Risk-Aware Model Predictive Control Enabled by Bayesian Learning

Yingke Li, Yifan Lin, Enlu Zhou and Fumin Zhang

Abstract— **The performance of a model predictive controller depends on the accuracy of the objective and prediction model of the system. Although significant efforts have been dedicated to improving the robustness of model predictive control (MPC), they typically do not take a risk-averse perspective. In this paper, we propose a risk-aware MPC framework, which estimates the underlying parameter distribution using online Bayesian learning and derives a risk-aware control policy by reformulating classical MPC problems as Bayesian Risk Optimization (BRO) problems. The consistency of the Bayesian estimator and the convergence of the control policy are rigorously proved. Furthermore, we investigate the consistency requirement and propose a risk monitoring mechanism to guarantee the satisfaction of the consistency requirement. Simulation results demonstrate the effectiveness of the proposed approach.**

I. Introduction

For general, complex, and safety-critical control problems, model predictive control (MPC) [1] techniques have shown significant impact on both industrial and research-driven applications. Driven by the advances in the field of machine learning, many learning-based techniques have been developed to facilitate the MPC controller design [2], [3]. Bayesian MPC [4], [5] has been proposed to utilize reinforcement learning techniques for offline or episodic learning tasks. However, for online, single-execution learning tasks, few approaches provide principled mechanisms to monitor the risk during the online learning process and guarantee the convergence of the MPC controller.

In this paper, we propose a risk-aware MPC framework for online, single-execution learning tasks based on Bayesian learning. We estimate the posterior distribution of the unknown parameters according to the sequentially collected data based on online Bayesian learning algorithms. Taking advantage of the estimated posterior distribution of the unknown parameters, we propose to take the distributional uncertainty into consideration and enhance classical MPC problems by reformulating them as Bayesian Risk Optimization (BRO) problems. BRO [6], [7] is a stochastic optimization framework dealing with parametric uncertainty in the underlying distribution, which optimizes a risk functional applied to the posterior distribution of the unknown distributional parameter. The risk-aware control policy is derived by considering the worst-case and coherent risk measures (see [8] for introduction to coherent risk measures) in BRO. We design a Bayesian risk-aware MPC algorithm based on sequential Monte Carlo sampling (particle filter) [9] of the parameter distribution, which is practically feasible.

The consistency of Bayesian estimator with i.i.d. data has been well studied [10]. However, online Bayesian MPC deals with conditionally independent data. In our previous work [11], we have proved Bayesian consistency under conditionally independent observations, where the measurement distribution is conditionally independent with respect to the state. In this paper, we prove Bayesian consistency under conditionally independent transitions, where the transition kernel is conditionally independent with respect to the state and action. We also extend the consistency proof from finite parameter space in our previous work [11] to infinitely countable parameter space. We further prove the convergence of the risk-aware control policy based on the consistency of Bayesian estimator.

We investigate the consistency requirement for Bayesian estimation and propose a risk monitoring mechanism to guarantee the satisfaction of the consistency requirement. The risk monitoring mechanism utilizes the credible interval of the parameter distribution as an indicator of risk, and adds extra excitation into the control policy when the risk cannot be efficiently decreased. Simulation results are presented to explain the consistency requirement intuitively and justify the effectiveness of the risk monitoring mechanism.

Finally, we reveal some connections between the consistency requirement of Bayesian estimation and the persistent excitation condition of adaptive control. We make a comparison with some adaptive control methodologies, such as Model Reference Adaptive Control [12], and Concurrent Learning Adaptive Control [13], and explore the underlying similarity between those conditions.

II. Problem Formulation

Consider a nonlinear system with unknown parameter:

$$x_{k+1} = f(x_k, u_k, \theta^*) + w_k,$$

where $x_k \in \mathcal{X}^n$ is the state at time k, $u_k \in \mathcal{U}^p$ is the control input at time k, $\theta^* \in \Theta^q$ is the unknown parameter, and $w_k \in \mathcal{X}^n$ is the zero-mean independent and identically distributed (i.i.d.) noise.

For any feedback control policy $\psi = (\psi^{(0)}, \dots, \psi^{(T-1)})$ where each $\psi^{(k)} : \mathcal{X}^n \to \mathcal{U}^p$ maps from the state x_k to the control input u_k, its performance can be characterized by a

Yingke Li and Fumin Zhang are with School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA, USA yli3225,fumin@gatech.edu

Yifan Lin and Enlu Zhou are with School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA, USA ylin429,ezhou30@gatech.edu

The research work is supported by ONR grants N00014-19-1-2556 and N00014-19-1-2266; AFOSR grant FA9550-19-1-0283; NSF grants CNS-1828678, S&AS-1849228, DMS-2053489 and GCR-1934836; NOAA grant NA16NOS0120028; and NSF grant DMS-2053489.

978-1-6654-5197-0/$31.00 ©2022 AACC

finite-horizon cost

$$J_{\theta^*}^{\psi} = \frac{1}{T}\left[\sum_{k=0}^{T-1} c(x_k,u_k) + c_T(x_T)\right]$$
$$\text{s.t.}\quad u_k = \psi^{(k)}(x_k),$$
$$x_{k+1} = f(x_k,u_k,\theta^*) + w_k,$$

where $c(x_k,u_k)$ is the stage cost function for state x_k and control input u_k, and $c_T(x_T)$ is the terminal cost function.

Let W be the joint probability distribution of the random noise process $[w_0, w_1, \cdots, w_{T-1}]$, then the stochastic MPC problem is defined as

$$\min_{\psi} \mathbb{E}_W[J_{\theta^*}^{\psi}]. \tag{1}$$

Due to the existence of the unknown parameter θ^* in the nonlinear system, simultaneous identification and control are needed to solve this problem. Taking the Bayesian estimation approach, the unknown parameter θ^* is estimated by a random vector $\theta \in \Theta^q$ which obeys $\pi : \Theta^q \to [0,1]$ such that $\pi(\theta^*)$ is the probability that θ is the true parameter θ^*.

Viewing the unknown parameter θ^* as a random vector introduces extra parameter uncertainty into this stochastic MPC problem. To explore the middle ground between optimistically ignoring the distributional uncertainty of the parameter and pessimistically fixating on the worst-case scenario, we take the distribution of the unknown parameter $\pi(\theta)$ into consideration and reformulate this stochastic MPC problem using the BRO framework:

$$\min_{\psi} \mathcal{R}_{\pi(\theta)}\left[\mathbb{E}_W[J_{\theta}^{\psi}]\right], \tag{2}$$

where $\mathcal{R}_{\pi(\theta)}$ is a risk functional taken with respect to $\pi(\theta)$, which accounts for the uncertainty in the estimation of the unknown parameter θ^*, for example, mean-variance, Value at Risk (VaR), and Conditional Value at Risk (CVaR) [6].

Remark. Different from the nested formulation in general multi-stage optimization problems (see [14]), where the risk functional is taken for each stage, here we only take the risk functional for the whole horizon. The reason that we can simplify this problem and reduce the computational complexity is owed to the receding horizon property of MPC. MPC uses current estimated model to optimize the finite-horizon performance but only the first stage action is executed. Then new observations will be used to refine the estimation. We take advantage of this property to avoid the update of estimation at each stage in the nested formulation.

We propose to utilize Bayesian learning to estimate the distribution of the unknown parameter $\pi(\theta)$ online and design a risk-aware control policy based on the reformulated stochastic MPC problem.

III. Bayesian Risk-Aware MPC

A. Bayesian Update

Let $x_{0:k}$ be the states from time 0 to k, and $u_{0:k}$ be the corresponding control inputs from time 0 to k. Since the system is Markovian,

$$\Pr(x_k|\theta, x_{0:k-1}, u_{0:k-1}) = \Pr(x_k|\theta, x_{k-1}, u_{k-1}).$$

We define $\pi_k(\theta) = \Pr(\theta|x_{0:k}, u_{0:k})$ as the posterior distribution of θ at time k. Therefore,

$$\pi_k(\theta) = \Pr(\theta|x_{0:k}, u_{0:k}) = \Pr(\theta|x_{0:k}, u_{0:k-1}),$$

followed by the fact that the control input u_k does not affect the information on θ until an observation of the new state x_{k+1} is taken.

According to the Bayesian rule,

$$\begin{aligned}
&\Pr(\theta|x_{0:k}, u_{0:k}) = \Pr(\theta|x_{0:k}, u_{0:k-1})\\
=&\frac{\Pr(\theta, x_k|x_{0:k-1}, u_{0:k-1})}{\Pr(x_k|x_{0:k-1}, u_{0:k-1})}\\
=&\frac{\Pr(x_k|\theta, x_{0:k-1}, u_{0:k-1})\Pr(\theta|x_{0:k-1}, u_{0:k-1})}{\Pr(x_k|x_{0:k-1}, u_{0:k-1})}\\
=&\frac{\Pr(x_k|\theta, x_{k-1}, u_{k-1})\Pr(\theta|x_{0:k-1}, u_{0:k-1})}{\int \Pr(x_k|\theta, x_{k-1}, u_{k-1})\Pr(\theta|x_{0:k-1}, u_{0:k-1})d\theta}.
\end{aligned}$$

We define $q(\cdot)$ as the probability density function that satisfies $\int_A q(x_k;\theta,x_{k-1},u_{k-1})dx_k = \Pr(A|\theta, x_{k-1}, u_{k-1})$, where A is an arbitrary measurable set. Then $q(\cdot)$ is the transition kernel of the nonlinear system, which is determined by the transition function $f(\cdot)$ and noise w_{k-1}. Therefore, the posterior distribution of θ is updated as

$$\pi_k(\theta) = \frac{q(x_k;\theta,x_{k-1},u_{k-1})}{\int q(x_k;\theta,x_{k-1},u_{k-1})\pi_{k-1}(\theta)d\theta}\pi_{k-1}(\theta). \tag{3}$$

For the majority of the systems in control field, the noise is assumed to be Gaussian white noise, which naturally satisfies the following assumption.

Assumption III.1. *The transition kernel $q(x';\theta,x,u)$ is continuously differentiable in x' and has bounded first order derivative in x'.*

With the above assumption, we can easily show that the transition kernel $q(x';\theta,x,u)$ is bounded, then the integration in (3) is finite since $\int \pi(\theta)d\theta = 1$. Thus the posterior distribution is well-defined.

Since the parameter space Θ is continuous, the posterior distribution of θ in general does not have an analytical form. In practice, approximations are necessary. Therefore, we utilize some sampling-based approaches, such as particle filter, to approximate the Bayesian update in (3). Let there be N_s equally-weighted particles approximating the distribution $\pi_{k-1}(\theta)$. We update the weights based on the transition kernel $q(x_k;\theta,x_{k-1},u_{k-1})$, and resample with those updated weights to obtain N_s new equally-weighted particles to approximate the distribution $\pi_k(\theta)$.

B. Risk-Aware Control Policy

Based on the posterior distribution $\pi_k(\theta)$, we can design a risk-aware control policy according to the reformulated stochastic MPC problem.

If the risk functional is chosen as the worst-case measure, VaR, or a coherent risk measure such as CVaR, (2) can be rewritten as a Distributionally Robust Optimization (DRO) [15] problem with an appropriately chosen ambiguity set $\mathcal{D}$:

$$\min_{\psi} \max_{\theta \in \mathcal{D}} \mathbb{E}_W[J_{\theta}^{\psi}]. \tag{4}$$

The dual of the minimax problem (4) is obtained by interchanging the 'min' and 'max' operators, which is

$$\max_{\theta\in\mathcal{D}}\min_{\psi}\mathbb{E}_W[J_\theta^\psi]. \tag{5}$$

The optimal value of dual problem (5) is always less than or equal to the optimal value of the primal problem. And under certain regularity conditions, the optimal values of problems (4) and (5) are equal to each other.

Assumption III.2. *The state space $\mathcal{X}$, input space $\mathcal{U}$ and parameter space Θ are compact. The objective function $\mathbb{E}_W[J_\theta^\psi]$ is convex in ψ and concave in θ.*

With the above assumption, it is possible to establish the desirable no duality gap property according to *Sion's Minimax Theorem* [16]. For detailed derivation of the minimax theorem applied to multistage stochastic programming, Markov Decision Processes or stochastic control, we refer the readers to [17], [18] and [19]. Therefore, the solution of problem (2) is equivalent to the solution of problem (5).

We consider to utilize *Credible Interval* to construct the ambiguity set $\mathcal{D}$. The *Credible Interval* (CI) $\mathcal{C}^\gamma$ of a posterior probability distribution is a continuous subset within the space of a random vector. The probability that the value of the random vector falls with that subset is $\gamma \in [0,1]$, i.e., $\Pr(\theta \in \mathcal{C}^\gamma) = \gamma$. In practice, we can take the $(1-\gamma)/2$ and $(1+\gamma)/2$ quantile of $\pi(\theta)$ as the lower and upper end of $\mathcal{C}^\gamma$. More specifically, suppose we rank the samples $\theta_1, \cdots, \theta_{N_s}$ from $\pi(\theta)$ in ascending order. The lower end of $\mathcal{C}^\gamma$ is $\theta_{\lceil(1-\gamma)N_s/2\rceil}$, and the upper end of $\mathcal{C}^\gamma$ is $\theta_{\lceil(1+\gamma)N_s/2\rceil}$.

Now we can utilize $\mathcal{C}^\gamma$ to construct the ambiguity set $\mathcal{D}$. For example, for worst-case measure we could set $\mathcal{D} = \mathcal{C}^1$; for VaR with α risk level, we could set $\mathcal{D} = \mathcal{C}^\alpha$. For coherent risk measures, the construction of the ambiguity set is more involved; we refer the readers to [20] for the equivalence between optimizing a coherent risk measure and constructing the corresponding DRO.

Our control policy is chosen to minimize the worst case cost over all the parameters θ in the ambiguity set $\mathcal{D}$. For each $\theta \in \mathcal{D}$, we first solve a stochastic MPC problem: $\min_\psi \mathbb{E}_W[J_\theta^\psi]$. There are many principled approaches, such as scenario tree search and dynamic programming, that can find the optimal solution ψ_θ^* for each θ. Then we find the value θ^r that gives the worst performance among all $\theta \in \mathcal{D}$. Our *risk-aware control policy* ψ^r is then selected as

$$\psi^r = \psi_{\theta^r}^*. \tag{6}$$

Since MPC has receding horizon, we only apply the first stage control policy as the control input $u_k = (\psi^r)^{(0)}(x_k)$.

It is worth noticing that this risk-averse approach naturally incorporates randomness and robustness into the designed control policy, which provides both excitation required for system parameter identification and robustness for system performance. Different from typical MPC control laws which additionally add randomness as an "exploration" term into the estimated "exploitation" term, our proposed approach can internally leverage exploration and exploitation. The Bayesian posterior distribution quantifies the uncertainty in the estimation of the unknown parameter. The policy starts with relatively large randomness to enable more exploration of the system parameter space, and gradually decreases the randomness as the parameter estimation becomes more accurate.

The introduction of CI also gives us the ability to monitor the risk of the control policy online. Let $\mathcal{C}^\gamma$ be the set $\{\theta \in \Theta^q : \theta_{\lceil(1-\gamma)N/2\rceil} \le \theta \le \theta_{\lceil(1+\gamma)N/2\rceil}\}$. Then its volume $|\mathcal{C}^\gamma| = \prod_{i=1}^q |\theta^i_{\lceil(1+\gamma)N/2\rceil} - \theta^i_{\lceil(1-\gamma)N/2\rceil}|$ (θ^i represents the ith dimension of θ), can be used as an indicator of how concentrated the posterior distribution is on θ^*. This also provides us the potential to design a risk monitoring mechanism to ensure Bayesian consistency based on $|\mathcal{C}^\gamma|$, which will be shown in Section IV-C.

We design a Bayesian risk-aware MPC algorithm based on particle filter, which is presented as Algorithm 1.

Algorithm 1: Bayesian Risk-Aware MPC

```
Initialize prior $\pi_0(\theta)$ with uniform probability
  distribution over the parameter space $\Theta$
Create $N_s$ particles $S_0^{1:N_s}$ with equal weights $\frac{1}{N_s}$,
  which samples from $\pi_0(\theta)$, i.e. $S_0^i \sim \pi_0(\theta)$
Initialize state $x_0$, input $u_0 = 0$, iterator $k = 1$
while $|\mathcal{C}^\gamma| > \epsilon$ do
    Take an observation of $x_k$
    Compute the weights
      $W_k^i \propto q(x_k; S_{k-1}^i, x_{k-1}, u_{k-1})$
    Resample $\{W_k^i, S_{k-1}^i\}$ to obtain $N_s$ new
      equally-weighted particles $\{\frac{1}{N_s}, S_k^i\}$
    Take the $(1-\gamma)/2$ and $(1+\gamma)/2$ quantile of the
      emipirical distribution (particles) to form $\mathcal{C}^\gamma$
    Calculate the control policy according to (6)
    Decide the control input $u_k = (\psi^r)^{(0)}(x_k)$
    $k := k + 1$
end
```

IV. Convergence Analysis

Due to technical challenges of analyzing the continuous parameter space, we analyze the consistency and convergence of our proposed approach by assuming that the parameter space Θ is discrete but consists of infinite number of candidates, which can approximate the continuous parameter space with arbitrary precision. We also assume that the approximation error of particle filter is negligible.

A. Consistency of the Bayesian Estimator

Definition IV.1. *The Bayesian estimator is* (strongly) consistent *if $\pi_k(\theta)$, the posterior distribution of θ, converges to the degenerated distribution $\delta_{\theta^*}(\theta)$ that concentrates on the true parameter value θ^*, with probability 1.*

To prove the consistency of the Bayesian estimator, we first prove Lemma IV.1 based on Assumption IV.1.

Assumption IV.1. *The prior distribution $\pi_0(\theta)$ has non-zero probability at θ^*.*

Lemma IV.1. *The marginal transition kernel* $\hat{q}(x_k; x_{k-1}, u_{k-1}) = \sum_\theta \pi_{k-1}(\theta) q(x_k; \theta, x_{k-1}, u_{k-1})$ *converges to the true transition kernel* $q(x_k; \theta^*, x_{k-1}, u_{k-1})$, *i.e.,* $\lim_{k\to\infty} \hat{q}(x_k; x_{k-1}, u_{k-1}) = q(x_k; \theta^*, x_{k-1}, u_{k-1})$, *with probability 1.*

Proof. Let $\mathcal{F}_k = \sigma\{(x_{s-1}, x_s), s \leq k\}$ be the $\sigma-$filtration generated by the past observed states. According to (3), the estimated probability of the true parameter satisfies the following equation,

$$\log \pi_k(\theta^*) = \log \pi_{k-1}(\theta^*) + \log \frac{q(x_k; \theta^*, x_{k-1}, u_{k-1})}{\hat{q}(x_k; x_{k-1}, u_{k-1})}.$$

Taking expectation on both sides, we have that

$$\begin{aligned}
&\mathbb{E}[\log \pi_k(\theta^*)] \\
=&\mathbb{E}[\log \pi_{k-1}(\theta^*)] + \mathbb{E}\left[\log \frac{q(x_k; \theta^*, x_{k-1}, u_{k-1})}{\hat{q}(x_k; x_{k-1}, u_{k-1})}\right] \\
=&\mathbb{E}[\log \pi_{k-1}(\theta^*)] \\
&+ \mathbb{E}\left[\mathbb{E}\left[\log \frac{q(x_k; \theta^*, x_{k-1}, u_{k-1})}{\hat{q}(x_k; x_{k-1}, u_{k-1})} | x_{k-1}, u_{k-1}, \mathcal{F}_{k-1}\right]\right] \\
=&\mathbb{E}[\log \pi_{k-1}(\theta^*)] + \mathbb{E}[d_{k-1}]
\end{aligned}$$

where $d_{k-1} = DL(q(x_k; \theta^*, x_{k-1}, u_{k-1}) || \hat{q}(x_k; x_{k-1}, u_{k-1}))$ is the *relative entropy* (*Kullback–Leibler divergence*) [21] between $q(x_k; \theta^*, x_{k-1}, u_{k-1})$ and $\hat{q}(x_k; x_{k-1}, u_{k-1})$. Then the expectation of d_{k-1} can be represented as follows,

$$\mathbb{E}[d_{k-1}] = \mathbb{E}[\log \pi_k(\theta^*)] - \mathbb{E}[\log \pi_{k-1}(\theta^*)].$$

For any n, taking summation over k from 1 to n on both sides, we have

$$\sum_{k=1}^{n} \mathbb{E}[d_{k-1}] = \mathbb{E}[\log \pi_n(\theta^*)] - \log \pi_0(\theta^*) \leq -\log \pi_0(\theta^*) < \infty,$$

where the last inequality holds according to Assumption IV.1.

Therefore, take $n \to \infty$ and we get

$$\sum_{k=1}^{\infty} \mathbb{E}[d_{k-1}] \leq -\log \pi_0(\theta^*) < \infty.$$

By *Markov Inequality*, we know that for any $\epsilon > 0$,

$$\sum_{k=0}^{\infty} \Pr[d_k \geq \epsilon] \leq \frac{1}{\epsilon} \sum_{k=0}^{\infty} \mathbb{E}[d_k] < \infty.$$

We can then apply *Borel-Cantelli Lemma* and show that $P(d_k \geq \epsilon, i.o.) = 0$, which further implies $\lim_{k\to\infty} d_k = 0$, with probability 1.

Moreover, since $d_k \geq 0$, by *Tonelli's Theorem*, we have

$$\mathbb{E}\left[\sum_{k=0}^{\infty} d_k\right] = \sum_{k=0}^{\infty} \mathbb{E}[d_k] \leq -\log \pi_0(\theta^*).$$

Since $\sum_{k=0}^{\infty} d_k$ has bounded expectation, it must be finite with probability 1.

Note that the *total variation distance* between two distributions is related to the *relative entropy* by *Pinsker's Inequality*:

$$||q(x_k; \theta^*, x_{k-1}, u_{k-1}) - \hat{q}(x_k; x_{k-1}, u_{k-1})||_{TV} \leq \sqrt{2d_{k-1}},$$

where

$$\begin{aligned}
&||q(x_k; \theta^*, x_{k-1}, u_{k-1}) - \hat{q}(x_k; x_{k-1}, u_{k-1})||_{TV} = \\
&\sup_{x_k} |q(x_k; \theta^*, x_{k-1}, u_{k-1}) - \hat{q}(x_k; x_{k-1}, u_{k-1})|.
\end{aligned}$$

Letting $k \to \infty$, by the convergence of d_k, we have

$$\lim_{k\to\infty} \int_{x_k} |q(x_k; \theta^*, x_{k-1}, u_{k-1}) - \hat{q}(x_k; x_{k-1}, u_{k-1})| dx_k = 0,$$

with probability 1.

According to *Dominated Convergence Theorem*, we further have

$$\int_{x_k} \lim_{k\to\infty} |q(x_k; \theta^*, x_{k-1}, u_{k-1}) - \hat{q}(x_k; x_{k-1}, u_{k-1})| dx_k = 0.$$

Moreover, since

$$|q(x_k; \theta^*, x_{k-1}, u_{k-1}) - \hat{q}(x_k; x_{k-1}, u_{k-1})| \geq 0$$

and $q(x_k; \theta^*, x_{k-1}, u_{k-1}) - \hat{q}(x_k; x_{k-1}, u_{k-1})$ is continuous in x_k, then for any x_k,

$$\lim_{k\to\infty} |q(x_k; \theta^*, x_{k-1}, u_{k-1}) - \hat{q}(x_k; x_{k-1}, u_{k-1})| = 0,$$

which means

$$\lim_{k\to\infty} \hat{q}(x_k; x_{k-1}, u_{k-1}) = q(x_k; \theta^*, x_{k-1}, u_{k-1}), \quad (7)$$

with probability 1. □

Now we prove that the posterior distribution of θ converges to the distribution $\delta_{\theta^*}(\theta)$ based on Assumption IV.2.

Assumption IV.2. *For any* (x_k, x_{k-1}, u_{k-1}) *and* $K \subseteq \mathbb{N}$, $\{q(x_k; \theta_i, x_{k-1}, u_{k-1})\}_{i\in K}$ *are linearly independent, i.e., if*

$$\sum_{i\in K} c_i q(x_k; \theta_i, x_{k-1}, u_{k-1}) = 0$$

holds for all (x_k, x_{k-1}, u_{k-1}), *then* $c_i = 0$ *for all* $i \in K$.

Theorem IV.2. *The posterior distribution of* θ *converges to the distribution* $\delta_{\theta^*}(\theta)$, *i.e.* $\lim_{k\to\infty} \pi_k(\theta) = \delta_{\theta^*}(\theta)$, *with probability 1.*

Proof. Note that

$$\begin{aligned}
&q(x_k; \theta^*, x_{k-1}, u_{k-1}) - \hat{q}(x_k; x_{k-1}, u_{k-1}) \\
=&[1 - \pi_{k-1}(\theta^*)] q(x_k; \theta^*, x_{k-1}, u_{k-1}) \\
&- \sum_{\theta\neq\theta^*} \pi_{k-1}(\theta) q(x_k; \theta, x_{k-1}, u_{k-1}).
\end{aligned} \quad (8)$$

Note that for any $t > 0$, $(\pi_t(\theta_1), \pi_t(\theta_2), \cdots)$ is infinitely dimensional bounded vector with all components sum up to 1, we can take a subsequence $\{\pi_{t_k}\}$ such that for each component j, $\{\pi_{t_k}(\theta_j)\}$ converges to a limit which is denoted by $\pi_\infty(\theta_j)$, which is also known as weak convergence (of a deterministic sequence).

Next, we will show that $\pi_\infty(\theta)$ is a normalized vector. Note that for any $j \in \mathbb{N}$, $\lim_{t_k\to\infty} \pi_{t_k}(\theta_j) = \pi_\infty(\theta_j)$, which is equivalent to

$$\forall \epsilon_j > 0, \exists N \in \mathbb{N}, s.t. \forall n \geq N, |\pi_\infty(\theta_j) - \pi_n(\theta_j)| \leq \epsilon.$$

Therefore, we have

$$-\epsilon_j < \pi_\infty(\theta_j) - \pi_n(\theta_j) < \epsilon_j, j = 1, 2, \cdots \tag{9}$$

According to the Bayesian update rule, we know $\sum_{j=1}^{\infty} \pi_n(\theta_j) = 1$. It then follows that $\forall \epsilon > 0$, take $\epsilon_j = \frac{\epsilon}{2^j}$ and sum over (9) for all $j \in \mathbb{N}$, we get

$$-(\frac{\epsilon}{2^1} + \frac{\epsilon}{2^2} + \cdots) < \sum_{j=1}^{\infty} \pi_\infty(\theta_j) - 1 < (\frac{\epsilon}{2^1} + \frac{\epsilon}{2^2} + \cdots),$$

which indicates $\forall \epsilon > 0$, $|\sum_{j=1}^{\infty} \pi_\infty(\theta_j) - 1| < \epsilon$, and it implies that $\sum_{j=1}^{\infty} \pi_\infty(\theta_j) = 1$. So the limit is also a valid probability simplex.

Since every weakly convergent sequence in L^1 is strongly convergent (cf. Chapter 2 in [22]), we can take any convergent subsequence $\{\pi_{t_k}\}$ with limit $(p_1^*, p_2^*, \cdots)$.

Since $\mathcal{X}$ and $\mathcal{U}$ are also compact, from this subsequence, we could take a further subsequence $\{\pi_{\tau_k}\}$ such that $\{x_{\tau_k}\}$ converges to x_∞, and u_{τ_k} converges to u_∞. Then take limit over (8) along $\{\tau_1, \tau_2, \cdots\}$, and by (7), we have

$$[1 - \pi_\infty(\theta^*)]q(\cdot; \theta^*, x_\infty, u_\infty) - \sum_{\theta \neq \theta^*} \pi_\infty(\theta) q(\cdot; \theta, x_\infty, u_\infty) = 0,$$

with probability 1.

According to Assumption IV.2, for any convergent subsequence, $1 - \pi_\infty(\theta^*) = 0, \pi_\infty(\theta) = 0 \;\; \forall \theta \neq \theta^*$, which further implies

$$\lim_{k \to \infty} \pi_k(\theta) = \delta_{\theta^*}(\theta),$$

with probability 1. □

Therefore, the strong consistency of the Bayesian estimator is proved.

Remark. Note that the above strong consistency of the Bayesian estimator indicates that for almost every sample path of observations, the posterior distribution $\pi_k(\theta)$ converges to δ_{θ^*}. Even though the Bayesian consistency with non i.i.d. data has been studied in [23], the assumptions in these general results are often abstract (such as existence of testing function sequence) and hard to verify in practice. On the other hand, our Bayesian consistency result is built on assumptions (in particular Assumption IV.2) that are easy to verify and have a nice interpretation, which will be discussed in details in Section IV-C.

B. Convergence of the Control Policy

Based on the consistency guarantee of Theorem IV.2, we can prove the convergence of our proposed risk-aware control policy.

Theorem IV.3. *The risk-aware control policy converges to the true optimal control policy, i.e.* $\lim_{k \to \infty} \psi_k^r = \psi_{\theta^*}^*$*, with probability 1.*

Proof. Since $\lim_{k \to \infty} \pi_k(\theta) = \delta_{\theta^*}(\theta)$ w.p. 1 according to Theorem IV.2, for any $\epsilon > 0$ and $M > 0$, there must exist $K > 0$, such that for $k > K$, we have

$$|\pi_k(\theta^*) - \delta_{\theta^*}(\theta^*)| \leq \frac{\epsilon}{M}, \text{w.p. } 1.$$

Then we have $1 - \frac{\epsilon}{M} \leq \pi_k(\theta^*) \leq 1$. For any confidence level $0 < \gamma < 1$, we can find a small enough $\epsilon > 0$ such that $\gamma < 1 - \frac{\epsilon}{M}$. Therefore, the CI in Algorithm 1 only contains θ^*, i.e. $\mathcal{C}^\gamma = \{\theta^*\}$. According to control policy (6),

$$\psi^r = \psi_{\theta^r}^* = \psi_{\theta^*}^*.$$

Therefore, $\lim_{k \to \infty} \psi_k^r = \psi_{\theta^*}^*$, with probability 1. □

Remark. The stability of the closed-loop system is closely related with the stability of the nominal MPC problem. The analysis of the stability properties of the closed-loop system is beyond the scope of this paper and will appear in our later works.

C. Consistency Requirement

The hidden consistency requirement in Assumption IV.2 can be interpreted as: (1) The system is distinguishable, i.e., the candidate parameters θ_i must be distinguishable to ensure that the transition kernels $q(\cdot; \theta_i, x, u)$ are linearly independent for the entire space of (x, u). (2) The choice of x_k and u_k is informative, i.e., the data space (x_k, u_k) expands enough to ensure that the transition kernels $q(\cdot; \theta_i, x, u)$ are linearly independent for the subspace (x_k, u_k).

We propose a risk monitoring mechanism to guarantee the satisfaction of the consistency requirement. As mentioned in Section III-B, we use the volume of CI $|\mathcal{C}^\gamma|$ to indicate the risk of parameter estimation. When we monitor that $|\mathcal{C}^\gamma|$ does not decrease for several iterations, we may infer that the algorithm gets "stuck" in some local region. Therefore, the algorithm needs more "excitation" to leave that region and explore more informative data.

Note that for a general continuous non-constant and non-periodic function, the values of the function are often distinguishable if the distance of the variables is large enough. Thus, we can choose to add an extra noise, which can be chosen as a Gaussian white noise, into the calculated control input to provide enough "excitation".

Therefore when the risk cannot be sufficiently decreased, we utilize a risk monitoring mechanism to replace the control input in Algorithm 1 as $u_k = (\psi^r)^{(0)}(x_k) + v_k$, where v_k is an i.i.d. Gaussian white noise.

V. Simulation Results and Discussion

Consider the following system:

$$x_{k+1} = \cos(\theta^*)x_k + \sin(\theta^*)u_k + w_k,$$

where $\theta^* = \frac{\pi}{4}$ is the unknown system parameter, and w_k is i.i.d Gaussian white noise with variance 1.

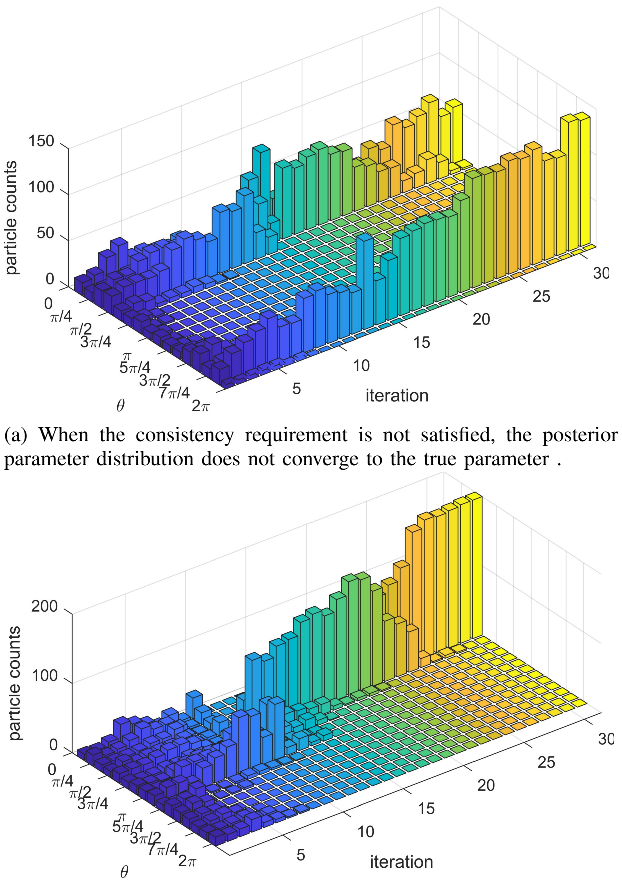
Assuming the control objective is to make the system state reach a predefined location a, consider a one-step MPC problem, where the cost function is chosen as

$$\text{VaR}_{\pi(\theta)}\left[\mathbb{E}_W[J_\theta^\psi]\right] = \text{VaR}_{\pi(\theta)}\left[(\cos(\theta)x + \sin(\theta)u - a)^2 + u^2\right].$$

We solve this control problem using Algorithm 1. We choose 16 candidates of θ ranging from 0 to 2π with equal interval, the number of particles $N_s = 200$ and $\gamma = 0.8$.

Fig.1(a) shows a case when the posterior parameter distribution does not converge to the true parameter. The reason of this failure is due to the fact that the consistency requirement is not satisfied. The control input u_k lies very closely to 0, which provides little excitation to the term $\sin(\theta)$ of the system. θ is not distinguishable if only the evaluation of the term $\cos(\theta)$ is provided.

Fig.1(b) shows that the posterior parameter distribution converges to the true parameter given that the consistency requirement is satisfied. The risk monitor triggers additional excitation to the control input when $|\mathcal{C}^\gamma|$ remains large for several iterations, which ensures the satisfaction of the consistency requirement.



(a) When the consistency requirement is not satisfied, the posterior parameter distribution does not converge to the true parameter .

(b) When the consistency requirement is satisfied, the posterior parameter distribution converges to the true parameter.

Fig. 1: The influence of the consistency requirement to the convergence of the posterior parameter distribution.

We compare the consistency requirement of Bayesian risk-aware MPC with some adaptive control methodologies, such as Model Reference Adaptive Control (MRAC), and Concurrent Learning Adaptive Control (CLAC), and explore the underlying similarity between those conditions. MRAC adaptive laws can guarantee parameter consistency if and only if the plant states are Persistently Exciting (PE) [12]. For CLAC, a verifiable condition on the linear independence of the recorded data is sufficient to guarantee parameter consistency [13]. Those conditions focus on either system state or recorded data. Different from those approaches, Bayesian risk-aware MPC requires that the transition kernels for different parameters, which are determined by the transition function and noise, are linearly independent in the data space.

Despite different perspectives, the hidden explanation of all those conditions is that, the collected data for identification should be expanded enough toward the parameter space to provide “sufficient exploration” of the parameter space.

References

[1] Manfred Morari and Jay H. Lee. Model predictive control: Past, present and future. *Computers and Chemical Engineering*, 23(4-5):667–682, 1999.

[2] Ali Mesbah. Stochastic model predictive control with active uncertainty learning: A Survey on dual control. *Annual Reviews in Control*, 45:107–117, 2018.

[3] Lukas Hewing, Kim P. Wabersich, Marcel Menner, and Melanie N. Zeilinger. Learning-Based Model Predictive Control: Toward Safe Learning in Control. *Annual Review of Control, Robotics, and Autonomous Systems*, 3(1):269–296, 2020.

[4] Kim P. Wabersich and Melanie N. Zeilinger. Bayesian model predictive control: Efficient model exploration and regret bounds using posterior sampling. *arXiv*, 120:1–10, 2020.

[5] Kim P. Wabersich and Melanie N. Zeilinger. Performance and safety of Bayesian model predictive control: Scalable model-based RL with guarantees. *arXiv*, pages 1–11, 2020.

[6] Di Wu, Helin Zhu, and Enlu Zhou. A Bayesian risk approach to data-driven stochastic optimization: Formulations and asymptotics. *SIAM Journal on Optimization*, 28(2):1588–1612, 2018.

[7] Sait Cakmak, Raul Astudillo Marban, Peter Frazier, and Enlu Zhou. Bayesian Optimization of Risk Measures. In *Advances in Neural Information Processing Systems*, pages 20130–20141, 2020.

[8] Alexander Shapiro, Darinka Dentcheva, and Andrzej Ruszczyński. *Lectures on Stochastic Programming: Modeling and Theory, Second Edition*. 2014.

[9] Arnaud Doucet and Adam M Johansen. A tutorial on particle filtering and smoothing: fifteen years later. *Handbook of Nonlinear Filtering*, (December):4–6, 2010.

[10] Persi Diaconis and David Freedman. On the Consistency of Bayes Estimates. *The Annals of Statistics*, 14(1):1–26, 1986.

[11] Yingke Li, Tianyi Liu, Enlu Zhou, and Fumin Zhang. Bayesian Learning Model Predictive Control for Process-Aware Source Seeking. *IEEE Control Systems Letters*, 6:692–697, 2022.

[12] Stephen Boyd and S. S. Sastry. Necessary and sufficient conditions for parameter convergence in adaptive control. *Automatica*, 22(6):629–639, 1986.

[13] Girish Chowdhary, Maximilian Mühlegg, and Eric Johnson. Exponential parameter and tracking error convergence guarantees for adaptive controllers without persistency of excitation. *International Journal of Control*, 87(8):1583–1603, 2014.

[14] Alexander Shapiro. Tutorial on risk neutral, distributionally robust and risk averse multistage stochastic programming. *European Journal of Operational Research*, 288(1):1–13, 2021.

[15] Hamed Rahimian and Sanjay Mehrotra. Distributionally Robust Optimization: A Review. 2019.

[16] Maurice Sion. On general minimax theorems. *Pacific Journal of Mathematics*, 8(1):171–176, 1958.

[17] Alexander Shapiro. Stochastic programming approach to optimization under uncertainty. *Math. Program., Ser. B*, 112:183–220, 2008.

[18] D. V. Lindley. Dynamic Programming and Decision Theory. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 10(1):39–51, 3 1961.

[19] Nicole Bäuerle and Ulrich Rieder. Markov Decision Processes under Ambiguity. *Banach Center Publications*, 122:25–39, 7 2019.

[20] Andrzej Ruszczyński and Alexander Shapiro. Optimization of convex risk functions. *Mathematics of Operations Research*, 31(3):433–452, 8 2006.

[21] Achim Klenke. *Probability Theory: A Comprehensive Course*. 2014.

[22] Gert K. Pedersen. *Analysis now*, volume 118. 2012.

[23] Subhashis Ghosal and Aad Van Der Vaart. Convergence rates of posterior distributions for noniid observations. *Annals of Statistics*, 35(1):192–223, 2 2007.