

# Minimum-Information LQG Control

## Part II: Retentive Controllers

Roy Fox<sup>†</sup> and Naftali Tishby<sup>†</sup>

**Abstract—Retentive (memory-utilizing) sensing-acting agents may operate under limitations on the communication between their sensing, memory and acting components, requiring them to trade off the external cost that they incur with the capacity of their communication channels. In this paper we formulate this problem as a sequential rate-distortion problem of minimizing the rate of information required for the controller’s operation under a constraint on its external cost. We reduce this bounded retentive control problem to the memoryless one, studied in Part I of this work [1], by viewing the memory reader as one more sensor and the memory writer as one more actuator. We further investigate the structure of the resulting optimal solution and demonstrate its interesting phenomenology.**

### I. INTRODUCTION

In a feedback-control system, the internal state of the agent interacts with the external state of the world through sensors that pay attention to the agent’s environment and actuators that apply intention to it, in a perception-action cycle [2]. This interaction is limited by external constraints on observability and controllability, as well as internal constraints on the information-processing resources available to the controller.

In Part I of this work [1], we focused on memoryless controllers that have no internal memory and can only attend to their most recent input observation. We discussed how the communication from the sensor to the actuator is central to the agent’s ability to act upon the perceived information. The degree of this attention, measured by the amount of Shannon information about the input observation that is utilized in the output control, is a lower bound on the required capacity of the communication channel between the controller’s sensor and its actuator. When this capacity for internal communication is limited, the agent needs to trade off some external cost for reducing the rate at which it transmits information.

A related but often overlooked resource is memory bandwidth. We can think of memory as a communication channel from the past internal state of the controller to its future internal state. When memory resources are remote, communication constraints apply to them as well. Even local memory is limited by its capacity to store information and by the capacity of the internal communication channels to and from the memory components. This limitation is evidenced by the hierarchical design of memory in modern digital

computers, which places larger capacity on the channels to closer but smaller cache memory components [3].

When the controller is retentive (memory-utilizing), it does maintain an internal memory state which can have information on more than the most recent observation. As in Part I, our guiding principle in this work is to measure the information complexity of the controller’s internal representation by asking “*How much information does the controller have on the past?*”. The retentive controller receives information of the past through both memory and sensory channels (Figure 2) and the amount of information that it keeps of the past is a lower bound on the total capacity of both these channels [4].

In a sense, we can consider the reader of the memory state to be one more sensor and the writer of the memory state to be one more actuator. This suggests a reduction from the retentive case to the memoryless case, in which the memory state is considered external and part of the world state [5], [6]. This memory component is fully observable, fully controllable, has no process noise and incurs no cost. Rather than redevelop our results for the retentive controllers similarly to Part I, this reduction allows us to reuse those results and underlines the structure of the solution.

In this paper we make two contributions. First, we present a method for the design of controllers that are optimal under a constraint on both their memory and sensory channel capacity. To our knowledge, this is the first explicit treatment of the channel capacity of the memory process in the context of continuous state-space systems.

Second, we provide a reduction from the problem of bounded retentive control to the problem of bounded memoryless control. This reduction is conceptually convenient and constructive, allowing us to treat both problems using the same framework and providing insight into the structure of the optimal retentive controller.

In Section II we define the LQG task and restate the results of Part I. In Section III we present the retentive control model, its reduction to memoryless control and the structure of the resulting optimal solution. In Section IV we illustrate our results with an example.

### II. PRELIMINARIES

#### A. Control task

We consider the same closed-loop control problem detailed in Part I [1, Section II]. In time  $t$ , a plant in state  $x_t \in \mathbb{R}^n$  emits an observation  $y_t \in \mathbb{R}^k$ , takes in a control input  $u_t \in \mathbb{R}^\ell$  and undergoes a stochastic state transition. We focus on discrete-time systems with linear dynamics,

<sup>†</sup>School of Computer Science and Engineering, The Hebrew University, {royf,tishby}@cs.huji.ac.il

\*This work was supported by the DARPA MSEE Program, the Gatsby Charitable Foundation, the Israel Science Foundation and the Intel ICRI-CI Institute

Gaussian noise and quadratic cost rate (LQG). For simplicity, all elements are taken to be homogeneous, i.e. centered at the origin, and time-invariant. We note that all our results hold without these assumptions, with the appropriate adjustments, as usual in LQG problems [7].

*Definition 1:* A linear-Gaussian time-invariant (LTI) plant  $\langle A, B, C, \Sigma_\xi, \Sigma_\epsilon \rangle$  has state dynamics

$$x_{t+1} = Ax_t + Bu_t + \xi_t; \quad \xi_t \sim \mathcal{N}(0, \Sigma_\xi), \quad (1)$$

where  $A \in \mathbb{R}^{n \times n}$ ,  $B \in \mathbb{R}^{n \times \ell}$ ,  $0 \preceq \Sigma_\xi \in \mathbb{S}_+^n$  and  $\xi_t$  is independent of  $(x^t, y^t, u^t)$ . The observation dynamics are

$$y_t = Cx_t + \epsilon_t; \quad \epsilon_t \sim \mathcal{N}(0, \Sigma_\epsilon), \quad (2)$$

where  $C \in \mathbb{R}^{k \times n}$ ,  $\Sigma_\epsilon \in \mathbb{S}_+^k$  and  $\epsilon_t$  is independent of  $(y^{t-1}, u^{t-1}, x^t)$ , where we denote  $x^t = \{x_\tau\}_{\tau \leq t}$ , etc.

*Definition 2:* A linear-quadratic-Gaussian (LQG) task  $\langle A, B, C, \Sigma_\xi, \Sigma_\epsilon, Q, R \rangle$  involves a LTI plant and the cost rate

$$\mathcal{J}_t = \frac{1}{2}(x_t^\top Q x_t + u_t^\top R u_t),$$

where  $Q \in \mathbb{S}_+^n$  and  $R \in \mathbb{S}_+^\ell$ . The task is to achieve a low long-term average expected cost rate, with respect to the distribution induced by the plant and the controller  $\pi$

$$\mathcal{J}_\pi = \limsup_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T \mathbb{E}_\pi[\mathcal{J}_t].$$

As motivated in Part I, we are particularly interested in linear-Gaussian time-invariant (LTI) controllers, which induce, jointly with a LTI plant, a stationary Gaussian process, independent of any initial conditions. With  $\Sigma_x \in \mathbb{S}_+^n$ ,  $\Sigma_y \in \mathbb{S}_+^k$  and  $\Sigma_u \in \mathbb{S}_+^\ell$ , respectively the stationary covariances of the state, the observation and the control, we have

$$\Sigma_y = C \Sigma_x C^\top + \Sigma_\epsilon,$$

and the reverse relation

$$x_t = Ky_t + \kappa_t; \quad \kappa_t \sim \mathcal{N}(0, \Sigma_\kappa)$$

$$K = \Sigma_x C^\top \Sigma_y^\dagger$$

$$\Sigma_\kappa = \Sigma_x - \Sigma_x C^\top \Sigma_y^\dagger C \Sigma_x,$$

with  $\cdot^\dagger$  the Moore-Penrose pseudoinverse. Assuming that the process has mean 0, the stationary expected cost rate is

$$\mathcal{J}_\pi = \frac{1}{2}(\text{tr}(Q \Sigma_x) + \text{tr}(R \Sigma_u)).$$

### B. Bounded memoryless control

In this section we restate the main result of Part I [1, Section IV].

*Definition 3:* A memoryless linear-Gaussian time-invariant (LTI) controller has control law of the form

$$u_t = Hy_t + \eta_t; \quad \eta_t \sim \mathcal{N}(0, \Sigma_\eta), \quad (3)$$

where  $H \in \mathbb{R}^{\ell \times k}$ ,  $\Sigma_\eta \in \mathbb{S}_+^\ell$  and  $\eta_t$  is independent of  $(u^{t-1}, x^t, y^t)$ .

The controller is bounded and operates under limitations on its capacity to process the observation and produce the control. Namely, with the Shannon information rate

$$\mathcal{I}_t = \mathbb{I}[y_t; u_t] = \mathbb{E} \left[ \log \frac{f(y_t, u_t)}{f(y_t)f(u_t)} \right], \quad (4)$$

where  $f$  denotes the various probability density functions, as indicated by their arguments, we are interested in a LTI controller  $\pi$  that minimizes the long-term average rate

$$\mathcal{I}_\pi = \limsup_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T \mathcal{I}_t, \quad (5)$$

under the constraint that it achieves some guarantee level  $c$  of expected cost rate.

*Problem 1:* Given a LQG task, the bounded memoryless LTI controller optimization problem is

$$\begin{aligned} \min_{\pi} \quad & \mathcal{I}_\pi \\ \text{s.t.} \quad & \mathcal{J}_\pi \leq c, \end{aligned}$$

with  $\mathcal{I}_\pi$  as in (5), where  $\mathcal{I}_t = \mathbb{I}[y_t; u_t]$ , and with  $u_t$  as in (3).

To solve the optimization problem, we consider the minimum mean square error (MMSE) estimators

$$\hat{x}_{y_t} = \mathbb{E}[x_t | y_t] = Ky_t$$

$$\hat{x}_{u_t} = \mathbb{E}[x_t | u_t] = \Sigma_{x;u} \Sigma_u^\dagger u_t,$$

respectively for the state given the observation and the control. Since  $\hat{x}_{u_t}$  is a sufficient statistic of  $u_t$  for  $x_t$ , we can reverse their causality, basing  $u_t$  on  $\hat{x}_{u_t}$  instead of vice versa. This puts the control law in the form

$$\hat{x}_{y_t} = Ky_t$$

$$\hat{x}_{u_t} = W \hat{x}_{y_t} + \omega_t; \quad \omega_t \sim \mathcal{N}(0, \Sigma_\omega)$$

$$u_t = L \hat{x}_{u_t}.$$

The optimal memoryless controller satisfies the conditions of Theorem 1 in Part I, Section IV-A, restated below in algorithmic form. To numerically find the optimal solution, we can interpret these conditions as update equations, which we apply iteratively until a fixed point is reached.

We split the equations into three parts, a forward iteration (Algorithm 1) updating the marginal distributions, a backward iteration (Algorithm 2) updating the cost-to-go and the control policy, and an eigenvalue decomposition (EVD) for finding the control-based estimator covariance (Algorithm 3). We can alternate between Algorithms 1, 2 and 3, iterating until the solution converges to a fixed point of the equations.

## III. BOUNDED RETENTIVE CONTROLLERS

### A. Control model

In this section we discuss retentive (memory-utilizing) controllers with bounded communication resources. A retentive controller has an internal memory state  $z_t$  in some space  $\mathcal{Z}$ . The memory allows the controller to output a control that indirectly depends on past input observations rather than only on the most recent observation. The controller takes as input an observation  $y_t$  and outputs a control  $u_t$ , while also making a memory state transition from  $z_{t-1}$  to  $z_t$ . Thus, in each time step, there are two inputs,  $z_{t-1}$  and  $y_t$ , and two outputs,  $z_t$  and  $u_t$ .

*Definition 4:* A controller is retentive if it satisfies the following independence properties:

---

**Algorithm 1** Forward iteration

---

**function** FORWARD( $\Sigma_x, \Sigma_{\hat{x}_u}, L$ )

Update

$$\begin{aligned}\Sigma_x &\leftarrow (A + BL) \Sigma_{\hat{x}_u} (A + BL)^\top \\ &\quad + A(\Sigma_x - \Sigma_{\hat{x}_u})A^\top + \Sigma_\xi \\ \Sigma_y &\leftarrow C \Sigma_x C^\top + \Sigma_\epsilon \\ K &\leftarrow \Sigma_x C^\top \Sigma_y^\dagger \\ \Sigma_{\hat{x}_y} &\leftarrow K \Sigma_y K^\top\end{aligned}$$

**end function**

---

---

**Algorithm 2** Backward iteration

---

**function** BACKWARD( $\Sigma_{\hat{x}_y}, \Sigma_{\hat{x}_u}, K, S; \beta$ )

Update

$$\begin{aligned}M &\leftarrow \beta^{-1} C^\top K^\top (\Sigma_{\hat{x}_y | \hat{x}_u}^\dagger - \Sigma_{\hat{x}_y}^\dagger) K C \\ S &\leftarrow Q + A^\top S A - M \\ L &\leftarrow -(R + B^\top S B)^\dagger B^\top S A \\ N &\leftarrow L^\top (R + B^\top S B) L\end{aligned}$$

**end function**

---

---

**Algorithm 3** Activation of control-based estimator modes

---

**function** ACTIVATION( $\Sigma_{\hat{x}_y}, N; \beta$ )

Update

$$V, \Lambda \leftarrow \text{EVD}(\Sigma_{\hat{x}_y}^{1/2} N \Sigma_{\hat{x}_y}^{1/2})$$

with  $n - \text{rank}(\Sigma_{\hat{x}_y})$  columns of  $V$  spanning  $\ker(\Sigma_{\hat{x}_y})$ 

$$D \leftarrow \text{diag} \begin{cases} 1 - \beta^{-1} \lambda_i^{-1} & \lambda_i > \beta^{-1} \\ 0 & \lambda_i \leq \beta^{-1} \end{cases}$$

$$\Sigma_{\hat{x}_u} \leftarrow \Sigma_{\hat{x}_y}^{1/2} V D V^\top \Sigma_{\hat{x}_y}^{1/2}$$

**end function**

---

- 1) The memory state depends only on the previous memory state and the current observation; that is,  $z_t$  is independent of  $(z^{t-2}, y^{t-1}, u^{t-1}, x^t)$  given  $z_{t-1}$  and  $y_t$ .
- 2) The control depends only on the memory state; that is,  $u_t$  is independent of  $(z^{t-1}, u^{t-1}, x^t, y^t)$  given  $z_t$ .

A system including a retentive controller satisfies the Bayesian network in Figure 1.

As motivated in Part I for the memoryless case, we are particularly interested in controllers where both the memory state update and the control are linear-Gaussian and time-invariant (LTI), since they are easier to optimize and implement. Linear controllers with limited memory are known not to be optimal for all control problems [8], [9]. The conditions under which such controllers are optimal for our bounded control problem are beyond our current scope.

*Definition 5:* A retentive linear-Gaussian time-invariant (LTI) controller has memory state space that is a vector space

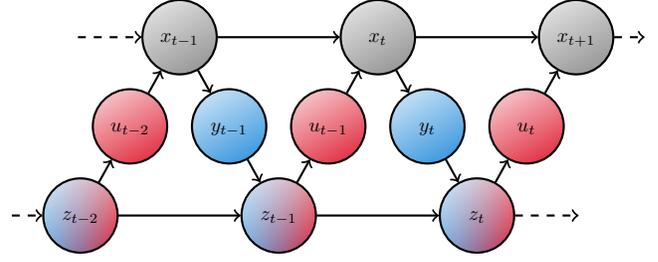


Fig. 1. Bayesian network of retentive control

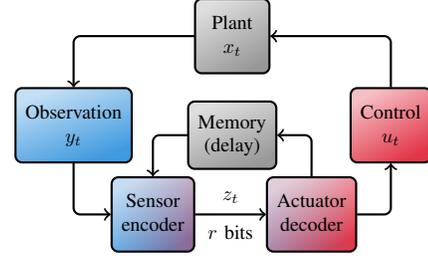


Fig. 2. Block diagram of a closed-loop retentive control system, with a communication channel from the sensor-reader to the actuator-writer

 $\mathcal{Z} = \mathbb{R}^d$  and control law of the form

$$z_t = F z_{t-1} + G y_t + \zeta_t; \quad \zeta_t \sim \mathcal{N}(0, \Sigma_\zeta) \quad (6a)$$

$$u_t = L z_t + \nu_t; \quad \nu_t \sim \mathcal{N}(0, \Sigma_\nu) \quad (6b)$$

where  $F \in \mathbb{R}^{d \times d}$ ,  $G \in \mathbb{R}^{d \times k}$ ,  $\Sigma_\zeta \in \mathbb{S}_+^d$ ,  $L \in \mathbb{R}^{m \times d}$ ,  $\Sigma_\nu \in \mathbb{S}_+^m$ ,  $\zeta_t$  is independent of  $(z_{t-1}, y_t)$  and  $\nu_t$  is independent of  $z_t$ .

We are interested in reducing the information complexity of implementing this controller. To measure this complexity, we consider the capacity of a memoryless communication channel from the sensor-reader to the actuator-writer (Figure 2). The encoder and the decoder themselves are memoryless, but the memory component has perfect fidelity, making everything written by the actuator available for the sensor to read in the next step.

We could use  $\mathcal{Z} = \{0, 1\}^r$ , the set of  $r$ -bit strings, instead of the vector space  $\mathbb{R}^d$ , to indicate that the controller can process at most  $r$  bits of information per time step

$$\mathbb{I}[z_{t-1}, y_t; z_t, u_t] = \mathbb{I}[z_{t-1}, y_t; z_t] \leq \mathbb{H}[z_t] \leq r \log 2.$$

As in the memoryless case (Part I [1, Section III-B]), the information rate is generally not a tight lower bound on the capacity of a discrete memory, but here again, if the controller is LTI, there exists a perfectly matched memoryless additive Gaussian noise channel. As shown in the Supplementary Material<sup>1</sup> (SM), Appendix I, the capacity of this channel optimally equals the information rate  $\mathbb{I}[z_{t-1}, y_t; z_t, u_t]$  and a constraint on the information rate is equivalent to a constraint on the power available for transmission on the channel.

The retentive controller optimization problem is therefore similar to Problem 1, but with the information rate including both the memory and the sensory channels.

<sup>1</sup>Available at <https://arxiv.org/abs/1606.01947>

*Problem 2:* Given a LQG task, the bounded retentive LTI controller optimization problem is

$$\begin{aligned} \min_{\pi} \quad & \mathcal{I}_{\pi} \\ \text{s.t.} \quad & \mathcal{J}_{\pi} \leq c, \end{aligned}$$

with  $\mathcal{I}_{\pi}$  as in (5), where

$$\mathcal{I}_t = \mathbb{I}[z_{t-1}, y_t; z_t, u_t], \quad (7)$$

and with  $z_t$  and  $u_t$  as in (6).

Note that here there is no additional constraint or cost on the precision of  $u_t$  given  $z_t$ , implying that optimally  $\Sigma_{\nu} = 0$ .

There is an interesting connection between the retentive information rate  $\mathcal{I}_{\pi}$  and the long-term average of the directed information rate [10], [11], defined by

$$\begin{aligned} \mathbb{I}[\{y_t\} \rightarrow \{z_t\}] &= \limsup_{T \rightarrow \infty} \frac{1}{T} \mathbb{I}[y^T \rightarrow z^T] \\ &= \limsup_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T \mathbb{I}[y^t; z_t | z^{t-1}]. \end{aligned}$$

By the independence properties of the retentive controller and by the chain rule for information [12], we have

$$\begin{aligned} \mathbb{I}[z_{t-1}, y_t; z_t, u_t] &= \mathbb{I}[z_{t-1}, y_t; z_t] \\ &= \mathbb{I}[z^{t-1}, y^t; z_t] \\ &= \mathbb{I}[z^{t-1}; z_t] + \mathbb{I}[y^t; z_t | z^{t-1}]. \end{aligned}$$

We can thus define the following extension of the concept of directed information.

*Definition 6:* The retentive directed information from the sequence of observations  $y^T$  to the sequence of memory states  $z^T$  is

$$\mathbb{I}[y^T \rightarrow z^T] = \sum_{t=1}^T \mathbb{I}[z^{t-1}, y^t; z_t].$$

Since  $\mathbb{I}[y^T \rightarrow z^T] \geq \mathbb{I}[y^T \rightarrow z^T]$ , the retentive directed information rate is always a tighter lower bound on the capacity of the channel in Figure 2. Despite the apparent similarity to Figure 2 in [11], notice that their encoder and decoder have unlimited memory of  $z^t$  and  $u^t$ . This justifies their use of directed information, regardless of the residual term  $\mathbb{I}[z^{t-1}; z_t]$  being infinite in their optimal controller.

Some further properties of the retentive directed information can be found in the SM, Appendix VI.

### B. Reduction to memoryless controllers

We can analyze the bounded retentive control problem (Problem 2) directly using the same tools developed in Part I [1, Section IV-A] for Problem 1. Fortunately, there is no need to repeat that entire treatment, since a simple and insightful reduction will allow us to reuse the results obtained there.

We start by reformulating the problem. The following relaxation and Lemma 1 that shows its equivalence to the original problem allow us to reverse the causality between

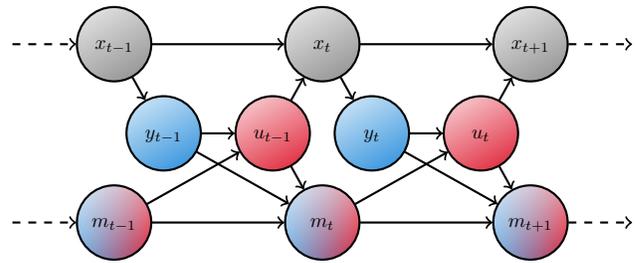


Fig. 3. Bayesian network of relaxed retentive control

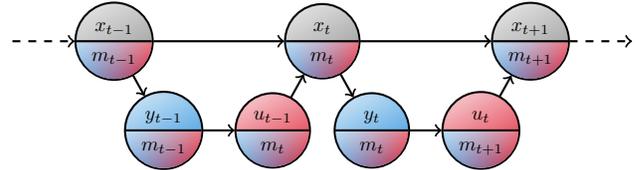


Fig. 4. Bayesian network of relaxed retentive control, redrawn in the form of memoryless control

$u_t$  and  $z_t$ . We need a new notation for the resulting time-shifted memory state sequence and define for each  $t$

$$m_t = z_{t-1}.$$

*Definition 7:* A retentive controller is relaxed if  $u_t$  is not required to be independent of  $(m_t, y_t)$  given  $m_{t+1}$ . Thus the relaxed controller satisfies the Bayesian network in Figure 3 and its control law is given by  $\pi(u_t, m_{t+1} | m_t, y_t)$ .

*Lemma 1:* The relaxed controller optimization problem is equivalent to the original Problem 2.

*Proof:* The following proof does not assume that the controller is linear-Gaussian and holds for the LTI controller as a special case.

Let  $\pi$  be a controller satisfying the Bayesian network in Figure 3. We construct a controller  $\tilde{\pi}$  with  $\tilde{z}_t = (u_t, m_{t+1})$  for each  $t$ , such that

$$\begin{aligned} \tilde{\pi}(\tilde{z}_t | \tilde{z}_{t-1}, y_t) &= \pi(u_t, m_{t+1} | m_t, y_t) \\ \tilde{\pi}(u_t | \tilde{z}_t) &= \delta_{\tilde{z}_t = (u_t, \cdot)}. \end{aligned}$$

This controller satisfies the Bayesian network in Figure 1 and

$$\begin{aligned} \mathbb{I}_{\tilde{\pi}}[\tilde{z}_{t-1}, y_t; \tilde{z}_t, u_t] &= \mathbb{I}_{\pi}[(u_{t-1}, m_t), y_t; (u_t, m_{t+1})] \\ &= \mathbb{I}_{\pi}[m_t, y_t; u_t, m_{t+1}]. \end{aligned}$$

Thus the controller  $\tilde{\pi}$  is feasible for the unrelaxed Problem 2 and has the same performance as the relaxed controller  $\pi$ , since it induces a stochastic process with the same distribution and information rate.  $\square$

The structure in Figure 3 can now be redrawn as in Figure 4. Comparing this Bayesian network to the one in Part I, Figure 2, we have clearly reduced the bounded retentive control problem to a special case of the bounded memoryless control problem, as stated formally in the following lemma.

*Lemma 2:* The bounded retentive LTI controller optimization problem (Problem 2) for the LQG task  $\langle A_x, B_{x;u}, C_{y;x}, \Sigma_{\xi}, \Sigma_{\epsilon}, Q_x, R_u \rangle$  is equivalent to the

bounded memoryless LTI controller optimization problem (Problem 1) for the LQG task  $\langle A, B, C, \Sigma_{\tilde{\xi}}, \Sigma_{\tilde{\epsilon}}, Q, R \rangle$ , where

$$A = \begin{bmatrix} A_x & 0 \\ 0 & 0 \end{bmatrix}; \quad B = \begin{bmatrix} B_{x;u} & 0 \\ 0 & I \end{bmatrix}; \quad C = \begin{bmatrix} C_{y;x} & 0 \\ 0 & I \end{bmatrix}$$

$$\Sigma_{\tilde{\xi}} = \begin{bmatrix} \Sigma_{\xi} & 0 \\ 0 & 0 \end{bmatrix}; \quad \Sigma_{\tilde{\epsilon}} = \begin{bmatrix} \Sigma_{\epsilon} & 0 \\ 0 & 0 \end{bmatrix}$$

$$Q = \begin{bmatrix} Q_x & 0 \\ 0 & 0 \end{bmatrix}; \quad R = \begin{bmatrix} R_u & 0 \\ 0 & 0 \end{bmatrix}.$$

Here all matrices are extended by  $d$  rows and  $d$  columns.

*Proof:* Given the retentive control stochastic process  $\{x_t, m_t, y_t, u_t\}$ , we consider the memoryless control stochastic process  $\{\tilde{x}_t, \tilde{y}_t, \tilde{u}_t\}$  with

$$\tilde{x}_t = \begin{bmatrix} x_t \\ m_t \end{bmatrix}; \quad \tilde{y}_t = \begin{bmatrix} y_t \\ m_t \end{bmatrix}; \quad \tilde{u}_t = \begin{bmatrix} u_t \\ m_{t+1} \end{bmatrix}.$$

The dynamics for this process can easily be seen to be given by (1), (2), with  $A, B, C, \Sigma_{\tilde{\xi}}$  and  $\Sigma_{\tilde{\epsilon}}$  as in the lemma. The cost rate applies only to the  $x_t$  and  $u_t$  parts

$$\mathcal{J}_t = \frac{1}{2} \left( \begin{bmatrix} x_t \\ m_t \end{bmatrix}^\top \begin{bmatrix} Q_x & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} x_t \\ m_t \end{bmatrix} + \begin{bmatrix} u_t \\ m_{t+1} \end{bmatrix}^\top \begin{bmatrix} R_u & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} u_t \\ m_{t+1} \end{bmatrix} \right).$$

The information rate is

$$\mathcal{I}_t = \mathbb{I}[\tilde{y}_t; \tilde{u}_t] = \mathbb{I}[m_t, y_t; u_t, m_{t+1}],$$

where the left-hand side is taken as in (4) and the right-hand side as in (7), as required.  $\square$

### C. Structure of the optimal solution

We can substitute the form of the reduction in Lemma 2 into the optimal solution in Section II-B, to study more explicitly the structure of the optimal solution in the retentive case. The detailed derivations can be found in the SM, Appendix VII.

For the backward process, it is useful to borrow notation from the forward process and denote

$$S = \begin{bmatrix} S_x & S_{x;m} \\ S_{m;x} & S_m \end{bmatrix}$$

$$S_{x|m} = S_x - S_{x;m} S_m^\dagger S_{m;x}$$

$$S_{u|m} = R + B^\top S_{x|m} B.$$

Then we can find the feedback gain

$$L = -(R + B^\top S B)^\dagger B^\top S A$$

$$= \begin{bmatrix} L_{u;x|m} & 0 \\ -S_m^\dagger S_{m;x} (A_x + B_{x;u} L_{u;x|m}) & 0 \end{bmatrix}, \quad (8)$$

with a memory-conditioned form of the classic feedback gain

$$L_{u;x|m} = -S_{u|m}^\dagger B_{x;u}^\top S_{x|m} A_x.$$

The memory-conditioned cost reduction matrix is

$$N = L^\top (R + B^\top S B) L = \begin{bmatrix} N_{x|m} & 0 \\ 0 & 0 \end{bmatrix},$$

with

$$N_{x|m} = A_x^\top (S_x - S_{x|m} + S_{x|m} B_{x;u} S_{u|m}^\dagger B_{x;u}^\top S_{x|m}) A_x.$$

Thus  $\text{rank}(D) \leq \text{rank}(N) \leq n$ , with  $D$  the mode activation matrix (see Algorithm 3), implying that at most  $n$  modes can be active.

The  $d$  rightmost columns in (8) are 0, implying that  $\tilde{u}_t$  depends only on the state estimator  $\hat{x}_{\tilde{u}_t} = \mathbb{E}[x_t | \tilde{u}_t]$  of  $x_t$  and not on an estimator of the memory component  $m_t$ . Since  $\hat{x}_{\tilde{y}_t} = \mathbb{E}[x_t | \tilde{y}_t]$  is a sufficient statistic of  $\tilde{y}_t$  for  $x_t$ , we also have the Markov chain

$$x_t \text{ --- } \hat{x}_{\tilde{y}_t} \text{ --- } \tilde{y}_t \text{ --- } \hat{\tilde{x}}_{\tilde{y}_t} \text{ --- } \hat{\tilde{x}}_{\tilde{u}_t} \text{ --- } \hat{x}_{\tilde{u}_t} \text{ --- } \tilde{u}_t,$$

with

$$\hat{\tilde{x}}_{\tilde{y}_t} = \mathbb{E}[\tilde{x}_t | \tilde{y}_t] = \mathbb{E} \left[ \begin{bmatrix} x_t \\ m_t \end{bmatrix} \middle| \begin{bmatrix} y_t \\ m_t \end{bmatrix} \right]$$

and similarly for  $\hat{\tilde{x}}_{\tilde{u}_t}$ . This implies that we need only consider the first component  $\hat{x}_{\tilde{y}_t}$  of  $\hat{\tilde{x}}_{\tilde{y}_t}$ , which is obtained from the observation  $\tilde{y}_t$  using

$$K = \Sigma_x C^\top \Sigma_{\tilde{y}}^\dagger$$

$$= [K_{x;y|m} \quad (I - K_{x;y|m} C_{y;x}) \Sigma_{x;m} \Sigma_m^\dagger],$$

where

$$K_{x;y|m} = \Sigma_{x|m} C_{y;x}^\top \Sigma_{y|m}^\dagger$$

is the Kalman gain that performs optimal inference in the classic LQG task [7].

Crucially, we see that  $\hat{x}_{\tilde{y}_t}$  depends on  $m_t$  only through

$$\hat{x}_{m_t} = \mathbb{E}[x_t | m_t] = \Sigma_{x;m} \Sigma_m^\dagger m_t.$$

This implies that, for a controller  $\pi$ , we can design an equivalent controller  $\pi'$  whose memory state is the MMSE estimator  $m'_t = \hat{x}_{m_t}$ . The feedback gain for  $\pi'$  is

$$L' = \begin{bmatrix} I & 0 \\ 0 & \Sigma_{x;m} \Sigma_m^\dagger \end{bmatrix} L.$$

Note that, since  $m'_t$  is a sufficient statistic of  $m_t$  for  $x_t$ , we have  $\Sigma_{x|m'} = \Sigma_{x|m}$  and  $K_{x;y|m'} = K_{x;y|m}$ . Thus

$$K' = [K_{x;y|m} \quad I - K_{x;y|m} C_{y;x}],$$

with  $\Sigma_{x;m'} \Sigma_{m'}^\dagger = \Sigma_{m'} \Sigma_{m'}^\dagger$  in the second component omitted due to its redundancy.

The controllers  $\pi$  and  $\pi'$  generate the same control  $u_t$  and thus incur the same external cost. At the same time, since  $m'_t$  is a function of  $m_t$ , by the data-processing inequality the information rate of  $\pi'$  is at most that of  $\pi$ . Thus any controller can be converted into a MMSE controller without loss of performance, allowing us to consider the MMSE controller canonical. In particular, this proves again that  $d = n$  is always sufficient for representing the memory state.

We now diverge from the solution given in Section II-B, which has freedom in its choice of memory representation, and is therefore not guaranteed to be a MMSE controller. Instead, we explicitly constrain the controller to be MMSE, which in return enables us to relax some of the conditions

given in Section II-B, which are now not necessary (and indeed do not hold at the optimum), as discussed below.

Constraining the controller to be MMSE imposes the structure

$$\Sigma_{\hat{x}} = \begin{bmatrix} \Sigma_{x|m} + \Sigma_m & \Sigma_m \\ \Sigma_m & \Sigma_m \end{bmatrix},$$

parameterized by  $\Sigma_{x|m}$  and  $\Sigma_m$ . The reduced number of independent parameters leaves  $M$  overparameterized (see SM, Appendix VII) and we can choose, without loss of performance, the structure

$$M = \begin{bmatrix} M_{x|m} + M_m & -M_m \\ -M_m & M_m \end{bmatrix}$$

with

$$\begin{aligned} M_{x|m} &= \beta^{-1} Z \\ M_m &= \beta^{-1} (C_{y;x}^\top K_{x;y|m}^\top Z K_{x;y|m} C_{y;x} - Z), \end{aligned}$$

where  $Z = \Sigma_{\hat{x}_y|\hat{x}_u}^\dagger - \Sigma_{\hat{x}_y}^\dagger$  is the signal-to-noise-ratio (SNR) matrix for the channel  $\hat{x}_{\hat{y}_t} \rightarrow \hat{x}_{\hat{u}_t}$ . Due to the shrinkage effect of  $K_{x;y|m} C_{y;x}$

$$M_m \leq 0 \leq M_{x|m} + M_m.$$

The Hessian of the cost-to-go now has the form

$$\begin{aligned} S &= Q + A^\top S A - M \\ &= \begin{bmatrix} Q_x + A_x^\top S_x A_x - M_{x|m} - M_m & M_m \\ M_m & -M_m \end{bmatrix} \end{aligned}$$

and the second-order expansion of the cost-to-go, at the optimum, has the form

$$\begin{aligned} \tilde{x}_t^\top S \tilde{x}_t &= x_t^\top (Q_x + A_x^\top S_x A_x - \beta^{-1} Z) x_t \\ &\quad - (m_t - x_t)^\top M_m (m_t - x_t). \end{aligned}$$

The first term measures the divergence of the state  $x_t$  from 0 and the second the divergence of the controller's estimator  $m_t$  from the true state  $x_t$ , which is the expected form for a MMSE controller. Both terms link the SNR matrix  $Z$  to the cost reduction. In this form,  $S$  is again positive semidefinite, while now  $M$  is generally not.

Finally, when  $\beta = \infty$ , we can recover the classic LQG results. Similarly to Part I [1, Section IV-B], we can substitute  $N_{x|m}$  for  $\beta^{-1} Z$ , to recover the algebraic Riccati equation

$$\begin{aligned} S_{x|m} &= Q_x + A_x^\top S_x A_x - N_{x|m} \\ &= Q_x + A_x^\top (S_{x|m} - S_{x|m} B_{x;u} S_{u|m}^\dagger B_{x;u}^\top S_{x|m}) A_x. \end{aligned}$$

#### IV. EXAMPLE

As a simple example, consider the double mass-spring-damper system in Figure 5, adapted from [13]. The

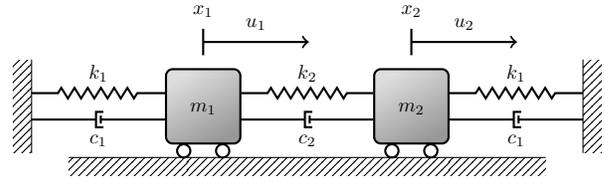


Fig. 5. Double mass-spring-damper system; masses:  $m_1 = 5$  kg,  $m_2 = \sqrt{15}$  kg; spring constants:  $k_1 = 1$  N/m,  $k_2 = 0.5$  N/m; damping coefficients:  $c_1 = c_2 = 1$  N·sec/m

continuous-time dynamics of this system are given by

$$\begin{aligned} A &= \begin{bmatrix} 0 & 1 & 0 & 0 \\ -\frac{k_1+k_2}{m_1} & -\frac{c_1+c_2}{m_1} & \frac{k_2}{m_1} & \frac{c_2}{m_1} \\ 0 & 0 & 0 & 1 \\ \frac{k_2}{m_2} & \frac{c_2}{m_2} & -\frac{k_1+k_2}{m_2} & -\frac{c_1+c_2}{m_2} \end{bmatrix} \\ B &= \begin{bmatrix} 0 & 0 \\ \frac{1}{m_1} & 0 \\ 0 & 0 \\ 0 & \frac{1}{m_2} \end{bmatrix} \quad C = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}, \end{aligned}$$

with  $m_1 = 5$  kg,  $m_2 = \sqrt{15}$  kg,  $k_1 = 1$  N/m,  $k_2 = 0.5$  N/m and  $c_1 = c_2 = 1$  N·sec/m. We discretize the time using the Tustin transformation with sampling frequency 20Hz and consider the isotropic noises and cost rates

$$\Sigma_\xi = I \quad \Sigma_\epsilon = I \quad Q = I \quad R = I.$$

For the memoryless control problem, we initialize a solution with  $\Sigma_x = S = 0$ . For the retentive control problem, we apply the reduction in Lemma 2 to obtain a reduced plant and then initialize a solution using the classic LQG controller, as described in Section III-C. To the initial solution, we apply the forward-backward iterations of Section II-B, with fixed  $\beta$ , until convergence to a fixed point, suspected as a global optimum. To improve running time, we employ a reverse-annealing scheme, decreasing  $\beta$  gradually over its range and using the fixed point for one value of  $\beta$  to initialize the iterations for the next value of  $\beta$ .

Figures 6 and 7 show, respectively, the resulting cost-log-beta and cost-information curves, demonstrating that even this simple example exhibits interesting phenomenology.

We see that both the memoryless (blue) and the retentive (green) controllers undergo phase transitions as  $\beta$  increases. The system is controllable and observable, allowing the retentive controller to undergo 4 phase transitions, until it fully remembers and controls all modes of the system. However, the rank-2 matrices  $B$  and  $C$  only allow the memoryless controller to undergo 2 phase transitions and reach order  $d = 2$ .

In the first phase transition, the controllers begin controlling a single mode, in order to reduce the external cost, at the expense of communication resources. This is not depicted in the cost-information plot (Figure 7), since below this critical point the information is 0 and the cost is fixed.

The second and fourth phase transitions involve memory and only occur in the retentive controller. Below these critical

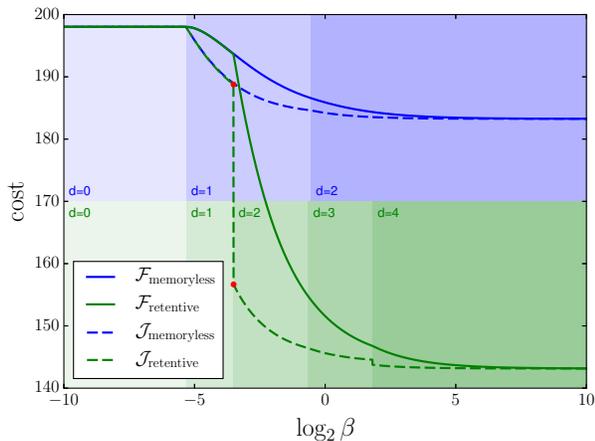


Fig. 6. Cost-log-beta curve for the double mass-spring-damper problem. Memoryless control (blue) generally incurs higher cost than retentive control (green). The Lagrangian  $\mathcal{F}$  (solid) is continuous, whereas the external cost  $\mathcal{J}$  (dashed) is discontinuous in the retentive case in phase transitions 2 (red dots) and 4. Background shades indicate the controller order  $d$ , with boundaries at critical points.

point, a hypothetical order-2 retentive controller is worse than the order-1 controller, in terms of the target  $\mathcal{F}$ , the total external and internal cost-to-go it incurs. At the critical point, the order-2 controller overtakes the order-1 controller, already with a significantly reduced cost rate and a significant information rate (see red dots in Figures 6 and 7). The critical point is where the ratio between these costs is  $\beta^{-1}$  (see (12) in Part I [1, Section IV-B]).

The third phase transition is again common to the memoryless and the retentive controllers, although by now the retentive controller has committed to memory much valuable information, reducing the cost much beyond the capabilities of the memoryless controller.

## V. DISCUSSION

In this paper we introduce the problem of optimal LQG control with bounded channel capacity in both the memory and the sensory channels. We show how to reduce this problem to that of bounded memoryless LQG control, study the structure of the resulting solution and illustrate its interesting phenomenology with a simple example.

One aspect of this phenomenology that merits further study is the existence of suboptimal fixed points of the iterative algorithm (Section II-B). For example, around the second critical point in the double mass-spring-damper system (Section IV), both an order-1 controller and a retentive order-2 controller are fixed points. Before the phase transition, one of these solutions is stable, while the other is metastable and suboptimal, and at the phase transition they switch. This resembles well-studied phenomena in statistical physics.

LQG control with constraints on the sensory channel capacity has now been studied in the regime of unlimited memory [11], no memory (Part I of this work [1]) and in this paper, a shared channel capacity for sensing and memory. More generally, the memory and the sensory channels can

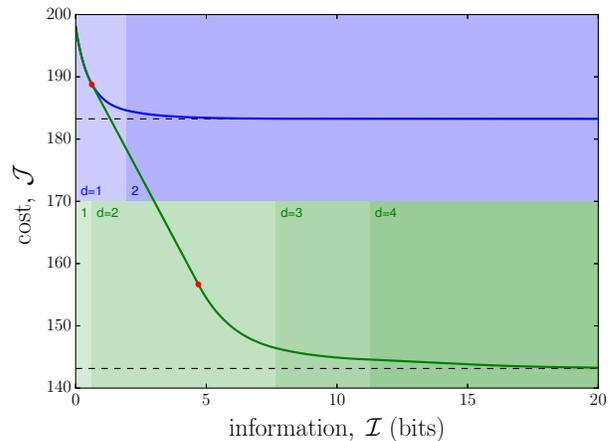


Fig. 7. Cost-information curve for the double mass-spring-damper problem. Memoryless control (blue) incurs higher cost than retentive control (green) after phase transition 2 (red dots). The asymptotic costs at  $\beta = \infty$  (dashed black) can be approximated with very little information and a reduced order.

be separate, with their relative costs ranging from 0 (no memory) to 1 (shared capacity) to  $\infty$  (unlimited memory) including any intermediate value. This memory-sensory trade-off has been studied in the context of finite-state systems [4] and further insight can be gained from studying this more general problem in the LQG context.

## REFERENCES

- [1] R. Fox and N. Tishby, "Minimum-information LQG control — Part I: Memoryless controllers," in *Proceedings of the 55th IEEE Conference on Decision and Control (CDC)*, 2016. [Online]. Available: <https://arxiv.org/abs/1606.01946>
- [2] V. Cutsuridis, A. Hussain, and J. G. Taylor, *Perception-Action Cycle: Models, Architectures, and Hardware*, ser. Cognitive and Neural Systems. Springer, 2011.
- [3] D. A. Patterson, "Latency lags bandwidth," *Communications of the ACM*, vol. 47, no. 10, pp. 71–75, 2004.
- [4] R. Fox and N. Tishby, "Bounded planning in passive POMDPs," in *Proceedings of the 29th International Conference on Machine Learning*, 2012, pp. 1775–1782.
- [5] —, "Optimal selective attention in reactive agents," *arXiv preprint arXiv:1512.08575*, 2015.
- [6] N. Meuleau, K.-E. Kim, L. P. Kaelbling, and A. R. Cassandra, "Solving POMDPs by searching the space of finite policies," in *Proceedings of the 15th conference on Uncertainty in Artificial Intelligence (UAI)*. Morgan Kaufmann Publishers Inc., 1999, pp. 417–426.
- [7] D. P. Bertsekas, *Dynamic Programming and Optimal Control*. Athena Scientific, 1995, vol. I.
- [8] H. S. Witsenhausen, "A counterexample in stochastic optimum control," *SIAM Journal on Control*, vol. 6, no. 1, pp. 131–147, 1968.
- [9] S. Mitter and A. Sahai, "Information and control: Witsenhausen revisited," in *Learning, Control and Hybrid Systems*, ser. Lecture Notes in Control and Information Sciences. Springer, 1999, vol. 241, pp. 281–293.
- [10] J. L. Massey, "Causality, feedback and directed information," in *Proceedings of the International Symposium on Information Theory and Its Applications*, 1990, pp. 303–305.
- [11] T. Tanaka, P. M. Esfahani, and S. K. Mitter, "LQG control with minimal information: Three-stage separation principle and SDP-based solution synthesis," *arXiv preprint arXiv:1510.04214*, submitted for publication.
- [12] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. Wiley, 2012.
- [13] F. Creutzig, A. Globerson, and N. Tishby, "Past-future information bottleneck in dynamical systems," *Physical Review E*, vol. 79, no. 4, p. 041925, 2009.