Chin. Ann. Math. Ser. B
47(5), 2026, 833–854
DOI: 10.1007/s11401-026-0027-6

Chinese Annals of
Mathematics, Series B

Optimal Control with Learning on the Fly from Finite to Infinite-Dimensional Systems*

Feifei MIAO[1]

Abstract The main aim of this paper is to extend the Bayesian approach to finding quadratic optimal control to a wider range of stochastic linear systems. These systems involve an unknown parameter in the drift term, which is observed through a noisy linear channel. The author demonstrates the effectiveness of the Bayesian strategy by comparing the cost it incurs with that of an optimal control that possesses complete knowledge of the parameter. The findings reveal that the Bayesian strategy minimizes the worst-case multiplicative regret. Furthermore, the author provides a proof that the corresponding adaptive scheme is optimal when the unknown system parameter belongs to an infinite space. This result further validates the effectiveness of the Bayesian strategy in handling systems with unknown parameters. In summary, this paper contributes to the generalization of the Bayesian strategy for the optimal control in stochastic linear systems with unknown parameters. It also establishes a theoretical basis for its optimality.

Keywords Bayesian strategy, Kalman-Bucy filter, Optimal control, Adaptive control, Multiplicative regret
2020 MR Subject Classification 93C40, 93C41, 93E11

1 Introduction

The study of optimal control problems for linear systems with quadratic cost functionals, also known as LQ problems, has a long history. In these problems, the system is described by a linear ordinary differential equation and the cost functional is quadratic, making it one of the simplest nontrivial optimal control problems. As a result, there have been extensive studies and excellent results on the topic of linear quadratic optimal control problems, particularly in the case where the model parameters are known and the state is fully observed (cf. [4, 24, 27]).

This paper focuses on the case where the model parameters are unknown. In this setting, the control actions not only affect the cost but also influence the rate at which the system dynamics are learned. This adds a layer of complexity to the problem. The objective now is to minimize the regret of the controller, which is defined as the difference between the cost of the learned controller and the cost of the optimal controller. The focus is on developing strategies that can adapt to unknown model parameters and learn the dynamics of the system over time. By minimizing the regret, the proposed adaptive control strategies aim to achieve performance comparable to that of the optimal controller, even in the presence of uncertainty.

Manuscript received September 6, 2023. Revised August 21, 2024.
[1]School of Mathematics, Sichuan University, Chengdu 610065, China. E-mail: feifeim369@163.com
*This work was supported by the National Natural Science Foundation of China (Nos. 12025105, 11971334).

A common strategy for controlling a system with uncertainties is to initially estimate the model parameters and then design an optimal controller based on the estimated model. This learning-based approach was first introduced from a learning-theoretic perspective in [1]. Learning-based control has been widely applied in various fields, as evidenced by books such as [7, 11, 19, 25] and recent papers like [2–3]. The classical papers [5, 22] and recent surveys [8] also discuss the well-known "multi-armed bandit" problem. However, in safety-critical scenarios where uninterrupted control is required, there is no opportunity for a dedicated learning phase. In such cases, there is a strong motivation to find an optimal control strategy within a finite timescale, without distinct learning and control phases. A related approach is to divide the available time into epochs, where the model is refined and the controller is updated within each epoch. Papers such as [1, 9–10, 12, 15, 20] present this epoch-based methodology. All of these results strive for optimality in the asymptotic limit of large time.

Recently, the Bayesian strategy has gained widespread popularity and demonstrated excellent performance in scenarios where the unknown parameter appears in the drift term and is independent of the state variable. Unlike traditional approaches that are divided into distinct phases of exploration and exploitation, the Bayesian strategy takes into account the history at every moment, making it more adaptive. Notably, these results are optimal for a fixed time interval rather than for large time scales.

The core idea behind the Bayesian strategy is to apply Bayesian statistics to transform the original problem into a linear quadratic Gaussian (LQG for short) control problem. In the LQG framework, the process noise and observation noise are assumed to follow Gaussian distributions, enabling the optimal solution to be split into a Kalman filter and a linear quadratic regulator. For a deeper understanding, Speyer and Chung [26] provide a detailed introduction to linear quadratic Gaussian control. This idea can be traced back to the principle of separation of estimation and control, which has been extensively studied (cf. [6, 14, 17, 23]). Specifically, when the system state is one-dimensional and can be measured without any sensor noise, the Bayesian strategy, compared with the scenario with fully known parameter values, is shown to minimize the worst-case regret with a specific fixed prior, as demonstrated in [16]. Moreover, when the state is multi-dimensional and the measurements are noisy, the work in [16] is extended in [18], leading to the same conclusion.

If we incorporate the influence of the state variable itself into the drift term, we obtain a well-known stochastic differential equation called the Ornstein-Uhlenbeck process. This process is often used to model phenomena such as the motion of particles subject to friction in physics and stock price distributions in financial mathematics. Furthermore, in the multi-dimensional case, considering the realistic factors, the number of control variables we apply is often expected to be less than the number of state variables, and usually we can only observe part of the information of the state variables. Thus, it motivates us to consider more complex models, and the model in [18] can be considered as a special case. More details can be found in Remarks 2.2–2.3. Another interesting phenomenon worth noting is that in the numerical example, it is found that due to the dependence of the drift term on the state, the challenge of short-time learning and control of dynamical systems with unknown parameters no longer arises at time

$t = 0$, as proposed in [16, 18]. Please refer to Figure 1(a) and Figure 1(b) for more details.

Afterwards, we make an effort to extend this highly practical result to infinite-dimensional spaces. This is because numerous real-life phenomena can be represented by partial differential equation models, such as fluid motion, derivatives pricing theories, the Prato problem, and so on. Fortunately, it is possible to achieve these extensions within the same evaluation framework, namely multiplicative regret. Consequently, we are able to determine the optimal agnostic strategy for our model. The specific results and detailed proofs are provided below.

This article is divided into three parts. The first two parts discuss Bayesian optimal control problems in both finite- and infinite-dimensions. Each part has three subsections that address different aspects such as formulating the problem, deriving optimal costs, and computing the optimal cost with full knowledge of the parameters. In the last part of the article, a Bayesian strategy with a fixed prior on the drift parameter is shown to minimize the worst-case multiplicative regret. This result holds true for both finite- and infinite-dimensional systems.

2 Finite-Dimensional Systems

2.1 Problem formulation

Consider the following control system:

$$\begin{cases} \mathrm{d}\boldsymbol{q}(t) = (A\boldsymbol{q} + \boldsymbol{b} + B\boldsymbol{u}(t))\mathrm{d}t + \mathrm{d}\boldsymbol{w}(t), & t \in [0, T], \\ \boldsymbol{q}(0) = \boldsymbol{q}_0, \\ \mathrm{d}\boldsymbol{y}(t) = C\boldsymbol{q}\mathrm{d}t + \mathrm{d}\boldsymbol{v}(t), & t \in [0, T], \\ \boldsymbol{y}(0) = \boldsymbol{0}, \end{cases} \tag{2.1}$$

where $\boldsymbol{q}(t) \in \mathbb{R}^n$ denotes the state of the system, observed through a noisy measurement $\boldsymbol{y} \in \mathbb{R}^p$ $(p \le n)$; $A \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{n\times m}$, and $C \in \mathbb{R}^{n\times p}$ are given time-invariant matrices. $\boldsymbol{q}_0$ is an $\mathbb{R}^n$-valued Gaussian random variable, independent of $\boldsymbol{w}(t)$ and $\boldsymbol{v}(t)$, with zero mean and covariance matrix $\Sigma_{\boldsymbol{q}_0}$. $\boldsymbol{w}(t)$ and $\boldsymbol{v}(t)$ are independent Wiener processes on $\mathbb{R}^n$ and $\mathbb{R}^p$ with covariance matrices $\Sigma_{\boldsymbol{w}}$ and $\Sigma_{\boldsymbol{v}}$, respectively.

Given some $T_0 \in [0, T)$, we are allowed to observe the system for $t \in [0, T_0]$ without applying any control, after which we control the system for $t \in [T_0, T]$. The objective is to find the optimal control strategy $\boldsymbol{u}(t) \in \mathbb{R}^m$ $(m \le n)$, depending on the measurements $\{\boldsymbol{y}(\tau) : 0 \le \tau < t\}$ and taking values in $\mathbb{R}^p$, that minimizes the cost functional:

$$J(\boldsymbol{q}, \boldsymbol{u}; \boldsymbol{b}) = \mathbb{E}\int_{T_0}^{T} (\boldsymbol{q}^{\mathrm{T}} Q \boldsymbol{q} + \boldsymbol{u}^{\mathrm{T}} R \boldsymbol{u})\mathrm{d}t. \tag{2.2}$$

The symmetric matrices Q and R are chosen as positive semi-definite and positive definite, respectively.

If the drift parameter $\boldsymbol{b}$ is known, then the expectation in (2.2) is well-defined, and the task reduces to a classical optimal control problem. However, when $\boldsymbol{b}$ is unknown, the objective is to find the optimal agnostic strategy, as explored in a previous study in [18], that minimizes the worst-case multiplicative regret. In what follows we will show that the Bayesian strategy is indeed this optimal agnostic strategy.

In the Bayesian version, a prior belief is assumed on $\boldsymbol{b}$, that is, $\boldsymbol{b}$ is chosen at random according to a probability measure $\mu(\boldsymbol{b})$. The expected cost associated with $\boldsymbol{u}$ can then be computed as

$$\mathcal{J}(\boldsymbol{q},\boldsymbol{u};\mu)=\mathbb{E}_{\boldsymbol{b}}[J(\boldsymbol{q},\boldsymbol{u};\boldsymbol{b})]:=\int_{\mathbb{R}^n}J(\boldsymbol{q},\boldsymbol{u};\boldsymbol{b})\mathrm{d}\mu(\boldsymbol{b}). \tag{2.3}$$

In this case, the objective is to find a control strategy $\boldsymbol{u}(t)$ that minimizes the expected value in (2.3) given the prior belief $\mu(\boldsymbol{b})$.

2.2 A Bayesian strategy for unknown $\boldsymbol{b}$

This section demonstrates the optimal strategy for the Bayesian problem given a fixed prior belief $\mu(\boldsymbol{b})$. We restrict ourselves to Gaussian prior beliefs only, i.e., $\mathrm{d}\mu(\boldsymbol{b})=\rho(\boldsymbol{b})\mathrm{d}\boldsymbol{b}$, where $\rho(\boldsymbol{b})=\mathcal{N}(\boldsymbol{b};\boldsymbol{0},\Sigma)$ for some covariance matrix $\Sigma\succeq 0$. To approach this problem in a more traditional control-theoretic way, the system (2.1) is rewritten as follows:

$$\begin{cases}\mathrm{d}\boldsymbol{x}(t)=(F\boldsymbol{x}+D\boldsymbol{u}(t))\mathrm{d}t+G\mathrm{d}\boldsymbol{w}(t), & t\in[0,T],\\ \mathrm{d}\boldsymbol{y}(t)=H\boldsymbol{x}\mathrm{d}t+\mathrm{d}\boldsymbol{v}(t), & t\in[0,T],\end{cases} \tag{2.4}$$

where $\boldsymbol{x}=\left[\begin{smallmatrix}\boldsymbol{q}\\ \boldsymbol{b}\end{smallmatrix}\right]\in\mathbb{R}^{2n}, F=\left[\begin{smallmatrix}A & \boldsymbol{I}_{n\times n}\\ \boldsymbol{0}_{n\times n} & \boldsymbol{0}_{n\times n}\end{smallmatrix}\right], D=\left[\begin{smallmatrix}B\\ \boldsymbol{0}_{n\times m}\end{smallmatrix}\right], G=\left[\begin{smallmatrix}\boldsymbol{I}_{n\times n}\\ \boldsymbol{0}_{n\times n}\end{smallmatrix}\right]$, and $H=[C\quad \boldsymbol{0}_{p\times n}]$. The cost (2.3) can equivalently be written as

$$\mathcal{J}(\boldsymbol{x},\boldsymbol{u})=\mathbb{E}\Big[\int_{T_0}^{T}(\boldsymbol{x}^{\mathrm{T}}\widetilde{Q}\boldsymbol{x}+\boldsymbol{u}^{\mathrm{T}}R\boldsymbol{u})\mathrm{d}t\Big] \tag{2.5}$$

with $\widetilde{Q}=\left[\begin{smallmatrix}Q & \boldsymbol{0}_{n\times n}\\ \boldsymbol{0}_{n\times n} & \boldsymbol{0}_{n\times n}\end{smallmatrix}\right]$. Now the problem of minimizing (2.5) subject to the dynamics (2.4) becomes a standard linear-quadratic-Gaussian optimal control problem with partial information.

Let $\widehat{\boldsymbol{x}}(t)=\left[\begin{smallmatrix}\widehat{\boldsymbol{q}}(t)\\ \widehat{\boldsymbol{b}}(t)\end{smallmatrix}\right]=\mathbb{E}[\boldsymbol{x}(t)\mid\boldsymbol{x}(0),\boldsymbol{y}(\tau):\tau\in[0,t)]$ be the Bayesian estimate of $\boldsymbol{q}(t)$ and $\boldsymbol{b}$ at time t. Then

$$\mathrm{d}\widehat{\boldsymbol{x}}(t)=(F\widehat{\boldsymbol{x}}(t)+D\boldsymbol{u})\mathrm{d}t+K\mathrm{d}\boldsymbol{I}(t), \tag{2.6}$$

where the innovations process, a zero-mean white-noise process, is

$$\mathrm{d}\boldsymbol{I}(t)=\mathrm{d}\boldsymbol{y}(t)-H\widehat{\boldsymbol{x}}(t)\mathrm{d}t,\quad E[\mathrm{d}\boldsymbol{I}(t)\mathrm{d}(\boldsymbol{I}(t))^{\mathrm{T}}]=\Sigma_{\boldsymbol{v}}\mathrm{d}t,$$

and the Kalman gain is

$$K=P(t)H^{\mathrm{T}}\Sigma_{\boldsymbol{v}}^{-1}$$

with $P(t)=\left[\begin{smallmatrix}P_{11} & P_{12}\\ P_{12}^{\mathrm{T}} & P_{22}\end{smallmatrix}\right]=\mathbb{E}[(\boldsymbol{x}(t)-\widehat{\boldsymbol{x}}(t))(\boldsymbol{x}(t)-\widehat{\boldsymbol{x}}(t))^{\mathrm{T}}]$ given by the Riccati ODEs,

$$\begin{cases}\dot{P}(t)=FP+PF^{\mathrm{T}}-PH^{\mathrm{T}}\Sigma_{\boldsymbol{v}}^{-1}HP+G\Sigma_{\boldsymbol{w}}G^{\mathrm{T}},\\ P(0)=\begin{bmatrix}\Sigma_{\boldsymbol{q}_0} & 0\\ 0 & \Sigma\end{bmatrix}.\end{cases} \tag{2.7}$$

The cost to go at time t is

$$\begin{aligned}\mathcal{J}(t,\boldsymbol{q},\boldsymbol{y};\mu) &= \mathbb{E}\Big[\int_t^T (\boldsymbol{x}^{\mathrm{T}}\widetilde{Q}\boldsymbol{x}+\boldsymbol{u}^{\mathrm{T}}R\boldsymbol{u})\mathrm{d}\tau\Big]\\ &= \mathbb{E}\Big[\int_t^T (\widehat{\boldsymbol{x}}^{\mathrm{T}}\widetilde{Q}\widehat{\boldsymbol{x}}+\boldsymbol{u}^{\mathrm{T}}R\boldsymbol{u})\mathrm{d}\tau\Big]+\int_t^T \mathrm{Tr}(P(\tau)\widetilde{Q})\mathrm{d}\tau. \end{aligned}\tag{2.8}$$

The partial information LQG problem has been reduced to finding the control law $\boldsymbol{u}$, which minimizes (2.8) subject to (2.6).

Mathematically speaking, we have the same problem as when we have full information. The only differences are additional terms involving $P(\tau)$ in the cost function and $\boldsymbol{v}(t)$ in the dynamic equation. However, neither of these terms is a function of the control, $\boldsymbol{u}$, nor will they affect the final result when we optimize the cost with respect to $\boldsymbol{u}$. Thus, the optimal control looks much the same as before.

Remark 2.1 Note that we have placed the unknown parameter $\boldsymbol{b}$ in the position of the state, so in order to ensure optimality and performance of the Kalman-Bucy filter, it is necessary to initially assume that the prior distribution of $\boldsymbol{b}$ is Gaussian.

To solve this problem, assume that the optimal value $\mathcal{J}^0(\widehat{\boldsymbol{x}},t)$ of the cost function has the following form:

$$\mathcal{J}^0(\widehat{\boldsymbol{x}},t)=\widehat{\boldsymbol{x}}^{\mathrm{T}}S(t)\widehat{\boldsymbol{x}}+\alpha(t),$$

and substitute this solution into the Hamilton-Jacobi-Bellman equation:

$$\begin{aligned}0&=\frac{\partial\mathcal{J}^0}{\partial t}+\min_{\boldsymbol{u}\in\mathcal{U}}\Big[\frac{\partial\mathcal{J}^0}{\partial\widehat{\boldsymbol{x}}}(F\widehat{\boldsymbol{x}}+D\boldsymbol{u})+\frac{1}{2}\mathrm{Tr}(K\Sigma_{\boldsymbol{v}}K^{\mathrm{T}})\frac{\partial^2\mathcal{J}^0}{\partial\widehat{\boldsymbol{x}}^2}+(\widehat{\boldsymbol{x}}^{\mathrm{T}}\widetilde{Q}\widehat{\boldsymbol{x}}+\boldsymbol{u}^{\mathrm{T}}R\boldsymbol{u})\Big]\\ &=\widehat{\boldsymbol{x}}^{\mathrm{T}}\dot{S}\widehat{\boldsymbol{x}}+\dot{\alpha}+\min_{\boldsymbol{u}\in\mathcal{U}}[2\widehat{\boldsymbol{x}}^{\mathrm{T}}S(F\widehat{\boldsymbol{x}}+D\boldsymbol{u})+\mathrm{Tr}(K\Sigma_{\boldsymbol{v}}K^{\mathrm{T}}S)+(\widehat{\boldsymbol{x}}^{\mathrm{T}}\widetilde{Q}\widehat{\boldsymbol{x}}+\boldsymbol{u}^{\mathrm{T}}R\boldsymbol{u})],\end{aligned}$$

where $\mathcal{U}$ is $\mathbb{R}^m$. Now, carry out the minimization with respect to $\boldsymbol{u}$, to obtain that

$$\widehat{\boldsymbol{x}}^{\mathrm{T}}SD+\boldsymbol{u}^{\mathrm{T}}R=0,$$

which implies that

$$\boldsymbol{u}^0(\widehat{\boldsymbol{x}},t)=-R^{-1}D^{\mathrm{T}}S\widehat{\boldsymbol{x}}=-R^{-1}B^{\mathrm{T}}(S_{11}\widehat{\boldsymbol{q}}+S_{12}\widehat{\boldsymbol{b}}).$$

Substituting this optimal $\boldsymbol{u}^0$ into the Hamilton-Jacobi-Bellman equation, we get

$$0=\widehat{\boldsymbol{x}}^{\mathrm{T}}\dot{S}\widehat{\boldsymbol{x}}+\dot{\alpha}+\widehat{\boldsymbol{x}}^{\mathrm{T}}(SF+F^{\mathrm{T}}S)\widehat{\boldsymbol{x}}-\widehat{\boldsymbol{x}}^{\mathrm{T}}SDR^{-1}D^{\mathrm{T}}S\widehat{\boldsymbol{x}}+\widehat{\boldsymbol{x}}^{\mathrm{T}}\widetilde{Q}\widehat{\boldsymbol{x}}+\mathrm{Tr}(K\Sigma_{\boldsymbol{v}}K^{\mathrm{T}}S)$$

for every $\widehat{\boldsymbol{x}}\in\mathbb{R}^{2n}$. The above is identically satisfied if the matrix $S=\left[\begin{smallmatrix}S_{11}&S_{12}\\ S_{12}^{\mathrm{T}}&S_{22}\end{smallmatrix}\right]$ and scalar-valued function α satisfy the control Riccati equations:

$$\begin{cases}-\dot{S}=SF+F^{\mathrm{T}}S+\widetilde{Q}-SDR^{-1}D^{\mathrm{T}}S,\\ S(T)=0,\\ -\dot{\alpha}=\mathrm{Tr}(PH^{\mathrm{T}}\Sigma_{\boldsymbol{v}}^{-1}HPS),\\ \alpha(T)=0.\end{cases}\tag{2.9}$$

2.3 Optimal cost for a Bayesian strategy with unknown $\boldsymbol{b}$

The total cost incurred for $\boldsymbol{b}$ by the Bayesian strategy arising from a prior μ is

$$\mathcal{J}(\boldsymbol{q},\boldsymbol{b};\mu)=\mathbb{E}\Big[\int_{T_0}^{T}(\boldsymbol{q}^{\mathrm{T}}Q\boldsymbol{q}+(\boldsymbol{u}^0)^{\mathrm{T}}R\boldsymbol{u}^0)\mathrm{d}t\Big], \tag{2.10}$$

where $\boldsymbol{u}^0 = -R^{-1}B^{\mathrm{T}}(S_{11}\widehat{\boldsymbol{q}}+S_{12}\widehat{\boldsymbol{b}})$. It is known that $\widehat{\boldsymbol{q}}(t)$ and $\widehat{\boldsymbol{b}}(t)$ satisfy the following continuous-time Kalman filter equations:

$$\begin{aligned}\mathrm{d}\widehat{\boldsymbol{q}}(t)&=(A\widehat{\boldsymbol{q}}(t)+\widehat{\boldsymbol{b}}+B\boldsymbol{u})\mathrm{d}t+P_{11}(t)C^{\mathrm{T}}\Sigma_{\boldsymbol{v}}^{-1}(\mathrm{d}\boldsymbol{y}-C\widehat{\boldsymbol{q}}\mathrm{d}t),\\ \mathrm{d}\widehat{\boldsymbol{b}}(t)&=P_{12}^{\mathrm{T}}(t)C^{\mathrm{T}}\Sigma_{\boldsymbol{v}}^{-1}(\mathrm{d}\boldsymbol{y}-C\widehat{\boldsymbol{q}}\mathrm{d}t)\end{aligned}$$

from (2.6). With control $\boldsymbol{u}^0$, the stochastic differential equations describing $\boldsymbol{q}$, $\widehat{\boldsymbol{q}}$, and $\widehat{\boldsymbol{b}}$ become

$$\mathrm{d}\boldsymbol{q}=[A\boldsymbol{q}+\boldsymbol{b}-BR^{-1}B^{\mathrm{T}}(S_{11}\widehat{\boldsymbol{q}}+S_{12}\widehat{\boldsymbol{b}})]\mathrm{d}t+\mathrm{d}\boldsymbol{w}, \tag{2.11}$$

$$\mathrm{d}\widehat{\boldsymbol{q}}=[A\widehat{\boldsymbol{q}}+\widehat{\boldsymbol{b}}-BR^{-1}B^{\mathrm{T}}(S_{11}\widehat{\boldsymbol{q}}+S_{12}\widehat{\boldsymbol{b}})]\mathrm{d}t+P_{11}C^{\mathrm{T}}\Sigma_{\boldsymbol{v}}^{-1}[C(\boldsymbol{q}-\widehat{\boldsymbol{q}})\mathrm{d}t+\mathrm{d}\boldsymbol{v}], \tag{2.12}$$

$$\mathrm{d}\widehat{\boldsymbol{b}}=P_{12}^{\mathrm{T}}C^{\mathrm{T}}\Sigma_{\boldsymbol{v}}^{-1}[C(\boldsymbol{q}-\widehat{\boldsymbol{q}})\mathrm{d}t+\mathrm{d}\boldsymbol{v}]. \tag{2.13}$$

Let $\overline{\boldsymbol{b}}(t)\triangleq\mathbb{E}[\widehat{\boldsymbol{b}}(t)],\overline{\boldsymbol{q}}(t)\triangleq\mathbb{E}[\widehat{\boldsymbol{q}}(t)]$, and $\widetilde{\boldsymbol{q}}(t)\triangleq\mathbb{E}[\boldsymbol{q}(t)]$ with the associated covariances $\Sigma_{\widehat{\boldsymbol{b}}\widehat{\boldsymbol{b}}}$, $\Sigma_{\widehat{\boldsymbol{q}}\widehat{\boldsymbol{q}}}$, $\Sigma_{\widehat{\boldsymbol{q}}\widehat{\boldsymbol{b}}}$, $\Sigma_{\boldsymbol{q}\boldsymbol{q}}$, $\Sigma_{\boldsymbol{q}\widehat{\boldsymbol{q}}}$, and $\Sigma_{\boldsymbol{q}\widehat{\boldsymbol{b}}}$.

Theorem 2.1 *The optimal cost incurred by a specific Bayesian strategy can be expressed as*

$$\mathcal{J}(\boldsymbol{q},\boldsymbol{b};\mu)=\boldsymbol{b}^{\mathrm{T}}X(\Sigma,\Sigma_{\boldsymbol{v}},T_0,T)\boldsymbol{b}+Y(\Sigma,\Sigma_{\boldsymbol{v}},T_0,T), \tag{2.14}$$

where

$$X(\Sigma,\Sigma_{\boldsymbol{v}},T_0,T)=\int_{T_0}^{T}[C_1^{\mathrm{T}}QC_1+(S_{11}C_2+S_{12}C_3)^{\mathrm{T}}BR^{-1}B^{\mathrm{T}}(S_{11}C_2+S_{12}C_3)]\mathrm{d}t$$

and

$$\begin{aligned}Y(\Sigma,\Sigma_{\boldsymbol{v}},T_0,T)=\int_{T_0}^{T}&\mathrm{Tr}(Q\Sigma_{\boldsymbol{q}\boldsymbol{q}}+S_{11}^{\mathrm{T}}BR^{-1}B^{\mathrm{T}}S_{11}\Sigma_{\widehat{\boldsymbol{q}}\widehat{\boldsymbol{q}}}\\&+2S_{11}^{\mathrm{T}}BR^{-1}B^{\mathrm{T}}S_{12}\Sigma_{\widehat{\boldsymbol{q}}\widehat{\boldsymbol{b}}}+S_{12}^{\mathrm{T}}BR^{-1}B^{\mathrm{T}}S_{12}\Sigma_{\widehat{\boldsymbol{b}}\widehat{\boldsymbol{b}}})\mathrm{d}t\end{aligned}$$

for time-varying functions $C_1(t),C_2(t)$, and $C_3(t)$. All of the quantities appearing here can be found by solving the following system of ODEs:

$$\begin{aligned}\dot{C}_1&=AC_1+I-BR^{-1}B^{\mathrm{T}}(S_{11}C_2+S_{12}C_3),\\ \dot{C}_2&=AC_2+C_3-BR^{-1}B^{\mathrm{T}}(S_{11}C_2+S_{12}C_3)+P_{11}C^{\mathrm{T}}\Sigma_{\boldsymbol{v}}^{-1}C(C_1-C_2),\\ \dot{C}_3&=P_{12}^{\mathrm{T}}C^{\mathrm{T}}\Sigma_{\boldsymbol{v}}^{-1}C(C_1-C_2),\\ \dot{\Sigma}_{\widehat{\boldsymbol{b}}\widehat{\boldsymbol{b}}}&=(\Sigma_{\boldsymbol{q}\widehat{\boldsymbol{b}}}-\Sigma_{\widehat{\boldsymbol{q}}\widehat{\boldsymbol{b}}})^{\mathrm{T}}C^{\mathrm{T}}\Sigma_{\boldsymbol{v}}^{-1}CP_{12}+P_{12}^{\mathrm{T}}C^{\mathrm{T}}\Sigma_{\boldsymbol{v}}^{-1}C(\Sigma_{\boldsymbol{q}\widehat{\boldsymbol{b}}}-\Sigma_{\widehat{\boldsymbol{q}}\widehat{\boldsymbol{b}}})+P_{12}^{\mathrm{T}}C^{\mathrm{T}}\Sigma_{\boldsymbol{v}}^{-1}CP_{12},\end{aligned}$$

$$\begin{aligned}
\dot{\Sigma}_{\boldsymbol{q}\widehat{\boldsymbol{b}}} &= (\Sigma_{\boldsymbol{q}\boldsymbol{q}} - \Sigma_{\boldsymbol{q}\widehat{\boldsymbol{q}}})C^{\mathrm{T}}\Sigma_{\boldsymbol{v}}^{-1}CP_{12} - BR^{-1}B^{\mathrm{T}}(S_{11}\Sigma_{\widehat{\boldsymbol{q}}\widehat{\boldsymbol{b}}} + S_{12}\Sigma_{\widehat{\boldsymbol{q}}\widehat{\boldsymbol{q}}}) + A\Sigma_{\boldsymbol{q}\widehat{\boldsymbol{b}}},\\
\dot{\Sigma}_{\boldsymbol{q}\boldsymbol{q}} &= \Sigma_{\boldsymbol{w}} - (\Sigma_{\boldsymbol{q}\widehat{\boldsymbol{q}}}S_{11}^{\mathrm{T}} + \Sigma_{\boldsymbol{q}\widehat{\boldsymbol{b}}}S_{12}^{\mathrm{T}})BR^{-1}B^{\mathrm{T}}\\
&\quad - BR^{-1}B^{\mathrm{T}}(\Sigma_{\boldsymbol{q}\widehat{\boldsymbol{q}}}S_{11}^{\mathrm{T}} + \Sigma_{\boldsymbol{q}\widehat{\boldsymbol{b}}}S_{12}^{\mathrm{T}})^{\mathrm{T}} + \Sigma_{\boldsymbol{q}\boldsymbol{q}}A^{\mathrm{T}} + A\Sigma_{\boldsymbol{q}\boldsymbol{q}},\\
\dot{\Sigma}_{\boldsymbol{q}\widehat{\boldsymbol{q}}} &= (\Sigma_{\boldsymbol{q}\boldsymbol{q}} - \Sigma_{\boldsymbol{q}\widehat{\boldsymbol{q}}})C^{\mathrm{T}}\Sigma_{\boldsymbol{v}}^{-1}CP_{11}^{\mathrm{T}} - \Sigma_{\boldsymbol{q}\widehat{\boldsymbol{q}}}S_{11}^{\mathrm{T}}BR^{-1}B^{\mathrm{T}} + \Sigma_{\boldsymbol{q}\widehat{\boldsymbol{b}}}(I - BR^{-1}B^{\mathrm{T}}S_{12})^{\mathrm{T}}\\
&\quad - BR^{-1}B^{\mathrm{T}}(S_{11}\Sigma_{\widehat{\boldsymbol{q}}\widehat{\boldsymbol{q}}} + S_{12}\Sigma_{\widehat{\boldsymbol{q}}\widehat{\boldsymbol{b}}}^{\mathrm{T}}) + \Sigma_{\boldsymbol{q}\widehat{\boldsymbol{q}}}A^{\mathrm{T}} + A\Sigma_{\boldsymbol{q}\widehat{\boldsymbol{q}}},\\
\dot{\Sigma}_{\widehat{\boldsymbol{q}}\widehat{\boldsymbol{q}}} &= P_{11}C^{\mathrm{T}}\Sigma_{\boldsymbol{v}}^{-1}CP_{11}^{\mathrm{T}} + \Sigma_{\widehat{\boldsymbol{q}}\widehat{\boldsymbol{b}}}(I - BR^{-1}B^{\mathrm{T}}S_{12})^{\mathrm{T}} + (I - BR^{-1}B^{\mathrm{T}}S_{12})\Sigma_{\widehat{\boldsymbol{q}}\widehat{\boldsymbol{b}}}^{\mathrm{T}}\\
&\quad + (\Sigma_{\boldsymbol{q}\widehat{\boldsymbol{q}}} - \Sigma_{\widehat{\boldsymbol{q}}\widehat{\boldsymbol{q}}})^{\mathrm{T}}C^{\mathrm{T}}\Sigma_{\boldsymbol{v}}^{-1}CP_{11}^{\mathrm{T}} + P_{11}C^{\mathrm{T}}\Sigma_{\boldsymbol{v}}^{-1}C(\Sigma_{\boldsymbol{q}\widehat{\boldsymbol{q}}} - \Sigma_{\widehat{\boldsymbol{q}}\widehat{\boldsymbol{q}}}) - \Sigma_{\widehat{\boldsymbol{q}}\widehat{\boldsymbol{q}}}S_{11}^{\mathrm{T}}BR^{-1}B^{\mathrm{T}}\\
&\quad - BR^{-1}B^{\mathrm{T}}S_{11}\Sigma_{\widehat{\boldsymbol{q}}\widehat{\boldsymbol{q}}} + \Sigma_{\widehat{\boldsymbol{q}}\widehat{\boldsymbol{q}}}A^{\mathrm{T}} + A\Sigma_{\widehat{\boldsymbol{q}}\widehat{\boldsymbol{q}}},\\
\dot{\Sigma}_{\widehat{\boldsymbol{q}}\widehat{\boldsymbol{b}}} &= (\Sigma_{\boldsymbol{q}\widehat{\boldsymbol{q}}} - \Sigma_{\widehat{\boldsymbol{q}}\widehat{\boldsymbol{q}}})^{\mathrm{T}}C^{\mathrm{T}}\Sigma_{\boldsymbol{v}}^{-1}CP_{12} + (I - BR^{-1}B^{\mathrm{T}}S_{12})\Sigma_{\widehat{\boldsymbol{b}}\widehat{\boldsymbol{b}}} - BR^{-1}B^{\mathrm{T}}S_{12}\Sigma_{\widehat{\boldsymbol{q}}\widehat{\boldsymbol{b}}}\\
&\quad + P_{11}C^{\mathrm{T}}\Sigma_{\boldsymbol{v}}^{-1}C(\Sigma_{\boldsymbol{q}\widehat{\boldsymbol{b}}} - \Sigma_{\widehat{\boldsymbol{q}}\widehat{\boldsymbol{b}}} + P_{12}) + A\Sigma_{\widehat{\boldsymbol{q}}\widehat{\boldsymbol{b}}}.
\end{aligned}$$

Proof The cost (2.10) can be written as

$$\begin{aligned}
\mathcal{J}(\boldsymbol{q},\boldsymbol{b};\mu) &= \mathbb{E}\Big[\int_{T_0}^{T}(\boldsymbol{q}^{\mathrm{T}}Q\boldsymbol{q} + (\boldsymbol{u}^0)^{\mathrm{T}}R\boldsymbol{u}^0)\mathrm{d}t\Big]\\
&= \int_{T_0}^{T}[\widetilde{\boldsymbol{q}}^{\mathrm{T}}Q\widetilde{\boldsymbol{q}} + (S_{11}\overline{\boldsymbol{q}} + S_{12}\overline{\boldsymbol{b}})^{\mathrm{T}}BR^{-1}B^{\mathrm{T}}(S_{11}\overline{\boldsymbol{q}} + S_{12}\overline{\boldsymbol{b}})]\mathrm{d}t\\
&\quad + \int_{T_0}^{T}\mathrm{Tr}(Q\Sigma_{\boldsymbol{q}\boldsymbol{q}} + S_{11}^{\mathrm{T}}BR^{-1}B^{\mathrm{T}}S_{11}\Sigma_{\widehat{\boldsymbol{q}}\widehat{\boldsymbol{q}}} + 2S_{11}^{\mathrm{T}}BR^{-1}B^{\mathrm{T}}S_{12}\Sigma_{\widehat{\boldsymbol{q}}\widehat{\boldsymbol{b}}}\\
&\quad + S_{12}^{\mathrm{T}}BR^{-1}B^{\mathrm{T}}S_{12}\Sigma_{\widehat{\boldsymbol{b}}\widehat{\boldsymbol{b}}})\mathrm{d}t. \qquad (2.15)
\end{aligned}$$

To compute this cost, the ODEs for the aforementioned mean and covariances need to be derived from the SDEs (2.11)–(2.13). This can be done by computing the expectations for the discretized dynamics with a time step Δt and taking the limit as $\Delta t \to 0$. For example,

$$\begin{aligned}
\Delta\overline{\boldsymbol{b}} = \mathbb{E}[\Delta\widehat{\boldsymbol{b}}] &= \mathbb{E}[P_{12}^{\mathrm{T}}C^{\mathrm{T}}\Sigma_{\boldsymbol{v}}^{-1}[C(\boldsymbol{q} - \widehat{\boldsymbol{q}})\Delta t + \Delta\boldsymbol{v}]]\\
&= P_{12}^{\mathrm{T}}C^{\mathrm{T}}\Sigma_{\boldsymbol{v}}^{-1}[C(\widetilde{\boldsymbol{q}} - \overline{\boldsymbol{q}})\Delta t],
\end{aligned}$$

and therefore,

$$\dot{\overline{\boldsymbol{b}}} = P_{12}^{\mathrm{T}}C^{\mathrm{T}}\Sigma_{\boldsymbol{v}}^{-1}C(\widetilde{\boldsymbol{q}} - \overline{\boldsymbol{q}}).$$

Similarly, we get

$$\begin{aligned}
\dot{\widetilde{\boldsymbol{q}}} &= A\widetilde{\boldsymbol{q}} + \boldsymbol{b} - BR^{-1}B^{\mathrm{T}}(S_{11}\overline{\boldsymbol{q}} + S_{12}\overline{\boldsymbol{b}}),\\
\dot{\overline{\boldsymbol{q}}} &= A\overline{\boldsymbol{q}} + \overline{\boldsymbol{b}} - BR^{-1}B^{\mathrm{T}}(S_{11}\overline{\boldsymbol{q}} + S_{12}\overline{\boldsymbol{b}}) + P_{11}C^{\mathrm{T}}\Sigma_{\boldsymbol{v}}^{-1}C(\widetilde{\boldsymbol{q}} - \overline{\boldsymbol{q}}).
\end{aligned}$$

The ODEs for $\widetilde{\boldsymbol{q}}$, $\overline{\boldsymbol{q}}$, and $\overline{\boldsymbol{b}}$ yield $\widetilde{\boldsymbol{q}} = C_1(t)\boldsymbol{b}$, $\overline{\boldsymbol{q}} = C_2(t)\boldsymbol{b}$, and $\overline{\boldsymbol{b}} = C_3(t)\boldsymbol{b}$ for some C_1, C_2, $C_3 : [0,T] \to \mathbb{R}^{n\times n}$ that satisfy the same set of ODEs:

$$\dot{C}_1 = AC_1 + I - BR^{-1}B^{\mathrm{T}}(S_{11}C_2 + S_{12}C_3), \qquad (2.16)$$

$$\dot{C}_2 = AC_2 + C_3 - BR^{-1}B^{\mathrm{T}}(S_{11}C_2 + S_{12}C_3) + P_{11}C^{\mathrm{T}}\Sigma_{\boldsymbol{v}}^{-1}C(C_1 - C_2), \tag{2.17}$$

$$\dot{C}_3 = P_{12}^{\mathrm{T}}C^{\mathrm{T}}\Sigma_{\boldsymbol{v}}^{-1}C(C_1 - C_2). \tag{2.18}$$

Substituting (2.16)–(2.18) into the first term of the last equation of (2.15), we get

$$\int_{T_0}^{T}[\widetilde{\boldsymbol{q}}^{\mathrm{T}}Q\widetilde{\boldsymbol{q}} + (S_{11}\overline{\boldsymbol{q}} + S_{12}\overline{\boldsymbol{b}})^{\mathrm{T}}BR^{-1}B^{\mathrm{T}}(S_{11}\overline{\boldsymbol{q}} + S_{12}\overline{\boldsymbol{b}})]\mathrm{d}t = \boldsymbol{b}^{\mathrm{T}}X(\Sigma, \Sigma_{\boldsymbol{v}}, T_0, T)\boldsymbol{b}.$$

Again,

$$\begin{aligned}\Sigma_{\widehat{\boldsymbol{b}}\widehat{\boldsymbol{b}}}(t+\Delta t) &= \mathbb{E}[(\widehat{\boldsymbol{b}}(t+\Delta t) - \overline{\boldsymbol{b}}(t+\Delta t))(\widehat{\boldsymbol{b}}(t+\Delta t) - \overline{\boldsymbol{b}}(t+\Delta t))^{\mathrm{T}}]\\ &= \mathbb{E}[\Theta\Theta^{\mathrm{T}}]\\ &= \Sigma_{\widehat{\boldsymbol{b}}\widehat{\boldsymbol{b}}}(t) + [(\Sigma_{\boldsymbol{q}\widehat{\boldsymbol{b}}} - \Sigma_{\widehat{\boldsymbol{q}}\widehat{\boldsymbol{b}}})^{\mathrm{T}}C^{\mathrm{T}}\Sigma_{\boldsymbol{v}}^{-1}CP_{12}\\ &\quad + P_{12}^{\mathrm{T}}C^{\mathrm{T}}\Sigma_{\boldsymbol{v}}^{-1}C(\Sigma_{\boldsymbol{q}\widehat{\boldsymbol{b}}} - \Sigma_{\widehat{\boldsymbol{q}}\widehat{\boldsymbol{b}}}) + P_{12}^{\mathrm{T}}C^{\mathrm{T}}\Sigma_{\boldsymbol{v}}^{-1}CP_{12}]\Delta t,\end{aligned}$$

where $\Theta = \widehat{\boldsymbol{b}} - \overline{\boldsymbol{b}} + P_{12}^{\mathrm{T}}C^{\mathrm{T}}\Sigma_{\boldsymbol{v}}^{-1}C[(\boldsymbol{q} - \widetilde{\boldsymbol{q}}) - (\widehat{\boldsymbol{q}} - \overline{\boldsymbol{q}})]\Delta t + P_{12}^{\mathrm{T}}C^{\mathrm{T}}\Sigma_{\boldsymbol{v}}^{-1}\Delta\boldsymbol{v}$. Therefore,

$$\dot{\Sigma}_{\widehat{\boldsymbol{b}}\widehat{\boldsymbol{b}}} = (\Sigma_{\boldsymbol{q}\widehat{\boldsymbol{b}}} - \Sigma_{\widehat{\boldsymbol{q}}\widehat{\boldsymbol{b}}})^{\mathrm{T}}C^{\mathrm{T}}\Sigma_{\boldsymbol{v}}^{-1}CP_{12} + P_{12}^{\mathrm{T}}C^{\mathrm{T}}\Sigma_{\boldsymbol{v}}^{-1}C(\Sigma_{\boldsymbol{q}\widehat{\boldsymbol{b}}} - \Sigma_{\widehat{\boldsymbol{q}}\widehat{\boldsymbol{b}}}) + P_{12}^{\mathrm{T}}C^{\mathrm{T}}\Sigma_{\boldsymbol{v}}^{-1}CP_{12}.$$

Similarily,

$$\begin{aligned}\dot{\Sigma}_{\boldsymbol{q}\boldsymbol{q}} &= \Sigma_{\boldsymbol{w}} - (\Sigma_{\boldsymbol{q}\widehat{\boldsymbol{q}}}S_{11}^{\mathrm{T}} + \Sigma_{\boldsymbol{q}\widehat{\boldsymbol{b}}}S_{12}^{\mathrm{T}})BR^{-1}B^{\mathrm{T}}\\ &\quad - BR^{-1}B^{\mathrm{T}}(\Sigma_{\boldsymbol{q}\widehat{\boldsymbol{q}}}S_{11}^{\mathrm{T}} + \Sigma_{\boldsymbol{q}\widehat{\boldsymbol{b}}}S_{12}^{\mathrm{T}})^{\mathrm{T}} + \Sigma_{\boldsymbol{q}\boldsymbol{q}}A^{\mathrm{T}} + A\Sigma_{\boldsymbol{q}\boldsymbol{q}},\\ \dot{\Sigma}_{\boldsymbol{q}\widehat{\boldsymbol{q}}} &= (\Sigma_{\boldsymbol{q}\boldsymbol{q}} - \Sigma_{\boldsymbol{q}\widehat{\boldsymbol{q}}})C^{\mathrm{T}}\Sigma_{\boldsymbol{v}}^{-1}CP_{11}^{\mathrm{T}} - \Sigma_{\boldsymbol{q}\widehat{\boldsymbol{q}}}S_{11}^{\mathrm{T}}BR^{-1}B^{\mathrm{T}} + \Sigma_{\boldsymbol{q}\widehat{\boldsymbol{b}}}(I - BR^{-1}B^{\mathrm{T}}S_{12})^{\mathrm{T}}\\ &\quad - BR^{-1}B^{\mathrm{T}}(S_{11}\Sigma_{\widehat{\boldsymbol{q}}\widehat{\boldsymbol{q}}} + S_{12}\Sigma_{\widehat{\boldsymbol{q}}\widehat{\boldsymbol{b}}}^{\mathrm{T}}) + \Sigma_{\boldsymbol{q}\widehat{\boldsymbol{q}}}A^{\mathrm{T}} + A\Sigma_{\boldsymbol{q}\widehat{\boldsymbol{q}}},\\ \dot{\Sigma}_{\boldsymbol{q}\widehat{\boldsymbol{b}}} &= (\Sigma_{\boldsymbol{q}\boldsymbol{q}} - \Sigma_{\boldsymbol{q}\widehat{\boldsymbol{q}}})C^{\mathrm{T}}\Sigma_{\boldsymbol{v}}^{-1}CP_{12} - BR^{-1}B^{\mathrm{T}}(S_{11}\Sigma_{\widehat{\boldsymbol{q}}\widehat{\boldsymbol{b}}} + S_{12}\Sigma_{\widehat{\boldsymbol{b}}\widehat{\boldsymbol{b}}}) + A\Sigma_{\boldsymbol{q}\widehat{\boldsymbol{b}}},\\ \dot{\Sigma}_{\widehat{\boldsymbol{q}}\widehat{\boldsymbol{q}}} &= P_{11}C^{\mathrm{T}}\Sigma_{\boldsymbol{v}}^{-1}CP_{11}^{\mathrm{T}} + \Sigma_{\widehat{\boldsymbol{q}}\widehat{\boldsymbol{b}}}(I - BR^{-1}B^{\mathrm{T}}S_{12})^{\mathrm{T}} + (I - BR^{-1}B^{\mathrm{T}}S_{12})\Sigma_{\widehat{\boldsymbol{q}}\widehat{\boldsymbol{b}}}^{\mathrm{T}}\\ &\quad + (\Sigma_{\boldsymbol{q}\widehat{\boldsymbol{q}}} - \Sigma_{\widehat{\boldsymbol{q}}\widehat{\boldsymbol{q}}})^{\mathrm{T}}C^{\mathrm{T}}\Sigma_{\boldsymbol{v}}^{-1}CP_{11}^{\mathrm{T}} + P_{11}C^{\mathrm{T}}\Sigma_{\boldsymbol{v}}^{-1}C(\Sigma_{\boldsymbol{q}\widehat{\boldsymbol{q}}} - \Sigma_{\widehat{\boldsymbol{q}}\widehat{\boldsymbol{q}}}) - \Sigma_{\widehat{\boldsymbol{q}}\widehat{\boldsymbol{q}}}S_{11}^{\mathrm{T}}BR^{-1}B^{\mathrm{T}}\\ &\quad - BR^{-1}B^{\mathrm{T}}S_{11}\Sigma_{\widehat{\boldsymbol{q}}\widehat{\boldsymbol{q}}} + \Sigma_{\widehat{\boldsymbol{q}}\widehat{\boldsymbol{q}}}A^{\mathrm{T}} + A\Sigma_{\widehat{\boldsymbol{q}}\widehat{\boldsymbol{q}}},\\ \dot{\Sigma}_{\widehat{\boldsymbol{q}}\widehat{\boldsymbol{b}}} &= (\Sigma_{\boldsymbol{q}\widehat{\boldsymbol{q}}} - \Sigma_{\widehat{\boldsymbol{q}}\widehat{\boldsymbol{q}}})^{\mathrm{T}}C^{\mathrm{T}}\Sigma_{\boldsymbol{v}}^{-1}CP_{12} + (I - BR^{-1}B^{\mathrm{T}}S_{12})\Sigma_{\widehat{\boldsymbol{b}}\widehat{\boldsymbol{b}}} - BR^{-1}B^{\mathrm{T}}S_{12}\Sigma_{\widehat{\boldsymbol{q}}\widehat{\boldsymbol{b}}}\\ &\quad + P_{11}C^{\mathrm{T}}\Sigma_{\boldsymbol{v}}^{-1}C(\Sigma_{\boldsymbol{q}\widehat{\boldsymbol{b}}} - \Sigma_{\widehat{\boldsymbol{q}}\widehat{\boldsymbol{b}}} + P_{12}) + A\Sigma_{\widehat{\boldsymbol{q}}\widehat{\boldsymbol{b}}}.\end{aligned}$$

Remark 2.2 Note that if we set $A = 0$, $m = n$, $B = I_{n\times n}$, and $p = n$, $C = I_{n\times n}$, then the equations of C_1, C_2, C_3, $\Sigma_{\widehat{\boldsymbol{b}}\widehat{\boldsymbol{b}}}$, $\Sigma_{\widehat{\boldsymbol{q}}\widehat{\boldsymbol{q}}}$, $\Sigma_{\widehat{\boldsymbol{q}}\widehat{\boldsymbol{b}}}$, $\Sigma_{\boldsymbol{q}\boldsymbol{q}}$, $\Sigma_{\boldsymbol{q}\widehat{\boldsymbol{q}}}$, and $\Sigma_{\boldsymbol{q}\widehat{\boldsymbol{b}}}$ are the same as the equations of C_1, C_2, C_3, $\Sigma_{\widehat{\boldsymbol{a}}\widehat{\boldsymbol{a}}}$, $\Sigma_{\widehat{\boldsymbol{q}}\widehat{\boldsymbol{q}}}$, $\Sigma_{\widehat{\boldsymbol{q}}\widehat{\boldsymbol{a}}}$, $\Sigma_{\boldsymbol{q}\boldsymbol{q}}$, $\Sigma_{\boldsymbol{q}\widehat{\boldsymbol{q}}}$, and $\Sigma_{\boldsymbol{q}\widehat{\boldsymbol{a}}}$ in [18, Theorem 3], respectively.

2.4 Optimal cost for a Bayesian strategy with known $\boldsymbol{b}$

If $\boldsymbol{b}$ is known, then $\widehat{\boldsymbol{b}}(t) \equiv \overline{\boldsymbol{b}}(t) \equiv \boldsymbol{b}$, $\Sigma_{\widehat{\boldsymbol{b}}\widehat{\boldsymbol{b}}}(t) \equiv 0$, $\Sigma_{\widehat{\boldsymbol{q}}\widehat{\boldsymbol{b}}}(t) \equiv 0$, and $\Sigma_{\boldsymbol{q}\widehat{\boldsymbol{b}}}(t) \equiv 0$. We use similar notation as above, denoting all of the analogous quantities for this case with a superscript #. Hence the Riccati equation (2.7) becomes

$$\dot{P}_{11}^{\#} = AP_{11}^{\#} + P_{11}^{\#}A^{\mathrm{T}} - P_{11}^{\#}C^{\mathrm{T}}\Sigma_{\boldsymbol{v}}^{-1}CP_{11}^{\#} + \Sigma_{\boldsymbol{w}}, P_{11}^{\#}(0) = \Sigma_{\boldsymbol{q}_0}$$

and $P_{12}^{\#}(t) = P_{22}^{\#}(t) \equiv 0$. The estimation equation (2.6) becomes

$$\mathrm{d}\widehat{\boldsymbol{q}}^{\#}(t) = (A\widehat{\boldsymbol{q}}^{\#} + \boldsymbol{b} + B\boldsymbol{u})\mathrm{d}t + P_{11}^{\#}(t)C^{\mathrm{T}}\Sigma_{\boldsymbol{v}}^{-1}(\mathrm{d}\boldsymbol{y} - C\widehat{\boldsymbol{q}}^{\#}\mathrm{d}t).$$

This, in turn, yields the optimal control

$$\boldsymbol{u}^{\#}(t) = -R^{-1}B^{\mathrm{T}}(S_{11}\widehat{\boldsymbol{q}}^{\#}(t) + S_{12}\boldsymbol{b}),$$

where S_{11} and S_{12} are elements of S defined identically to (2.9).

With this control, the stochastic differential equations describing $\boldsymbol{q}^{\#}$ and $\widehat{\boldsymbol{q}}^{\#}$ become

$$\begin{aligned}
\mathrm{d}\boldsymbol{q}^{\#} &= [A\boldsymbol{q}^{\#} + \boldsymbol{b} - BR^{-1}B^{\mathrm{T}}(S_{11}\widehat{\boldsymbol{q}}^{\#} + S_{12}\boldsymbol{b})]\mathrm{d}t + \mathrm{d}\boldsymbol{w},\\
\mathrm{d}\widehat{\boldsymbol{q}}^{\#} &= [A\widehat{\boldsymbol{q}}^{\#} + \boldsymbol{b} - BR^{-1}B^{\mathrm{T}}(S_{11}\widehat{\boldsymbol{q}}^{\#} + S_{12}\boldsymbol{b})]\mathrm{d}t + P_{11}^{\#}(t)C^{\mathrm{T}}\Sigma_{\boldsymbol{v}}^{-1}[C(\boldsymbol{q}^{\#} - \widehat{\boldsymbol{q}}^{\#})\mathrm{d}t + \mathrm{d}\boldsymbol{v}].
\end{aligned}$$

Let $\overline{\boldsymbol{q}}^{\#}(t) \triangleq \mathbb{E}[\widehat{\boldsymbol{q}}^{\#}(t)]$, and $\widetilde{\boldsymbol{q}}^{\#}(t) \triangleq \mathbb{E}[\boldsymbol{q}^{\#}(t)]$ with the associated covariances $\Sigma_{\widehat{\boldsymbol{q}}\widehat{\boldsymbol{q}}}^{\#}, \Sigma_{\boldsymbol{q}\boldsymbol{q}}^{\#}$, and $\Sigma_{\boldsymbol{q}\widehat{\boldsymbol{q}}}^{\#}$.

Theorem 2.2 *The optimal cost for the Bayesian strategy with known* $\boldsymbol{b}$ *can be expressed as*

$$J(\boldsymbol{q}, \boldsymbol{b}) = \boldsymbol{b}^{\mathrm{T}}X^{\#}(\Sigma_{\boldsymbol{v}}, T_0, T)\boldsymbol{b} + Y^{\#}(\Sigma_{\boldsymbol{v}}, T_0, T), \tag{2.19}$$

where

$$X^{\#}(\Sigma_{\boldsymbol{v}}, T_0, T) = \int_{T_0}^{T} [C_1^{\#}QC_1^{\#} + (S_{11}C_2^{\#} + S_{12})^{\mathrm{T}}BR^{-1}B^{\mathrm{T}}(S_{11}C_2^{\#} + S_{12})]\mathrm{d}t$$

and

$$Y^{\#}(\Sigma_{\boldsymbol{v}}, T_0, T) = \int_{T_0}^{T} \mathrm{Tr}(Q\Sigma_{\boldsymbol{q}\boldsymbol{q}}^{\#} + S_{11}^{\mathrm{T}}BR^{-1}B^{\mathrm{T}}S_{11}\Sigma_{\widehat{\boldsymbol{q}}\widehat{\boldsymbol{q}}}^{\#})\mathrm{d}t$$

with the time-varying functions $C_1^{\#}(t), C_2^{\#}(t), \Sigma_{\boldsymbol{q}\boldsymbol{q}}^{\#}$, *and* $\Sigma_{\widehat{\boldsymbol{q}}\widehat{\boldsymbol{q}}}^{\#}$ *described by the ODEs*:

$$\begin{aligned}
\dot{C}_1^{\#} &= AC_1^{\#} + I - BR^{-1}B^{\mathrm{T}}(S_{11}C_2^{\#} + S_{12}),\\
\dot{C}_2^{\#} &= AC_2^{\#} + I - BR^{-1}B^{\mathrm{T}}(S_{11}C_2^{\#} + S_{12}) + P_{11}^{\#}C^{\mathrm{T}}\Sigma_{\boldsymbol{v}}^{-1}C(C_1^{\#} - C_2^{\#}),\\
\dot{\Sigma}_{\boldsymbol{q}\boldsymbol{q}}^{\#} &= \Sigma_{\boldsymbol{w}} - \Sigma_{\boldsymbol{q}\widehat{\boldsymbol{q}}}^{\#}S_{11}^{\mathrm{T}}BR^{-1}B^{\mathrm{T}} - BR^{-1}B^{\mathrm{T}}(\Sigma_{\boldsymbol{q}\widehat{\boldsymbol{q}}}^{\#}S_{11}^{\mathrm{T}})^{\mathrm{T}} + \Sigma_{\boldsymbol{q}\boldsymbol{q}}^{\#}A^{\mathrm{T}} + A\Sigma_{\boldsymbol{q}\boldsymbol{q}}^{\#},\\
\dot{\Sigma}_{\widehat{\boldsymbol{q}}\widehat{\boldsymbol{q}}}^{\#} &= P_{11}^{\#}C^{\mathrm{T}}\Sigma_{\boldsymbol{v}}^{-1}C(P_{11}^{\#})^{\mathrm{T}} + (\Sigma_{\boldsymbol{q}\widehat{\boldsymbol{q}}}^{\#} - \Sigma_{\widehat{\boldsymbol{q}}\widehat{\boldsymbol{q}}}^{\#})^{\mathrm{T}}C^{\mathrm{T}}\Sigma_{\boldsymbol{v}}^{-1}C(P_{11}^{\#})^{\mathrm{T}} + P_{11}^{\#}C^{\mathrm{T}}\Sigma_{\boldsymbol{v}}^{-1}C(\Sigma_{\boldsymbol{q}\widehat{\boldsymbol{q}}}^{\#} - \Sigma_{\widehat{\boldsymbol{q}}\widehat{\boldsymbol{q}}}^{\#})\\
&\quad - \Sigma_{\widehat{\boldsymbol{q}}\widehat{\boldsymbol{q}}}^{\#}S_{11}^{\mathrm{T}}BR^{-1}B^{\mathrm{T}} - BR^{-1}B^{\mathrm{T}}S_{11}\Sigma_{\widehat{\boldsymbol{q}}\widehat{\boldsymbol{q}}}^{\#} + \Sigma_{\widehat{\boldsymbol{q}}\widehat{\boldsymbol{q}}}^{\#}A^{\mathrm{T}} + A\Sigma_{\widehat{\boldsymbol{q}}\widehat{\boldsymbol{q}}}^{\#},\\
\dot{\Sigma}_{\boldsymbol{q}\widehat{\boldsymbol{q}}}^{\#} &= (\Sigma_{\boldsymbol{q}\boldsymbol{q}}^{\#} - \Sigma_{\boldsymbol{q}\widehat{\boldsymbol{q}}}^{\#})C^{\mathrm{T}}\Sigma_{\boldsymbol{v}}^{-1}C(P_{11}^{\#})^{\mathrm{T}} - \Sigma_{\boldsymbol{q}\widehat{\boldsymbol{q}}}^{\#}S_{11}^{\mathrm{T}}BR^{-1}B^{\mathrm{T}}\\
&\quad - BR^{-1}B^{\mathrm{T}}(S_{11}\Sigma_{\widehat{\boldsymbol{q}}\widehat{\boldsymbol{q}}}^{\#}) + \Sigma_{\boldsymbol{q}\widehat{\boldsymbol{q}}}^{\#}A^{\mathrm{T}} + A\Sigma_{\boldsymbol{q}\widehat{\boldsymbol{q}}}^{\#}.
\end{aligned}$$

Proof The cost (2.2) can be written as

$$\begin{aligned}
J(\boldsymbol{q}, \boldsymbol{b}) &= \mathbb{E}\Big[\int_{T_0}^{T}((\boldsymbol{q}^{\#})^{\mathrm{T}}Q\boldsymbol{q}^{\#} + (\boldsymbol{u}^{\#})^{\mathrm{T}}R\boldsymbol{u}^{\#})\mathrm{d}t\Big]\\
&= \int_{T_0}^{T}[(\widetilde{\boldsymbol{q}}^{\#})^{\mathrm{T}}Q\widetilde{\boldsymbol{q}}^{\#} + (S_{11}\overline{\boldsymbol{q}}^{\#} + S_{12}\boldsymbol{b})^{\mathrm{T}}BR^{-1}B^{\mathrm{T}}(S_{11}\overline{\boldsymbol{q}}^{\#} + S_{12}\boldsymbol{b})]\mathrm{d}t
\end{aligned}$$

$$+\int_{T_0}^{T}\mathrm{Tr}(Q\Sigma_{\boldsymbol{q}\boldsymbol{q}}^{\#}+S_{11}^{\mathrm{T}}BR^{-1}B^{\mathrm{T}}S_{11}\Sigma_{\widehat{\boldsymbol{q}}\widehat{\boldsymbol{q}}}^{\#})\mathrm{d}t. \tag{2.20}$$

By the same calculations as in Theorem 2.1, we can get

$$\begin{aligned}
\dot{\widetilde{\boldsymbol{q}}}^{\#} &= A\widetilde{\boldsymbol{q}}^{\#}+\boldsymbol{b}-BR^{-1}B^{\mathrm{T}}(S_{11}\overline{\boldsymbol{q}}^{\#}+S_{12}\boldsymbol{b}),\\
\dot{\overline{\boldsymbol{q}}}^{\#} &= A\overline{\boldsymbol{q}}^{\#}+\boldsymbol{b}-BR^{-1}B^{\mathrm{T}}(S_{11}\overline{\boldsymbol{q}}^{\#}+S_{12}\boldsymbol{b})+P_{11}^{\#}C^{\mathrm{T}}\Sigma_{\boldsymbol{v}}^{-1}C(\widetilde{\boldsymbol{q}}^{\#}-\overline{\boldsymbol{q}}^{\#}),\\
\dot{\Sigma}_{\boldsymbol{q}\boldsymbol{q}}^{\#} &= \Sigma_{\boldsymbol{w}}-\Sigma_{\boldsymbol{q}\widehat{\boldsymbol{q}}}^{\#}S_{11}^{\mathrm{T}}BR^{-1}B^{\mathrm{T}}-BR^{-1}B^{\mathrm{T}}(\Sigma_{\boldsymbol{q}\widehat{\boldsymbol{q}}}^{\#}S_{11}^{\mathrm{T}})^{\mathrm{T}}+\Sigma_{\boldsymbol{q}\boldsymbol{q}}^{\#}A^{\mathrm{T}}+A\Sigma_{\boldsymbol{q}\boldsymbol{q}}^{\#},\\
\dot{\Sigma}_{\widehat{\boldsymbol{q}}\widehat{\boldsymbol{q}}}^{\#} &= P_{11}^{\#}C^{\mathrm{T}}\Sigma_{\boldsymbol{v}}^{-1}C(P_{11}^{\#})^{\mathrm{T}}+(\Sigma_{\boldsymbol{q}\widehat{\boldsymbol{q}}}^{\#}-\Sigma_{\widehat{\boldsymbol{q}}\widehat{\boldsymbol{q}}}^{\#})^{\mathrm{T}}C^{\mathrm{T}}\Sigma_{\boldsymbol{v}}^{-1}C(P_{11}^{\#})^{\mathrm{T}}+P_{11}^{\#}\Sigma_{\boldsymbol{v}}^{-1}C(\Sigma_{\boldsymbol{q}\widehat{\boldsymbol{q}}}^{\#}-\Sigma_{\widehat{\boldsymbol{q}}\widehat{\boldsymbol{q}}}^{\#})\\
&\quad -\Sigma_{\widehat{\boldsymbol{q}}\widehat{\boldsymbol{q}}}^{\#}S_{11}^{\mathrm{T}}BR^{-1}B^{\mathrm{T}}-BR^{-1}B^{\mathrm{T}}S_{11}\Sigma_{\widehat{\boldsymbol{q}}\widehat{\boldsymbol{q}}}^{\#}+\Sigma_{\widehat{\boldsymbol{q}}\widehat{\boldsymbol{q}}}^{\#}A^{\mathrm{T}}+A\Sigma_{\widehat{\boldsymbol{q}}\widehat{\boldsymbol{q}}}^{\#},\\
\dot{\Sigma}_{\boldsymbol{q}\widehat{\boldsymbol{q}}}^{\#} &= (\Sigma_{\boldsymbol{q}\boldsymbol{q}}^{\#}-\Sigma_{\boldsymbol{q}\widehat{\boldsymbol{q}}}^{\#})C^{\mathrm{T}}\Sigma_{\boldsymbol{v}}^{-1}C(P_{11}^{\#})^{\mathrm{T}}-\Sigma_{\boldsymbol{q}\widehat{\boldsymbol{q}}}^{\#}S_{11}^{\mathrm{T}}BR^{-1}B^{\mathrm{T}}\\
&\quad -BR^{-1}B^{\mathrm{T}}(S_{11}\Sigma_{\widehat{\boldsymbol{q}}\widehat{\boldsymbol{q}}}^{\#})+\Sigma_{\boldsymbol{q}\widehat{\boldsymbol{q}}}^{\#}A^{\mathrm{T}}+A\Sigma_{\boldsymbol{q}\widehat{\boldsymbol{q}}}^{\#}.
\end{aligned}$$

The ODEs for $\widetilde{\boldsymbol{q}}^{\#}$ and $\overline{\boldsymbol{q}}^{\#}$ yield $\widetilde{\boldsymbol{q}}^{\#}=C_1^{\#}(t)\boldsymbol{b}$ and $\overline{\boldsymbol{q}}^{\#}=C_2^{\#}(t)\boldsymbol{b}$ for some $C_1^{\#},C_2^{\#}:[0,T]\to\mathbb{R}^{n\times n}$ that satisfy the same set of ODEs:

$$\dot{C}_1^{\#}=AC_1^{\#}+I-BR^{-1}B^{\mathrm{T}}(S_{11}C_2^{\#}+S_{12}), \tag{2.21}$$

$$\dot{C}_2^{\#}=AC_2^{\#}+I-BR^{-1}B^{\mathrm{T}}(S_{11}C_2^{\#}+S_{12})+P_{11}^{\#}C^{\mathrm{T}}\Sigma_{\boldsymbol{v}}^{-1}C(C_1^{\#}-C_2^{\#}). \tag{2.22}$$

Substituting (2.21)–(2.22) into (2.20) completes the proof of Theorem 2.2.

Remark 2.3 Again if we set $A=0$, $m=n$, $B=I_{n\times n}$, and $p=n$, $C=I_{n\times n}$, the equations of $C_1^{\#}$, $C_2^{\#}$, $\Sigma_{\widehat{\boldsymbol{q}}\widehat{\boldsymbol{q}}}^{\#},\Sigma_{\boldsymbol{q}\boldsymbol{q}}^{\#}$, and $\Sigma_{\boldsymbol{q}\widehat{\boldsymbol{q}}}^{\#}$ are the same as the equations of C_1^*, C_2^*, $\Sigma_{\widehat{\boldsymbol{q}}\widehat{\boldsymbol{q}}}^{*},\Sigma_{\boldsymbol{q}\boldsymbol{q}}^{*}$, and $\Sigma_{\boldsymbol{q}\widehat{\boldsymbol{q}}}^{*}$ in [18, Theorem 4], respectively. Therefore, combining with Remark 2.2, the model in [18] can be seen as a specific case.

3 Infinite-Dimensional Systems

3.1 Preliminaries

Let $\mathscr{X}$, $\mathscr{Y}$, and $\mathscr{U}$ be real separable Hilbert spaces. In what follows, for simplicity of notation, when there is no confusion, we shall use $\langle\cdot,\cdot\rangle$ for inner products in possibly different Hilbert spaces and $|\cdot|$ for norms of elements and operators in different Hilbert spaces. Denote the spaces of bounded linear operators by $\mathscr{L}(\cdot)$ and $\mathscr{L}(\cdot,\cdot)$, for example, $\mathscr{L}(\mathscr{X})$, $\mathscr{L}(\mathscr{X},\mathscr{Y})$.

3.2 Problem formulation

Consider the following infinite-dimensional linear system:

$$\begin{cases}
\mathrm{d}y(t)=(\mathcal{A}y(t)+\theta)\mathrm{d}t+\mathcal{B}u\mathrm{d}t+\mathrm{d}w(t), & t\in[0,T],\\
y(0)=y_0,\\
\mathrm{d}z(t)=\mathcal{C}y(t)\mathrm{d}t+\mathrm{d}v(t), & t\in[0,T),\\
z(0)=0,
\end{cases} \tag{3.1}$$

where $\mathcal{A}$ is a linear closed operator on $\mathscr{X}$ which generates a contraction semigroup $\mathscr{S}(t)$, $\mathcal{B} \in \mathscr{L}(\mathscr{U}, \mathscr{X})$, $\mathcal{C} \in \mathscr{L}(\mathscr{X}, \mathscr{Y})$, y_0 is an $\mathscr{X}$-valued Gaussian random variable independent of $w(t)$ and $v(t)$, with zero expectation and covariance operator $\mathcal{R}_{y_0}$. $w(t)$ and $v(t)$ are independent Wiener processes on $\mathscr{X}$ and $\mathscr{Y}$ with covariance operators $\mathcal{R}_w$ and $\mathcal{R}_v$, respectively. Then (3.1) has the unique mild solution $y(t)$; the definition of the mild solution and the well-posedness of the solution can be found in [24, Definition 3.9, Theorem 3.13].

We assume that the observation space, $\mathscr{Y} = \mathbb{R}^m$, is finite-dimensional. $\theta \in \mathscr{X}$ is the unknown parameter. For admissible controls we take $u(t)$ which are adapted to $\sigma(w(s), 0 \leq s \leq t)$ and satisfy $\int_0^T \mathbb{E}|u(t)|^2 \mathrm{d}t < \infty$.

Given a $T_0 \in [0, T)$, we are allowed to observe the system for $t \in [0, T_0]$ without applying any control, after which we control the system on $t \in [T_0, T]$. Our objective is to find the optimal control strategy $u(t) \in \mathscr{U}$, depending on the measurements $\{z(\tau) : 0 \leq \tau < t\}$ and taking values in $\mathscr{Y}$, that minimizes the cost function

$$\mathcal{V}(y, u; \theta) = \int_{T_0}^{T} \mathbb{E}\{\langle \mathcal{M}y(t), y(t)\rangle + \langle \mathcal{N}u(t), u(t)\rangle\}\mathrm{d}t, \tag{3.2}$$

where $0 \leq \mathcal{M} \in \mathscr{L}(\mathscr{X})$ and $0 < \mathcal{N} \in \mathscr{L}(\mathscr{U})$ with $\mathcal{N}^{-1} \in \mathscr{L}(\mathscr{U})$. For the sake of brevity, we let $\mathcal{M}$, $\mathcal{N}$ be self-adjoint operators.

In the Bayesian version, a prior belief is assumed on θ, that is, θ is chosen at random according to a probability measure $\mu(\theta)$. The expected cost associated with u can then be computed as

$$\mathscr{V}(y, u; \mu) = \mathbb{E}_\theta[\mathcal{V}(y, u; \theta)] = \int_{\mathscr{X}} \mathcal{V}(y, u; \theta)\mathrm{d}\mu(\theta). \tag{3.3}$$

In this case, the objective is to find a control strategy $u(t)$ that minimizes the expected value in (3.3) given the prior belief $\mu(\theta)$.

3.3 A Bayesian strategy for unknown θ

This subsection demonstrates the optimal strategy for the Bayesian problem given a fixed prior belief $\mu(\theta)$. As in the finite-dimensional case, we restrict ourselves to Gaussian prior beliefs only, and assume the covariance operator is $\mathcal{R} \succeq 0$. We set θ as an element of states. Then the original system (3.1) can be rewritten as

$$\begin{cases} \mathrm{d}x(t) = \mathcal{F}x(t)\mathrm{d}t + \mathcal{D}u\mathrm{d}t + \mathcal{G}\mathrm{d}w(t), & t \in [0, T], \\ x(0) = x_0, & \\ \mathrm{d}z(t) = \mathcal{H}x(t)\mathrm{d}t + \mathrm{d}v(t), & t \in [0, T), \\ z(0) = 0, & \end{cases} \tag{3.4}$$

where $x = \left[\begin{smallmatrix} y \\ \theta \end{smallmatrix}\right] \in \mathscr{X} \times \mathscr{X}$, $\mathcal{F} = \left[\begin{smallmatrix} \mathcal{A} & \mathcal{I} \\ 0 & 0 \end{smallmatrix}\right]$, $\mathcal{D} = \left[\begin{smallmatrix} \mathcal{B} \\ 0 \end{smallmatrix}\right]$, $\mathcal{G} = \left[\begin{smallmatrix} \mathcal{I} \\ 0 \end{smallmatrix}\right]$, $x_0 = \left[\begin{smallmatrix} y_0 \\ \theta \end{smallmatrix}\right]$, and $\mathcal{H} = [\mathcal{C} \quad 0]$, $\mathcal{I}$ is the identity map operator. Again, $\mathcal{F}$ is a linear closed operator on $\mathscr{X} \times \mathscr{X}$ which generates a contraction semigroup $\mathscr{E}(t)$. Then $x(t)$ has the unique mild solution

$$x(t) = \mathscr{E}(t)x_0 + \int_0^t \mathscr{E}(t-s)\mathcal{D}u(s)\mathrm{d}s + \int_0^t \mathscr{E}(t-s)\mathcal{G}\mathrm{d}w(s),$$

also we see that $z(t)$ is a well-defined stochastic process with differential given by (3.4).

Let $\widehat{x}(t) = (\widehat{y}(t), \widehat{\theta}(t))^{\mathrm{T}} = \mathbb{E}[x(t) \mid x(0), z(\tau) : \tau \in [0, t)]$ be the Bayesian estimators of $y(t)$ and θ at time t. Then $\widehat{x}(t)$ is the solution to the filtering problem and is the weak solution of the stochastic evolution equation

$$\mathrm{d}\widehat{x}(t) = \mathcal{F}(t)\widehat{x}(t)\mathrm{d}t + \mathcal{D}u(t)\mathrm{d}t + \mathcal{K}(t)(\mathrm{d}z(t) - \mathcal{H}\widehat{x}(t)\mathrm{d}t) \tag{3.5}$$

given by Curtain [13–14], where

$$\mathcal{K}(t) = \mathcal{P}(t)\mathcal{H}^*\mathcal{R}_v^{-1}.$$

For arbitrary $\varphi, \psi \in \mathscr{D}(\mathcal{F}^*)$ the operator-valued matrix $\mathcal{P} = \left[\begin{smallmatrix} \mathcal{P}_{11} & \mathcal{P}_{12} \\ \mathcal{P}_{12}^* & \mathcal{P}_{22} \end{smallmatrix}\right]$ satisfying the following inner product version of the infinite-dimensional Riccati equation:

$$\begin{cases} \langle \Psi(\mathcal{P})\varphi, \psi \rangle = 0, \\ \mathcal{P}(0) = \mathcal{P}_0 = \begin{bmatrix} \mathcal{R}_{y_0} & 0 \\ 0 & \mathcal{R} \end{bmatrix}, \end{cases} \tag{3.6}$$

where $\Psi(\mathcal{P}) = \frac{\mathrm{d}\mathcal{P}(t)}{\mathrm{d}t} - \mathcal{F}\mathcal{P}(t) - \mathcal{P}(t)\mathcal{F}^* - \mathcal{G}\mathcal{R}_w\mathcal{G}^* + \mathcal{P}(t)\mathcal{H}^*\mathcal{R}_v^{-1}\mathcal{H}\mathcal{P}(t)$.

Let $\widetilde{\mathcal{M}} = \left[\begin{smallmatrix} \mathcal{M} & 0 \\ 0 & 0 \end{smallmatrix}\right]$. The cost to go at time t is

$$\begin{aligned} \mathscr{V}(t, x, z; \mu) &= \mathbb{E}\Big[\int_t^T \langle \widetilde{\mathcal{M}}x(\tau), x(\tau)\rangle + \langle \mathcal{N}u(\tau), u(\tau)\rangle \mathrm{d}\tau\Big] \\ &= \mathbb{E}\Big[\int_t^T \langle \widetilde{\mathcal{M}}\widehat{x}(\tau), \widehat{x}(\tau)\rangle + \langle \mathcal{N}u(\tau), u(\tau)\rangle \mathrm{d}\tau\Big] + \int_t^T \mathrm{Tr}(\mathcal{P}(\tau)\widetilde{\mathcal{M}})\mathrm{d}\tau. \end{aligned}$$

The optimal value of $\mathscr{V}$ solves the Hamilton-Jacobi-Bellman equation and takes the form

$$\mathscr{V}(t) = \langle \mathcal{Q}(t)\widehat{x}, \widehat{x}\rangle + r(t).$$

Then for arbitrary $\varphi \in \mathscr{D}(\mathcal{F})$, we obtain an operator-valued matrix $\mathcal{Q} = \left[\begin{smallmatrix} \mathcal{Q}_{11} & \mathcal{Q}_{12} \\ \mathcal{Q}_{12}^* & \mathcal{Q}_{22} \end{smallmatrix}\right]$ and a scalar-valued function r that satisfy the control Riccati equations

$$\begin{cases} \dfrac{\mathrm{d}}{\mathrm{d}t}\langle \mathcal{Q}(t)\varphi, \varphi\rangle + 2\langle \mathcal{F}\varphi, \mathcal{Q}(t)\varphi\rangle + \langle \widetilde{\mathcal{M}}\varphi, \varphi\rangle - \langle \mathcal{Q}(t)\mathcal{D}\mathcal{N}^{-1}\mathcal{D}^*\mathcal{Q}(t)\varphi, \varphi\rangle = 0, \\ \mathcal{Q}(T) = 0, \\ r(t) = \displaystyle\int_t^T \mathrm{Tr}(\mathcal{K}^*\mathcal{Q}(\tau)\mathcal{K}\mathcal{R}_v)\mathrm{d}\tau, \\ r(T) = 0. \end{cases} \tag{3.7}$$

Following [21, Theorem 4.1], the feedback control is

$$\overline{u} = -\mathcal{N}^{-1}\mathcal{D}^*\mathcal{Q}(t)\widehat{x} = -\mathcal{N}^{-1}\mathcal{B}^*(\mathcal{Q}_{11}\widehat{y} + \mathcal{Q}_{12}\widehat{\theta}).$$

3.4 Optimal cost for a Bayesian strategy with unknown $\boldsymbol{\theta}$

The total cost incurred for θ by the Bayesian strategy arising from a prior μ is

$$\mathscr{V}(y,\theta;\mu) = \int_{T_0}^{T} \mathbb{E}\{\langle \mathcal{M}y(t), y(t)\rangle + \langle \mathcal{N}\overline{u}(t), \overline{u}(t)\rangle\}\mathrm{d}t, \tag{3.8}$$

where $\overline{u} = -\mathcal{N}^{-1}\mathcal{B}^*(\mathcal{Q}_{11}\widehat{y}+\mathcal{Q}_{12}\widehat{\theta})$. From (3.5), $\widehat{y}(t)$ and $\widehat{\theta}(t)$ respectively satisfy the continuous-time Kalman filter equations

$$\begin{aligned}
\mathrm{d}\widehat{y}(t) &= (\mathcal{A}\widehat{y}(t) + \widehat{\theta} + \mathcal{B}u)\mathrm{d}t + \mathcal{P}_{11}(t)\mathcal{C}^*\mathcal{R}_v^{-1}(\mathrm{d}z - \mathcal{C}\widehat{y}\mathrm{d}t),\\
\mathrm{d}\widehat{\theta}(t) &= \mathcal{P}_{12}^*(t)\mathcal{C}^*\mathcal{R}_v^{-1}(\mathrm{d}z - \mathcal{C}\widehat{y}\mathrm{d}t).
\end{aligned}$$

With control $\overline{u}$, the stochastic differential equations describing y, $\widehat{y}$, and $\widehat{\theta}$ become

$$\mathrm{d}y = [\mathcal{A}y + \theta - \mathcal{B}\mathcal{N}^{-1}\mathcal{B}^*(\mathcal{Q}_{11}\widehat{y} + \mathcal{Q}_{12}\widehat{\theta})]\mathrm{d}t + \mathrm{d}w, \tag{3.9}$$

$$\mathrm{d}\widehat{y} = [\mathcal{A}\widehat{y} + \widehat{\theta} - \mathcal{B}\mathcal{N}^{-1}\mathcal{B}^*(\mathcal{Q}_{11}\widehat{y} + \mathcal{Q}_{12}\widehat{\theta})]\mathrm{d}t + \mathcal{P}_{11}(t)\mathcal{C}^*\mathcal{R}_v^{-1}[\mathcal{C}(y-\widehat{y})\mathrm{d}t + \mathrm{d}v], \tag{3.10}$$

$$\mathrm{d}\widehat{\theta} = \mathcal{P}_{12}^*(t)\mathcal{C}^*\mathcal{R}_v^{-1}[\mathcal{C}(y-\widehat{y})\mathrm{d}t + \mathrm{d}v]. \tag{3.11}$$

Let $\overline{\theta}(t) \triangleq \mathbb{E}[\widehat{\theta}(t)], \overline{y}(t) \triangleq \mathbb{E}[\widehat{y}(t)]$, and $\widetilde{y}(t) \triangleq \mathbb{E}[y(t)]$ with the associated covariances $\mathcal{R}_{\widehat{\theta}\widehat{\theta}}, \mathcal{R}_{\widehat{y}\widehat{y}}$, $\mathcal{R}_{\widehat{y}\widehat{\theta}}, \mathcal{R}_{yy}, \mathcal{R}_{y\widehat{y}}$, and $\mathcal{R}_{y\widehat{\theta}}$.

Theorem 3.1 *The optimal cost incurred by a specific Bayesian strategy can be expressed as*

$$\mathscr{V}(y,\theta;\mu) = \langle \mathcal{X}(\mathcal{R},\mathcal{R}_v,T_0,T)\theta,\theta\rangle + \mathcal{Y}(\mathcal{R},\mathcal{R}_v,T_0,T), \tag{3.12}$$

where

$$\mathcal{X}(\mathcal{R},\mathcal{R}_v,T_0,T) = \int_{T_0}^{T} \mathcal{I}_1^*\mathcal{M}\mathcal{I}_1 + (\mathcal{I}_2^*\mathcal{Q}_{11}^* + \mathcal{I}_3^*\mathcal{Q}_{12}^*)\mathcal{B}\mathcal{N}^{-1}\mathcal{B}^*(\mathcal{Q}_{11}\mathcal{I}_2 + \mathcal{Q}_{12}\mathcal{I}_3)\mathrm{d}t$$

and

$$\begin{aligned}
&\mathcal{Y}(\mathcal{R},\mathcal{R}_v,T_0,T)\\
&= \int_{T_0}^{T} \mathrm{Tr}(\mathcal{M}\mathcal{R}_{yy} + \mathcal{Q}_{11}^*\mathcal{B}\mathcal{N}^{-1}\mathcal{B}^*\mathcal{Q}_{11}\mathcal{R}_{\widehat{y}\widehat{y}} + 2\mathcal{Q}_{11}^*\mathcal{B}\mathcal{N}^{-1}\mathcal{B}^*\mathcal{Q}_{12}\mathcal{R}_{\widehat{y}\widehat{\theta}}\mathcal{Q}_{12}^*\mathcal{B}\mathcal{N}^{-1}\mathcal{B}^*\mathcal{Q}_{12}\mathcal{R}_{\widehat{\theta}\widehat{\theta}})\mathrm{d}t
\end{aligned}$$

for time-varying functions $\mathcal{I}_1(t)$, $\mathcal{I}_2(t)$, and $\mathcal{I}_3(t)$. All of the quantities appearing here can be found by solving the following system of ODEs:

$$\dot{\mathcal{I}}_1 = \mathcal{A}\mathcal{I}_1 + \mathcal{I} - \mathcal{B}\mathcal{N}^{-1}\mathcal{B}^*(\mathcal{Q}_{11}\mathcal{I}_2 + \mathcal{Q}_{12}\mathcal{I}_3),$$

$$\begin{aligned}
\dot{\mathcal{I}}_2 &= \mathcal{A}\mathcal{I}_2 + \mathcal{I}_3 - \mathcal{B}\mathcal{N}^{-1}\mathcal{B}^*(\mathcal{Q}_{11}\mathcal{I}_2 + \mathcal{Q}_{12}\mathcal{I}_3) + \mathcal{P}_{11}\mathcal{C}^*\mathcal{R}_v^{-1}\mathcal{C}(\mathcal{I}_1 - \mathcal{I}_2),\\
\dot{\mathcal{I}}_3 &= \mathcal{P}_{12}^*\mathcal{C}^*\mathcal{R}_v^{-1}\mathcal{C}(\mathcal{I}_1 - \mathcal{I}_2),\\
\dot{\mathcal{R}}_{\widehat{\theta}\widehat{\theta}} &= (\mathcal{R}_{y\widehat{\theta}}^* - \mathcal{R}_{\widehat{y}\widehat{\theta}}^*)\mathcal{C}^*\mathcal{R}_v^{-1}\mathcal{C}\mathcal{P}_{12} + \mathcal{P}_{12}^*\mathcal{C}^*\mathcal{R}_v^{-1}\mathcal{C}(\mathcal{R}_{y\widehat{\theta}} - \mathcal{R}_{\widehat{y}\widehat{\theta}}) + \mathcal{P}_{12}^*\mathcal{C}^*\mathcal{R}_v^{-1}\mathcal{C}\mathcal{P}_{12},\\
\dot{\mathcal{R}}_{yy} &= \mathcal{R}_w - (\mathcal{R}_{y\widehat{y}}\mathcal{Q}_{11}^* + \mathcal{R}_{y\widehat{\theta}}\mathcal{Q}_{12}^*)\mathcal{B}\mathcal{N}^{-1}\mathcal{B}^*\\
&\quad - \mathcal{B}\mathcal{N}^{-1}\mathcal{B}^*(\mathcal{Q}_{11}\mathcal{R}_{y\widehat{y}}^* + \mathcal{Q}_{12}\mathcal{R}_{y\widehat{\theta}}^*) + \mathcal{R}_{yy}\mathcal{A}^* + \mathcal{A}\mathcal{R}_{yy},\\
\dot{\mathcal{R}}_{y\widehat{y}} &= (\mathcal{R}_{yy} - \mathcal{R}_{y\widehat{y}})\mathcal{C}^*\mathcal{R}_v^{-1}\mathcal{C}\mathcal{P}_{11}^* - \mathcal{R}_{y\widehat{y}}\mathcal{Q}_{11}^*\mathcal{B}\mathcal{N}^{-1}\mathcal{B}^* + \mathcal{R}_{y\widehat{\theta}}(\mathcal{I} - \mathcal{Q}_{12}^*\mathcal{B}\mathcal{N}^{-1}\mathcal{B})\\
&\quad - \mathcal{B}\mathcal{N}^{-1}\mathcal{B}^*(\mathcal{Q}_{11}\mathcal{R}_{\widehat{y}\widehat{y}} + \mathcal{Q}_{12}\mathcal{R}_{\widehat{y}\widehat{\theta}}^*) + \mathcal{R}_{y\widehat{y}}\mathcal{A}^* + \mathcal{A}\mathcal{R}_{y\widehat{y}},\\
\dot{\mathcal{R}}_{y\widehat{\theta}} &= (\mathcal{R}_{yy} - \mathcal{R}_{y\widehat{y}})\mathcal{C}^*\mathcal{R}_v^{-1}\mathcal{C}\mathcal{P}_{12} - \mathcal{B}\mathcal{N}^{-1}\mathcal{B}^*(\mathcal{Q}_{11}\mathcal{R}_{\widehat{y}\widehat{\theta}} + \mathcal{Q}_{12}\mathcal{R}_{\widehat{\theta}\widehat{\theta}}) + \mathcal{A}\mathcal{R}_{y\widehat{\theta}},\\
\dot{\mathcal{R}}_{\widehat{y}\widehat{y}} &= \mathcal{P}_{11}\mathcal{C}^*\mathcal{R}_v^{-1}\mathcal{C}\mathcal{P}_{11}^* + \mathcal{R}_{\widehat{y}\widehat{\theta}}(\mathcal{I} - \mathcal{Q}_{12}^*\mathcal{B}\mathcal{N}^{-1}\mathcal{B}) + (\mathcal{I} - \mathcal{B}\mathcal{N}^{-1}B^*\mathcal{Q}_{12})\mathcal{R}_{\widehat{y}\widehat{\theta}}^*\\
&\quad + (\mathcal{R}_{y\widehat{y}}^* - \mathcal{R}_{\widehat{y}\widehat{y}}^*)\mathcal{C}^*\mathcal{R}_v^{-1}\mathcal{C}\mathcal{P}_{11}^* + \mathcal{P}_{11}\mathcal{C}^*\mathcal{R}_v^{-1}\mathcal{C}(\mathcal{R}_{y\widehat{y}} - \mathcal{R}_{\widehat{y}\widehat{y}}) - \mathcal{R}_{\widehat{y}\widehat{y}}\mathcal{Q}_{11}^*\mathcal{B}\mathcal{N}^{-1}\mathcal{B}^*\\
&\quad - \mathcal{B}\mathcal{N}^{-1}\mathcal{B}^*\mathcal{Q}_{11}\mathcal{R}_{\widehat{y}\widehat{y}} + \mathcal{R}_{\widehat{y}\widehat{y}}\mathcal{A}^* + \mathcal{A}\mathcal{R}_{\widehat{y}\widehat{y}},\\
\dot{\mathcal{R}}_{\widehat{y}\widehat{\theta}} &= (\mathcal{R}_{y\widehat{y}}^* - \mathcal{R}_{\widehat{y}\widehat{y}}^*)\mathcal{C}^*\mathcal{R}_v^{-1}\mathcal{C}\mathcal{P}_{12} + (\mathcal{I} - \mathcal{B}\mathcal{N}^{-1}\mathcal{B}^*\mathcal{Q}_{12})\mathcal{R}_{\widehat{\theta}\widehat{\theta}} - \mathcal{B}\mathcal{N}^{-1}\mathcal{B}^*\mathcal{Q}_{12}\mathcal{R}_{\widehat{y}\widehat{\theta}}\\
&\quad + \mathcal{P}_{11}\mathcal{C}^*\mathcal{R}_v^{-1}\mathcal{C}(\mathcal{R}_{y\widehat{\theta}} - \mathcal{R}_{\widehat{y}\widehat{\theta}} + \mathcal{P}_{12}) + \mathcal{A}\mathcal{R}_{\widehat{y}\widehat{\theta}}.
\end{aligned}$$

Proof The cost (3.8) can be written as

$$\begin{aligned}
\mathscr{V}(y,\theta;\mu) &= \int_{T_0}^T \mathbb{E}[\langle \mathcal{M}y(t), y(t)\rangle + \langle \mathcal{N}\overline{u}(t), \overline{u}(t)\rangle]\mathrm{d}t\\
&= \int_{T_0}^T [\langle \mathcal{M}\widetilde{y}, \widetilde{y}\rangle + \langle \mathcal{Q}_{11}\overline{y} + \mathcal{Q}_{12}\overline{\theta}, \mathcal{B}\mathcal{N}^{-1}\mathcal{B}^*(\mathcal{Q}_{11}\overline{y} + \mathcal{Q}_{12}\overline{\theta})\rangle]\mathrm{d}t\\
&\quad + \int_{T_0}^T \mathrm{Tr}(\mathcal{M}\mathcal{R}_{yy} + \mathcal{Q}_{11}^*\mathcal{B}\mathcal{N}^{-1}\mathcal{B}^*\mathcal{Q}_{11}\mathcal{R}_{\widehat{y}\widehat{y}} + 2\mathcal{Q}_{11}^*\mathcal{B}\mathcal{N}^{-1}\mathcal{B}^*\mathcal{Q}_{12}\mathcal{R}_{\widehat{y}\widehat{\theta}}\\
&\quad + \mathcal{Q}_{12}^*\mathcal{B}\mathcal{N}^{-1}\mathcal{B}^*\mathcal{Q}_{12}\mathcal{R}_{\widehat{\theta}\widehat{\theta}})\mathrm{d}t. \qquad (3.13)
\end{aligned}$$

To compute this cost, we need to derive the ODEs for the aforementioned mean and covariances from the SDEs (3.9)–(3.11). This can be done by computing the expectations for the discretized dynamics with a time step Δt and taking the limit as $\Delta t \to 0$. For example,

$$\begin{aligned}
\Delta\overline{\theta} &= \mathbb{E}[\Delta\widehat{\theta}] = \mathbb{E}[\mathcal{P}_{12}^*(t)\mathcal{C}^*\mathcal{R}_v^{-1}[\mathcal{C}(y - \widehat{y})\Delta t + \Delta v]]\\
&= \mathcal{P}_{12}^*(t)\mathcal{C}^*\mathcal{R}_v^{-1}[\mathcal{C}(\widetilde{y} - \overline{y})\Delta t],
\end{aligned}$$

and therefore,

$$\dot{\overline{\theta}} = \mathcal{P}_{12}^*(t)\mathcal{C}^*\mathcal{R}_v^{-1}\mathcal{C}(\widetilde{y} - \overline{y}).$$

Similarly, we get

$$\dot{\widetilde{y}} = \mathcal{A}\widetilde{y} + \theta - \mathcal{B}\mathcal{N}^{-1}\mathcal{B}^*(\mathcal{Q}_{11}\overline{y} + \mathcal{Q}_{12}\overline{\theta}),$$

$$\dot{\overline{y}} = \mathcal{A}\overline{y} + \overline{\theta} - \mathcal{B}\mathcal{N}^{-1}\mathcal{B}^*(\mathcal{Q}_{11}\overline{y} + \mathcal{Q}_{12}\overline{\theta}) + \mathcal{P}_{11}(t)\mathcal{C}^*\mathcal{R}_v^{-1}(\widetilde{y} - \overline{y}).$$

The ODEs for $\widetilde{y}$, $\overline{y}$, and $\overline{\theta}$ yield $\widetilde{y} = \mathcal{I}_1(t)\theta$, $\overline{y} = \mathcal{I}_2(t)\theta$ and $\overline{\theta} = \mathcal{I}_3(t)\theta$ for some $\mathcal{I}_1$, $\mathcal{I}_2$, $\mathcal{I}_3 : [0,T] \to \mathscr{L}(\mathscr{X})$ that satisfy the same set of ODEs:

$$\dot{\mathcal{I}}_1 = \mathcal{A}\mathcal{I}_1 + \mathcal{I} - \mathcal{B}\mathcal{N}^{-1}\mathcal{B}^*(\mathcal{Q}_{11}\mathcal{I}_2 + \mathcal{Q}_{12}\mathcal{I}_3), \tag{3.14}$$

$$\dot{\mathcal{I}}_2 = \mathcal{A}\mathcal{I}_2 + \mathcal{I}_3 - \mathcal{B}\mathcal{N}^{-1}\mathcal{B}^*(\mathcal{Q}_{11}\mathcal{I}_2 + \mathcal{Q}_{12}\mathcal{I}_3) + \mathcal{P}_{11}\mathcal{C}^*\mathcal{R}_v^{-1}\mathcal{C}(\mathcal{I}_1 - \mathcal{I}_2), \tag{3.15}$$

$$\dot{\mathcal{I}}_3 = \mathcal{P}_{12}^*\mathcal{C}^*\mathcal{R}_v^{-1}\mathcal{C}(\mathcal{I}_1 - \mathcal{I}_2). \tag{3.16}$$

Substituting (3.14)–(3.16) into the first term of the last equation of (3.13), we can get

$$\begin{aligned}
&\int_{T_0}^{T} [\langle \mathcal{M}\widetilde{y}, \widetilde{y}\rangle + \langle \mathcal{Q}_{11}\overline{y} + \mathcal{Q}_{12}\overline{\theta}, \mathcal{B}\mathcal{N}^{-1}\mathcal{B}^*(\mathcal{Q}_{11}\overline{y} + \mathcal{Q}_{12}\overline{\theta})\rangle]\mathrm{d}t \\
&= \Big\langle \Big[\int_{T_0}^{T} \mathcal{I}_1^*\mathcal{M}\mathcal{I}_1 + (\mathcal{I}_2^*\mathcal{Q}_{11}^* + \mathcal{I}_3^*\mathcal{Q}_{12}^*)\mathcal{B}\mathcal{N}^{-1}\mathcal{B}^*\mathcal{Q}_{11}\mathcal{I}_2 + \mathcal{Q}_{12}\mathcal{I}_3)\mathrm{d}t\Big]\theta, \theta\Big\rangle.
\end{aligned}$$

Again,

$$\begin{aligned}
\mathcal{R}_{\widehat{\theta}\widehat{\theta}}(t + \Delta t) &= \mathbb{E}[\{\widehat{\theta}(t+\Delta t) - \overline{\theta}(t+\Delta t)\} \circ \{\widehat{\theta}(t+\Delta t) - \overline{\theta}(t+\Delta t)\}] \\
&= \mathbb{E}[\gamma \circ \gamma] \\
&= \mathcal{R}_{\widehat{\theta}\widehat{\theta}}(t) + [(\mathcal{R}_{y\widehat{\theta}}^* - \mathcal{R}_{\widehat{y}\widehat{\theta}}^*)\mathcal{C}^*\mathcal{R}_v^{-1}\mathcal{C}\mathcal{P}_{12} + \mathcal{P}_{12}^*\mathcal{C}^*\mathcal{R}_v^{-1}\mathcal{C}(\mathcal{R}_{y\widehat{\theta}} - \mathcal{R}_{\widehat{y}\widehat{\theta}}) \\
&\quad + \mathcal{P}_{12}^*\mathcal{C}^*\mathcal{R}_v^{-1}\mathcal{C}\mathcal{P}_{12}]\Delta t,
\end{aligned}$$

where "$\circ$" denotes the Cartesian product, $\gamma = \widehat{\theta} - \overline{\theta} + \mathcal{P}_{12}^*\mathcal{C}^*\mathcal{R}_v^{-1}\mathcal{C}[(y - \widetilde{y}) - (\widehat{y} - \overline{y})]\Delta t + \mathcal{P}_{12}^*\mathcal{C}^*\mathcal{R}_v^{-1}\Delta v$. Therefore,

$$\dot{\mathcal{R}}_{\widehat{\theta}\widehat{\theta}} = (\mathcal{R}_{y\widehat{\theta}}^* - \mathcal{R}_{\widehat{y}\widehat{\theta}}^*)\mathcal{C}^*\mathcal{R}_v^{-1}\mathcal{C}\mathcal{P}_{12} + \mathcal{P}_{12}^*\mathcal{C}^*\mathcal{R}_v^{-1}\mathcal{C}(\mathcal{R}_{y\widehat{\theta}} - \mathcal{R}_{\widehat{y}\widehat{\theta}}) + \mathcal{P}_{12}^*\mathcal{C}^*\mathcal{R}_v^{-1}\mathcal{C}\mathcal{P}_{12}.$$

Similarly, we have

$$\begin{aligned}
\dot{\mathcal{R}}_{yy} &= \mathcal{R}_w - (\mathcal{R}_{y\widehat{y}}\mathcal{Q}_{11}^* + \mathcal{R}_{y\widehat{\theta}}\mathcal{Q}_{12}^*)\mathcal{B}\mathcal{N}^{-1}\mathcal{B}^* \\
&\quad - \mathcal{B}\mathcal{N}^{-1}\mathcal{B}^*(\mathcal{Q}_{11}\mathcal{R}_{y\widehat{y}}^* + \mathcal{Q}_{12}\mathcal{R}_{y\widehat{\theta}}^*) + \mathcal{R}_{yy}\mathcal{A}^* + \mathcal{A}\mathcal{R}_{yy}, \\
\dot{\mathcal{R}}_{y\widehat{y}} &= (\mathcal{R}_{yy} - \mathcal{R}_{y\widehat{y}})\mathcal{C}^*\mathcal{R}_v^{-1}\mathcal{C}\mathcal{P}_{11}^* - \mathcal{R}_{y\widehat{y}}\mathcal{Q}_{11}^*\mathcal{B}\mathcal{N}^{-1}\mathcal{B}^* + \mathcal{R}_{y\widehat{\theta}}(\mathcal{I} - \mathcal{Q}_{12}^*\mathcal{B}\mathcal{N}^{-1}\mathcal{B}) \\
&\quad - \mathcal{B}\mathcal{N}^{-1}\mathcal{B}^*(\mathcal{Q}_{11}\mathcal{R}_{\widehat{y}\widehat{y}} + \mathcal{Q}_{12}\mathcal{R}_{\widehat{y}\widehat{\theta}}^*) + \mathcal{R}_{y\widehat{y}}\mathcal{A}^* + \mathcal{A}\mathcal{R}_{y\widehat{y}}, \\
\dot{\mathcal{R}}_{y\widehat{\theta}} &= (\mathcal{R}_{yy} - \mathcal{R}_{y\widehat{y}})\mathcal{C}^*\mathcal{R}_v^{-1}\mathcal{C}\mathcal{P}_{12} - \mathcal{B}\mathcal{N}^{-1}\mathcal{B}^*(\mathcal{Q}_{11}\mathcal{R}_{\widehat{y}\widehat{\theta}} + \mathcal{Q}_{12}\mathcal{R}_{\widehat{y}\widehat{y}}) + \mathcal{A}\mathcal{R}_{y\widehat{\theta}}, \\
\dot{\mathcal{R}}_{\widehat{y}\widehat{\theta}} &= (\mathcal{R}_{y\widehat{y}}^* - \mathcal{R}_{\widehat{y}\widehat{y}}^*)\mathcal{C}^*\mathcal{R}_v^{-1}\mathcal{C}\mathcal{P}_{12} + (\mathcal{I} - \mathcal{B}\mathcal{N}^{-1}\mathcal{B}^*\mathcal{Q}_{12})\mathcal{R}_{\widehat{\theta}\widehat{\theta}} - \mathcal{B}\mathcal{N}^{-1}\mathcal{B}^*\mathcal{Q}_{12}\mathcal{R}_{\widehat{y}\widehat{\theta}} \\
&\quad + \mathcal{P}_{11}\mathcal{C}^*\mathcal{R}_v^{-1}\mathcal{C}(\mathcal{R}_{y\widehat{\theta}} - \mathcal{R}_{\widehat{y}\widehat{\theta}} + \mathcal{P}_{12}) + \mathcal{A}\mathcal{R}_{\widehat{y}\widehat{\theta}}, \\
\dot{\mathcal{R}}_{\widehat{y}\widehat{y}} &= \mathcal{P}_{11}\mathcal{C}^*\mathcal{R}_v^{-1}\mathcal{C}\mathcal{P}_{11}^* + \mathcal{R}_{\widehat{y}\widehat{\theta}}(\mathcal{I} - \mathcal{Q}_{12}^*\mathcal{B}\mathcal{N}^{-1}\mathcal{B}) + (\mathcal{I} - \mathcal{B}\mathcal{N}^{-1}B^*\mathcal{Q}_{12})\mathcal{R}_{\widehat{y}\widehat{\theta}}^* \\
&\quad + (\mathcal{R}_{y\widehat{y}}^* - \mathcal{R}_{\widehat{y}\widehat{y}}^*)\mathcal{C}^*\mathcal{R}_v^{-1}\mathcal{C}\mathcal{P}_{11}^* + \mathcal{P}_{11}\mathcal{C}^*\mathcal{R}_v^{-1}\mathcal{C}(\mathcal{R}_{y\widehat{y}} - \mathcal{R}_{\widehat{y}\widehat{y}}) - \mathcal{R}_{\widehat{y}\widehat{y}}\mathcal{Q}_{11}^*\mathcal{B}\mathcal{N}^{-1}\mathcal{B}^* \\
&\quad - \mathcal{B}\mathcal{N}^{-1}\mathcal{B}^*\mathcal{Q}_{11}\mathcal{R}_{\widehat{y}\widehat{y}} + \mathcal{R}_{\widehat{y}\widehat{y}}\mathcal{A}^* + \mathcal{A}\mathcal{R}_{\widehat{y}\widehat{y}}.
\end{aligned}$$

3.5 Optimal cost for a Bayesian strategy with known θ

If θ is known, then $\widehat{\theta}(t) \equiv \overline{\theta}(t) \equiv \theta, \mathcal{R}_{\widehat{\theta}\widehat{\theta}}(t) \equiv 0, \mathcal{R}_{\widehat{y}\widehat{\theta}}(t) \equiv 0$, and $\mathcal{R}_{y\widehat{\theta}}(t) \equiv 0$. We use similar notation as above, denoting all of the analogous quantities for this case with a superscript #. Hence for arbitrary $\varphi, \psi \in \mathscr{D}(\mathcal{A}^*)$ the Riccati equation (3.6) becomes

$$\begin{cases} \langle \Phi(\mathcal{P}_{11}^{\#}(t))\varphi, \psi \rangle = 0, \\ \mathcal{P}_{11}^{\#}(0) = \mathcal{R}_{y_0}, \end{cases}$$

where $\Phi(\mathcal{P}_{11}^{\#}(t)) = \frac{\mathrm{d}\mathcal{P}_{11}^{\#}(t)}{\mathrm{d}t} - \mathcal{A}\mathcal{P}_{11}^{\#}(t) - \mathcal{P}_{11}^{\#}(t)\mathcal{A}^* - \mathcal{R}_w + \mathcal{P}_{11}^{\#}(t)\mathcal{C}^*\mathcal{R}_v^{-1}\mathcal{C}\mathcal{P}_{11}^{\#}(t)$, and $\mathcal{P}_{12}^{\#}(t) = \mathcal{P}_{22}^{\#}(t) \equiv 0$. The estimation equation (3.5) becomes

$$\mathrm{d}\widehat{y}^{\#}(t) = (\mathcal{A}\widehat{y}^{\#} + \theta + \mathcal{B}u)\mathrm{d}t + \mathcal{P}_{11}^{\#}(t)\mathcal{C}^*\mathcal{R}_v^{-1}(\mathrm{d}z - \mathcal{C}\widehat{y}^{\#}\mathrm{d}t).$$

This, in turn, yields the optimal control

$$\overline{u}^{\#}(t) = -\mathcal{N}^{-1}\mathcal{B}^*(\mathcal{Q}_{11}\widehat{y}^{\#}(t) + \mathcal{Q}_{12}\theta),$$

where $\mathcal{Q}_{11}$ and $\mathcal{Q}_{12}$ are the elements of the operator-valued matrix $\mathcal{Q} = \left[\begin{smallmatrix} \mathcal{Q}_{11} & \mathcal{Q}_{12} \\ \mathcal{Q}_{12}^* & \mathcal{Q}_{22} \end{smallmatrix}\right]$ defined identically to (3.7). With this control, the stochastic differential equations describing $y^{\#}$ and $\widehat{y}^{\#}$ become

$$\begin{aligned} \mathrm{d}y^{\#} &= [\mathcal{A}y^{\#} + \theta - \mathcal{B}\mathcal{N}^{-1}\mathcal{B}^*(\mathcal{Q}_{11}\widehat{y}^{\#} + \mathcal{Q}_{12}\theta)]\mathrm{d}t + \mathrm{d}w, \\ \mathrm{d}\widehat{y}^{\#} &= [\mathcal{A}\widehat{y}^{\#} + \theta - \mathcal{B}\mathcal{N}^{-1}\mathcal{B}^*(\mathcal{Q}_{11}\widehat{y}^{\#} + \mathcal{Q}_{12}\theta)]\mathrm{d}t + \mathcal{P}_{11}^{\#}(t)\mathcal{C}^*\mathcal{R}_v^{-1}[\mathcal{C}(y^{\#} - \widehat{y}^{\#}\mathrm{d}t) + \mathrm{d}v]. \end{aligned}$$

Let $\overline{y}^{\#}(t) \triangleq \mathbb{E}[\widehat{y}^{\#}(t)]$ and $\widetilde{y}^{\#}(t) \triangleq \mathbb{E}[y^{\#}(t)]$ with the associated covariances $\mathcal{R}_{\widehat{y}\widehat{y}}^{\#}, \mathcal{R}_{yy}^{\#}$, and $\mathcal{R}_{y\widehat{y}}^{\#}$.

Theorem 3.2 *The optimal cost for the Bayesian strategy with known θ can be expressed as*

$$\mathcal{V}(y, \theta) = \langle \mathcal{X}^{\#}(\mathcal{R}_v, T_0, T)\theta, \theta \rangle + \mathcal{Y}^{\#}(\mathcal{R}_v, T_0, T), \tag{3.17}$$

where

$$\mathcal{X}^{\#}(\mathcal{R}_v, T_0, T) = \int_{T_0}^{T} [\mathcal{I}_1^{\#}\mathcal{Q}\mathcal{I}_1^{\#} + (\mathcal{Q}_{11}\mathcal{I}_2^{\#} + \mathcal{Q}_{12})^*\mathcal{B}\mathcal{N}^{-1}\mathcal{B}^*(\mathcal{Q}_{11}\mathcal{I}_2^{\#} + \mathcal{Q}_{12})]\mathrm{d}t$$

and

$$\mathcal{Y}^{\#}(\mathcal{R}_v, T_0, T) = \int_{T_0}^{T} \mathrm{Tr}(\mathcal{M}\mathcal{R}_{yy}^{\#} + \mathcal{Q}_{11}^*\mathcal{B}\mathcal{N}^{-1}\mathcal{B}^*\mathcal{Q}_{11}\mathcal{R}_{\widehat{y}\widehat{y}}^{\#})\mathrm{d}t$$

with the time-varying functions $\mathcal{I}_1^{\#}(t), \mathcal{I}_2^{\#}(t), \mathcal{R}_{yy}^{\#}$ and $\mathcal{R}_{\widehat{y}\widehat{y}}^{\#}$ described by the ODEs

$$\begin{aligned} \dot{\mathcal{I}}_1^{\#} &= \mathcal{A}\mathcal{I}_1^{\#} + \mathcal{I} - \mathcal{B}\mathcal{N}^{-1}\mathcal{B}^*(\mathcal{Q}_{11}\mathcal{I}_2^{\#} + \mathcal{Q}_{12}), \\ \dot{\mathcal{I}}_2^{\#} &= \mathcal{A}\mathcal{I}_2^{\#} + \mathcal{I} - \mathcal{B}\mathcal{N}^{-1}\mathcal{B}^*(\mathcal{Q}_{11}\mathcal{I}_2^{\#} + \mathcal{Q}_{12}) + \mathcal{P}_{11}^{\#}\mathcal{C}^*\mathcal{R}_v^{-1}\mathcal{C}(\mathcal{I}_1^{\#} - \mathcal{I}_2^{\#}), \end{aligned}$$

$$\dot{\mathcal{R}}^{\#}_{yy} = \mathcal{R}_w - (\mathcal{R}^{\#}_{y\widehat{y}}\mathcal{Q}^*_{11})\mathcal{B}\mathcal{N}^{-1}\mathcal{B}^* - \mathcal{B}\mathcal{N}^{-1}\mathcal{B}^*\mathcal{Q}_{11}(\mathcal{R}^{\#}_{y\widehat{y}})^* + \mathcal{R}^{\#}_{yy}\mathcal{A}^* + \mathcal{A}\mathcal{R}^{\#}_{yy},$$

$$\dot{\mathcal{R}}^{\#}_{\widehat{y}\widehat{y}} = \mathcal{P}^{\#}_{11}\mathcal{C}^*\mathcal{R}_v^{-1}\mathcal{C}(\mathcal{P}^{\#}_{11})^* + (\mathcal{R}^{\#}_{y\widehat{y}} - \mathcal{R}^{\#}_{\widehat{y}\widehat{y}})^*\mathcal{C}^*\mathcal{R}_v^{-1}\mathcal{C}(\mathcal{P}^{\#}_{11})^* + \mathcal{P}^{\#}_{11}\mathcal{C}^*\mathcal{R}_v^{-1}\mathcal{C}(\mathcal{R}^{\#}_{y\widehat{y}} - \mathcal{R}^{\#}_{\widehat{y}\widehat{y}})$$
$$- \mathcal{R}^{\#}_{\widehat{y}\widehat{y}}\mathcal{Q}^*_{11}\mathcal{B}\mathcal{N}^{-1}\mathcal{B}^* - \mathcal{B}\mathcal{N}^{-1}\mathcal{B}^*\mathcal{Q}_{11}\mathcal{R}^{\#}_{\widehat{y}\widehat{y}} + \mathcal{R}^{\#}_{\widehat{y}\widehat{y}}\mathcal{A}^* + \mathcal{A}\mathcal{R}^{\#}_{\widehat{y}\widehat{y}},$$

$$\dot{\mathcal{R}}^{\#}_{y\widehat{y}} = (\mathcal{R}^{\#}_{yy} - \mathcal{R}^{\#}_{y\widehat{y}})\mathcal{C}^*\mathcal{R}_v^{-1}\mathcal{C}(\mathcal{P}^{\#}_{11})^* - \mathcal{R}^{\#}_{y\widehat{y}}\mathcal{Q}^*_{11}\mathcal{B}\mathcal{N}^{-1}\mathcal{B}^*$$
$$- \mathcal{B}\mathcal{N}^{-1}\mathcal{B}^*(\mathcal{Q}_{11}\mathcal{R}^{\#}_{\widehat{y}\widehat{y}}) + \mathcal{R}^{\#}_{y\widehat{y}}\mathcal{A}^* + \mathcal{A}\mathcal{R}^{\#}_{y\widehat{y}}.$$

Proof The cost (3.2) can be written as

$$\begin{aligned}\mathcal{V}(y,\theta) &= \int_{T_0}^{T} \mathbb{E}\{\langle \mathcal{M}y(t), y(t)\rangle + \langle \mathcal{N}\overline{u}^{\#}(t), \overline{u}^{\#}(t)\rangle\}\mathrm{d}t \\ &= \int_{T_0}^{T} \{\langle \mathcal{M}\widetilde{y}^{\#}, \widetilde{y}^{\#}\rangle + \langle \mathcal{Q}_{11}\overline{y}^{\#} + \mathcal{Q}_{12}\theta, \mathcal{B}\mathcal{N}^{-1}\mathcal{B}^*(\mathcal{Q}_{11}\overline{y}^{\#} + \mathcal{Q}_{12}\theta)\rangle\}\mathrm{d}t \\ &\quad + \int_{T_0}^{T} \mathrm{Tr}(\mathcal{M}\mathcal{R}^{\#}_{yy} + \mathcal{Q}^*_{11}\mathcal{B}\mathcal{N}^{-1}\mathcal{B}^*\mathcal{Q}_{11}\mathcal{R}^{\#}_{\widehat{y}\widehat{y}})\mathrm{d}t. \end{aligned} \tag{3.18}$$

By the same calculations as in Theorem 3.1, we can get

$$\dot{\widetilde{y}}^{\#} = \mathcal{A}\widetilde{y}^{\#} + \theta - \mathcal{B}\mathcal{N}^{-1}\mathcal{B}^*(\mathcal{Q}_{11}\overline{y}^{\#} + \mathcal{Q}_{12}\theta),$$

$$\dot{\overline{y}}^{\#} = \mathcal{A}\overline{y}^{\#} + \theta - \mathcal{B}\mathcal{N}^{-1}\mathcal{B}^*(\mathcal{Q}_{11}\overline{y}^{\#} + \mathcal{Q}_{12}\theta) + \mathcal{P}^{\#}_{11}\mathcal{C}^*\mathcal{R}_v^{-1}(\widetilde{y}^{\#} - \overline{y}^{\#}),$$

$$\dot{\mathcal{R}}^{\#}_{yy} = \mathcal{R}_w - (\mathcal{R}^{\#}_{y\widehat{y}}\mathcal{Q}^*_{11})\mathcal{B}\mathcal{N}^{-1}\mathcal{B}^* - \mathcal{B}\mathcal{N}^{-1}\mathcal{B}^*\mathcal{Q}_{11}(\mathcal{R}^{\#}_{y\widehat{y}})^* + \mathcal{R}^{\#}_{yy}\mathcal{A}^* + \mathcal{A}\mathcal{R}^{\#}_{yy},$$

$$\dot{\mathcal{R}}^{\#}_{\widehat{y}\widehat{y}} = \mathcal{P}^{\#}_{11}\mathcal{C}^*\mathcal{R}_v^{-1}\mathcal{C}(\mathcal{P}^{\#}_{11})^* + (\mathcal{R}^{\#}_{y\widehat{y}} - \mathcal{R}^{\#}_{\widehat{y}\widehat{y}})^*\mathcal{R}_v^{-1}\mathcal{C}(\mathcal{P}^{\#}_{11})^* + \mathcal{P}^{\#}_{11}\mathcal{C}^*\mathcal{R}_v^{-1}(\mathcal{R}^{\#}_{y\widehat{y}} - \mathcal{R}^{\#}_{\widehat{y}\widehat{y}})$$
$$- \mathcal{R}^{\#}_{\widehat{y}\widehat{y}}\mathcal{Q}^*_{11}\mathcal{B}\mathcal{N}^{-1}\mathcal{B}^* - \mathcal{B}\mathcal{N}^{-1}\mathcal{B}^*\mathcal{Q}_{11}\mathcal{R}^{\#}_{\widehat{y}\widehat{y}} + \mathcal{R}^{\#}_{\widehat{y}\widehat{y}}\mathcal{A}^* + \mathcal{A}\mathcal{R}^{\#}_{\widehat{y}\widehat{y}},$$

$$\dot{\mathcal{R}}^{\#}_{y\widehat{y}} = (\mathcal{R}^{\#}_{yy} - \mathcal{R}^{\#}_{y\widehat{y}})\mathcal{R}_v^{-1}\mathcal{C}(\mathcal{P}^{\#}_{11})^* - \mathcal{R}^{\#}_{y\widehat{y}}\mathcal{Q}^*_{11}\mathcal{B}\mathcal{N}^{-1}\mathcal{B}^*$$
$$- \mathcal{B}\mathcal{N}^{-1}\mathcal{B}^*(\mathcal{Q}_{11}\mathcal{R}^{\#}_{\widehat{y}\widehat{y}}) + \mathcal{R}^{\#}_{y\widehat{y}}\mathcal{A}^* + \mathcal{A}\mathcal{R}^{\#}_{y\widehat{y}}.$$

The ODEs for $\widetilde{y}^{\#}$ and $\overline{y}^{\#}$ yield $\widetilde{y}^{\#} = \mathcal{I}^{\#}_1(t)\theta$ and $\overline{y}^{\#} = \mathcal{I}^{\#}_2(t)\theta$ for some $\mathcal{I}^{\#}_1, \mathcal{I}^{\#}_2 : [0,T] \to \mathscr{L}(\mathscr{X})$ that satisfy the same set of ODEs

$$\dot{\mathcal{I}}^{\#}_1 = \mathcal{A}\mathcal{I}^{\#}_1 + \mathcal{I} - \mathcal{B}\mathcal{N}^{-1}\mathcal{B}^*(\mathcal{Q}_{11}\mathcal{I}^{\#}_2 + \mathcal{Q}_{12}), \tag{3.19}$$

$$\dot{\mathcal{I}}^{\#}_2 = \mathcal{A}\mathcal{I}^{\#}_2 + \mathcal{I} - \mathcal{B}\mathcal{N}^{-1}\mathcal{B}^*(\mathcal{Q}_{11}\mathcal{I}^{\#}_2 + \mathcal{Q}_{12}) + \mathcal{P}^{\#}_{11}\mathcal{C}^*\mathcal{R}_v^{-1}(\mathcal{I}^{\#}_1 - \mathcal{I}^{\#}_2). \tag{3.20}$$

Substituting (3.19)–(3.20) into (3.18) completes the proof of Theorem 3.2.

4 Performance of the Bayesian Strategy

The performance of a general Bayesian strategy needs to be quantified for each fixed true value of the parameter $\boldsymbol{b}$ and θ. In this paper, the following definition of regret is used to quantify the performance of the strategy.

Definition 4.1 (Multiplicative regret) (cf. [18]) *The multiplicative regret or competitive ratio, as it is more commonly called in the literature, is the ratio*:

$$\mathrm{MR}_{\mathcal{S}}(\mathbf{a}) \triangleq \frac{J_{\mathcal{S}}(\mathbf{a})}{J_o(\mathbf{a})} \geq 1,$$

where $J_0(\mathbf{a})$ *is the optimal cost with full knowledge of* $\mathbf{a}$. *A strategy that minimizes the worst-case multiplicative regret*

$$\mathrm{MR}_{\mathcal{S}}^{\star} = \sup_{\mathbf{a}} \mathrm{MR}_{\mathcal{S}}(\mathbf{a})$$

is deemed optimal.

Lemma 4.1 (cf. [18]) *Suppose the Bayesian strategy* $\mathcal{S}$ *is optimal under the prior* μ, *achieves constant regret* R *for all* $\mathbf{a}$ *in the support of* μ, *and achieves regret not exceeding* R *for all* $\mathbf{a}$ *not in the support of* μ. *Then* $\mathcal{S}$ *is an optimal agnostic strategy which achieves a worst-case regret of* R.

In the finite-dimensional systems, the optimal cost (2.19) with known $\boldsymbol{b}$ is

$$J(\boldsymbol{q}, \boldsymbol{b}) = \boldsymbol{b}^{\mathrm{T}} X^{\#}(\Sigma_{\boldsymbol{v}}, T_0, T)\boldsymbol{b} + Y^{\#}(\Sigma_{\boldsymbol{v}}, T_0, T).$$

Hence, the multiplicative regret for a Bayesian strategy with prior covariance Σ is

$$\mathrm{MR}(\boldsymbol{b}) = \frac{\boldsymbol{b}^{\mathrm{T}} X(\Sigma, \Sigma_{\boldsymbol{v}}, T_0, T)\boldsymbol{b} + Y(\Sigma, \Sigma_{\boldsymbol{v}}, T_0, T)}{\boldsymbol{b}^{\mathrm{T}} X^{\#}(\Sigma_{\boldsymbol{v}}, T_0, T)\boldsymbol{b} + Y^{\#}(T, T_0, \Sigma_{\boldsymbol{v}})}. \tag{4.1}$$

We want to find the optimal agnostic strategy which achieves a worst-case regret, which is equivalent to find the Σ such that (4.1) becomes independent of $\boldsymbol{b}$ from Lemma 4.1. This occurs precisely when

$$X(\Sigma, \Sigma_{\boldsymbol{v}}, T_0, T) = \frac{Y(\Sigma, \Sigma_{\boldsymbol{v}}, T_0, T)}{Y^{\#}(\Sigma_{\boldsymbol{v}}, T_0, T)} X^{\#}(\Sigma_{\boldsymbol{v}}, T_0, T). \tag{4.2}$$

Similarly, in the infinite-dimensional systems, the optimal cost (3.17) with known θ is

$$\mathcal{V}(y, \theta) = \langle \mathcal{X}^{\#}(\mathcal{R}_v, T_0, T)\theta, \theta\rangle + \mathcal{Y}^{\#}(\mathcal{R}_v, T_0, T).$$

Hence, the multiplicative regret for a Bayesian strategy with prior covariance $\mathcal{R}$ is

$$\mathrm{MR}(\theta) = \frac{\langle \mathcal{X}(\mathcal{R}, \mathcal{R}_v, T_0, T)\theta, \theta\rangle + \mathcal{Y}(\mathcal{R}, \mathcal{R}_v, T_0, T)}{\langle \mathcal{X}^{\#}(\mathcal{R}_v, T_0, T)\theta, \theta\rangle + \mathcal{Y}^{\#}(\mathcal{R}_v, T_0, T)}. \tag{4.3}$$

We want to find the optimal agnostic strategy which achieves a worst-case regret, which is equivalent to find the $\mathcal{R}$ such that (4.3) becomes independent of θ from Lemma 4.1. This occurs precisely when

$$\mathcal{X}(\mathcal{R}, \mathcal{R}_v, T_0, T) = \frac{\mathcal{Y}(\mathcal{R}, \mathcal{R}_v, T_0, T)}{\mathcal{Y}^{\#}(\mathcal{R}_v, T_0, T)} \mathcal{X}^{\#}(\mathcal{R}_v, T_0, T).$$

At optimality, the worst-case multiplicative regret is equal to both the multiplicative regret when

$$\boldsymbol{b} = \mathbf{0}$$

and

$$\theta = 0$$

(given by $\frac{Y}{Y^{\#}}$ or $\frac{\mathcal{Y}}{\mathcal{Y}^{\#}}$), and the multiplicative regret in the limit as $\|\boldsymbol{b}\| \to \infty$ or $|\theta| \to \infty$ (given by the constant ratio of matrices X and $X^{\#}$, or operators $\mathcal{X}$ and $\mathcal{X}^{\#}$). That is, the optimal agnostic strategy perfectly balances the regret for no drift and very large drift. Moreover, the equality of these multiplicative regrets is a sufficient condition for optimality.

5 Numerical Example

For illustration, we apply the multiplicative regret-minimizing Bayesian strategy to a scalar system described by dynamics (2.1). Our assumptions included

$$\Sigma_w = 1,$$

$$A = B = C = Q = R = 1,$$

and

$$T_0 = 0.$$

We computed the optimal prior variance Σ by numerically solving (4.2) using a gradient descent method.

Figures 1 and 2 display the relationships between the optimal prior variance Σ, sensor noise variance Σ_v, and the worst-case multiplicative regret MR. Due to the dependence of the drift term on the state, the phenomenon described in [18], where for small time horizons, there is very little time to learn the dynamics of the system, and thus the prior needs to be flexible for the controller to adapt quickly, no longer occurs. Instead, the optimal prior variance Σ peaks near

$$T = 1.$$

This effect becomes particularly pronounced in scenarios characterized by high sensor noise Σ_v (see Figure 1(a) and Figure 1(b)). No reasonable explanation can be given yet. In contrast, with longer time horizons, the controller can afford a more conservative approach, given the ample time available for system stabilization and the impact of poor performance at small times is only weakly penalized. Notably, as the time horizon T approaches infinity, the optimal prior variance Σ converges towards zero, see Figure 1(c).

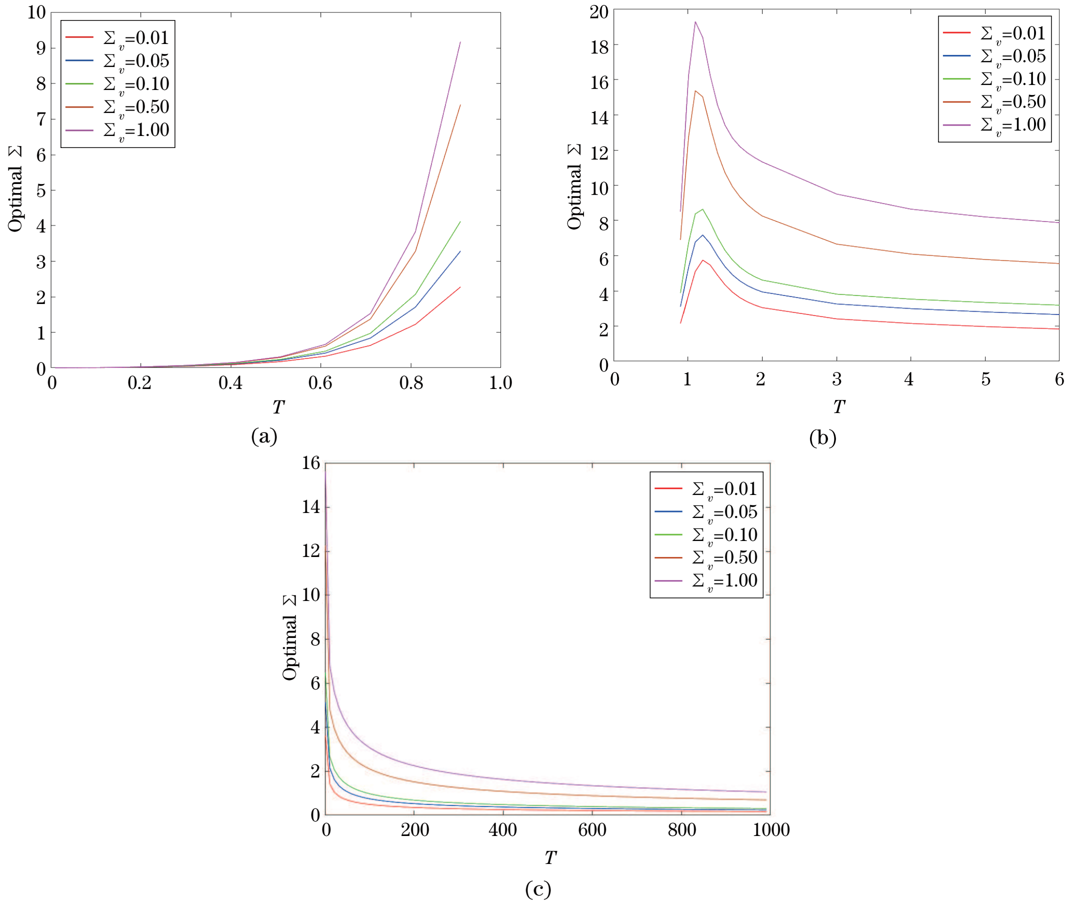
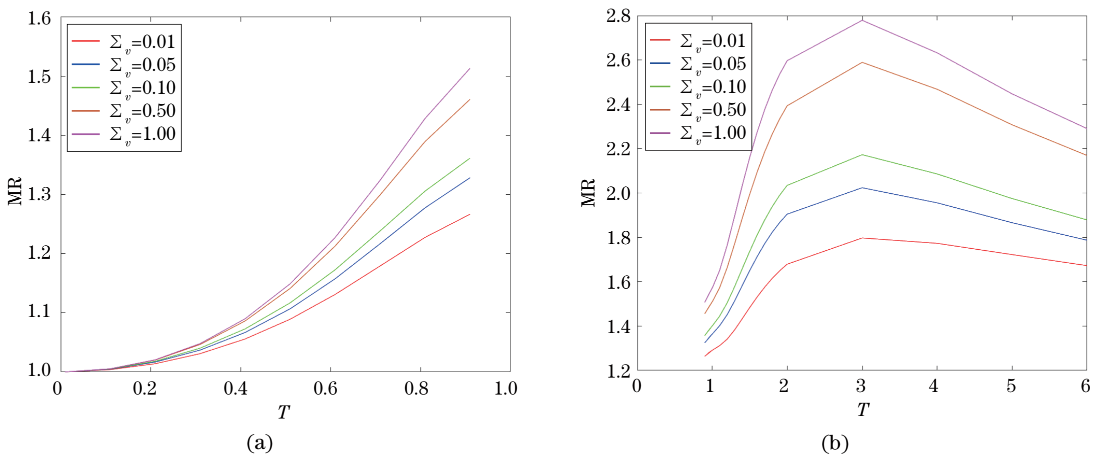
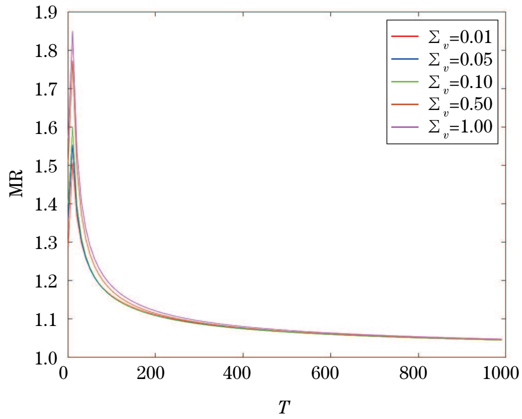


Figure 1 Bayesian strategies minimizing the worst-case multiplicative regret with varying standard deviation Σ_v of the sensor noise: Optimal variance Σ of the prior.

As anticipated, the worst-case regret increases with higher sensor noise, see Figure 2(a) and Figure 2(b). It peaks at $T \approx 3$ and tends to unity as T tends to either 0 or infinity (see Figure 2(a) and Figure 2(c)).





(c)

Figure 2 Bayesian strategies minimizing the worst-case multiplicative regret with varying standard deviation Σ_v of the sensor noise: Worst-case multiplicative regret MR.

Declarations

Conflicts of interest The authors declare no conflicts of interest.

References

[1] Abbasi-Yadkori, Y. and Szepesvári, C., Regret bounds for the adaptive control of linear quadratic systems, Proceedings of the 24th Annual Conference on Learning Theory, *J. Math. Learn. Res. Proc. Track.*, **19**, 2011, 1–26.

[2] Agarwal, N., Bullins, B., Hazan, E., et al., Online control with adversarial disturbances, 36th International Conference on Machine Learning, 2019, IMLS, California, 154–165.

[3] Agarwal, N., Hazan, E. and Singh, K., Logarithmic regret for online control, Advances in Neural Information Processing Systems, Curran Associates, Canada, **32**, 2019, 10175–10184.

[4] Anderson, B. and Moore, J., Optimal Control: Linear Quadratic Methods, Prentice Hall, NJ, 1989.

[5] Auer, P., Cesa-Bianchi, N. and Fischer, P., Finite-time analysis of the multiarmed bandit problem, *Machine learning*, **47**, 2002, 235–256.

[6] Bensoussan, A., Stochastic Control of Partially Observable Systems, Cambridge University Press, Cambridge, 1992.

[7] Bertsekas, D. P., Dynamic Programming and Optimal Control, Vol. 1, Athena Scientific, Belmont, MA, 2005.

[8] Bubeck, S. and Cesa-Bianchi, N., Regret analysis of stochastic and nonstochastic multi-armed bandit problems, *Found. Trends in Mach. Learn.*, **5**(1), 2012, 101–112.

[9] Carruth, J., Eggl, M. F., Fefferman, C., et al., Controlling unknown linear dynamics with bounded multiplicative regret, *Rev. Mat. Iberoam.*, **38**(7), 2022, 2185–2216.

[10] Cassel, A., Cohen, A. and Koren, T., Logarithmic regret for learning linear quadratic regulators efficiently, Proceedings of the 37th International Conference on Machine Learning, JMLR, **119**, New York, 2020, 1328–1337.

[11] Cesa-Bianchi, N. and Lugosi, G., Prediction, Learning, and Games, Cambridge University Press, Cambridge, 2006.

[12] Cohen, A., Koren, T. and Mansour, Y., Learning linear-quadratic regulators efficiently with only $\sqrt{T}$ Regret, International Conference on Machine Learning, PMLR, **97**, 2019, 1300–1309.

[13] Curtain, R. F., Infinite-dimensional filtering, *SIAM J. Control Optim.*, **13**(1), 1975, 89–104.

[14] Curtain, R. F. and Ichikawa, A., The separation principle for Stochastic evolution equations, *SIAM J. Control Optim.*, **15**(3), 1977, 367–383.

[15] Dean, S., Mania, H., Matni, N., et al., Regret bounds for robust adaptive control of the linear quadratic regulator, Advances in Neural Information Processing Systems, **31**, Curran Associates, Inc., New York, 2018, 4192–4201.

[16] Fefferman, C., Pegueroles, B. G., Rowley, C. W. and Weber, M., Optimal control with learning on the fly: A toy problem, *Rev. Mat. Iberoam.*, **38**(1), 2022, 175–187.

[17] Georgiou, T. T. and Lindquist, A., The separation principle in stochastic control, redux, *IEEE Trans. Automat. Control*, **58**(10), 2013, 2481–2494.

[18] Gurevich, D., Goswami, D., Fefferman, C. L. and Rowley, C. W., Optimal control with learning on the fly: System with unknown drift, Learning for Dynamics and Control Conference, PMLR, **168**, California, 2022, 870–880.

[19] Hazan, E., Introduction to Online Convex Optimization, MIT Press, California, 2022.

[20] Ibrahimi, M., Javanmard, A. and Roy, B., Efficient reinforcement learning for high dimensional linear quadratic systems, Advances in Neural Information Processing Systems, **25**, 2012, https://api.semanticscholar.org/CorpuslD:11732491.

[21] Ichikawa, A., Dynamic programming approach to stochastic evolution equations, *SIAM J. Control Optim.*, **17**(1), 1979, 152–174.

[22] Lai, T. L. and Robbins, H., Asymptotically efficient adaptive allocation rules, *Adv. Appl. Math.*, **6**(1), 1985, 4–22.

[23] Lindquist, A., On feedback control of linear stochastic systems, *SIAM J. Control*, **11**(2), 1973, 323–343.

[24] Lü, Q. and Zhang, X., Mathematical Control Theory for Stochastic Partial Differential Equations, Springer-Verlag, Charm, 2021.

[25] Powell, W. B., Approximate Dynamic Programming: Solving the Curses of Dimensionality, John Wiley & Sons, Hoboken, NJ, 2007.

[26] Speyer, J. L. and Chung, W. H., Stochastic Processes, Estimation, and Control, SIAM, PA, Philadelphia, 2008.

[27] Yong, J. and Zhou, X. Y., Stochastic Controls: Hamiltonian Systems and HJB Equations, Springer-Verlag, New York, 1999.