# Structural Results and Explicit Solution for Two-Player LQG Systems on a Finite Time Horizon

Laurent Lessard[1] Ashutosh Nayyar[2]

IEEE Conference on Decision and Control, pp. 6542–6549, 2013

## Abstract

It is well-known that linear dynamical systems with Gaussian noise and quadratic cost (LQG) satisfy a separation principle. Finding the optimal controller amounts to solving separate dual problems; one for control and one for estimation. For the discrete-time finite-horizon case, each problem is a simple forward or backward recursion. In this paper, we consider a generalization of the LQG problem with two controllers and a partially nested information structure. Each controller is responsible for one of two system inputs, but has access to different subsets of the available measurements. Our paper has three main contributions. First, we prove a fundamental structural result: sufficient statistics for the controllers can be expressed as conditional means of the global state. Second, we give explicit state-space formulae for the optimal controller. These formulae are reminiscent of the classical LQG solution with dual forward and backward recursions, but with the important difference that they are intricately coupled. Lastly, we show how these recursions can be solved efficiently, with computational complexity comparable to that of the centralized problem.

## 1 Introduction

With the advent of large systems operating on a global scale such as the internet or power networks, the past decade has seen a resurgence of interest in decentralized control. For such large systems, it is inevitable that some control decisions must be made using only local or partial information. Two natural questions that arise are:

1. Can the ever-growing information history be aggregated without compromising achievable performance? In other words, what are sufficient statistics for the decision-makers?

2. When and how can optimal decentralized policies be efficiently computed?

In this paper, we give complete answers to the above questions for a fundamental decentralized control problem: the two-player partially nested LQG problem.

Briefly, our problem consists of two linear Gaussian systems with their own local controllers. The systems are coupled. System 1 affects System 2 through its state and input but not vice-versa, and the controller for System 1 shares its measurement with the controller for System 2 but not vice versa. A formal description of the problem is given in Section 3.

It is believed that decentralized control problems are likely hard in general [1, 16]. However, partially-nested LQG problems admit an optimal controller that is linear [3]. In decentralized control problems, partial nestedness is typically manifested in two ways:

1. If a subsystem $i$ affects another subsystem $j$, then the controller at subsystem $i$ shares all information with the controller at subsystem $j$. In other words, the information flow obeys the *sparsity constraints* of the dynamics.

2. If subsystem $i$ affects subsystem $j$ after some delay $d$, then controller at subsystem $i$ shares its information with controller at subsystem $j$ with delay not exceeding $d$. In other words, the information flow obeys the *delay constraints* of the dynamics.

Several combinations of the above two manifestations of partial nestedness have been explored under state feedback assumptions. State feedback problems have been investigated in [15] under sparsity constraints, in [5] under delay constraints and in [6] under a mixture of delay and sparsity constraints. A state feedback problem where partial nestedness was captured by a partial order on subsystems was investigated in [14].

The problem considered in this paper is an output feedback problem where controllers observe noisy measurements of states. With only two controllers, it is perhaps the simplest output feedback problem. However, as we shall see, having noisy measurements introduces a nontrivial coupling between estimation and control and complicates the solution significantly. An explicit solution to the continuous-time version of the two-player problem as well as an extension to the broadcast case appeared in [7, 8, 9]. These works use a spectral factorization approach that is completely different from the common information approach used herein. Furthermore, they solve the problem over an infinite time horizon, which makes the coupling between estimation and control simpler due

[1] L. Lessard is with the Department of Mechanical Engineering at the University of California, Berkeley, CA 94720, USA. lessard@berkeley.edu

[2] A. Nayyar is with the Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, CA 94720, USA. anayyar@berkeley.edu

to the steady-state assumption. A particular case of output-feedback partially nested LQG problem namely the one-step delayed sharing problem was investigated in [13].

Our paper has three main contributions. In Section 4, we find sufficient statistics for the two-player LQG problem. Our result relies on a common information based approach developed in [10] and [11]. In Section 5, we give an explicit state-space solution to the two-player problem using dynamic programming. Lastly, in Section 6, we show how to efficiently compute the solution to the two-player problem, and show that it can be done with computational effort comparable to that required for the centralized version of the problem. Namely, computational effort is proportional to the length of the time horizon.

## 2 Notation

Real vectors and matrices are represented by lower- and upper-case letters respectively. Boldface symbols denote random vectors, and their non-boldface counterparts denote particular realizations. The probability density function of $\mathbf{x}$ evaluated at $x$ is denoted $\mathbb{P}(\mathbf{x} = x)$, and conditional densities are written as $\mathbb{P}(\mathbf{x} \,|\, \mathbf{y} = y)$. We write $\mathbf{x} = \mathcal{N}(\mu, \Sigma)$ when $\mathbf{x}$ is normally distributed with mean $\mu$ and variance $\Sigma$. This paper considers stochastic processes in discrete time over a finite time interval $[0, T]$. Time is indicated using subscripts, and we use the colon notation to denote ranges. For example: $x_{0:T-1} = \{x_0, x_1, \ldots, x_{T-1}\}$. In general, all symbols are time-varying. In an effort to present general results while keeping equations clear and concise, we introduce a new notation to represent a family of equations. We write

$$\mathbf{x}_+ \overset{t}{=} A\mathbf{x} + \mathbf{w}$$

to mean that $\mathbf{x}_{t+1} = A_t\mathbf{x}_t + \mathbf{w}_t$ holds for $0 \le t \le T-1$. The subscript "+" indicates that the associated symbol is incremented to $t+1$. We similarly overload summations:

$$\sum_t x^\mathsf{T} Q x \quad \text{instead of writing} \quad \sum_{t=0}^{T-1} x_t^\mathsf{T} Q_t x_t$$

Any time we use $t$ above a binary relation or below a summation, it is implied that $0 \le t \le T-1$. The same time horizon $T$ is used throughout this paper.

We denote subvectors by using superscripts so that they are not confused with time indices. For submatrices, which require double indexing, we will interchangeably use superscripts and subscripts to minimize clutter. For example, we write $P_+^{21}$ and $A_{22}^\mathsf{T}$ to avoid writing $P_{21,+}$ and $A^{22,\mathsf{T}}$ respectively. We also introduce matrices $E_1$ and $E_2$ to aid in the manipulation of $2 \times 2$ block matrices. We partition an identity matrix as $I = \begin{bmatrix} E_1 & E_2 \end{bmatrix}$ where the dimension of $E_i$ is inferred by context. For example, suppose $B \in \mathbb{R}^{(n_1+n_2)\times(m_1+m_2)}$, with the block-triangular structure given by

$$B = \begin{bmatrix} B_{11} & 0 \\ B_{21} & B_{22} \end{bmatrix} \qquad \text{where } B_{ij} \in \mathbb{R}^{n_i \times m_j}$$

then we may write $E_2^\mathsf{T} B E_1 = B_{21}$ and $E_1^\mathsf{T} B = B_{11} E_1^\mathsf{T}$. However, the same symbol may vary in dimension depending on context. In this case, either $E_i \in \mathbb{R}^{(n_1+n_2)\times n_i}$ or $E_i \in \mathbb{R}^{(m_1+m_2)\times m_i}$ depending on whether the symbol multiplies $B$ on the left or right, respectively.

## 3 Problem statement

Consider two interconnected linear systems with the following state update and measurement equations.

$$\begin{aligned} \begin{bmatrix} \mathbf{x}_+^1 \\ \mathbf{x}_+^2 \end{bmatrix} &\overset{t}{=} \begin{bmatrix} A_{11} & 0 \\ A_{21} & A_{22} \end{bmatrix} \begin{bmatrix} \mathbf{x}^1 \\ \mathbf{x}^2 \end{bmatrix} + \begin{bmatrix} B_{11} & 0 \\ B_{21} & B_{22} \end{bmatrix} \begin{bmatrix} \mathbf{u}^1 \\ \mathbf{u}^2 \end{bmatrix} + \begin{bmatrix} \mathbf{w}^1 \\ \mathbf{w}^2 \end{bmatrix} \\ \begin{bmatrix} \mathbf{y}^1 \\ \mathbf{y}^2 \end{bmatrix} &\overset{t}{=} \begin{bmatrix} C_{11} & 0 \\ C_{21} & C_{22} \end{bmatrix} \begin{bmatrix} \mathbf{x}^1 \\ \mathbf{x}^2 \end{bmatrix} + \begin{bmatrix} \mathbf{v}^1 \\ \mathbf{v}^2 \end{bmatrix} \end{aligned} \tag{1}$$

For brevity, we write the vector $(\mathbf{x}_t^1, \mathbf{x}_t^2)$ above as simply $\mathbf{x}_t$ and similarly for $\mathbf{y}_t$, $\mathbf{u}_t$, $\mathbf{w}_t$, $\mathbf{v}_t$. The random vectors in the collection

$$\left\{ \mathbf{x}_0, \begin{bmatrix} \mathbf{w}_0 \\ \mathbf{v}_0 \end{bmatrix}, \ldots, \begin{bmatrix} \mathbf{w}_{T-1} \\ \mathbf{v}_{T-1} \end{bmatrix} \right\} \tag{2}$$

are mutually independent and jointly Gaussian with the following known probability density functions.

$$\begin{aligned} \mathbf{x}_0 &= \mathcal{N}(0, \Sigma_{\text{init}}) \\ \begin{bmatrix} \mathbf{w} \\ \mathbf{v} \end{bmatrix} &\overset{t}{=} \mathcal{N}\left(0, \begin{bmatrix} W & U^\mathsf{T} \\ U & V \end{bmatrix}\right) \end{aligned} \tag{3}$$

There are two controllers, and the information available to each controller at time $t$ is

$$\begin{aligned} \hat{\mathbf{i}}_t &= \{\mathbf{y}_{0:t-1}^1, \mathbf{u}_{0:t-1}^1\} \\ \mathbf{i}_t &= \{\mathbf{y}_{0:t-1}^1, \mathbf{y}_{0:t-1}^2, \mathbf{u}_{0:t-1}^1, \mathbf{u}_{0:t-1}^2\} \end{aligned} \tag{4}$$

The controllers select actions according to control strategies $f^i := (f_0^i, f_1^i, \ldots, f_{T-1}^i)$ for $i = 1, 2$. That is,

$$\mathbf{u}_t^1 = f_t^1(\hat{\mathbf{i}}_t) \text{ and } \mathbf{u}_t^2 = f_t^2(\mathbf{i}_t) \quad \text{for } 0 \le t \le T-1 \tag{5}$$

The performance of control strategies $f^1, f^2$ is measured by the finite horizon expected quadratic cost given by

$$\hat{\mathcal{J}}_0(f^1, f^2) = \mathbb{E}^{f^1,f^2}\left( \sum_t \begin{bmatrix} \mathbf{x} \\ \mathbf{u} \end{bmatrix}^\mathsf{T} \begin{bmatrix} Q & S \\ S^\mathsf{T} & R \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ \mathbf{u} \end{bmatrix} + \mathbf{x}_T^\mathsf{T} P_{\text{final}} \mathbf{x}_T \right) \tag{6}$$

The expectation is taken with respect to the joint probability measure on $(\mathbf{x}_{0:T}, \mathbf{u}_{0:T-1})$ induced by the choice of $f^1$ and $f^2$. We are interested in the following problem.

**Problem 1** (Two-Player LQG)**.** *For the model* (1)–(5), *find control strategies $f^1, f^2$ that minimize the cost* (6).

A related and well-known problem is the centralized LQG problem. It is the special case of the two-player problem for which there is a single decision-maker.

**Problem 2** (Centralized LQG)**.** *Consider the model* (1)–(3)*, where* $A$*,* $B$*,* $C$ *are no longer required to be block-lower-triangular. Suppose* $\mathbf{u}_t = f_t(\mathbf{i}_t)$*, where* $\mathbf{i}_t := (\mathbf{y}_{0:t-1}, \mathbf{u}_{0:t-1})$*, and our goal is to choose* $f := f_{0:T-1}$ *such that we minimize*

$$\mathcal{J}_0(f) = \mathbb{E}^f \left( \sum_t \begin{bmatrix} \mathbf{x} \\ \mathbf{u} \end{bmatrix}^\mathsf{T} \begin{bmatrix} Q & S \\ S^\mathsf{T} & R \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ \mathbf{u} \end{bmatrix} + \mathbf{x}_T^\mathsf{T} P_{\text{final}} \mathbf{x}_T \right)$$

*The expectation is with respect to the joint probability measure on* $(\mathbf{x}_{0:T}, \mathbf{u}_{0:T-1})$ *induced by the choice of* $f$*.*

In both Problem 1 and Problem 2, the sizes of the various matrices and vectors may also vary with time. It is assumed that $\Sigma_{\text{init}}, P_{\text{final}}$, as well as the values of $A, B, C, Q, R, S, U, V, W$ for all $t$, are available to all decision-makers for any $t \geq 0$. We also clarify that while we often call the decision-making agents *players*, this is not a game. The players are cooperative and their strategies are to be jointly optimized.

## 4 Structural results

In Problem 1, the lower triangular nature of the state, control and observation matrices of (1) implies that Player 1's state and control actions affect Player 2's information but not vice versa. Further, any information available of Player 1 is also available to Player 2. Hence, Problem 1 is *partially nested* and the optimal strategies for the two players are linear functions of their respective information histories [3, Thm. 2].

In this section, we show that the information histories can be aggregated into sufficient statistics. In Sections 5 and 6, we will use this fact to derive a recursive finite-memory implementation of the optimal controller. We start with a well-known structural result for the centralized LQG problem (Problem 2).

**Lemma 3.** *In Problem 2, there exists an optimal control strategy of the form* $u \overset{t}{=} Kz$ *where* $z_t := \mathbb{E}(\mathbf{x}_t \,|\, \mathbf{i}_t = i_t)$*, and* $K_{0:T-1}$ *are fixed matrices of appropriate dimensions.*

We will also make use of some properties of conditional expectations, which apply because of the nested information. We state the result as a lemma.

**Lemma 4.** *Suppose* $\hat{\mathbf{i}}$ *and* $\mathbf{i}$ *are information sets that satisfy* $\hat{\mathbf{i}} \subset \mathbf{i}$*. Define conditional estimates* $z := \mathbb{E}(\mathbf{x} \,|\, \mathbf{i} = i)$ *and* $\hat{z} := \mathbb{E}(\mathbf{x} \,|\, \hat{\mathbf{i}} = \hat{i}\,)$*. Then*

*(i)* $\mathbb{E}(\hat{\mathbf{i}} \,|\, \mathbf{i} = i\,) = \hat{i}$ *(since* $\hat{i} \subset i$*)*

*(ii)* $\mathbb{E}(\mathbf{z} \,|\, \hat{\mathbf{i}} = \hat{i}\,) = \hat{z}$ *(smoothing property)*

### 4.1 Structural result for Player 2

We now turn our attention to Problem 1. Consider any arbitrary linear strategy for Player 1. Thus, Player 1's control actions are of the form

$$u^1 \overset{t}{=} G\,\hat{i} \tag{7}$$

where $G_{0:T-1}$ are fixed matrices of appropriate dimensions and $\hat{i}_t$ is the realization of Player 1's information. Given this strategy for Player 1, we want to find the optimal strategy for Player 2.

In the next two results, we show that once Player 1's strategy is fixed, finding the optimal strategy for Player 2 amounts to solving a centralized LQG problem. Thus we may apply the structural result presented in Lemma 3.

**Lemma 5.** *Consider Problem 1, and assume any fixed strategy for Player 1 given by* (7)*. Define* $\bar{\mathbf{x}}_t$ *as follows.*

$$\bar{\mathbf{x}}_t := \begin{bmatrix} \mathbf{x}_t \\ \hat{\mathbf{i}}_t \end{bmatrix}, \bar{\mathbf{y}}_t := \begin{bmatrix} \mathbf{y}_t \\ \hat{\mathbf{i}}_t \end{bmatrix} \quad \textit{for} \quad 0 \leq t \leq T$$

*Then, the following statements are true.*

*(i) There exist matrices* $\bar{A}_t, \bar{B}_t, \bar{C}_t, \bar{D}_t$ *such that*

$$\begin{aligned} \bar{\mathbf{x}}_0 &= \mathcal{N}(0, \Sigma_{\text{init}}) \\ \bar{\mathbf{x}}_+ &\overset{t}{=} \bar{A}\bar{\mathbf{x}} + \bar{B}\mathbf{u}^2 + \bar{D} \begin{bmatrix} \mathbf{w} \\ \mathbf{v} \end{bmatrix} \\ \bar{\mathbf{y}} &\overset{t}{=} \bar{C}\bar{\mathbf{x}} + \mathbf{v} \end{aligned}$$

*(ii) There exist matrices* $\bar{Q}_t, \bar{R}_t, \bar{S}_t, \bar{P}_{\text{final}}$ *such that the total expected cost can be written as*

$$\mathbb{E} \left( \sum_t \begin{bmatrix} \bar{\mathbf{x}} \\ \mathbf{u}^2 \end{bmatrix}^\mathsf{T} \begin{bmatrix} \bar{Q} & \bar{S} \\ \bar{S}^\mathsf{T} & \bar{R} \end{bmatrix} \begin{bmatrix} \bar{\mathbf{x}} \\ \mathbf{u}^2 \end{bmatrix} + \bar{\mathbf{x}}_T^\mathsf{T} \bar{P}_{\text{final}} \bar{\mathbf{x}}_T \right)$$

**Proof.** The proof follows from the definition of $\bar{\mathbf{x}}_t$, the state, observation, and cost equations of Problem 1, and the fixed strategy for Player 1 given by (7). ■

**Theorem 6.** *Consider Problem 1. For any choice of Player 1's strategy, the optimal strategy for Player 2 has the structure*

$$u^2 \overset{t}{=} H^1\,\hat{i} + H^2 z \tag{8}$$

*where* $z_t := \mathbb{E}(\mathbf{x}_t \,|\, \mathbf{i}_t = i_t)$ *and* $\hat{i}_t \subset i_t$ *is the realization of Player 1's information. Further,* $z_t$ *has a linear update equation that does not depend on the choice of Player 1's strategy. This linear update equation is the standard Kalman filter, given explicitly in* (15)–(16)*.*

**Proof.** Lemma 5 implies that when Player 1's strategy is fixed, the optimization problem for Player 2 is an instance of the centralized LQG problem (Problem 2) with $\bar{\mathbf{x}}_t$ as the state of the linear system, $\mathbf{y}_t$ as the observation, and $\mathbf{u}_t^2$ as the control action. Therefore, by Lemma 3, the optimal strategy for Player 2 is of the form $u_t^2 = H_t\,\mathbb{E}(\bar{\mathbf{x}}_t \,|\, \mathbf{i}_t = i_t)$ for some matrix $H_t$. Further, it follows from Lemma 4 that

$$\mathbb{E}(\bar{\mathbf{x}}_t \,|\, \mathbf{i}_t = i_t) = \begin{bmatrix} \mathbb{E}(\mathbf{x}_t \,|\, \mathbf{i}_t = i_t) \\ \mathbb{E}(\,\hat{\mathbf{i}}_t \,|\, \mathbf{i}_t = i_t) \end{bmatrix} = \begin{bmatrix} z_t \\ \hat{i}_t \end{bmatrix} \tag{9}$$

Therefore, the optimal strategy for Player 2 is of the form $u_t^2 = H_t\,\mathbb{E}(\bar{\mathbf{x}}_t \,|\, \mathbf{i}_t = i_t) = H_t^1\,\hat{i}_t + H_t^2 z_t$, as required. ■

### 4.2 Joint structural result

We may rewrite the result of Theorem 6 as

$$u^2 \overset{t}{=} \tilde{u}^2 + H^2 z \tag{10}$$

where $\tilde{u}_t^2$ is a linear function of Player 1's information. Note that $\tilde{u}_t^2$ and $u_t^1$ are linear functions of the same information. In order to further characterize the structure of optimal strategies, we consider a coordinated system where a coordinator knows the common information among the players (that is, $\hat{i}_t$) and selects both $\tilde{u}_t^2$ and $u_t^1$ based on this common information. Once the coordinator selects $\tilde{u}_t^2$, Player 2's control action is $u_t^2 = H_t^2 z_t + \tilde{u}_t^2$, for some $H_t^2$. It is clear that any strategy of the form (10) can be implemented in the coordinated system.

Given an arbitrary choice of $H_t^2$, we want to find the optimal strategy for the coordinator. As in Lemma 5, this can be formulated as a centralized LQG problem.

**Lemma 7.** *Consider Problem 1 where $u_t^2$ is given by* (10)*, and assume any fixed choice of $H_t^2$. Define $\tilde{\mathbf{x}}_t$ as follows.*

$$\tilde{\mathbf{x}}_t := \begin{bmatrix} \mathbf{x}_t \\ \mathbf{z}_t \end{bmatrix} \quad \textit{for} \quad 0 \le t \le T$$

*Then, the following statements are true.*

*(i) There exist matrices $\tilde{A}_t$, $\tilde{B}_t$, $\tilde{D}_t$, and $\tilde{\Sigma}_{\text{init}}$ such that*

$$\begin{aligned} \tilde{\mathbf{x}}_0 &= \mathcal{N}(0, \tilde{\Sigma}_{\text{init}}) \\ \tilde{\mathbf{x}}_+ &\overset{t}{=} \tilde{A}\tilde{\mathbf{x}} + \tilde{B} \begin{bmatrix} \mathbf{u}^1 \\ \tilde{\mathbf{u}}^2 \end{bmatrix} + \tilde{D} \begin{bmatrix} \mathbf{w} \\ \mathbf{v} \end{bmatrix} \\ \mathbf{y}^1 &\overset{t}{=} C_{11}\mathbf{x}^1 + \mathbf{v}^1 \end{aligned}$$

*(ii) There exist matrices $\Omega_t$ and $\Omega_{\text{final}}$ such that the total expected cost can be written as*

$$\mathbb{E}\left( \sum_t \begin{bmatrix} \tilde{\mathbf{x}} \\ \mathbf{u}^1 \\ \tilde{\mathbf{u}}^2 \end{bmatrix}^{\mathsf{T}} \Omega \begin{bmatrix} \tilde{\mathbf{x}} \\ \mathbf{u}^1 \\ \tilde{\mathbf{u}}^2 \end{bmatrix} + \tilde{\mathbf{x}}_T^{\mathsf{T}}\, \Omega_{\text{final}}\, \tilde{\mathbf{x}}_T \right)$$

**Proof.** The proof follows from the definition of $\tilde{\mathbf{x}}_t$, the state, observation, and cost equations of Problem 1, and the fact that $z_t$ has a linear update equation. This linear update equation is the standard Kalman filter, given explicitly in (15)–(16). ■

**Theorem 8.** *The optimal strategies for the two players in Problem 1 are of the form*

$$u_t^1 = G_t \hat{z}_t \qquad u_t^2 = H_t^1 \hat{z}_t + H_t^2 z_t \tag{11}$$

*where $\hat{z}_t := \mathbb{E}(\mathbf{x}_t \,|\, \hat{\mathbf{i}}_t = \hat{i}_t)$ and $z_t := \mathbb{E}(\mathbf{x}_t \,|\, \mathbf{i}_t = i_t)$.*

**Proof.** Lemma 7 implies that when $H_t^2$ is fixed in (10), the coordinator's optimization problem is an instance of Problem 2 with $\tilde{\mathbf{x}}_t$ as the state of the linear system and $(\mathbf{u}_t^1, \tilde{\mathbf{u}}_t^2)$ as the control action. By Lemma 3, we obtain

$$\begin{bmatrix} u_t^1 \\ \tilde{u}_t^2 \end{bmatrix} = \tilde{H}_t\, \mathbb{E}(\tilde{\mathbf{x}}_t \,|\, \hat{\mathbf{i}}_t = \hat{i}_t) = \tilde{H}_t\, \mathbb{E}\left( \begin{bmatrix} \mathbf{x}_t \\ \mathbf{z}_t \end{bmatrix} \middle|\, \hat{\mathbf{i}}_t = \hat{i}_t \right) \tag{12}$$

for some matrix $\tilde{H}_t$. The first component of the expectation is simply $\hat{z}_t$, and the second component is also $\hat{z}_t$ by Lemma 4. Therefore, $u_t^1$ and $\tilde{u}_t^2$ are linear functions of $\hat{z}_t$, Player 1's estimate. ■

Theorem 8 shows that for Problem 1, a sufficient statistic is the set of conditional means $(\hat{z}_t, z_t)$. Note that for a given realization of $\mathbf{i}_t$, Player 2's estimate $z_t$ does not depend on players' strategies. However, for a given realization of $\hat{\mathbf{i}}_t$, Player 1's estimate $\hat{z}_t$ depends on the choice of matrices $H_{0:t-1}^2$.

### 4.3 Extension to the POMDP case

In the centralized LQG problem (Problem 2), the fact that the optimal control action is a linear function of the conditional mean of the state is a consequence of the LQG assumptions. If state update and measurement equations are nonlinear with non-Gaussian noise, the centralized problem becomes a partially observable Markov decision process (POMDP). For POMDPs, optimal actions are functions of the *conditional probability density of the state* and not just the conditional mean. In other words:

$$\begin{aligned} &\text{LQG:} && u_t = K_t z_t && \text{where} && z_t = \mathbb{E}(\mathbf{x}_t \,|\, \mathbf{i}_t = i_t) \\ &\text{POMDP:} && u_t = \phi_t(\pi_t) && \text{where} && \pi_t = \mathbb{P}(\mathbf{x}_t \,|\, \mathbf{i}_t = i_t) \end{aligned}$$

where $\phi_t$ is a (possibly) nonlinear function, and $\pi_t$ is called the *belief state*. Note that $\pi_t$ is a probability density function while $z_t$ is simply a real vector.

For the two-player problem (Problem 1), the simple form of the structural results in Theorems 6 and 8 is a consequence of the triangular information structure and the LQG assumptions. We now investigate the corresponding two-player structural result for POMDPs.

A straightforward extension of Lemma 5 shows that for any choice of Player 1's strategy, Player 2's optimization problem is a POMDP and the optimal strategy has the form $u_t^2 = \gamma_t(\pi_t, \hat{i}_t)$ where $\gamma_t$ is a (possibly) nonlinear function. Because Player 2's optimal policy is no longer linear, the coordinator's problem of Section 4.2 is more complicated. The coordinator is now required to select a control action for Player 1 and a *function that maps Player 2's belief to Player 2's action.* The coordinator's problem can be viewed as a POMDP with $(\mathbf{x}_t, \boldsymbol{\pi}_t)$ as the state. Therefore, the associated structural result involves a belief on the pair $(\mathbf{x}_t, \boldsymbol{\pi}_t)$. Not only is the coordinator required to keep a belief on the state, it must also keep a *belief on Player 2's belief on the state.*

As shown in Section 4, the structures of the optimal strategies for the centralized and two-player problems in the LQG case are of comparable complexity. However, this is not the case for the nonlinear, non-Gaussian versions of these problems. The centralized case requires maintaining a belief on the system state, while the two-player case requires maintaining a belief on a belief. This is substantially more complicated object.

## 5 Explicit solution

In this section, we use the structural results of Section 4 to derive an explicit and efficiently computable state-space realization for the optimal controller for Problem 1.

To ensure a unique optimal controller with a recursively computable structure, we make some additional mild assumptions, which we list below.

**Main assumptions.** We assume the following.

$$\begin{bmatrix} W & U^\mathsf{T} \\ U & V \end{bmatrix} \overset{t}{\geq} 0, \qquad \Sigma_{\text{init}} \geq 0, \qquad \text{and} \qquad V \overset{t}{>} 0 \tag{13}$$

$$\begin{bmatrix} Q & S \\ S^\mathsf{T} & R \end{bmatrix} \overset{t}{\geq} 0, \qquad P_{\text{final}} \geq 0, \qquad \text{and} \qquad R \overset{t}{>} 0 \tag{14}$$

The assumptions that $V_t > 0$ and $R_t > 0$ are made for simplicity and can generally be relaxed. For example, it is only required that $C_t \Sigma_t C_t^\mathsf{T} + V_t > 0$, so as long as this holds, we can have $V_t \geq 0$.

The well-known solution to the centralized LQG problem (Problem 2) in given in the following lemma.

**Lemma 9.** *Consider Problem 2 and suppose the main assumptions* (13)–(14) *hold. The optimal policy is*

$$\begin{aligned} z_0 &= 0 \\ z_+ &\overset{t}{=} Az + Bu - L(y - Cz) \\ u &\overset{t}{=} Kz \end{aligned} \tag{15}$$

*where $L_{0:T-1}$ satisfies the forward recursion*

$$\begin{aligned} \Sigma_0 &= \Sigma_{\text{init}} \\ \Sigma_+ &\overset{t}{=} A\Sigma A^\mathsf{T} + L(C\Sigma A^\mathsf{T} + U) + W \\ L &\overset{t}{=} -(A\Sigma C^\mathsf{T} + U^\mathsf{T})(C\Sigma C^\mathsf{T} + V)^{-1} \end{aligned} \tag{16}$$

*and $K_{0:T-1}$ satisfies the backward recursion*

$$\begin{aligned} P_T &= P_{\text{final}} \\ P &\overset{t}{=} A^\mathsf{T} P_+ A + (A^\mathsf{T} P_+ B + S)K + Q \\ K &\overset{t}{=} -(B^\mathsf{T} P_+ B + R)^{-1}(B^\mathsf{T} P_+ A + S^\mathsf{T}) \end{aligned} \tag{17}$$

*For every $t$, the belief state has the distribution*

$$\mathbb{P}(\mathbf{x}_t \,|\, \mathbf{i}_t = i_t) = \mathcal{N}(z_t, \Sigma_t) \tag{18}$$

*and the optimal average cost is given by*

$$\begin{aligned} \mathcal{J}_0 &= \mathbf{tr}(P_0 \Sigma_{\text{init}}) \\ &\quad + \sum_t \Big( \mathbf{tr}(P_+ W) + \mathbf{tr}\big[\Sigma K^\mathsf{T}(B^\mathsf{T} P_+ B + R)K\big] \Big) \end{aligned} \tag{19}$$

**Proof.** See for example [4] or [12]. ■

Note that Lemma 9 holds in great generality. All system, cost, and covariance matrices may vary with time. The above formulae hold even in the case where the dimensions of the matrices are different at every timestep.

The main result of this section is a state-space solution to Problem 1, the two-player problem. The result, given below in Theorem 10, is similar in structure and generality to Lemma 9, but with one important difference. In Lemma 9, the gains $K_{0:T-1}$ and $L_{0:T-1}$ can be computed separately using different recursions. Thus, Lemma 9 provides both a solution and a recipe for its construction. In contrast, the recursions for $\hat{K}_{0:T-1}$ and $\hat{L}_{0:T-1}$ found in Theorem 10 are coupled. Therefore, Theorem 10 provides an implicitly defined solution, but no obvious construction method. In Section 6, we make our solution explicit by showing how the various gains described in Theorem 10 can be efficiently computed.

**Theorem 10.** *Consider Problem 1 and suppose the main assumptions* (13)–(14) *hold. The optimal policy is*

$$\begin{aligned} \hat{z}_0 &= 0 \\ \hat{z}_+ &\overset{t}{=} A\hat{z} + B\hat{u} - \hat{L}(y - C\hat{z}) \\ \hat{u} &\overset{t}{=} K\hat{z} \end{aligned} \tag{20}$$

$$\begin{aligned} z_0 &= 0 \\ z_+ &\overset{t}{=} Az + Bu - L(y - Cz) \\ u &\overset{t}{=} K\hat{z} + \hat{K}(z - \hat{z}) \end{aligned} \tag{21}$$

*where $L_{0:T-1}$ and $K_{0:T-1}$ satisfy* (16)–(17)*. If we define $\hat{A} :\overset{t}{=} A + B\hat{K} + \hat{L}C$, then $\hat{L}_{0:T-1}$ satisfies the recursion*

$$\begin{aligned} \hat{\Sigma}_0 &= \Sigma_{\text{init}} \\ \hat{\Sigma}_+ &\overset{t}{=} \Sigma_+ + \hat{A}(\hat{\Sigma} - \Sigma)\hat{A}^\mathsf{T} \\ &\qquad + (\hat{L} - L)(C\Sigma C^\mathsf{T} + V)(\hat{L} - L)^\mathsf{T} \\ \hat{L} &\overset{t}{=} -\big(A\hat{\Sigma}C^\mathsf{T} + U^\mathsf{T} + B\hat{K}(\hat{\Sigma} - \Sigma)C^\mathsf{T}\big) \\ &\qquad \times E_1(C_{11}\hat{\Sigma}^{11}C_{11}^\mathsf{T} + V_{11})^{-1}E_1^\mathsf{T} \end{aligned} \tag{22}$$

*and $\hat{K}_{0:T-1}$ satisfies the recursion*

$$\begin{aligned} \hat{P}_T &= P_{\text{final}} \\ \hat{P} &\overset{t}{=} P + \hat{A}^\mathsf{T}(\hat{P}_+ - P_+)\hat{A} \\ &\qquad + (\hat{K} - K)^\mathsf{T}(B^\mathsf{T} P_+ B + R)(\hat{K} - K) \\ \hat{K} &\overset{t}{=} -E_2(B_{22}^\mathsf{T}\hat{P}_+^{22}B_{22} + R_{22})^{-1}E_2^\mathsf{T} \\ &\qquad \times \big(B^\mathsf{T}\hat{P}_+ A + S^\mathsf{T} + B^\mathsf{T}(\hat{P}_+ - P_+)\hat{L}C\big) \end{aligned} \tag{23}$$

*where $E_i$ are matrices defined in Section 2. For every $t$, the belief states have the distributions*

$$\begin{aligned} \mathbb{P}(\mathbf{x}_t \,|\, \hat{\mathbf{i}}_t = \hat{i}_t) &= \mathcal{N}(\hat{z}_t, \hat{\Sigma}_t) \\ \mathbb{P}(\mathbf{x}_t \,|\, \mathbf{i}_t = i_t) &= \mathcal{N}(z_t, \Sigma_t) \end{aligned} \tag{24}$$

*and the optimal average cost is given by*

$$\begin{aligned} \hat{\mathcal{J}}_0 &= \mathbf{tr}(P_0 \Sigma_{\text{init}}) \\ &\quad + \sum_t \Big( \mathbf{tr}(P_+ W) + \mathbf{tr}\big[\Sigma K^\mathsf{T}(B^\mathsf{T} P_+ B + R)K\big] \\ &\quad + \mathbf{tr}\big[(\hat{\Sigma} - \Sigma)(\hat{K} - K)^\mathsf{T}(B^\mathsf{T} P_+ B + R)(\hat{K} - K)\big] \Big) \end{aligned} \tag{25}$$

**Proof.** Note that $\hat{L}E_2 = 0$ and $E_1^\mathsf{T}\hat{K} = 0$. That is, the second block-column of $\hat{L}$ and the first block-row of $\hat{K}$ are zero. The required triangular structure is therefore satisfied because $\hat{z}_+$ only depends on $y^1$ in (20) and $u^1$ only depends on $\hat{z}$ in (21).

Since Player 2 observes all measurements and control actions, its estimate of the state is the standard Kalman filter. Therefore, $z_t$ evolves according to (21) where $L_{0:T-1}$ satisfies (16).

The result of Theorem 8 implies that the optimal control vector can be expressed as $u \overset{t}{=} \tilde{u} + \hat{K}z$ where $\tilde{u}_t$ is chosen by the coordinator and $\hat{K}_t$ is a matrix whose first block-row is zero, so $E_1^\mathsf{T}\hat{K}_t = 0$. Our first step will be to fix $\hat{K}_t$ for all $t$ and to optimize for the coordinator's strategy. We begin by computing $\hat{z}_t = \mathbb{E}(\mathbf{x}_t \,|\, \hat{\mathbf{i}}_t = \hat{i}_t)$. To this end, we construct an equivalent centralized problem and appeal once again to Lemma 9. By Lemma 4, we have $\hat{z}_t = \mathbb{E}(\mathbf{z}_t \,|\, \hat{\mathbf{i}}_t = \hat{i}_t)$ so we may estimate $\mathbf{x}_t$ by estimating $\mathbf{z}_t$ instead. Substituting the definitions for $\mathbf{u}_t$ and $\mathbf{y}_t$ into (1) and (21), we obtain the state equations

$$\begin{bmatrix}\mathbf{z}_+\\ \mathbf{e}_+\end{bmatrix} \overset{t}{=} \begin{bmatrix}A + B\hat{K} & -LC\\ 0 & A + LC\end{bmatrix}\begin{bmatrix}\mathbf{z}\\ \mathbf{e}\end{bmatrix} + \begin{bmatrix}B\\ 0\end{bmatrix}\tilde{\mathbf{u}} + \begin{bmatrix}-L\mathbf{v}\\ \mathbf{w} + L\mathbf{v}\end{bmatrix}$$
$$\mathbf{y}^1 \overset{t}{=} \begin{bmatrix}E_1^\mathsf{T}C & E_1^\mathsf{T}C\end{bmatrix}\begin{bmatrix}\mathbf{z}\\ \mathbf{e}\end{bmatrix} + \mathbf{v}^1$$

where we have defined the error signal $\mathbf{e} :\overset{t}{=} \mathbf{x} - \mathbf{z}$. Apply Lemma 9 to compute the $\Sigma$-recursion. A straightforward induction argument shows that the covariance and gain matrices that satisfy (16) at time $t$ are given by

$$\begin{bmatrix}\hat{\Sigma}_t - \Sigma_t & 0\\ 0 & \Sigma_t\end{bmatrix} \quad \text{and} \quad \begin{bmatrix}\hat{L}_t E_1\\ 0\end{bmatrix}$$

where $\hat{\Sigma}_t$ and $\hat{L}_t$ satisfy (22). Computing the estimation equations (15), we find that the estimate of $\mathbf{e}_t$ is 0, and

$$\hat{\mathbf{z}}_+ \overset{t}{=} \hat{A}\hat{\mathbf{z}} + B\tilde{\mathbf{u}} - \hat{L}\mathbf{y} \tag{26}$$

where $\hat{A}$ is defined in the theorem statement. State and input split into conditionally independent parts

$$\begin{bmatrix}\mathbf{x}\\ \mathbf{u}\end{bmatrix} \overset{t}{=} \begin{bmatrix}\mathbf{z}\\ \tilde{\mathbf{u}} + \hat{K}\mathbf{z}\end{bmatrix} + \begin{bmatrix}\mathbf{e}\\ 0\end{bmatrix}$$

so the only relevant part of the cost (6) is

$$\mathbb{E}\left(\sum_t \begin{bmatrix}\mathbf{z}\\ \tilde{\mathbf{u}} + \hat{K}\mathbf{z}\end{bmatrix}^\mathsf{T}\begin{bmatrix}Q & S\\ S^\mathsf{T} & R\end{bmatrix}\begin{bmatrix}\mathbf{z}\\ \tilde{\mathbf{u}} + \hat{K}\mathbf{z}\end{bmatrix} + \mathbf{z}_T^\mathsf{T}P_{\text{final}}\mathbf{z}_T\right)$$

Applying the $P$-recursion (17) from Lemma 9 to solve for $\tilde{u}$, we find after some algebra that

$$\tilde{u} \overset{t}{=} (K - \hat{K})\hat{z} \tag{27}$$

where $K_{0:T-1}$ is the centralized gain given by (17). Substituting (27) into (26), we recover the desired form for Player 1's estimator (20).

We have shown thus far that for any fixed $\hat{K}_t$, the optimal Player 1 estimator has the form specified in (20), where (22), and (24) hold and the coordinator's action is given by (27). Note that since the first block-row of $\hat{K}_t$ is zero, it follows from (27) that $u^1 \overset{t}{=} E_1^\mathsf{T}K\hat{z}$ where $K_{0:T-1}$ is the centralized gain. Therefore, the optimal strategies for the two players have the following structure:

$$\begin{aligned}\hat{z}_0 &= 0\\ \hat{z}_+ &\overset{t}{=} (A + BK)\hat{z} - \hat{L}(y - C\hat{z})\\ u^1 &\overset{t}{=} E_1^\mathsf{T}K\hat{z}\end{aligned} \tag{28}$$

$$\begin{aligned}z_0 &= 0\\ z_+ &\overset{t}{=} Az + Bu - L(y - Cz)\\ u^2 &\overset{t}{=} E_2^\mathsf{T}K\hat{z} + E_2^\mathsf{T}\hat{K}(z - \hat{z})\end{aligned} \tag{29}$$

for some $\hat{L}_{0:T-1}$ and $\hat{K}_{0:T-1}$. Because our problem is partially nested, a strategy of the form (28)–(29) is globally optimal if and only if it is person by person optimal.

For a given $\hat{K}_{0:T-1}$, Player 1's strategy of the form (28) will coincide with the coordinator's best response to $\hat{K}_{0:T-1}$ if $\hat{L}_{0:T-1}$ satisfies (22). Therefore, a strategy of the form (28) with $\hat{L}_{0:T-1}$ satisfying (22) must be Player 1's best response to $\hat{K}_{0:T-1}$. For a given choice of $\hat{L}_{0:T-1}$, we now seek the best response of Player 2 of the form in (29). The combined control vector of the two players can be written as $u \overset{t}{=} K\hat{z} + E_2\bar{u}$, where we allow $\bar{u}$ to be a function of player 2's entire information. Gathering the state equations (1) and the estimator equations (20)–(21), we obtain

$$\begin{bmatrix}\mathbf{x}_+\\ \hat{\mathbf{e}}_+\end{bmatrix} \overset{t}{=} \begin{bmatrix}A{+}BK & -BK\\ 0 & A{+}\hat{L}C\end{bmatrix}\begin{bmatrix}\mathbf{x}\\ \hat{\mathbf{e}}\end{bmatrix} + \begin{bmatrix}BE_2\\ BE_2\end{bmatrix}\bar{\mathbf{u}} + \begin{bmatrix}\mathbf{w}\\ \mathbf{w} + \hat{L}\mathbf{v}\end{bmatrix}$$
$$\mathbf{y} \overset{t}{=} \begin{bmatrix}C & 0\end{bmatrix}\begin{bmatrix}\mathbf{x}\\ \hat{\mathbf{e}}\end{bmatrix} + \mathbf{v}$$

where we have defined the error signal $\hat{\mathbf{e}} :\overset{t}{=} \mathbf{x} - \hat{\mathbf{z}}$. The cost (6) is given by

$$\mathbb{E}\left(\sum_t \begin{bmatrix}\mathbf{x}\\ K\hat{\mathbf{z}} + E_2\bar{\mathbf{u}}\end{bmatrix}^\mathsf{T}\begin{bmatrix}Q & S\\ S^\mathsf{T} & R\end{bmatrix}\begin{bmatrix}\mathbf{x}\\ K\hat{\mathbf{z}} + E_2\bar{\mathbf{u}}\end{bmatrix} + \mathbf{x}_T^\mathsf{T}P_{\text{final}}\mathbf{x}_T\right)$$

where the correct coordinates can be obtained by by substituting $\hat{\mathbf{z}} \overset{t}{=} \mathbf{x} - \hat{\mathbf{e}}$. Now apply Lemma 9 to compute the $P$-recursion. A straightforward induction argument shows that the cost-to-go and gain matrices that satisfy (17) at time $t$ are given by

$$\begin{bmatrix}P_t & 0\\ 0 & \hat{P}_t - P_t\end{bmatrix} \quad \text{and} \quad \begin{bmatrix}0 & E_2^\mathsf{T}\hat{K}_t\end{bmatrix}$$

where $\hat{P}_t$ and $\hat{K}_t$ satisfy (23). It follows from Lemma 9 that the optimal input is

$$\bar{u}_t = \begin{bmatrix}0 & E_2^\mathsf{T}\hat{K}_t\end{bmatrix}\begin{bmatrix}\mathbb{E}(\mathbf{x}_t \,|\, \mathbf{i}_t = i_t)\\ \mathbb{E}(\hat{\mathbf{e}}_t \,|\, \mathbf{i}_t = i_t)\end{bmatrix} = E_2^\mathsf{T}\hat{K}_t(z_t - \hat{z}_t)$$

Despite allowing $\bar{u}_t$ to depend on the full measurement history $i_t$, we find that it only depends on $(z_t - \hat{z}_t)$, so that Player 2's best response is of the form (29).

Thus, $\hat{K}_{0:T-1}, \hat{L}_{0:T-1}$ satisfying (22) and (23) constitute a person-by-person optimal, and consequently, a globally optimal solution of our problem. ∎

**Corollary 11** (Nonzero initial state)**.** *If the initial state for Problems 1 and 2 is changed to* $\mathbf{x}_0 = \mathcal{N}(\mu_{\text{init}}, \Sigma_{\text{init}})$*, then Lemma 9 and Theorem 10 change as follows.*

*(i) Estimators initialized at* $z_0 = \hat{z}_0 = \mu_{\text{init}}$*.*

*(ii) Costs* $\mathcal{J}_0$ *and* $\hat{\mathcal{J}}_0$ *each increased by* $\mu_{\text{init}}^\mathsf{T} P_0 \mu_{\text{init}}$*.*

**Proof.** Add a pre-initial timestep $\mathbf{x}_{-1} = \mathcal{N}(0, I)$ with trivial dynamics $A_{-1} = C_{-1} = I$, $B_{-1} = 0$ and apply the zero-initial-mean result to the augmented system. ∎

## 6 Efficient computation

Theorem 10 provides a state-space realization for the two-player problem similar to Lemma 9, with an important difference. In Lemma 9, the recursions (16) and (17) can be solved independently by propagating time forward or backward respectively. However, the recursions (22)–(23) are coupled in an intricate way. Both $\hat{P}$ and $\hat{\Sigma}$ recursions contain $\hat{A}$, which depends on $\hat{K}$ and $\hat{L}$. Furthermore, the equations for $\hat{K}$ contains $\hat{L}$ and vice-versa.

Despite being nonlinear difference equations coupled across all timesteps, the recursions (22)–(23) can be solved efficiently. In the following theorem, we show that the equations for $\hat{\Sigma}, \hat{P}, \hat{L}, \hat{K}$ can be reduced to a *linear* two-point boundary-value problem and thereby solved as efficiently as (16)–(17).

**Theorem 12.** *In the solution to Problem 1 given by* (20)*–*(21)*, the gains* $\hat{L}_{0:T-1}$*,* $\hat{K}_{0:T-1}$ *are of the form*

$$\hat{L} \overset{t}{=} \begin{bmatrix} M & 0 \\ \hat{L}^{21} & 0 \end{bmatrix} \qquad \text{and} \qquad \hat{K} \overset{t}{=} \begin{bmatrix} 0 & 0 \\ \hat{K}^{21} & J \end{bmatrix}$$

*where* $M_{0:T-1}$ *satisfies the forward recursion*

$$\begin{aligned} \Gamma_0 &= \Sigma_{\text{init}}^{11} \\ \Gamma_+ &\overset{t}{=} A_{11}\Gamma A_{11}^\mathsf{T} + M(C_{11}\Gamma A_{11}^\mathsf{T} + U_{11}) + W_{11} \\ M &\overset{t}{=} -(A_{11}\Gamma C_{11}^\mathsf{T} + U_{11}^\mathsf{T})(C_{11}\Gamma C_{11}^\mathsf{T} + V_{11})^{-1} \end{aligned} \tag{30}$$

*and* $J_{0:T-1}$ *satisfies the backward recursion*

$$\begin{aligned} F_T &= P_{\text{final}}^{22} \\ F &\overset{t}{=} A_{22}^\mathsf{T} F_+ A_{22} + (A_{22}^\mathsf{T} F_+ B_{22} + S_{22})J + Q_{22} \\ J &\overset{t}{=} -(B_{22}^\mathsf{T} F_+ B_{22} + R_{22})^{-1}(B_{22}^\mathsf{T} F_+ A_{22} + S_{22}^\mathsf{T}) \end{aligned} \tag{31}$$

*Finally,* $\hat{\Sigma}_{0:T}^{21}$*,* $\hat{L}_{0:T}^{21}$*,* $\hat{P}_{0:T-1}^{21}$*,* $\hat{K}_{0:T-1}^{21}$ *satisfy the coupled forward and backward recursions* (34)*–*(35) *together with the definitions* $A_M :\overset{t}{=} A_{11} + MC_{11}$ *and* $A_J :\overset{t}{=} A_{22} + B_{22}J$*.*

**Proof.** This result follows from Theorem 10 and some straightforward algebra, so we omit the details. The recursions (30) and (31) are obtained by simplifying the 11 block of (22) and the 22 block of (23), respectively. Finally, the recursions (34) and (35) are obtained by simplifying the 21 blocks of (22) and (23) respectively. ∎

Theorem 12 reduces the coupled recursions found in Theorem 10 to a two-point linear boundary value problem. From a computational standpoint, computing the matrices $L_t$, $M_t$, $K_t$, $J_t$ using (16)–(17) and (30)–(31) requires recursing through the entire time horizon. This requires $\mathcal{O}(T)$ operations.

It turns out that $\hat{\Sigma}_t^{21}$, $\hat{P}_t^{21}$, $\hat{L}_t^{21}$, $\hat{K}_t^{21}$ (and consequently $\hat{L}_t$ and $\hat{K}_t$) can also be computed in $\mathcal{O}(T)$. To see why, note that (34)–(35) are of the form

$$\begin{aligned} \hat{\Sigma}_0^{21} &= \Sigma_{\text{init}}^{21} & \hat{P}_T^{21} &= P_{\text{final}}^{21} \\ \hat{\Sigma}_+^{21} &\overset{t}{=} g_1(\hat{\Sigma}^{21}, \hat{K}^{21}) & \hat{P}^{21} &\overset{t}{=} g_2(\hat{P}_+^{21}, \hat{L}^{21}) \\ \hat{L}^{21} &\overset{t}{=} g_3(\hat{\Sigma}^{21}, \hat{K}^{21}) & \hat{K}^{21} &\overset{t}{=} g_4(\hat{P}_+^{21}, \hat{L}^{21}) \end{aligned} \tag{32}$$

where $g_1, \dots, g_4$ are affine functions. Eliminating $\hat{L}_t^{21}$ and $\hat{K}_t^{21}$ from (32) using the last row of equations,

$$\begin{aligned} \hat{\Sigma}_0^{21} &= \Sigma_{\text{init}}^{21} & \hat{P}_T^{21} &= P_{\text{final}}^{21} \\ \hat{\Sigma}_+^{21} &\overset{t}{=} h_1(\hat{\Sigma}^{21}, \hat{P}_+^{21}) & \hat{P}^{21} &\overset{t}{=} h_2(\hat{\Sigma}^{21}, \hat{P}_+^{21}) \end{aligned} \tag{33}$$

where $h_1$ and $h_2$ are affine functions. Now let

$$\eta :\overset{t}{=} \begin{bmatrix} \mathbf{vec}\, \hat{P}^{21} \\ \mathbf{vec}\, \hat{\Sigma}_+^{21} \end{bmatrix}$$

where $\mathbf{vec}\, X$ is the vector obtained by stacking the columns of $X$. Then (33) is a block-tridiagonal system of the form

$$\begin{bmatrix} I & H_1 & & \\ G_1 & I & \ddots & \\ & \ddots & \ddots & H_{T-1} \\ & & G_{T-1} & I \end{bmatrix} \begin{bmatrix} \eta_0 \\ \eta_1 \\ \vdots \\ \eta_{T-1} \end{bmatrix} = \begin{bmatrix} c_0 \\ c_1 \\ \vdots \\ c_{T-1} \end{bmatrix}$$

for some constant matrices $G_{1:T-1}$ and $H_{1:T-1}$ and a constant vector $c_{0:T-1}$. Equations of this form can be solved in $\mathcal{O}(T)$ using for example block tridiagonal LU factorization [2, § 4.5.1].

Therefore, the optimal controller for the two-player problem presented in Theorem 10 can be computed with comparable effort to its centralized counterpart in Lemma 9.

Note that the infinite-horizon two-player problem can be solved by making suitable assumptions on the system parameters and taking limits in Theorem 12. The recursions (16)–(17) and (30)–(31) become algebraic Riccati equations, and the coupled recursions (32) become a small set of linear equations.

## 7 Concluding remarks

In this paper, we used a coordinator-based approach to derive a new structural result for a two-player partially nested LQG problem. Our results generalize those from

$$
\begin{aligned}
\hat{\Sigma}_0^{21} &= \Sigma_{\text{init}}^{21} \\
\hat{\Sigma}_+^{21} &\overset{t}{=} A_J\hat{\Sigma}^{21}A_M^\mathsf{T} + B_{22}\hat{K}^{21}(\Gamma - \Sigma^{11})A_M^\mathsf{T} + \big(A_{21}\Gamma - B_{22}J\Sigma^{21}\big)A_M^\mathsf{T} + U_{12}^\mathsf{T}M^\mathsf{T} + W_{21} && (34) \\
\hat{L}^{21} &\overset{t}{=} -\big(A_J\hat{\Sigma}^{21}C_{11}^\mathsf{T} + B_{22}\hat{K}^{21}(\Gamma - \Sigma^{11})C_{11}^\mathsf{T} + (A_{21}\Gamma - B_{22}J\Sigma^{21})C_{11}^\mathsf{T} + U_{12}^\mathsf{T}\big)(C_{11}\Gamma C_{11}^\mathsf{T} + V_{11})^{-1} \\
\hat{P}_T^{21} &= P_{\text{final}}^{21} \\
\hat{P}^{21} &\overset{t}{=} A_J^\mathsf{T}\hat{P}_+^{21}A_M + A_J^\mathsf{T}(F_+ - P_+^{22})\hat{L}^{21}C_{11} + A_J^\mathsf{T}\big(F_+A_{21} - P_+^{21}MC_{11}\big) + J^\mathsf{T}S_{12}^\mathsf{T} + Q_{21} && (35) \\
\hat{K}^{21} &\overset{t}{=} -(B_{22}^\mathsf{T}F_+B_{22} + R_{22})^{-1}\big(B_{22}^\mathsf{T}\hat{P}_+^{21}A_M + B_{22}^\mathsf{T}(F_+ - P_+^{22})\hat{L}^{21}C_{11} + B_{22}^\mathsf{T}(F_+A_{21} - P_+^{21}MC_{11}) + S_{12}^\mathsf{T}\big)
\end{aligned}
$$

classical LQG theory in a very intuitive way. Rather than maintaining a single estimate of the state, two different estimates must be maintained, to account for the two different sets of information available. As in the centralized case, finding the optimal two-player controller requires solving forward and backward recursions for estimation and control respectively. The key difference is that the recursions for the two-player case are coupled and must be solved together. We show that these recursions can be solved as efficiently as in the centralized case, with complexity proportional to the length of the time horizon.

An effort was made to express our results in a form that showcases the duality between estimation and control. This duality is apparent in (22)–(23), (30)–(35), and in the proof of Theorem 10. The extent of the duality observed in the solution is perhaps unexpected. Indeed, one might expect a greater burden on the second player since it receives more measurements and must correct for the estimation errors inevitably made by the first player. However, from a different perspective, one might expect a greater burden on the first player since it has more control authority and must act to influence states of the system that the second player cannot control. The second player's lack of control authority mirrors the first player's lack of estimation ability.

## References

[1] V. D. Blondel and J. N. Tsitsiklis. A survey of computational complexity results in systems and control. 36(9):1249–1274, 2000.

[2] G. H. Golub and C. F. Van Loan. *Matrix Computations.* Johns Hopkins University Press; fourth edition, 2012.

[3] Y.-C. Ho and K.-C. Chu. Team decision theory and information structures in optimal control problems—Part I. *IEEE Transactions on Automatic Control*, 17(1):15–22, 1972.

[4] P. Kumar and P. Varaiya. *Stochastic Systems: Estimation, Identification and Adaptive Control.* Prentice-Hall, 1986.

[5] A. Lamperski and J. C. Doyle. Dynamic programming solutions for decentralized state-feedback LQG problems with output feedback. In *American Control Conference*, pages 6322–6327, 2012.

[6] A. Lamperski and L. Lessard. Optimal state-feedback control under sparsity and delay constraints. In *IFAC Workshop on Distributed Estimation and Control in Networked Systems*, pages 204–209, 2012.

[7] L. Lessard. Decentralized LQG control of systems with a broadcast architecture. In *IEEE Conference on Decision and Control*, pages 6241–6246, 2012.

[8] L. Lessard and S. Lall. A state-space solution to the two-player decentralized optimal control problem. In *Allerton Conference on Communication, Control, and Computing*, pages 1559–1564, 2011.

[9] L. Lessard and S. Lall. Optimal controller synthesis for the decentralized two-player problem with output feedback. In *American Control Conference*, pages 6314–6321, 2012.

[10] A. Nayyar. *Sequential decision-making in decentralized systems.* PhD thesis, University of Michigan, 2011.

[11] A. Nayyar, A. Mahajan, and D. Teneketzis. Decentralized stochastic control with partial history sharing: A common information approach. *IEEE Transactions on Automatic Control*, 58(7):1644–1658, 2013.

[12] K. J. Åström. *Introduction to Stochastic Control Theory.* Dover Publications, 2006.

[13] N. Sandell, Jr. and M. Athans. Solution of some nonclassical LQG stochastic decision problems. *IEEE Transactions on Automatic Control*, 19(2):108–116, 1974.

[14] P. Shah and P. A. Parrilo. $\mathcal{H}_2$-optimal decentralized control over posets: A state space solution for state-feedback. In *IEEE Conference on Decision and Control*, pages 6722–6727, 2010.

[15] J. Swigart and S. Lall. An explicit dynamic programming solution for a decentralized two-player optimal linear-quadratic regulator. In *International Symposium on Mathematical Theory of Networks and Systems*, pages 1443–1447, 2010.

[16] H. S. Witsenhausen. A counterexample in stochastic optimum control. *SIAM Journal on Control*, 6(1):131–147, 1968.