0,1]})) \\
&\qquad\qquad \times Q(dy_2|x_2)\mathcal{T}(dx_3|x_2,\gamma(y_{[0,2]}))Q(dy_3|x_3)c(x_3,\gamma(y_{[0,3]}))\Big| \to 0.
\end{aligned}$$

Then, using Lemma A.2, it suffices to show that for any $x_1^n \to x_1$

$$\begin{aligned}
&\sup_{\gamma\in\Gamma}\Big|\int Q(dy_1|x_1^n)\mathcal{T}_n(dx_2|x_1^n,\gamma(y_{[0,1]}))Q(dy_2|x_2)\mathcal{T}_n(dx_3|x_2,\gamma(y_{[0,2]})) \\
&\qquad\qquad \times Q(dy_3|x_3)c(x_3,\gamma(y_{[0,3]})) \\
&\qquad - \int Q(dy_1|x_1)\mathcal{T}(dx_2|x_1,\gamma(y_{[0,1]}))Q(dy_2|x_2)\mathcal{T}(dx_3|x_2,\gamma(y_{[0,2]})) \\
&\qquad\qquad \times Q(dy_3|x_3)c(x_3,\gamma(y_{[0,3]}))\Big| \to 0
\end{aligned} \tag{A.4}$$

and

$$\begin{aligned}
&\sup_{\gamma\in\Gamma}\Big|\int Q(dy_1|x_1^n)\mathcal{T}(dx_2|x_1^n,\gamma(y_{[0,1]}))Q(dy_2|x_2)\mathcal{T}(dx_3|x_2,\gamma(y_{[0,2]})) \\
&\qquad\qquad \times Q(dy_3|x_3)c(x_3,\gamma(y_{[0,3]})) \\
&\qquad - \int Q(dy_1|x_1)\mathcal{T}(dx_2|x_1,\gamma(y_{[0,1]}))Q(dy_2|x_2)\mathcal{T}(dx_3|x_2,\gamma(y_{[0,2]})) \\
&\qquad\qquad \times Q(dy_3|x_3)c(x_3,\gamma(y_{[0,3]}))\Big| \to 0.
\end{aligned} \tag{A.5}$$

We only focus on the term (A.4); the analysis for the term (A.5) follows from identical

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 05/17/20 to 130.15.241.167. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Downloaded 05/17/20 to 130.15.241.167. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

steps. For (A.4), we write the following:

$$\begin{aligned}
&\sup_{\gamma\in\Gamma}\Bigg|\int Q(dy_1|x_1^n)\mathcal{T}_n(dx_2|x_1^n,\gamma(y_{[0,1]}))Q(dy_2|x_2)\mathcal{T}_n(dx_3|x_2,\gamma(y_{[0,2]}))\\
&\qquad\qquad\times Q(dy_3|x_3)c(x_3,\gamma(y_{[0,3]}))\\
&\quad-\int Q(dy_1|x_1)\mathcal{T}(dx_2|x_1,\gamma(y_{[0,1]}))Q(dy_2|x_2)\mathcal{T}(dx_3|x_2,\gamma(y_{[0,2]}))\\
&\qquad\qquad\times Q(dy_3|x_3)c(x_3,\gamma(y_{[0,3]}))\Bigg|\\
&\le\sup_{\gamma\in\Gamma}\Bigg|\int Q(dy_1|x_1^n)\mathcal{T}_n(dx_2|x_1^n,\gamma(y_{[0,1]}))Q(dy_2|x_2)\mathcal{T}_n(dx_3|x_2,\gamma(y_{[0,2]}))\\
&\qquad\qquad\times Q(dy_3|x_3)c(x_3,\gamma(y_{[0,3]}))\\
&\quad-\int Q(dy_1|x_1)\mathcal{T}_n(dx_2|x_1^n,\gamma(y_{[0,1]}))Q(dy_2|x_2)\mathcal{T}_n(dx_3|x_2,\gamma(y_{[0,2]}))\\
&\qquad\qquad\times Q(dy_3|x_3)c(x_3,\gamma(y_{[0,3]}))\Bigg|\\
&+\sup_{\gamma\in\Gamma}\Bigg|\int Q(dy_1|x_1)\mathcal{T}_n(dx_2|x_1^n,\gamma(y_{[0,1]}))Q(dy_2|x_2)\mathcal{T}_n(dx_3|x_2,\gamma(y_{[0,2]}))\\
&\qquad\qquad\times Q(dy_3|x_3)c(x_3,\gamma(y_{[0,3]}))\\
&\quad-\int Q(dy_1|x_1)\mathcal{T}(dx_2|x_1,\gamma(y_{[0,1]}))Q(dy_2|x_2)\mathcal{T}(dx_3|x_2,\gamma(y_{[0,2]}))\\
&\qquad\qquad\times Q(dy_3|x_3)c(x_3,\gamma(y_{[0,3]}))\Bigg|. \qquad\text{(A.6)}
\end{aligned}$$

The first term goes to 0 since the channel is continuous in total variation. For the second term, using Lemma A.2 and the total variation continuity of Q successively, it reduces to show that

$$\sup_{\gamma\in\Gamma}\left|\int Q(dy_3|x_3^n)c(x_3^n,\gamma(y_{[0,3]})-\int Q(dy_3|x_3)c(x_3,\gamma(y_{[0,3]})\right|\to 0.$$

To show this, we write the following:

$$\begin{aligned}
&\sup_{\gamma\in\Gamma}\left|\int Q(dy_3|x_3^n)c(x_3^n,\gamma(y_{[0,3]})-\int Q(dy_3|x_3)c(x_3,\gamma(y_{[0,3]})\right|\\
&\le\sup_{\gamma\in\Gamma}\left|\int Q(dy_3|x_3^n)c(x_3^n,\gamma(y_{[0,3]})-\int Q(dy_3|x_3)c(x_3^n,\gamma(y_{[0,3]})\right|\\
&\quad+\sup_{\gamma\in\Gamma}\int Q(dy_3|x_3)\big|c(x_3^n,\gamma(y_{[0,3]})-c(x_3,\gamma(y_{[0,3]})\big|.
\end{aligned}$$

The first term goes to 0 since Q is continuous in total variation and the second term goes to 0 since c is continuous in x uniformly over $\mathbb{U}$. Thus, (A.4) holds true. (A.5) also holds true with identical arguments; we use the convergence of $\mathcal{T}(\cdot|x_n,u)$ to $\mathcal{T}(\cdot|x,u)$ uniformly over $u\in\mathbb{U}$ whereas for (A.4), we use the convergence of $\mathcal{T}_n(\cdot|x_n,u)$ to $\mathcal{T}(\cdot|x,u)$ uniformly over $u\in\mathbb{U}$ at (A.6) with Lemma A.2. Therefore, (2.2) goes to 0 for the time step $t=3$. For a general finite time stage t, we can again use the iterative approach we used when $t=3$. Thus, we can generalize that for any $0<t<\infty$

$$\sup_{\gamma\in\Gamma}\left|E_P^{\mathcal{T}}\Big[c\big(X_t,\gamma(Y_{[0,t]})\big)\Big]-E_P^{\mathcal{T}_n}\Big[c\big(X_t,\gamma(Y_{[0,t]})\big)\Big]\right|\to 0.$$

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 05/17/20 to 130.15.241.167. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

A.2. Proof of (2.4) in Theorem 2.5. Before the proof we give a key lemma.

LEMMA A.3. *For a uniformly bounded family of functions* $\{f_n^\gamma : \mathbb{X} \to \mathbb{R}, n > 0, \gamma \in \Gamma\}$ *and* $\{f^\gamma : \mathbb{X} \to \mathbb{R}, \gamma \in \Gamma\}$ *if we have* $\sup_{\gamma\in\Gamma} \left|f_n^\gamma(x) - f^\gamma(x)\right| \to 0$, *then*

$$\sup_{\gamma\in\Gamma}\left|\int \mathcal{T}_n(dx_{t+1}|x_t,\gamma(y_{[0,t]}))f_n^\gamma(x_{t+1}) - \int \mathcal{T}(dx_{t+1}|x_t,\gamma(y_{[0,t]}))f^\gamma(x_{t+1})\right| \to 0$$

for a fixed observation realizations $y_{[0,t]} := \{y_0,\dots,y_t\}$ *and a fixed state* x_t, *under the following assumptions:*

(i) $\mathcal{T}_n$ *is such that for any sequence* $\{u_n\} \subset \mathbb{U}$ *converging to some* $u \in \mathbb{U}$, $\mathcal{T}_n(\cdot|x,u_n) \to \mathcal{T}(\cdot|x,u)$ *in total variation for all* $x \in \mathbb{X}$,

(ii) $\mathcal{T}(\cdot|x,u)$ *is continuous in total variation in* u *for every given* $x \in \mathbb{X}$.

Proof.

$$\begin{aligned}
&\sup_{\gamma\in\Gamma}\left|\int \mathcal{T}_n(dx_{t+1}|x_t,\gamma(y_{[0,t]}))f_n^\gamma(x_{t+1}) - \int \mathcal{T}(dx_{t+1}|x_t,\gamma(y_{[0,t]}))f^\gamma(x_{t+1})\right| \\
&\le \sup_{\gamma\in\Gamma}\left|\int \mathcal{T}_n(dx_{t+1}|x_t,\gamma(y_{[0,t]}))f_n^\gamma(x_{t+1}) - \int \mathcal{T}(dx_{t+1}|x_t,\gamma(y_{[0,t]}))f_n^\gamma(x_{t+1})\right| \\
&\qquad + \sup_{\gamma\in\Gamma}\left|\int \mathcal{T}(dx_{t+1}|x_t,\gamma(y_{[0,t]}))\big(f_n^\gamma(x_{t+1}) - f^\gamma(x_{t+1})\big)\right| \\
&\le \sup_{\gamma\in\Gamma}\|\mathcal{T}_n(dx_{t+1}|x_t,\gamma(y_{[0,t]})) - \mathcal{T}(dx_{t+1}|x_t,\gamma(y_{[0,t]}))\|_{TV} \\
&\qquad + \sup_{\gamma\in\Gamma}\left|\int \mathcal{T}(dx_{t+1}|x_t,\gamma(y_{[0,t]}))\big(f_n^\gamma(x_{t+1}) - f^\gamma(x_{t+1})\big)\right|.
\end{aligned}$$

Above, the first term goes to 0 as $\mathcal{T}_n(\cdot|x,u_n) \to \mathcal{T}(\cdot|x,u)$ in total variation and $\mathbb{U}$ is compact.

For the second term, first we use the assumption that $\mathcal{T}(\cdot|x,u)$ is continuous in u. For any $\epsilon > 0$, there exists a $\delta > 0$ such that $|u' - u| < \delta$ implies $\|\mathcal{T}(\cdot|x,u) - \mathcal{T}(\cdot|x,u')\|_{TV} < \epsilon$. Furthermore, by this assumption $\mathbb{U}$ is compact. Therefore, for the given δ, there exists a finite set $\{u_1,\dots,u_N\}$ such that for any $\gamma \in \Gamma$, we can find a u_i with $|u_i - \gamma(y_{[0,t]})| < \delta$.

Combining what we have; for any $\epsilon > 0$ and for any $\gamma \in \Gamma$, we can find a u_i such that $\|\mathcal{T}(\cdot|x,\gamma(y_{[0,t]})) - \mathcal{T}(\cdot|x,u_i)\|_{TV} < \epsilon$. Now we focus on the second term again:

$$\begin{aligned}
&\sup_{\gamma\in\Gamma}\left|\int \mathcal{T}(dx_{t+1}|x_t,\gamma(y_{[0,t]}))\big(f_n^\gamma(x_{t+1}) - f^\gamma(x_{t+1})\big)\right| \\
&\le \sup_{\gamma\in\Gamma}\left|\int \mathcal{T}(dx_{t+1}|x_t,\gamma(y_{[0,t]}))\big(f_n^\gamma(x_{t+1}) - f^\gamma(x_{t+1})\big)\right. \\
&\qquad\qquad \left. - \int \mathcal{T}(dx_{t+1}|x_t,u_i)\big(f_n^\gamma(x_{t+1}) - f^\gamma(x_{t+1})\big)\right| \\
&\qquad + \sup_{\gamma\in\Gamma}\int \mathcal{T}(dx_{t+1}|x_t,u_i)\left|f_n^\gamma(x_{t+1}) - f^\gamma(x_{t+1})\right| \\
&\le \|c\|_\infty \sup_{\gamma\in\Gamma}\|\mathcal{T}(\cdot|x,\gamma(y_{[0,t]})) - \mathcal{T}(\cdot|x,u_i)\|_{TV} \\
&\qquad\qquad + \sup_{\gamma\in\Gamma}\int \mathcal{T}(dx_{t+1}|x_t,u_i)\left|f_n^\gamma(x_{t+1}) - f^\gamma(x_{t+1})\right|,
\end{aligned}$$

where $\|c\|_\infty$ is a uniform bound of f_n and u_i is chosen according to the above dis-

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 05/17/20 to 130.15.241.167. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

cussion. Thus, the first term is less than $\|c\|_\infty \epsilon$ and the second term can be made arbitrarily small for large enough n since $\sup_{\gamma\in\Gamma}\left|f_n^\gamma(x) - f^\gamma(x)\right| \to 0$ by assumption. The result follows. ∎

Now we go back to (2.4). We will first study the case where $t = 3$.

$$\begin{aligned}
&\sup_{\gamma\in\Gamma}\left|E_P^{\mathcal{T}}\Big[c\big(X_3,\gamma(Y_{[0,3]})\big)\Big] - E_P^{\mathcal{T}_n}\Big[c\big(X_3,\gamma(Y_{[0,3]})\big)\Big]\right| \\
&= \sup_{\gamma\in\Gamma}\Bigg|\int P(dx_0)Q(dy_0|x_0)\mathcal{T}(dx_1|x_0,\gamma(y_0))Q(dy_1|x_1)\mathcal{T}(dx_2|x_1,\gamma(y_{[0,1]})) \\
&\qquad\qquad \times Q(dy_2|x_2)\mathcal{T}(dx_3|x_2,\gamma(y_{[0,2]}))Q(dy_3|x_3)c(x_3,\gamma(y_{[0,3]})) \\
&\qquad - \int P(dx_0)Q(dy_0|x_0)\mathcal{T}_n(dx_1|x_0,\gamma(y_0))Q(dy_1|x_1)\mathcal{T}_n(dx_2|x_1,\gamma(y_{[0,1]})) \\
&\qquad\qquad \times Q(dy_2|x_2)\mathcal{T}_n(dx_3|x_2,\gamma(y_{[0,2]}))Q(dy_3|x_3)c(x_3,\gamma(y_{[0,3]}))\Bigg|.
\end{aligned}$$

Using Lemma A.3 it suffices to show that

$$\begin{aligned}
&\sup_{\gamma\in\Gamma}\Bigg|\int\Big(Q(dy_1|x_1)\mathcal{T}_n(dx_2|x_1,\gamma(y_{[0,1]}))Q(dy_2|x_2)\mathcal{T}_n(dx_3|x_2,\gamma(y_{[0,2]})) \\
&\qquad - Q(dy_1|x_1)\mathcal{T}(dx_2|x_1,\gamma(y_{[0,1]}))Q(dy_2|x_2)\mathcal{T}(dx_3|x_2,\gamma(y_{[0,2]}))\Big) \\
&\qquad\qquad \times Q(dy_3|x_3)c(x_3,\gamma(y_{[0,3]}))\Bigg| \to 0.
\end{aligned}$$

Following the same procedure and using Lemma A.3 successively, it reduces to show that

$$\begin{aligned}
&\sup_{\gamma\in\Gamma}\Bigg|\int \mathcal{T}_n(dx_3|x_2,\gamma(y_{[0,2]}))Q(dy_3|x_3)c(x_3,\gamma(y_{[0,3]})) \\
&\qquad - \int \mathcal{T}(dx_3|x_2,\gamma(y_{[0,2]}))Q(dy_3|x_3)c(x_3,\gamma(y_{[0,3]}))\Bigg| \\
&\le \|c\|_\infty \sup_{\gamma\in\Gamma}\|\mathcal{T}_n(dx_3|x_2,\gamma(y_{[0,2]})) - \mathcal{T}(dx_3|x_2,\gamma(y_{[0,2]}))\|_{TV} \to 0,
\end{aligned}$$

which holds true by the assumptions, i.e., since the action space $\mathbb{U}$ is compact and $\mathcal{T}_n$ is such that for any sequence $\{u_n\} \subset \mathbb{U}$ converging to some $u \in \mathbb{U}$, $\mathcal{T}_n(\cdot|x,u_n) \to \mathcal{T}(\cdot|x,u)$ in total variation for all $x \in \mathbb{X}$. This argument can be applied to any time step $t < \infty$.

A.3. Proof of (2.6) in Theorem 2.6. First, we provide the analysis for $k = 2$.

$$\begin{aligned}
&\|P_{\mathcal{T}_n}^\gamma(d(x,y,u)_{[0,2]}) - P_{\mathcal{T}}^\gamma(d(x,y,u)_{[0,2]})\|_{TV} \\
&= \sup_{||f||_\infty\le 1}\Bigg|\int P(\mathrm{d}x_0)Q(\mathrm{d}y_0|x_0)1_{\{\gamma(y_0)\in \mathrm{d}u_0\}}\mathcal{T}_n(\mathrm{d}x_1|x_0,u_0)Q(\mathrm{d}y_1|x_1)1_{\{\gamma(y_0,y_1)\in \mathrm{d}u_1\}} \\
&\qquad\qquad \times \mathcal{T}_n(\mathrm{d}x_2|x_1,u_1)Q(\mathrm{d}y_2|x_2)1_{\{\gamma(y_0,y_1,y_2)\in \mathrm{d}u_2\}}f(x,y,u)_{[0,2]} \\
&\qquad - \int P(\mathrm{d}x_0)Q(\mathrm{d}y_0|x_0)1_{\{\gamma(y_0)\in \mathrm{d}u_0\}}\mathcal{T}(\mathrm{d}x_1|x_0,u_0)Q(\mathrm{d}y_1|x_1)1_{\{\gamma(y_0,y_1)\in \mathrm{d}u_1\}} \\
&\qquad\qquad \times \mathcal{T}(\mathrm{d}x_2|x_1,u_1)Q(\mathrm{d}y_2|x_2)1_{\{\gamma(y_0,y_1,y_2)\in \mathrm{d}u_2\}}f(x,y,u)_{[0,2]}\Bigg|
\end{aligned}$$

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

$$
\begin{aligned}
&\leq \sup_{||f||_\infty \leq 1} \int P^\gamma_{\mathcal{T}_n}(d(x_{[0,1]}, y_{[0,1]}, u_{[0,1]})) \\
&\qquad \times \Bigg| \int \mathcal{T}_n(\mathrm{d}x_2|x_1,u_1) Q(\mathrm{d}y_2|x_2) 1_{\{\gamma(y_{[0,2]})\in \mathrm{d}u_2\}} f(x,y,u)_{[0,2]} \\
&\qquad\qquad - \int \mathcal{T}(\mathrm{d}x_2|x_1,u_1) Q(\mathrm{d}y_2|x_2) 1_{\{\gamma(y_{[0,2]})\in \mathrm{d}u_2\}} f(x,y,u)_{[0,2]} \Bigg| \\
&+ \sup_{||f||_\infty \leq 1} \int P^\gamma_{\mathcal{T}}(d(x_0,y_0,u_0)) \\
&\qquad \times \Bigg| \int \mathcal{T}_n(\mathrm{d}x_1|x_0,u_0) Q(\mathrm{d}y_1|x_1) 1_{\{\gamma(y_0,y_1)\in \mathrm{d}u_1\}} P^\gamma_{\mathcal{T}}(d(x_2,y_2,u_2)) f(x,y,u)_{[0,2]} \\
&\qquad - \int \mathcal{T}(\mathrm{d}x_1|x_0,u_0) Q(\mathrm{d}y_1|x_1) 1_{\{\gamma(y_0,y_1)\in \mathrm{d}u_1\}} P^\gamma_{\mathcal{T}}(d(x_2,y_2,u_2)) f(x,y,u)_{[0,2]} \Bigg| \\
&\leq 2 \sup_{x\in\mathbb{X}, u\in\mathbb{U}} \|\mathcal{T}_n(.|x,u) - \mathcal{T}(.|x,u)\|_{TV}.
\end{aligned}
$$

Now, we do the same analysis for a general time step k:

$$
\begin{aligned}
&\|P^\gamma_{\mathcal{T}_n}(d(x,y,u)_{[0,k]}) - P^\gamma_{\mathcal{T}}(d(x,y,u)_{[0,k]})\|_{TV} \\
&= \sup_{||f||_\infty \leq 1} \Bigg| \int f(x,y,u)_{[0,k]} P^\gamma_{\mathcal{T}_n}(d(x,y,u)_{[0,k]}) - \int f(x,y,u)_{[0,k]} P^\gamma_{\mathcal{T}}(d(x,y,u)_{[0,k]}) \Bigg| \\
&\leq \sup_{||f||_\infty \leq 1} \Bigg| \int P^\gamma_{\mathcal{T},\mathcal{T}_n}(dx_0, dy_0, du_0) \\
&\qquad \times \Bigg[\int \mathcal{T}(dx_1|x_0,u_0) \int f(x,y,u)_{[0,k]} P^\gamma_{\mathcal{T},\mathcal{T}_n}(dx_{[2,k]}, dy_{[1,k]}, du_{[1,k]}) \\
&\qquad\qquad - \int \mathcal{T}_n(dx_1|x_0,u_0) \int f(x,y,u)_{[0,k]} P^\gamma_{\mathcal{T},\mathcal{T}_n}(dx_{[2,k]}, dy_{[1,k]}, du_{[1,k]}) \Bigg] \Bigg| \\
&+ \Bigg| \int P^\gamma_{\mathcal{T},\mathcal{T}_n}(dx_{[0,1]}, dy_{[0,1]}, du_{[0,1]}) \\
&\qquad \times \Bigg[\int \mathcal{T}(dx_2|x_1,u_1) \int f(x,y,u)_{[0,k]} P^\gamma_{\mathcal{T},\mathcal{T}_n}(dx_{[3,k]}, dy_{[2,k]}, du_{[3,k]}) \\
&\qquad\qquad - \int \mathcal{T}_n(dx_2|x_1,u_1) \int f(x,y,u)_{[0,k]} P^\gamma_{\mathcal{T},\mathcal{T}_n}(dx_{[3,k]}, dy_{[2,k]}, du_{[2,k]}) \Bigg] \Bigg| \\
&\cdots + \Bigg| \int P^\gamma_{\mathcal{T},\mathcal{T}_n}(dx_{[0,k-1]}, dy_{[0,k-1]}, du_{[0,k-1]}) \Bigg[\int \mathcal{T}(dx_k|x_{k-1},u_{k-1}) f(x,y,u)_{[0,k]} \\
&\qquad\qquad\qquad - \int \mathcal{T}_n(dx_k|x_{k-1},u_{k-1}) f(x,y,u)_{[0,k]} \Bigg] \Bigg| \\
&\leq \int P^\gamma_{\mathcal{T},\mathcal{T}_n}(dx_0, dy_0, du_0) \|\mathcal{T}(\cdot|x_0,u_0) - \mathcal{T}_n(\cdot|x_0,u_0)\|_{TV} \\
&\quad + \int P^\gamma_{\mathcal{T},\mathcal{T}_n}(dx_{[0,1]}, dy_{[0,1]}, du_{[0,1]}) \|\mathcal{T}(\cdot|x_1,u_1) - \mathcal{T}_n(\cdot|x_1,u_1)\|_{TV} \\
&\cdots + \int P^\gamma_{\mathcal{T},\mathcal{T}_n}(dx_{[0,k-1]}, dy_{[0,k-1]}, du_{[0,k-1]})) \|\mathcal{T}(\cdot|x_{k-1},u_{k-1}) - \mathcal{T}_n(\cdot|x_{k-1},u_{k-1})\|_{TV} \\
&\leq k \sup_{x\in\mathbb{X}, u\in\mathbb{U}} \|\mathcal{T}(.|x,u) - \mathcal{T}_n(.|x,u)\|_{TV}.
\end{aligned}
$$

Downloaded 05/17/20 to 130.15.241.167. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 05/17/20 to 130.15.241.167. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

In the argument above $P^{\gamma}_{\mathcal{T},\mathcal{T}_n}$ denotes a strategic measure that uses $\mathcal{T}$ or $\mathcal{T}_n$ at various steps. The terms are arranged so that at every term the applied strategic measures coincide.

A.4. Proof for Remark 4.3. We note that if the family where we search for policies is restricted to an equicontinuous family of functions, robustness can also be achieved. Let Γ_{eq} be the family of equicontinuous policies so that for any given $x_0 \in \mathbb{X}$ and $\epsilon > 0$, there exists a $\delta > 0$ such that $|\gamma(x) - \gamma(x_0)| \leq \epsilon$ for all $\gamma \in \Gamma_{eq}$ and for every x such that $|x - x_0| \leq \delta$.

We show that for all $t < \infty$

$$\sup_{\gamma\in\Gamma_{eq}} \left| E^{\mathcal{T}}\Big[c\big(X_t, \gamma(X_t)\big)\Big] - E^{\mathcal{T}_n}\Big[c\big(X_t, \gamma(X_t)\big)\Big] \right| \to 0.$$

For ease of notation we will first study the case where $t = 2$.

$$\begin{aligned}
&\sup_{\gamma\in\Gamma} \Big| E^{\mathcal{T}}\Big[c\big(X_2, \gamma(X_2)\big)\Big] - E^{\mathcal{T}_n}\Big[c\big(X_2, \gamma(X_2)\big)\Big] \Big| \\
&= \sup_{\gamma\in\Gamma} \Big| \int \mathcal{T}(dx_1|x_0, \gamma(x_0))\mathcal{T}(dx_2|x_1, \gamma(x_1))c(x,\gamma(x_2)) \\
&\quad - \int \mathcal{T}_n(dx_1|x_0, \gamma(x_0))\mathcal{T}_n(dx_2|x_1, \gamma(x_1))c(x_2, \gamma(x_2)) \Big|.
\end{aligned}$$

To show that the above term goes to 0, we use a lemma parallel to Lemma A.2 in the paper.

LEMMA A.4. *Suppose we have a uniformly bounded family of functions* $\{f^{\gamma}_n : \mathbb{X} \to \mathbb{R}, \gamma \in \Gamma_{eq}, n > 0\}$ *such that* $\|f^{\gamma}_n\|_\infty < C$ *for all* $\gamma \in \Gamma_{eq}$ *and for all* $n > 0$ *for some* $C < \infty$.

Further suppose we have another uniformly bounded family of functions $\{f^{\gamma} : \mathbb{X} \to \mathbb{R}, \gamma \in \Gamma_{eq}\}$ *such that* $\|f^{\gamma}\|_\infty < C$ *for all* $\gamma \in \Gamma_{eq}$ *for some* $C < \infty$ *under the following assumptions:*

(i) *For any* $x_n \to x$

$$\sup_{\gamma\in\Gamma_{eq}} \big|f^{\gamma}_n(x_n) - f^{\gamma}(x)\big| \to 0, \tag{A.7}$$

$$\sup_{\gamma\in\Gamma_{eq}} \big|f^{\gamma}(x_n) - f^{\gamma}(x)\big| \to 0. \tag{A.8}$$

(ii) $\mathcal{T}_n(\cdot|x_n, u_n)$ *converges weakly to* $\mathcal{T}(\cdot|x, u)$ *for any* $(x_n, u_n) \to (x, u)$.

(iii) $\mathcal{T}(\cdot|x, u)$ *is weakly continuous in* (x, u).

(iv) $\mathbb{U}$ *is compact.*

Then for some $x^n_t \to x_t$

$$\sup_{\gamma\in\Gamma_{eq}} \left| \int \mathcal{T}_n(dx_{t+1}|x^n_t, \gamma(x^n_t))f^{\gamma}_n(x_{t+1}) - \int \mathcal{T}(dx_{t+1}|x_t, \gamma(x_t))f^{\gamma}(x_{t+1}) \right| \to 0. \tag{A.9}$$

Proof. Using the same steps as in Lemma A.2 we can show that for any given compact set $K \subset \mathbb{X}$, and $\epsilon > 0$ there is a finite set of continuous functions $\mathbb{F} := \{f^1, \ldots, f^N\}$ so that for any γ, there is $f^i \in \mathbb{F}$ with

$$\sup_{x\in K} |f^{\gamma}(x) - f^i(x)| \leq \epsilon.$$

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

For the same $\epsilon > 0$, the same $f^i \in \mathbb{F}$, and the chosen compact set K, we can also make $\sup_{x\in K} |f_n^\gamma(x) - f^i(x)| \leq 3\epsilon/2$ for large enough n and for any $\gamma \in \Gamma_{eq}$.

We also have that the set of measures $S := \cup_{\gamma\in\Gamma_{eq}} S_\gamma$ is weakly compact where

$$S_\gamma = \{\mathcal{T}_n(\cdot|x_t^n, \gamma(x_t^n)) : \mathcal{T}_n(\cdot|x_t^n, \gamma(x_t^n)) \to \mathcal{T}(\cdot|x_t, \gamma(x_t))\}.$$

Hence, for any $\epsilon > 0$, there exists a compact set K_ϵ such that for all n and uniformly for all $\gamma \in \Gamma$,

$$\int_{K_\epsilon} \mathcal{T}_n(dx_1|x_t^n, \gamma(x_t^n)) \geq 1 - \epsilon.$$

Using these, we again follow the same steps as in the proof of Lemma A.2:

$$\begin{aligned}
&\sup_{\gamma\in\Gamma_{eq}} \Big| \int \mathcal{T}_n(dx_{t+1}|x_t^n, \gamma(x_t^n)) f_n^\gamma(x_{t+1}) - \int \mathcal{T}(dx_{t+1}|x_t, \gamma(x_t)) f^\gamma(x_{t+1}) \Big| \\
&\leq \sup_{\gamma\in\Gamma_{eq}} \Big| \int_{\mathbb{X}\setminus K_\epsilon} \mathcal{T}_n(dx_{t+1}|x_t^n, \gamma(x_t^n)) f_n^\gamma(x_{t+1}) - \int_{\mathbb{X}\setminus K_\epsilon} \mathcal{T}(dx_{t+1}|x_t, \gamma(x_t)) f^\gamma(x_{t+1}) \Big| \\
&\qquad + \sup_{\gamma\in\Gamma_{eq}} \Big| \int_{K_\epsilon} \mathcal{T}_n(dx_{t+1}|x_t^n, \gamma(x_t^n)) f_n^\gamma(x_{t+1}) - \int_{K_\epsilon} \mathcal{T}(dx_{t+1}|x_t, \gamma(x_t)) f^\gamma(x_{t+1}) \Big| \\
&\leq 2\epsilon C + \sup_{\gamma\in\Gamma_{eq}} \Big| \int_{K_\epsilon} \mathcal{T}_n(dx_{t+1}|x_t^n, \gamma(x_t^n)) \big(f_n^\gamma(x_{t+1}) - f^i(x_{t+1})\big) \\
&\qquad\qquad + \int_{K_\epsilon} \mathcal{T}_n(dx_{t+1}|x_t^n, \gamma(x_t^n)) f^i(x_{t+1}) - \int_{K_\epsilon} \mathcal{T}(dx_{t+1}|x_t, \gamma(x_t)) f^i(x_{t+1}) \\
&\qquad\qquad\qquad + \int_{K_\epsilon} \mathcal{T}(dx_{t+1}|x_t, \gamma(x_t)) \big(f^i(x_{t+1}) - f^\gamma(x_{t+1})\big) \Big| \\
&\leq 2\epsilon C + \sup_{\gamma\in\Gamma_{eq}} \Big| \int_{K_\epsilon} \mathcal{T}_n(dx_{t+1}|x_t^n, \gamma(x_t^n)) f^i(x_{t+1}) \\
&\qquad\qquad\qquad - \int_{K_\epsilon} \mathcal{T}(dx_{t+1}|x_t, \gamma(x_t)) f^i(x_{t+1}) \Big| + 5\epsilon/2 \leq 4\epsilon C + 7\epsilon/2,
\end{aligned}$$

where C is the uniform bound of f_n^γ and $f^i(x_{t+1})$ is chosen according to the discussion above such that f^i is ϵ close to $f^\gamma(x_{t+1})$ and the same f^i is $3\epsilon/2$ close to $f_n^\gamma(x_{t+1})$.

At the last step, we used the fact that $\mathcal{T}_n(dx_{t+1}|x_t^n, \gamma(x_t^n))$ converges weakly to $\mathcal{T}(dx_{t+1}|x_t, \gamma(x_t))$ uniformly over Γ_{eq} as Γ_{eq} is equicontinuous. As ϵ is arbitrary, the result follows. ∎

Now we go back to

$$\begin{aligned}
&\sup_{\gamma\in\Gamma} \Big| E^{\mathcal{T}}\Big[c\big(X_2, \gamma(X_2)\big)\Big] - E^{\mathcal{T}_n}\Big[c\big(X_2, \gamma(X_2)\big)\Big] \Big| \\
&= \sup_{\gamma\in\Gamma} \Big| \int \mathcal{T}(dx_1|x_0, \gamma(x_0)) \mathcal{T}(dx_2|x_1, \gamma(x_1)) c(x_2, \gamma(x_2)) \\
&\qquad - \int \mathcal{T}_n(dx_1|x_0, \gamma(x_0)) \mathcal{T}_n(dx_2|x_1, \gamma(x_1)) c(x_2, \gamma(x_2)) \Big|.
\end{aligned}$$

The previous lemma can be used to show that this term converges to 0.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 05/17/20 to 130.15.241.167. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Downloaded 05/17/20 to 130.15.241.167. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

A.5. Proof for (4.5). We focus on the discounted optimality equation for $\mathcal{T}_n$ for some initial point x_0^n, where $x_0^n \to x_0$:

$$J_\beta^*(\mathcal{T}_n, x_0^n) = c(x_0^n, \gamma_n^*(x_0^n)) + \beta \int J_\beta^*(\mathcal{T}_n, x_1)\mathcal{T}_n(dx_1|x_0^n, \gamma_n^*(x_0^n)). \tag{A.10}$$

Since the action space $\mathbb{U}$ is compact, we can find a subsequence n_k such that $\gamma_{n_k}^*(x_0^{n_k}) \to u^*$ for some $u^* \in \mathbb{U}$. Taking the limit $k \to \infty$ in (A.10) and using Theorem 4.2 (continuity) we get

$$J_\beta^*(\mathcal{T}, x_0) = c(x_0, u^*) + \beta \int J_\beta^*(\mathcal{T}, x_1)\mathcal{T}(dx_1|x_0, u^*). \tag{A.11}$$

Hence u^* satisfies DCOE for $\mathcal{T}$ and is an optimal action for x_0. In particular, any convergent subsequence of $\gamma_n^*(x_0^n)$ converges to an optimal action for x_0. With this observation, we claim that $T_n^k(J_\beta^*(\mathcal{T}))(x_n) \to J_\beta^*(\mathcal{T}, x))$ for any fixed $k < \infty$ as $n \to \infty$ for some $x_n \to x$, where T_n^k denotes that the operator T_n is applied k consecutive times. To show this, we follow an inductive approach. For $k = 1$, we write

$$T_n(J_\beta^*(\mathcal{T}))(x_n) = c(x_n, \gamma_n^*(x_n)) + \beta \int J_\beta^*(\mathcal{T}, x_1)\mathcal{T}(dx_1|x_n, \gamma_n^*(x_n)).$$

Suppose $\lim T_n(J_\beta^*(\mathcal{T}))(x_n) \neq J_\beta^*(\mathcal{T}, x)$ so that there exists a subsequence n_m and an $\epsilon > 0$ for which $|T_{n_m}(J_\beta^*(\mathcal{T}))(x_{n_m}) - J_\beta^*(\mathcal{T}, x)| > \epsilon$ for all m. Since $\mathbb{U}$ is compact, there exists a further subsequence $n_{m'}$ such that $\gamma_{n_{m'}}^*(x_{n_{m'}}) \to u$ for some $u \in \mathbb{U}$ and as we observed before u is an optimal action for x under the kernel $\mathcal{T}$. Hence

$$\begin{aligned}
&\lim_{m' \to \infty} T_{n_{m'}}(J_\beta^*(\mathcal{T}))(x_{n_{m'}}) \\
&= \lim_{m' \to \infty} c(x_{n_{m'}}, \gamma_{n_{m'}}^*(x_{n_{m'}})) + \beta \int J_\beta^*(\mathcal{T}, x_1)\mathcal{T}(dx_1|x_{n_{m'}}, \gamma_{n_{m'}}^*(x_{n_{m'}}) \\
&= c(x, u) + \beta \int J_\beta^*(\mathcal{T}, x_1)\mathcal{T}(dx_1|x, u) = J_\beta^*(\mathcal{T}, x),
\end{aligned}$$

where the last step follows from the observation that u is optimal for x under $\mathcal{T}$. Thus, we reach a contradiction and can conclude that $T_n(J_\beta^*(\mathcal{T}))(x_n) \to J_\beta^*(\mathcal{T}, x)$. Now assume that it also holds for $k - 1$ so that $T_n^{k-1}(J_\beta^*(\mathcal{T}))(x_n) \to J_\beta^*(\mathcal{T}, x)$. We write

$$T_n^k(J_\beta^*(\mathcal{T}))(x_n) = c(x_n, \gamma_n^*(x_n)) + \beta \int T_n^{k-1}(J_\beta^*(\mathcal{T}))(x_1)\mathcal{T}(dx_1|x_n, \gamma_n^*(x_n)).$$

Following a similar contradiction argument with the fact that $T_n^{k-1}(J_\beta^*(\mathcal{T}))(x_n) \to J_\beta^*(\mathcal{T}, x)$ and using [35, Theorem 3.5] or [49, Theorem 3.5] (weak convergence with varying functions), we can conclude that

$$\lim_{n \to \infty} T_n^k(J_\beta^*(\mathcal{T}))(x_n) = J_\beta^*(\mathcal{T}, x) \quad \forall k < \infty. \tag{A.12}$$

A.6. Proof for (4.7). We give a proof sketch building on section A.5. Define the discounted optimality equation for $\mathcal{T}_n$ for some initial point x_0.

$$J_\beta^*(\mathcal{T}_n, x_0) = c(x_0, \gamma_n^*(x_0)) + \beta \int J_\beta^*(\mathcal{T}_n, x_1)\mathcal{T}_n(dx_1|x_0, \gamma_n^*(x_0)). \tag{A.13}$$

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 05/17/20 to 130.15.241.167. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Since the action space $\mathbb{U}$ is compact, we can find a subsequence n_k such that $\gamma^*_{n_k}(x_0) \to u^*$ for some $u^* \in \mathbb{U}$. Taking the limit $k \to \infty$ in (A.13) and using Theorem 4.6 (continuity) we get

$$J^*_\beta(\mathcal{T}, x_0) = c(x_0, u^*) + \beta \int J^*_\beta(\mathcal{T}, x_1)\mathcal{T}(dx_1|x_0, u^*). \tag{A.14}$$

Hence u^* satisfies DCOE for $\mathcal{T}$ and is an optimal action for x_0. In particular, any convergent subsequence of $\gamma^*_n(x_0)$ converges to an optimal action for x_0. With this observation, we claim that $T^k_n(J^*_\beta(\mathcal{T}))(x) \to J^*_\beta(\mathcal{T}, x))$ for any fixed $k < \infty$ as $n \to \infty$ for any x, where T^k_n denotes the operator T_n applied k consecutive times. This can be shown by the same technique we use in section A.5 using [42, Theorem 20] (setwise convergence with varying functions).

REFERENCES

[1] A. Almudevar and E. F. Arruda, *Optimal approximation schedules for a class of iterative algorithms, with an application to multigrid value iteration*, IEEE Trans. Automat. Control, 57 (2012), pp. 3132–3146.

[2] E. F. Arruda, F. Ourique, J. Lacombe, and A. Almudevar, *Accelerating the convergence of value iteration by using partial transition functions*, European J. Oper. Res., 229 (2013), pp. 190–198.

[3] T. Başar and P. Bernhard, *H-infinity Optimal Control and Related Minimax Design Problems: A Dynamic Game Approach*, Birkhäuser, Boston, 1995.

[4] A. Benavoli and L. Chisci, *Robust stochastic control based on imprecise probabilities*, IFAC Proc. Vol., 44 (2011), pp. 4606–4613.

[5] D. P. Bertsekas and S. Shreve, *Stochastic Optimal Control: The Discrete Time Case*, Academic Press, New York, 1978.

[6] P. Billingsley, *Statistical methods in Markov chains*, Ann. Math. Stat., 32 (1961), pp. 12–40.

[7] P. Billingsley, *Convergence of Probability Measures*, Wiley, New York, 1968.

[8] P. Billingsley, *Probability and Measure*, 3rd ed. Wiley, New York, 1995.

[9] J. Blanchet and K. Murthy, *Quantifying distributional model risk via optimal transport*, Math. Oper. Res., 44 (2019), pp. 565–600.

[10] R. K. Boel, M. R. James, and I. R. Petersen, *Robustness and risk-sensitive filtering*, IEEE Trans. Automat. Control, 47 (2002), pp. 451–461.

[11] V. S. Borkar, *White-noise representations in stochastic realization theory*, SIAM J. Control Optim., 31 (1993), pp. 1093–1102.

[12] S. Dean, H. Mania, N. Matni, B. Recht, and S. Tu, *On the sample complexity of the linear quadratic regulator*, Found. Comput. Math., (2017), pp. 1–47.

[13] L. Devroye and L. Györfi, *Non-Parametric Density Estimation: The L_1 View*, Wiley, New York, 1985.

[14] R. M. Dudley, *Real Analysis and Probability*, 2nd ed., Cambridge University Press, Cambridge, 2002.

[15] P. Dupuis, M. R. James, and I. Petersen, *Robust properties of risk-sensitive control*, Math. Control Signals Systems, 13 (2000), pp. 318–332.

[16] P. K. E. Feinberg and M. Zgurovsky, *On Continuity of Transition Probabilities in Belief MDPs with General State and Action Spaces*, preprint, https://arxiv.org/abs/1903.11629, 2019.

[17] E. Erdoğan and G. N. Iyengar, *Ambiguous chance constrained problems and robust optimization*, Math. Program., 107 (2005), pp. 37–61.

[18] P. M. Esfahani and D. Kuhn, *Data-driven distributionally robust optimization using the Wasserstein metric: Performance guarantees and tractable reformulations*, Math. Program., 171 (2018), pp. 115–166.

[19] E. Feinberg, P. Kasyanov, and M. Zgurovsky, *Partially Observable Total-Cost Markov Decision Process with Weakly Continuous Transition Probabilities*, preprint, https://arxiv.org/abs/1401.2168, 2014.

[20] J. K. Ghosh and R. V. Ramamoorthi, *Bayesian Nonparametrics*, Springer, New York, 2003.

[21] I. I. Gihman and A. V. Skorohod, *Controlled Stochastic Processes*, Springer, New York, 2012.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

[22] O. Gossner and T. Tomala, *Entropy bounds on Bayesian learning*, J. Math. Econom., 44 (2008), pp. 24–32.
[23] R. M. Gray, *Entropy and Information Theory*, Springer, New York, 1990.
[24] L. Györfi and M. Kohler, *Nonparametric estimation of conditional distributions*, IEEE Trans. Inform. Theory, 53 (2007), pp. 1872–1879.
[25] B. Hajek, *Random Processes for Engineers*, Cambridge University Press, Cambridge, 2015.
[26] L. P. Hansen and T. J. Sargent, *Robust control and model uncertainty*, Amer. Econ. Rev., 91 (2001), pp. 60–66.
[27] O. Hernandez-Lerma and J. Lasserre, *Discrete-time Markov control processes*, Springer, New York, 1996.
[28] O. Hernández-Lerma and J. B. Lasserre, *Further topics on discrete-time Markov control processes*, Springer, New York, 1999.
[29] G. N. Iyengar, *Robust dynamic programming*, Math. Oper. Res., 30 (2005), pp. 257–280.
[30] D. Jacobson, *Optimal stochastic linear systems with exponential performance criteria and their relation to deterministic differential games*, IEEE Trans. Automat. Control, 18 (1973), pp. 124–131.
[31] A. D. Kara, N. Saldi, and S. Yüksel, *Weak Feller property of non-linear filters*, Systems Control Lett., 134 (2019), 104512.
[32] A. D. Kara and S. Yüksel, *Robustness to incorrect priors in partially observed stochastic control*, SIAM J. Control Optim., 57 (2019), pp. 1929–1964.
[33] M. L. Kleptsyna and A. Y. Veretennikov, *On robustness of discrete time optimal filters*, Math. Methods Statist., 25 (2016), pp. 207–218.
[34] H. Lam, *Robust sensitivity analysis for stochastic systems*, Math. Oper. Res., 41 (2016), pp. 1248–1275.
[35] H. Langen, *Convergence of dynamic programming models*, Math. Oper. Res., 6 (1981), pp. 493–512.
[36] A. Müller, *How does the value function of a Markov decision process depend on the transition probabilities?*, Math. Oper. Res., 22 (1997), pp. 872–885.
[37] A. Nilim and L. E. Ghaoui, *Robust control of Markov decision processes with uncertain transition matrices*, Oper. Res., 53 (2005), pp. 780–798.
[38] B. Øksendal and A. Sulem, *Forward-backward stochastic differential games and stochastic control under model uncertainty*, J. Optim. Theory Appl., 161 (2014), pp. 22–55.
[39] K. Parthasarathy, *Probability Measures on Metric Spaces*, AMS, Providence, RI, 1967.
[40] I. Petersen, M. R. James, and P. Dupuis, *Minimax optimal control of stochastic uncertain systems with relative entropy constraints*, IEEE Trans. Automat. Control, 45 (2000), pp. 398–412.
[41] P. D. Pra, L. Meneghini, and W. J. Runggaldier, *Connections between stochastic control and dynamic games*, Math. Control Signals Systems, 9 (1996), pp. 303–326.
[42] H. L. Royden, *Real Analysis*, Macmillan, New York, 1968.
[43] N. Saldi, T. Başar, and M. Raginsky, *Markov–Nash equilibria in mean-field games with discounted cost*, SIAM J. Control Optim., 56 (2018), pp. 4256–4287.
[44] N. Saldi, T. Linder, and S. Yuksel, *Finite Approximations in Discrete-Time Stochastic Control: Quantized Models and Asymptotic Optimality.*, Birkhäuser, Boston, 2018.
[45] N. Saldi, S. Yüksel, and T. Linder, *Near optimality of quantized policies in stochastic control under weak continuity conditions*, J. Math. Anal. Appl., 435 (2016), pp. 321–337.
[46] N. Saldi, S. Yüksel, and T. Linder, *Asymptotic optimality of finite approximations to Markov decision processes with Borel spaces*, Math. Oper. Res., 42 (2017), pp. 1–34.
[47] A. V. Savkin and I. R. Petersen, *Robust control of uncertain systems with structured uncertainty*, J. Math. Syst. Estimation Control, 6 (1996), pp. 1–14.
[48] M. Schäl, *Conditions for optimality in dynamic programming and for the limit of n-stage optimal policies to be optimal*, Z. Wahr., 32 (1975), pp. 179–296.
[49] R. Serfozo, *Convergence of Lebesgue integrals with varying measures*, Sankhya A, 41 (1982), pp. 380–402.
[50] H. Sun and H. Xu, *Convergence analysis for distributionally robust optimization and equilibrium problems*, Math. Oper. Res., 41 (2015), pp. 377–401.
[51] I. Tzortzis, C. Charalambous, and T. Charalambous, *Dynamic programming subject to total variation distance ambiguity*, SIAM J. Control Optim., 53 (2015), pp. 2040–2075.
[52] V. A. Ugrinovskii, *Robust H-infinity control in the presence of stochastic uncertainty*, Internat. J. Control, 71 (1998), pp. 219–237.
[53] C. Villani, *Optimal Transport: Old and New*, Springer, New York, 2008.
[54] R. L. Wheeden and A. Zygmund, *Measure and Integral*, Marcel Dekker, New York, 1977.
[55] W. Wiesemann, D. Kuhn, and B. Rustem, *Robust Markov decision processes*, Math. Oper. Res., 38 (2012), pp. 153–183.

Downloaded 05/17/20 to 130.15.241.167. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

[56] H. Xu and S. Mannor, *Distributionally robust Markov decision processes*, in Proceedings of Advances in Neural Information Processing Systems 23, J. D. Lafferty, C. K. I. Williams, J. Shawe-Taylor, R. S. Zemel, and A. Culotta, eds., Curran Associates, Red Hook, NY, 2010, pp. 2505–2513, also available online from http://papers.nips.cc/paper/3927-distributionally-robust-markov-decision-processes.pdf.

[57] S. Yüksel and T. Linder, *Optimization and convergence of observation channels in stochastic control*, SIAM J. Control Optim., 50 (2012), pp. 864–887.

[58] K. Zhou, J. C. Doyle, and K. Glover, *Robust and Optimal Control*, Prentice-Hall, Englewood Cliffs, NJ, 1996.

Downloaded 05/17/20 to 130.15.241.167. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.