An Explicit Dynamic Programming Solution for a Decentralized Two-Player Optimal Linear-Quadratic Regulator

John Swigart[1] Sanjay Lall[2]

Abstract

We develop optimal controller synthesis algorithms for decentralized control problems, in which individual subsystems are connected over a network. We consider a simple information structure, consisting of two interconnected linear systems, and construct the optimal controller subject to a decentralization constraint via a novel dynamic programming method. We provide explicit state-space formulae for the optimal controller, and show that each player has to do more than simply estimate the states that they cannot observe. In other words, the simplest separation principle does not hold for this decentralized control problem.

1 Introduction

Recently, attention has been focused on the problem of decentralized control. It has been known for some time that decentralized problems pose a challenge unlike their centralized counterparts. While linear, centralized systems provide tractable linear solutions, it has been shown that even the simplest linear, decentralized problems can prove intractable [2], or have nonlinear optimal solutions [10]. Nevertheless, the use of distributed systems is growing rapidly. Examples include the internet, the power grid, or satellites in formation. Systems of this nature contain so much information that centralization becomes infeasible.

Consequently, recent research has been aimed at trying to characterize those decentralized systems for which optimization is still tractable [3, 4, 1, 5]. Among the simplest decentralized problems to consider are systems connected over a graph. Conditions for the tractability of these problems were found in [8]. While these results reduce the nonlinear optimization problem to a convex one, these convex problems are, in general, still infinite-dimensional. Efficient methods are still needed to find the optimal policies.

In this paper, we consider the simplest of these decentralized systems, consisting of only two players. Player 1 may affect player 2's dynamics and shares his state information with player 2 but not vice versa. This system is known to have an optimal linear solution [3, 5, 9]. In fact, the optimal solution was found in [7], via a spectral factorization method.

In classical treatments of optimal control, there exist three major approaches: spectral factorization, dynamic programming, and semi-definite programming. The spectral factorization version of this problem was tackled in [7]. A semi-definite programming approach was considered in [6] for this problem. In this paper, we provide the dynamic programming version of the result. As in [7], we are able to provide an explicit state-space solution for the optimal control policies and provide intuition behind this solution.

2 Problem Formulation

We consider two interconnected systems (players), in which the dynamics of system 1 may affect the dynamics of system 2, but not vice versa. For each $t \in \{0, 1, \ldots, N\}$, system 1 has states $x_1(t) \in \mathbb{R}^{n_1}$ and inputs $u_1(t) \in \mathbb{R}^{m_1}$, and system 2 has states $x_2(t) \in \mathbb{R}^{n_1}$ and inputs $u_2(t) \in \mathbb{R}^{m_1}$. The sequence of states represent a random Markov process which evolves according to the overall dynamics:

$$\begin{bmatrix} x_1(t+1) \\ x_2(t+1) \end{bmatrix} = \begin{bmatrix} A_{11} & 0 \\ A_{21} & A_{22} \end{bmatrix} \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} + \begin{bmatrix} B_{11} & 0 \\ B_{21} & B_{22} \end{bmatrix} \begin{bmatrix} u_1(t) \\ u_2(t) \end{bmatrix} + \begin{bmatrix} w_1(t) \\ w_2(t) \end{bmatrix} \quad (1)$$

for $t = 0, \ldots, N-1$. The initial conditions $x_1(0), x_2(0)$ and exogenous noise $w_1(t), w_2(t)$ are assumed independent random variables, with Gaussian distributions

$$\begin{array}{l} x_i(0) \sim \mathcal{N}(0, \Sigma_i) \\ w_i(t) \sim \mathcal{N}(0, \Sigma_i) \end{array} \qquad i = 1, 2 \text{ and } t = 0, \ldots, N-1$$

For this system, a particular realization for the random variables $x_1(t), x_2(t)$, and $u_1(t), u_2(t)$ is denoted z_1^t, z_2^t,

[1] J. Swigart is with the Department of Aeronautics and Astronautics, Stanford University, Stanford, CA 94305, USA. jswigart@stanford.edu

[2] S. Lall is with the Department of Electrical Engineering and Department of Aeronautics and Astronautics, Stanford University, Stanford, CA 94305, USA. lall@stanford.edu

ISBN 978-963-311-370-7

and a_1^t, a_2^t, respectively. Since we have to index variables both spatially (by player) and temporally, we will use subscripts to denote the spatial index, and we label the temporal index as arguments (for random variables) and with superscripts (for realizations and functions). By dropping the subscripts, we group the variables spatially as

$$x(t) = (x_1(t), x_2(t))$$

so that $x(t) \in \mathbb{R}^n$, $u(t) \in \mathbb{R}^m$, where $n = n_1 + n_2$ and $m = m_1 + m_2$. Similarly, we label the matrices in (1) as A and B. In addition, to denote the temporal grouping of variables, we use the notation

$$x_1(0:t) = (x_1(0), \ldots, x_1(t))$$

Lastly, for each $t \in \{0, \ldots N-1\}$, it will be convenient to let $h_1(t) = (x_1(0:t), u_1(0:t-1))$ represent the history of system 1, $h_2(t) = (x_2(0:t), u_2(0:t-1))$ be the history of system 2, and $h(t) = (h_1(t), h_2(t))$. These random variables have realizations i_1^t, i_2^t, and i^t, respectively.

For each $t \in \{0, \ldots, N-1\}$, let $\mathcal{X}_1^t$ be the $(t+1)$-ary Cartesian product $\mathbb{R}^{n_1} \times \cdots \times \mathbb{R}^{n_1}$, and similarly define $\mathcal{X}_2^t$. Also, $\mathcal{U}_1^t$ is the t-ary Cartesian product $\mathbb{R}^{m_1} \times \cdots \times \mathbb{R}^{m_1}$, and similarly for $\mathcal{U}_2^t$. Moreover, we define $H_1^t = \mathcal{X}_1^t \times \mathcal{U}_1^t$, and $H_2^t = \mathcal{X}_2^t \times \mathcal{U}_2^t$. Also, let $H^t = H_1^t \times H_2^t$. In other words, $h_1(t) \in H_1^t$ and $h_2(t) \in H_2^t$.

Let L_1^t be the space of measurable functions mapping $H_1^t \mapsto \mathbb{R}^{m_1}$, and let L_2^t be the space of measurable functions mapping $H^t \mapsto \mathbb{R}^{m_2}$. Our objective is to find control policies $\gamma_1^t \in L_1^t$ and $\gamma_2^t \in L_2^t$, such that $u_1^t = \gamma_1^t(h_1(t))$ is a function of $x_1(0:t), u_1(0:t-1)$, and $u_2(t) = \gamma_2^t(h(t))$ is a function of $x(0:t), u(0:t-1)$. That is, player 1 has access only to his own states and actions, whereas player 2 can measure all states and actions. It was shown in [5] and [8] that this system has a property called *quadratic invariance*, which makes this problem amenable to convex optimization. However, finding analytic solutions greatly reduces computational complexity and provides significant insight into the optimal policies.

We note that (1) and $\gamma^{0:N-1} \in \prod_{t=0}^{N-1} L_1^t \times L_2^t$ induce probability measures μ^t on $x(0:t), u(0:t-1)$, for each $t \in \{0, \ldots, N\}$. These probability measures have a corresponding probability density function (pdf), denoted $p^{x(0:t),u(0:t-1)}(z^{0:t}, a^{0:t-1})$. We can also construct conditional probability densities, e.g. $p^{x_2(t)|x_1(0:t)=z_1^{0:t}}(z_2^t)$. We will compress this notation to $p(z^{0:t}, a^{0:t-1})$ for marginal distributions and $p(z_2^t | z_1^{0:t})$ for conditional distributions. Though this overloads our notation for p, the desired usage should be clear from the context.

The cost to be minimized by our choice of policies is the function $\mathcal{J} : \prod_{t=0}^{N-1} L_1^t \times L_2^t \to \mathbb{R}$, given by

$$\mathcal{J}(\gamma^{0:N-1}) = \mathbb{E}\left(\sum_{t=0}^{N-1} x(t)^T Q x(t) + u(t)^T R u(t) + x(N)^T Q^N x(N) \right) \quad (2)$$

where $Q, Q^N \geq 0$ and $R > 0$.

3 Dynamic Program

Since system 1 cannot measure the state of system 2, it will become necessary to compute an estimate of $x_2(t)$, given the information available to system 1. To this end, for each $t \in \{0, \ldots, N\}$, let $\hat{x}_2^t(i_1^t) = \mathbb{E}(x_2(t) \,|\, h_1(t) = i_1^t)$. These estimates can be computed recursively, as follows.

Lemma 1. *Suppose* $\gamma^{0:N-1} \in \prod_{t=0}^{N-1} L_1^t \times L_2^t$. *Then, for each* $t \in \{0, \ldots, N-1\}$,

$$\hat{x}_2^{t+1}\left(i_1^{t+1}\right) = A_{21} z_1^t + A_{22} \hat{x}_2^t(i_1^t) + B_{21} \gamma_1^t(i_1^t) + B_{22} \hat{\gamma}_2^t(i_1^t) \quad (3)$$

where

$$\hat{\gamma}_2^t(i_1^t) = \mathbb{E}\left(u_2(t) \,|\, h_1(t) = i_1^t\right) = \int a_2^t p\left(a_2^t \,|\, i_1^t\right) \, da_2^t$$

Proof. This result follows from manipulations of the conditional probabilities, using Bayes law.

$$\begin{aligned}
&\mathbb{E}(x_2(t+1) \,|\, h_1(t+1) = i_1^{t+1}) && (4)\\
&\quad = \int z_2^{t+1} p(z_2^{t+1} \,|\, i_1^{t+1}) \, dz_2^{t+1} \\
&\quad = \int z_2^{t+1} p(i_2^{t+1} \,|\, i_1^{t+1}) \, di_2^{t+1} \\
&\quad = \int \left(\int z_2^{t+1} p(z_2^{t+1} \,|\, z^t, a^t) \, dz_2^{t+1} \right) && (5)\\
&\qquad\qquad \times p(z_2^{0:t}, a_2^{0:t} \,|\, i_1^{t+1}) \, dz_2^{0:t} da_2^{0:t} \\
&\quad = \int \left(A_{21} z_1^t + A_{22} z_2^t + B_{21} a_1^t + B_{22} a_2^t \right) && (6)\\
&\qquad\qquad \times p(z_2^t, a_2^t \,|\, i_1^t) \, dz_2^t da_2^t \\
&\quad = A_{21} z_1^t + A_{22} \hat{x}_2^t(i_1^t) + B_{21} \gamma_1^t(i_1^t) + B_{22} \hat{\gamma}_2^t(i_1^t) && (7)
\end{aligned}$$

■

While (3) is certainly intuitive, it is actually a rather subtle result, which relies on the structure imposed on this system. In fact, it is this nice recursive form for the estimator which allows our dynamic programming scheme to work.

Let $q(t)$ be the quadratic form

$$q(t) = \begin{bmatrix} x(t) \\ u(t) \end{bmatrix}^T \begin{bmatrix} Q & 0 \\ 0 & R \end{bmatrix} \begin{bmatrix} x(t) \\ u(t) \end{bmatrix}$$

Before we state the result, consider the 2 time step case of the centralized problem

$$\min_{\gamma^0, \gamma^1} \mathbb{E}(q(0) + q(1) + q(2)) = \min_{\gamma^0, \gamma^1} \mathbb{E}(q(0) + q(1) + V^2(i^2))$$

where we've let $V^2(i^2) = q(2)$. We can equivalently write this problem as

$$\min_{\gamma^0, \gamma^1} f(\gamma^0) + g(\gamma^0, \gamma^1)$$

where $f(\gamma^0) = \mathbb{E}(q(0))$ and $g(\gamma^0, \gamma^1) = \mathbb{E}(q(1) + V^2(i^2))$. Since γ^1 receives the same measurements (and more) as γ^0, then for any γ^0, the optimal γ^1 is the one which minimizes g. If we plug back in the optimal γ^{1*}, then we can minimize $f(\gamma^0)+g(\gamma^0, \gamma^{1*})$ over γ^0. Translating this to our problem,

$$\gamma^{1*} = \arg\min_{\gamma^1} \mathbb{E}(q(1) + V^2(i^2)) \tag{8}$$

Plugging back in, we will find that $g(\gamma^0, \gamma^{1*}) = \mathbb{E}(V^1(i^1))$ so we now have

$$\min_{\gamma^0} = \mathbb{E}(q(0) + V^1(i^1))$$

The extension to more time steps is straightforward.

This is a standard result of dynamic programming and provides the framework for our dynamic programming solution. While the value functions essentially provide a solution to the problem, the difficulty lies in choosing a suitable form for the value functions, such that the optimization problem remains tractable as we progress recursively. The result is stated in the following theorem.

Theorem 2. *Suppose* $\mathcal{J} : \prod_{t=0}^{N-1} L_1^t \times L_2^t \to \mathbb{R}$ *is given by* (2) *and define* $V^0, \ldots, V^N$ *as*

$$V^t(i^t) = \begin{bmatrix} z^t \\ z_2^t - \hat{x}_2^t \end{bmatrix}^T \begin{bmatrix} X^t & 0 \\ 0 & Y^t - X_{22}^t \end{bmatrix} \begin{bmatrix} z^t \\ z_2^t - \hat{x}_2^t \end{bmatrix} + s^{t+1} \tag{9}$$

where $X^t \in \mathbb{R}^{n\times n}$, $Y^t \in \mathbb{R}^{n_2\times n_2}$, *and* $s^t \in \mathbb{R}$ *satisfy the following recursions*

$$X^t = Q + A^T X^{t+1} A \tag{10}$$
$$- A^T X^{t+1} B(R + B^T X^{t+1} B)^{-1} B^T X^{t+1} A \tag{11}$$
$$Y^t = Q_{22} + A_{22}^T Y^{t+1} A_{22} \tag{12}$$
$$- A_{22}^T Y^{t+1} B_{22}(R_{22} + B_{22}^T Y^{t+1} B_{22})^{-1} B_{22}^T Y^{t+1} A_{22} \tag{13}$$
$$s^t = s^{t+1} + \operatorname{trace}\left(X_{11}^t \Sigma_1\right) + \operatorname{trace}\left(Y^t \Sigma_2\right) \tag{14}$$

with $X^N = Q^N$, $Y^N = Q_{22}^N$, *and* $s^{N+1} = 0$. *Then, for each* $t \in \{0, \ldots, N-1\}$,

$$\min_{\gamma^{0:N-1}} \mathcal{J}(\gamma^{0:N-1}) = \min_{\gamma^{0:t}} \mathbb{E}\Bigg(\sum_{k=0}^{t} \begin{bmatrix} x(k) \\ u(k) \end{bmatrix}^T \begin{bmatrix} Q & 0 \\ 0 & R \end{bmatrix} \begin{bmatrix} x(k) \\ u(k) \end{bmatrix} + V^{t+1}\left(h(t+1)\right) \Bigg) \tag{15}$$

Moreover,

$$\min_{\gamma^{0:N-1}} \mathcal{J}(\gamma^{0:N-1}) = s^0$$

where an optimizing policy is given by

$$\gamma_1^t(h_1(t)) = K_{11}^t x_1(t) + K_{12}^t \hat{x}_2^t(h_1(t))$$
$$\gamma_2^t(h(t)) = K_{21}^t x_1(t) + K_{22}^t \hat{x}_2^t(h_1(t)) + J^t\left(x_2(t) - \hat{x}_2^t(h(t))\right)$$

with $K^t \in \mathbb{R}^{m\times n}$ *and* $J^t \in \mathbb{R}^{m_2\times n_2}$ *satisfying*

$$K^t = -(R + B^T X^{t+1} B)^{-1} B^T X^{t+1} A$$
$$J^t = -(R_{22} + B_{22}^T Y^{t+1} B_{22})^{-1} B_{22}^T Y^{t+1} A_{22}$$

Proof. We will show this by induction. Clearly, (15) holds for $t = N-1$, when $X^N = Q^N$, $Y^N = Q_{22}^N$, and $s^{N+1} = 0$. For the inductive step, assume that (15) holds for some $t \in \{1, \ldots, N-1\}$. Then, we have

$$\min_{\gamma^{0:N-1}} \mathcal{J}(\gamma^{0:N-1}) = \min_{\gamma^{0:t}} \int \left(\sum_{k=0}^{t} \begin{bmatrix} z^k \\ a^k \end{bmatrix}^T \begin{bmatrix} Q & 0 \\ 0 & R \end{bmatrix} \begin{bmatrix} z^k \\ a^k \end{bmatrix} + V^{t+1}(i^{t+1}) \right) d\mu^{t+1} \tag{16}$$

From our previous discussion, to find the optimal γ_2^t, we must minimize

$$\min_{\gamma_2^t} \int \left(\begin{bmatrix} z^t \\ a^t \end{bmatrix}^T \begin{bmatrix} Q & 0 \\ 0 & R \end{bmatrix} \begin{bmatrix} z^t \\ a^t \end{bmatrix} + V^{t+1}(i^{t+1}) \right) d\mu^{t+1} \tag{17}$$

To this end, using Lemma 1, we get

$$\begin{bmatrix} x(t+1) \\ x_2(t+1) - \hat{x}_2^{t+1} \end{bmatrix} = \begin{bmatrix} A & B & 0 & 0 \\ 0 & 0 & A_{22} & B_{22} \end{bmatrix} \begin{bmatrix} x(t) \\ u(t) \\ x_2(t) - \hat{x}_2^t \\ u_2(t) - \hat{\gamma}_2^t \end{bmatrix} + \begin{bmatrix} w(t) \\ w_2(t) \end{bmatrix}$$

For notational convenience, let $W^{t+1} = Y^{t+1} - X_{22}^{t+1}$, and define

$$M = \begin{bmatrix} Q + A^T X^{t+1} A & A^T X^{t+1} B \\ B^T X^{t+1} A & R + B^T X^{t+1} B \end{bmatrix}$$
$$P = \begin{bmatrix} A_{22}^T W^{t+1} A_{22} & A_{22}^T W^{t+1} B_{22} \\ B_{22}^T W^{t+1} A_{22} & B_{22}^T W^{t+1} B_{22} \end{bmatrix}$$

Also, let E_1 and E_2 be defined as

$$E_1 = \begin{bmatrix} I \\ 0 \end{bmatrix} \qquad E_2 = \begin{bmatrix} 0 \\ I \end{bmatrix}$$

where the dimensions will be implied by the context. Then, (17) becomes

$$s^{t+2} + \operatorname{trace}(X_{11}^{t+1}\Sigma_1) + \operatorname{trace}(Y^{t+1}\Sigma_2) + \min_{\gamma_2^t} \int \left(\begin{bmatrix} z^t \\ \gamma^t(i^t) \\ z_2^t - \hat{x}_2^t(i_1^t) \\ \gamma_2^t(i^t) - \hat{\gamma}_2^t(i_1^t) \end{bmatrix}^T \begin{bmatrix} M & 0 \\ 0 & P \end{bmatrix} \times \begin{bmatrix} z^t \\ \gamma^t(i^t) \\ z_2^t - \hat{x}_2^t(i_1^t) \\ \gamma_2^t(i^t) - \hat{\gamma}_2^t(i_1^t) \end{bmatrix} \right) d\mu^t \tag{18}$$

Let $g(\gamma_2^t)$ represent the integral term in this expression. Thus, let us optimize player 2's policy by taking a functional derivative of (18) with respect to γ_2^t. Fix $\gamma_1^{0:t}, \gamma_2^{0:t-1}$ and consider the perturbation of $\gamma_2^t + \varepsilon\Gamma_2$, for some $\varepsilon > 0$ and $\Gamma_2 \in L_2^t$. We obtain

$$\begin{aligned} &g(\gamma_2^t + \varepsilon\Gamma_2) = g(\gamma_2^t) \\ &+2\varepsilon \int \begin{bmatrix} \Gamma_2(i^t) \\ \Gamma_2(i^t) - \hat{\Gamma}_2(i_1^t) \end{bmatrix}^T \begin{bmatrix} E_2^T M_{21} & E_2^T M_{22} & 0 & 0 \\ 0 & 0 & P_{21} & P_{22} \end{bmatrix} \\ &\qquad \times \begin{bmatrix} z^t \\ \gamma^t(i^t) \\ z_2^t - \hat{x}_2^t \\ \gamma_2^t(i^t) - \hat{\gamma}_2^t(i_1^t) \end{bmatrix} d\mu^t + O(\varepsilon^2) \end{aligned}$$

where $\hat{\Gamma}_2(i_1^t) = \mathbb{E}(\Gamma_2(h(t)) \,|\, h_1(t) = i_1^t)$. Consequently, γ_2^t is optimal if and only if the term linear in ε equals zero for all $\Gamma_2 \in L_2^t$. Thus,

$$\begin{aligned} 0 &= \int \begin{bmatrix} \Gamma_2 \\ \Gamma_2 - \hat{\Gamma}_2 \end{bmatrix}^T \begin{bmatrix} E_2^T M_{21} & E_2^T M_{22} & 0 & 0 \\ 0 & 0 & P_{21} & P_{22} \end{bmatrix} \\ &\qquad \times \begin{bmatrix} x^t \\ \gamma^t \\ x_2^t - \hat{x}_2^t \\ \gamma_2^t - \hat{\gamma}_2^t \end{bmatrix} d\mu^t \\ &= \int \Gamma_2^T \begin{bmatrix} E_2^T M_{21} & E_2^T M_{22} & P_{21} & P_{22} \end{bmatrix} \begin{bmatrix} x^t \\ \gamma^t \\ x_2^t - \hat{x}_2^t \\ \gamma_2^t - \hat{\gamma}_2^t \end{bmatrix} d\mu^t \\ &\qquad - \int \hat{\Gamma}_2(i_1^t)^T \begin{bmatrix} P_{21} & P_{22} \end{bmatrix} \begin{bmatrix} x_2^t - \hat{x}_2^t \\ \gamma_2^t - \hat{\gamma}_2^t \end{bmatrix} d\mu^t \end{aligned}$$

Notice that since $\hat{\Gamma}_2$ is only function of i_1^t, the second term above is

$$\begin{aligned} &\int \hat{\Gamma}_2(i_1^t)^T \begin{bmatrix} P_{21} & P_{22} \end{bmatrix} \\ &\qquad \times \left(\int \begin{bmatrix} x_2^t - \hat{x}_2^t \\ \gamma_2^t - \hat{\gamma}_2^t \end{bmatrix} p(i_2^t \,|\, i_1^t)\, di_2^t \right) p(i_1^t)\, di_1^t = 0 \end{aligned}$$

from the definitions of $\hat{x}_2^t$ and $\hat{\gamma}_2^t$. Thus, γ_2^t is optimal if

$$\begin{aligned} &E_2^T M_{21} x(t) + E_2^T M_{22} \gamma^t(i^t) \\ &\qquad + P_{21}(x_2(t) - \hat{x}_2^t) + P_{22}(\gamma_2^t(i^t) - \hat{\gamma}_2^t) = 0 \end{aligned} \quad (19)$$

Taking the expectation of (19) conditioned on $h_1(t) = i_1^t$, we obtain

$$\begin{aligned} &E_2^T M_{21} z^t + E_2^T M_{22} \gamma^t(i^t) \\ &- E_2^T M_{21} E_2 (z_2^t - \hat{x}_2^t) - E_2^T M_{22} E_2 (\gamma_2^t(i^t) - \hat{\gamma}_2^t) = 0 \end{aligned}$$

Combining this with (19), we have

$$\begin{aligned} &(P_{22} + E_2^T M_{22} E_2)(\gamma_2^t(h(t)) - \hat{\gamma}_2^t) \\ &\qquad = -(P_{21} + E_2^T M_{21} E_2)(x_2(t) - \hat{x}_2^t) \end{aligned}$$

Defining $J^t \in \mathbb{R}^{m_2 \times n_2}$ by

$$\begin{aligned} J^t &= -(P_{22} + E_2^T M_{22} E_2)^{-1}(P_{21} + E_2^T M_{21} E_2) \\ &= -(R_{22} + B_{22}^T Y^{t+1} B_{22})^{-1} B_{22}^T Y^{t+1} A_{22} \end{aligned}$$

this reduces to $\gamma_2^t(h(t)) - \hat{\gamma}_2^t = J^t(x_2(t) - \hat{x}_2^t)$. Plugging back into (19), we obtain

$$E_2^T M_{22} \gamma^t = -E_2^T M_{21} x^t - (P_{21} + P_{22} J^t)(x_2^t - \hat{x}_2^t) \quad (20)$$

Similarly, to minimize (18) over γ_1^t, consider the permutation $\gamma_1^t + \varepsilon\Gamma_1$, for some $\varepsilon > 0$ and $\Gamma_1 \in L_1^t$. The resulting optimality condition is found to be

$$E_1^T M_{21} \begin{bmatrix} x_1(t) \\ \hat{x}_2^t \end{bmatrix} + E_1^T M_{22} \begin{bmatrix} \gamma_1(h_1(t)) \\ \hat{\gamma}_2^t \end{bmatrix} = 0 \quad (21)$$

Define $K^t \in \mathbb{R}^{m \times n}$ as

$$\begin{aligned} K^t &= -M_{22}^{-1} M_{21} \\ &= -(R + B^T X^{t+1} B)^{-1} B^T X^{t+1} A \end{aligned}$$

Combining (20) and (21), we find the optimal policy γ^t is given by

$$\begin{bmatrix} \gamma_1^t \\ \gamma_2^t \end{bmatrix} = \begin{bmatrix} K_{11}^t & K_{12}^t & 0 \\ K_{21}^t & K_{22}^t & J^t \end{bmatrix} \begin{bmatrix} x_1(t) \\ \hat{x}_2(t) \\ x_2(t) - \hat{x}_2^t \end{bmatrix} \quad (22)$$

Having established the form of the optimal controller, the last step is to substitute back to find V^t. This can be done in a straightforward manner to find that (9) satisfies (15), decremented by one time step, where X^t and Y^t satisfy the Riccati recursions in (10–12), and s^t satisfies the recursion in (14). By induction, (15) is satisfied for all t. In particular, at the last step in the recursion, we have

$$\min_{\gamma^{0:N-1}} J(\gamma^{0:N-1}) = \mathbb{E}(V^0(i^0))$$

Since the $x_1(0)$ and $x_2(0)$ are independent, then $\hat{x}_2^0(i_1^0) = 0$, which implies that

$$\begin{aligned} \min_{\gamma^{0:N-1}} J(\gamma^{0:N-1}) &= \mathbb{E}\left(x(0)^T \begin{bmatrix} X_{11}^0 & X_{12}^0 \\ X_{21}^0 & Y^0 \end{bmatrix} x(0) \right) + s^1 \\ &= \operatorname{trace}(X_{11}^0 \Sigma_1) + \operatorname{trace}(Y^0 \Sigma_2) + s^1 \\ &= s^0 \end{aligned}$$

which completes the proof. ∎

A common heuristic solution to this problem is the following:

$$\begin{aligned} \gamma_1^t &= K_{11}^t x_1(t) + K_{12}^t \hat{x}_2(t) \\ \gamma_2^t &= K_{21}^t x_1(t) + K_{22}^t x_2(t) \end{aligned}$$

In other words, the two players use the optimal centralized gains, except that player 1 uses his estimate of $x_2(t)$, since he can't measure it directly. However, it can be

shown that this heuristic policy can be arbitrarily suboptimal.

In fact, player 2 needs to estimate its own state (despite knowing it exactly), and cannot simply perform the centralized control policy. The intuition here is that player 2 must take into account the effects of player 1's estimation. This result matches the optimal policies obtained in [7]. However, the dynamic programming method taken here differs significantly from the spectral factorization approach there, and offers additional insight into the nature of optimal policies for decentralized control problems over networks. In particular, the optimal value of the objective, s^0, is readily obtained via the above method; this result is not obvious when solved via spectral factorization.

Another important note is the form of the value functions. One of the main drawbacks of the dynamic programming approach, in general, is the fact that a suitable value function must be guessed a priori. The rest of the work is then spent testing whether or not it turns out to be reasonable guess. Unfortunately, for more complicated systems, finding an appropriate form for the value function becomes tricky. In fact, the form chosen here was discovered after establishing the solution to the problem via spectral factorization in [7]. This difficulty is not encountered in the spectral factorization approach, which makes that method more attractive for more general systems.

Lastly, the approach taken here made the implicit assumption that the policies are deterministic. In general networked problems, mixed policies may be optimal. However, it can be shown that deterministic policies perform just as well as mixed policies. This fact can be determined separately, or with more care, the proof method here could be extended to allow for mixed policies. Since deterministic policies turn out to be optimal, the approach here was taken for the sake of clarity.

4 Conclusion

In this paper, we provided a dynamic programming approach to find explicit state-space solutions for a two-player decentralized problem. This approach provides an alternate method for arriving at the results of [7]. Interestingly, the optimal policy requires that player 2 perform an estimate of his own state, despite knowing it exactly. This contradicts the popular notion that player 2, who knows both states, would not need to perform any estimation. In fact, though stability was not a primary concern in this finite-horizon problem, it will be shown in future work that omitting the estimator in system 2 can actually destabilize the system.

As noted above, this paper and [7] provide the first steps in finding analytical state-space solutions for decentralized control problems. Our future work will extend these results to the infinite-horizon, continuous-time, output-feedback, and general graph cases.

References

[1] B. Bamieh and P. G. Voulgaris. Optimal distributed control with distributed delayed measurements. *Proceedings of the IFAC World Congress*, 2002.

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

[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] X. Qi, M. Salapaka, P. Voulgaris, and M. Khammash. Structured optimal and robust control with multiple criteria: A convex solution. *IEEE Transactions on Automatic Control*, 49(10):1623–1640, 2004.

[5] M. Rotkowitz and S. Lall. A characterization of convex problems in decentralized control. *IEEE Transactions on Automatic Control*, 51(2):274–286, 2002.

[6] C.W. Scherer. Structured finite-dimensional controller design by convex optimization. *Linear Algebra and its Applications*, 351(352):639–669, 2002.

[7] J. Swigart and S. Lall. An explicit state-space solution for a decentralized two-player optimal linear-quadratic regulator.

[8] J. Swigart and S. Lall. A graph-theoretic approach to distributed control over networks. In *Proceedings of the IEEE Conference on Decision and Control*, 2009.

[9] P. Vouglaris. Control of nested systems. In *Proceedings of the American Control Conference*, volume 6, pages 4442–4445, 2000.

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