2009 American Control Conference
Hyatt Regency Riverfront, St. Louis, MO, USA
June 10-12, 2009

Spectral Factorization of Non-Classical Information Structures Under Feedback

John Swigart[1] Sanjay Lall[2]

Abstract

We consider linear systems under feedback. We restrict our attention to non-classical information structures for which the optimal control policies can be found via a convex optimization problem. The first step to analytically solving such control problems is performing a spectral factorization to solve the optimality condition. In this paper we discuss two classes of information structures, for which such spectral factorizations can be found. In the first structure, the only constraint is that the controller can remember previous inputs that it has received. In the second structure, we consider a controller which is allowed to forget previous observations.

1 Introduction

The goal of control engineers has always been to find control policies of feedback systems which produce system behavior that is guaranteed optimal or suboptimal, in some sense. The natural starting point is to consider linear, time-invariant systems with full information available to the controller. Under specific assumptions about the objective function that we try to minimize and the statistics of the exogenous disturbance, this problem becomes the classical LQG problem (linear dynamics, quadratic cost, Gaussian noise), which has been solved for quite some time.

In 1968, Witsenhausen showed that a very simple system, which only differed from the classical system by a slight change to the information structure, could be intractable when trying to find the optimal control policy [9]. Since then, control theorists have tried to uncover which feedback structures could be optimally solved. Early on, Ho and Chu [4] were able to show that information structures which they called *partially nested* produced optimal linear control policies for the dynamic team problem. Within that framework, some specific problems have been solved [6, 11, 2].

More recently, it was shown that feedback systems whose information structure was *quadratically invariant* also produced tractable solutions which could be solved via convex optimization [8]. However, convex optimization loses much of the intuition associated with the control policies (separation, etc.). Thus, we desire state-space solutions to these problems.

In the classical, full information case, a number of key steps are made in the process of finding the optimal solution. In one particular method, the first step is re-expressing the objective function as a convex program. The next step is to analytically solve the resulting optimality condition via a spectral factorization. In the classical case, this spectral factorization has been previously established [5]. In this paper we consider two additional classes of information structures, for which a spectral factorization also exists.

When studying distributed controllers, it is important to note the connection between centralized controllers (a single controller), and distributed controllers (multiple, independent controllers). For instance, a fully decentralized structure could be represented as either multiple, independent decision makers, or equivalently as a single controller that has a particular block diagonal structure. Thus, a distributed control problem can be expressed as a single controller with a particular structure imposed upon it. We take the latter point of view in this paper.

This paper is organized as follows. In Section 2, we make use of the Youla parametrization. The dual optimization problem and resulting optimality condition that we will attempt to solve is discussed in Section 3. In order to introduce the reader to our method of solution, Section 4 discusses two important factorizations which can be used to solve the classical LQG problem. In Section 5, we will extend these factorizations to what has been termed a full information sharing structure [6]. In Section 6, we apply our method to a distributed control problem which, to our knowledge, has not been previously solved. Lastly, Section 7 provides an example of this factorization method for a simple distributed control problem.

We consider the following linear dynamical state space system on the finite horizon $t \in [0, N-1]$:

$$\begin{aligned} x(t+1) &= A_t x(t) + B_t u(t) + w(t) \\ y(t) &= C_t x(t) + v(t) \\ z(t) &= V_t x(t) + W_t u(t) \end{aligned} \tag{1}$$

where $x(t) \in \mathbb{R}^n$ is the system state which can be affected by the input $u(t) \in \mathbb{R}^m$ and process noise $w(t) \in \mathbb{R}^n$. The outputs of the system are the observations $y(t) \in \mathbb{R}^p$ which have been corrupted by sensor noise $v(t) \in \mathbb{R}^p$, and $z(t)$ which is a vector that we would like minimize in some sense.

[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

978-1-4244-4524-0/09/$25.00 ©2009 AACC

As usual, $x(0), w(t), v(t)$ are IID Gaussian random variables with zero mean and identity covariances. For convenience, we will also assume that $V_t^T V_t = Q_t \geq 0$, $W_t^T W_t = R_t > 0$, and $V_t^T W_t = 0$.

We define an information structure as the information available for making decision $u(t)$.

$$Y_t(y(0), \ldots, y(t)) \subseteq \{y(0), \ldots, y(t)\} \tag{2}$$
$$U_t(u(0), \ldots, u(t-1)) \subseteq \{u(0), \ldots, u(t-1)\} \tag{3}$$

In other words, the information (Y_t, U_t) available to the controller at time t is a subset of all observations and decisions previously made. In the classical, full information case, we have that Y_t and U_t are identity functions. Using the notation $\zeta = (x(0), w(0), \ldots, v(0), \ldots)$, $u = (u(0), \ldots, u(N-1))$, and similarly for y and z, this system can be represented by the block diagram in Figure 1. As a result, our controller can be written as the matrix K which maps $y \mapsto u$, where the information structure for the controller is represented by sparsity constraints on K. Our goal is to choose such a K which minimizes the expected 2-norm of z.

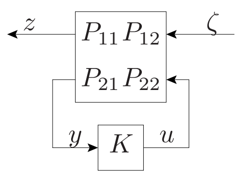

Figure 1: General Feedback System

2 Youla Parametrization

We are interested in the map of $\zeta \mapsto z$. We define Z to be

$$Z = \begin{bmatrix} 0 & & & \\ I_n & 0 & & \\ & \ddots & \ddots & \\ & & I_n & 0 \end{bmatrix}$$

with dimensions appropriate for the context in which it is used. We let $A = \operatorname{diag}(A_0, A_1, \ldots)$, and similarly construct the block diagonal matrices B, C, Q, R, V, and W. We can then write the map of $\zeta \mapsto z$ as

$$z = (P_{11} + P_{12}(I - KP_{22})^{-1} KP_{21})\zeta$$

where, using our notation above, the P matrices are

$$P_{11} = \begin{bmatrix} V(I - ZA)^{-1} & 0_{(n+m)N \times pN} \end{bmatrix} \tag{4}$$
$$P_{12} = W + V(I - ZA)^{-1} ZB \tag{5}$$
$$P_{21} = \begin{bmatrix} C(I - ZA)^{-1} & I_{pN} \end{bmatrix} \tag{6}$$
$$P_{22} = C(I - ZA)^{-1} ZB \tag{7}$$

Using the above definitions, we can finally recast our problem as an optimization problem in terms of K.

$$\begin{array}{ll} \min & \|(P_{11} + P_{12}(I - KP_{22})^{-1} KP_{21})\|_F^2 \\ \text{s.t.} & K \in \mathcal{S} \end{array}$$

where $\mathcal{S}$ represents the information structure of the problem as a sparsity constraint for our matrix K. For instance, in the classical, full information case, we have $\mathcal{S} = \{K \mid K_{ij} = 0 \quad \text{if} \quad i < j\}$. We will also define the complementary subspace $\mathcal{S}^\perp$, which for the classical case would be $\mathcal{S}^\perp = \{F \mid F_{ij} = 0 \quad \text{if} \quad i \geq j\}$.

For more general information structures $\mathcal{S}$, very few solutions to this problem are known, since this formulation is not convex. However, there are a class of information structures, those which are quadratically invariant [8], for which the above problem can be re-expressed as a convex optimization problem. We will restrict our discussion now to information structures which satisfy this property.

Let $\hat{Q} = (I - KP_{22})^{-1} K$. By making use of the quadratic invariance property of our information structure, we can make this simple substitution, which allows us to solve a convex optimization problem in terms of $\hat{Q}$. Using this change of variables, we arrive at the Youla parametrization of the problem which we would like to solve analytically.

$$\begin{array}{ll} \min & \|(P_{11} + P_{12}\hat{Q}P_{21})\|_F^2 \\ \text{s.t.} & \hat{Q} \in \mathcal{S} \end{array} \tag{8}$$

3 Lagrange Dual Problem

In order to analytically solve the optimization problem in (8), we make use of duality.

Lemma 1. *The optimal $\hat{Q}$ in (8) satisfies*

$$-\frac{1}{2}\Lambda = P_{12}^T P_{11} P_{21}^T + P_{12}^T P_{12} \hat{Q} P_{21} P_{21}^T \tag{9}$$

for some $\Lambda \in \mathcal{S}^\perp$.

Proof. Since our problem has a quadratic cost function and linear constraints, we know that strong duality will hold for this problem [1]. Thus, solving the dual problem will produce the optimal solution for our primal problem. To solve the dual problem, we must optimize the Lagrangian function over $\hat{Q}$.

$$L = \|(P_{11} + P_{12}\hat{Q}P_{21})\|_F^2 + \sum_{i,j} \Lambda_{ij} \hat{Q}_{ij}$$

where $\Lambda \in \mathcal{S}^\perp$.

Taking the derivative of this expression with respect to the non-zero components of $\hat{Q}$, we immediately arrive at the optimality condition in (9), where $\Lambda \in \mathcal{S}^\perp$ and $\hat{Q} \in \mathcal{S}$. ∎

Observing the optimality condition in (9), we notice that $\hat{Q}$ and Λ have complimentary structures. Our goal is then to efficiently solve this expression for $\hat{Q}$ independently of Λ.

We note that both sides of (9) must be in $\mathcal{S}^\perp$, since $\Lambda \in \mathcal{S}^\perp$. Thus, our goal in solving for $\hat{Q}$ will be to simplify (9) while preserving the complementary structures of $\hat{Q}$ and Λ. To this end, we will look for factorizations of $P_{12}^T P_{12}$ and $P_{21} P_{21}^T$ which preserve these structures. We will now show, for two classes of non-classical information structures, that such factorizations exist.

4 Two Useful Factorizations

Before proceeding to our special information structures, we remind the reader of two well-known factorizations for general symmetric, positive definite matrices. Namely, for such a matrix A, there exist lower triangular matrices L_1, L_2 and diagonal matrices D_1, D_2, of appropriate dimensions, such that $A = L_1^T D_1 L_1 = L_2 D_2 L_2^T$.

For our case, where we are factorizing $P_{12}^T P_{12}$ and $P_{21} P_{21}^T$, these factorizations can be found via the Riccati recursion.

Lemma 2. *Suppose there exists a matrix $P \in \mathbb{R}^{nN \times nN}$ which satisfies the Riccati recursion*

$$\begin{aligned} P &= Q + A^T Z^T P Z A \\ &\quad - A^T Z^T P Z B (R + B^T Z^T P Z B)^{-1} B^T Z^T P Z A \end{aligned} \tag{10}$$

then $P_{12}^T P_{12}$ can be factorized into $L^T D L$ where $L \in \mathbb{R}^{mN \times mN}$ is a block lower triangular matrix given by

$$L = I + (R + B^T Z^T P Z B)^{-1} B^T Z^T P Z A (I - ZA)^{-1} Z B$$

and $D \in \mathbb{R}^{mN \times mN}$ is a block diagonal matrix given by

$$D = R + B^T Z^T P Z B$$

Proof. This result follows directly from algebraic manipulation of (10) and is omitted here due to space constraints. ■

For the factorization of $P_{21} P_{21}^T$, the development parallels that of Lemma 2. We will simply state the factorization here.

Lemma 3. *Suppose there exists a matrix $S \in \mathbb{R}^{nN \times nN}$ which satisfies the Riccati recursion*

$$\begin{aligned} S &= I + Z A S A^T Z^T \\ &\quad - Z A S C^T (I + C S C^T)^{-1} C S A^T Z^T \end{aligned} \tag{11}$$

then $P_{21} P_{21}^T$ can be factorized into $L D L^T$ where $L \in \mathbb{R}^{pN \times pN}$ is a block lower triangular matrix given by

$$L = I + C (I - ZA)^{-1} Z A S C^T (I + C S C^T)^{-1}$$

and $D \in \mathbb{R}^{pN \times pN}$ is a block diagonal matrix given by

$$D = I + C S C^T$$

5 Vertical/Temporal Skyline Case

We now look at what we call the temporal, or vertical, skyline information structure, defined as follows.

Definition 4. *Consider the set of N integers $\mathcal{I}_1, \ldots, \mathcal{I}_N$, where $\mathcal{I}_j$ satisfies $j \leq \mathcal{I}_j \leq N+1$, for each $j = 1, \ldots, N$. We say that the controller K has a temporal skyline (TS) information structure if at time t it has as its information variables the observation set $Y_t = \{y(j-1) \mid \mathcal{I}_j \leq t+1\}$. We define the set of matrices with TS structure by $\mathcal{TS} = \{K \mid K_{ij} = 0 \quad \text{if} \quad i < \mathcal{I}_j\}$. We also define the complementary structure $\mathcal{TS}^\perp = \{K \mid K_{ij} = 0 \quad \text{if} \quad i \geq \mathcal{I}_j\}$.*

In simpler terms, the controller can receive any observation at any time, subject to causality, and remembers the observations in the future. When we consider our controller as a matrix K, we see that this TS structure implies that $K \in \mathcal{TS}$. It has been previously shown that in a causal system the TS structure produces a quadratically invariant controller [8]. This fact allows the direct use of the previously obtained optimality condition (9), with the dual variable $\Lambda \in \mathcal{TS}^\perp$.

Since $\mathcal{TS}$ is clearly a closed subspace, then any $A \in \mathbb{R}^{n \times n}$ can be decomposed into $B \in \mathcal{TS}$ and $C \in \mathcal{TS}^\perp$ such that $A = B + C$. We define $A^{TS} = B$ as the temporal skyline part of A.

Analogous to the full information case, having established the optimality condition, we now attempt to solve for $\hat{Q}$ while maintaining the complementary TS structure of (9). To this end, we need to define another sparsity pattern.

Definition 5. *We define the set of matrices $\mathcal{SA}$ by $\mathcal{SA} = \{A \mid A_{ij} = 0 \quad \text{if} \quad A_{ji} \neq 0, \text{ and } \quad A_{ii} = I\}$. We will call matrices in $\mathcal{SA}$* sparse antisymmetric.

Definition 6. *Given $\mathcal{I}_1, \ldots, \mathcal{I}_N$, we define the set of matrices $\mathcal{SA}^{(\mathcal{I},\Lambda)}$ such that $A \in \mathcal{SA}^{(\mathcal{I},\Lambda)}$ if and only if $A \in \mathcal{SA}$ and $A_{ij} = 0$ whenever $\mathcal{I}_i > \mathcal{I}_j$, or $\mathcal{I}_i = \mathcal{I}_j$ and $i > j$.*

We similarly define the set of matrices $\mathcal{SA}^{(\mathcal{I},Q)}$ such that $A \in \mathcal{SA}^{(\mathcal{I},Q)}$ if and only if $A \in \mathcal{SA}$ and $A_{ij} = 0$ whenever $\mathcal{I}_i < \mathcal{I}_j$, or $\mathcal{I}_i = \mathcal{I}_j$ and $i < j$.

Using this notation, we can solve for the optimal $\hat{Q}$ matrix in (9). The result, stated here, will be proved at the end of this section.

Theorem 7. *If $\mathcal{S} = \mathcal{TS}$, then the solution of (9) is*

$$\hat{Q} = -(D_C L_C)^{-1} (L_C^{-T} P_{12}^T P_{11} P_{21}^T \widehat{U}_E^{-1})^{TS} (\widehat{L}_E \widehat{D}_E)^{-1} \tag{12}$$

where $P_{12}^T P_{12} = L_C^T D_C L_C$ from Lemma 2, and $P_{21} P_{21}^T = \widehat{L}_E \widehat{D}_E \widehat{U}_E$, where $\widehat{L}_E \in \mathcal{SA}^{(\mathcal{I},Q)}$, $\widehat{U}_E \in \mathcal{SA}^{(\mathcal{I},\Lambda)}$, and $\widehat{D}_E$ is diagonal.

Before proving Theorem 7, we need to show which sets of matrices preserve the sparsity structure of $\mathcal{TS}^\perp$.

Lemma 8. *Suppose $\Lambda \in \mathcal{TS}^\perp$, $A \in \mathcal{SA}^{(\mathcal{I},\Lambda)}$, and U is an upper triangular matrix. Then, $U \Lambda A \in \mathcal{TS}^\perp$.*

Proof. Looking at the product of $U\Lambda A$, we have,

$$(U\Lambda A)_{ij} = \sum_{k,m} U_{ik} \Lambda_{km} A_{mj} = \sum_{\substack{i \leq k \\ k < \mathcal{I}_m \\ \mathcal{I}_m \leq \mathcal{I}_j}} U_{ik} \Lambda_{km} A_{mj}$$

Looking at the summation, we see that there will be no terms to sum over whenever $i \geq \mathcal{I}_j$, meaning that $(U\Lambda A)_{ij} = 0$ whenever that condition holds. Hence, $U\Lambda A \in \mathcal{TS}^\perp$. ■

This lemma is helpful since it is shows that left multiplication by upper triangular matrices preserves the structure of $\mathcal{TS}^\perp$, which is identical to the classical case and will allow us to use the factorization from Lemma 2. Moreover, the set $\mathcal{SA}^{(\mathcal{I},\Lambda)}$ preserves the structure of $\mathcal{TS}^\perp$ under right multiplication.

In the same manner as Lemma 8, we can show how to preserve the sparsity structure of $\hat{Q}$.

Lemma 9. *If $\hat{Q} \in \mathcal{TS}$, $A \in \mathcal{SA}^{(\mathcal{I},Q)}$, and L is a lower triangular matrix, then $L\hat{Q}A \in \mathcal{TS}$.*

Proof. This fact can be proven by explicit computation of the entries of $L\hat{Q}A$, in the same fashion as Lemma 8. ■

With Lemmas 8 and 9, the differences between the classic case and our present TS case become apparent. In our full information case, we could left and right multiply the optimality condition by any upper triangular matrices and preserve the strict upper triangular structure of the optimality condition. In our present case, we see that left multiplying by an upper triangular matrix will still work, but right multiplying will require matrices in $\mathcal{SA}^{(\mathcal{I},\Lambda)}$.

In the classical case, we could show that the inverse of any upper triangular matrix was another upper triangular matrix. A similar result holds for matrices in $\mathcal{SA}^{(\mathcal{I},\Lambda)}$.

Lemma 10. *Consider the matrix $A \in \mathcal{SA}^{(\mathcal{I},\Lambda)}$. There exists a permutation matrix J such that $J^T AJ$ is an upper triangular matrix.*

Proof. The proof follows from the definition of $\mathcal{SA}^{(\mathcal{I},\Lambda)}$ and is omitted here due to space constraints. ■

Lemma 11. *Suppose $A \in \mathcal{SA}^{(\mathcal{I},\Lambda)}$. Then, $A^{-1} \in \mathcal{SA}^{(\mathcal{I},\Lambda)}$.*

Proof. Using Lemma 10, we can find the permutation matrix J such that $U = J^T AJ$ is an upper triangular matrix with identities on the diagonal. Since U and J are invertible, then A^{-1} exists. Also, $U^{-1} = J^T A^{-1} J$ is upper triangular which implies that $A^{-1} \in \mathcal{SA}^{(\mathcal{I},\Lambda)}$. ■

As a result of these Lemmas, what we have shown is that in order to solve the optimality condition (9), we need to find a factorization for $P_{21}P_{21}^T$ whose factors are in $\mathcal{SA}$. Fortunately, the following lemma guarantees the existence of just such a factorization.

Lemma 12. *For any symmetric, positive definite matrix A, there exist matrices $\widehat{L} \in \mathcal{SA}^{(\mathcal{I},Q)}$, $\widehat{U} \in \mathcal{SA}^{(\mathcal{I},\Lambda)}$, and a diagonal matrix $\widehat{D}$ such that $A = \widehat{L}\widehat{D}\widehat{U}$.*

Proof. For the integers $\mathcal{I}$, we can find the permutation matrix J, as established in Lemma 10. Since we already know that an LDU factorization exists for any positive definite matrix, then we have

$$A = J(LDU)J^T = (JLJ^T)(JDJ^T)(JUJ^T) = \widehat{L}\widehat{D}\widehat{U}$$

where LDU is the LDU factorization for $J^T AJ$. Also, from Lemma 10, we know that $\widehat{U} = JUJ^T \in \mathcal{SA}^{(\mathcal{I},\Lambda)}$ and $\widehat{L} = JLJ^T \in \mathcal{SA}^{(\mathcal{I},Q)}$. Lastly, since J is a permutation matrix, then $\widehat{D} = JDJ^T$ is also a diagonal matrix. ■

With Lemma 12, we now have the means to solve this problem. We simply need to LDU factorize $J^T P_{21} P_{21}^T J$ and then compute $\widehat{U} = JUJ^T$, $\widehat{L} = JLJ^T$, and $\widehat{D} = JDJ^T$.

Proof of Theorem 7. Using Lemma 12 for the factorization of $P_{21}P_{21}^T$, the optimality condition (9) becomes

$$-\frac{1}{2}\Lambda = P_{12}^T P_{11} P_{21}^T + L_C^T D_C L_C \hat{Q} \widehat{L}_E \widehat{D}_E \widehat{U}_E$$

By making use of Lemmas 8, 9 and 11, we know that $D_C L_C \hat{Q} \widehat{L}_E \widehat{D}_E \in \mathcal{TS}$ and $L_C^{-T} \Lambda \widehat{U}_E^{-1} \in \mathcal{TS}^\perp$. Hence, we can rewrite the above expression as

$$-\frac{1}{2} L_C^{-T} \Lambda \widehat{U}_E^{-1} = L_C^{-T} P_{12}^T P_{11} P_{21}^T \widehat{U}_E^{-1} + D_C L_C \hat{Q} \widehat{L}_E \widehat{D}_E$$

This expression preserves structure of $\mathcal{TS}^\perp$. Moreover, the term affine in our variable $\hat{Q}$ now has the same TS structure of $\hat{Q}$. Thus, we have separated the solution of $\hat{Q}$ from the solution of Λ. In other words, if we take the TS part of the optimality condition, we have

$$0 = (L_C^{-T} P_{12}^T P_{11} P_{21}^T \widehat{U}_E^{-1})^{TS} + D_C L_C \hat{Q} \widehat{L}_E \widehat{D}_E$$

As a result, the solution for the optimal $\hat{Q}$ in (8) with a TS sparsity constraint is given by (12). ■

As a result of Theorem 7, we have performed the first step in finding an analytic solution for the optimal control policy. Namely, we have found a spectral factorization which allows us to solve (9) for $\hat{Q}$, independently of Λ.

We now turn our attention to a different non-classical information structure, for which we can find another spectral factorization.

6 Horizontal/Spatial Skyline Case

In the previous section we extended the results from the classic information structure to solve the control problem which had a temporal skyline (TS) information structure. Fortunately, the above analysis can also be utilized for other information structures. One such information structure is the horizontal skyline structure which we shall now define in a manner analogous to our TS structure.

Definition 13. *Consider the set of integers $\mathcal{J}_1, \ldots, \mathcal{J}_N$, where $\mathcal{J}_i$ satisfies $i \geq \mathcal{J}_i \geq 0$, for each $i = 1, \ldots, N$. We say that the controller K has a spatial skyline (SS) information structure if at time t it has as its information variables the observation set $Y_t = \{y(j) \mid j + 1 \leq \mathcal{J}_{t+1}\}$. We define the set of matrices with SS structure by $\mathcal{SS} = \{K \mid K_{ij} = 0 \quad if \quad j > \mathcal{J}_i\}$. We also define the complementary structure $\mathcal{SS}^\perp = \{K \mid K_{ij} = 0 \quad if \quad j \leq \mathcal{J}_i\}$.*

When we consider our controller as a matrix K, we see that this SS structure implies that $K \in \mathcal{SS}$. This sparsity structure is where it gets its alternate name as a horizontal skyline structure. While the TS and SS structures are functionally very similar, it is important to note that these two structures are physically very different. In the TS case, we can think of the system as a single decision maker who receives information at arbitrary times but remembers everything he receives. On the other hand, in the SS case, we are forced to think of the problem as a multiple player system where each player makes decisions in turn based on different sets of information.

Since $\mathcal{SS}$ is also a closed subspace, then any $A \in \mathbb{R}^{n \times n}$ can be decomposed into $B \in \mathcal{SS}$ and $C \in \mathcal{SS}^{\perp}$ such that $A = B + C$. We define $A^{SS} = B$ as the spatial skyline part of the matrix A.

Example 14. *Suppose $\mathcal{J} = (1, 2, 1, 3)$. Then, at time $t = 0$, player 0 has $\{y(0)\}$ available to make decision $u(0)$. At time $t = 1$, player 1 has $\{y(0), y(1)\}$ available for making decision $u(1)$. However, player 2 only has $\{y(0)\}$ available to make decision $u(2)$, while player 3 has $\{y(0), y(1), y(2)\}$ available to make decision $u(2)$.*

Our first step in analyzing this information structure is to show that it is quadratically invariant.

Lemma 15. *The set $\mathcal{SS}$ is quadratically invariant under any causal plant P.*

Proof. In order for the SS structure to be quadratically invariant, we must show that

$$KP_{22}K \in \mathcal{SS} \qquad \text{for all } K \in \mathcal{SS}$$

To lighten notation, let $G = P_{22}$ be lower triangular (causal). Then, we have

$$(KGK)_{ij} = \sum_{k,m} K_{ik} G_{km} K_{mj} = \sum_{\substack{k \le \mathcal{J}_i \\ m \le k \\ j \le \mathcal{J}_m}} K_{ik} G_{km} K_{mj}$$

Using the fact that $m \geq \mathcal{J}_m$, by definition, we must have $(KGK)_{ij} = 0$ whenever $j > \mathcal{J}_i$. Hence, $KGK \in \mathcal{SS}$, so $\mathcal{SS}$ is quadratically invariant under G. ■

With Lemma 15, we can make use of our previous analysis to arrive at the same optimality condition (9), where $\hat{Q} \in \mathcal{SS}$ and $\Lambda \in \mathcal{SS}^{\perp}$. Having established the optimality condition, we now try to solve it as we did in the TS information case, by maintaining the complementary SS structure of (9) while solving for $\hat{Q}$.

Given $\mathcal{J}_1, \ldots, \mathcal{J}_N$, we construct $\mathcal{SA}^{(\mathcal{J},\Lambda)}$ and $\mathcal{SA}^{(\mathcal{J},Q)}$ from Definition 6. Similarly to the TS case, these are the sets of sparse antisymmetric matrices which preserve the sparsity structure of Λ and $\hat{Q}$, respectively, under multiplication.

In a completely analogous manner to the TS case, we will state the following lemmas.

Lemma 16. *Suppose $\Lambda \in \mathcal{SS}^{\perp}$, $A \in \mathcal{SA}^{(\mathcal{J},\Lambda)}$, and U is an upper triangular matrix. Then, $A\Lambda U \in \mathcal{SS}^{\perp}$.*

Lemma 17. *For any symmetric, positive definite matrix A and integers $\mathcal{J}_1, \ldots, \mathcal{J}_N$, there exist matrices $\widehat{L} \in \mathcal{SA}^{(\mathcal{J},Q)}$, $\widehat{U} \in \mathcal{SA}^{(\mathcal{J},\Lambda)}$, and a diagonal matrix $\widehat{D}$ such that $A = \widehat{U}\widehat{D}\widehat{L}$.*

The above lemmas can be proved in the same manner as our analysis in the previous section. Thus, we see that for the SS case we simply need to perform the same factorizations that we did in the TS case, just for opposite sides of $\hat{Q}$ in (9).

$$\begin{bmatrix} \times & & & & & \\ & & & & & \\ \times & \times & \times & & & \\ \times & \times & & & & \\ \times & \times & \times & \times & \times & \\ \times & \times & \times & \times & & \end{bmatrix}$$

(a) Spatial Skyline Structure

$$\left[\begin{array}{cc|cc|cc} \times & \times & & & & \\ & \times & & & & \\ \hline \times & \times & \times & \times & & \\ \times & \times & & \times & & \\ \hline \times & \times & \times & \times & \times & \times \\ \times & \times & \times & \times & & \times \end{array}\right]$$

(b) Sparsity of $\mathcal{SA}^{(\mathcal{J},Q)}$

Figure 2: Sparsity patterns

Theorem 18. *If $\mathcal{S} = \mathcal{SS}$, then the solution of (9) is*

$$\hat{Q} = -(\widehat{D}_C \widehat{L}_C)^{-1} (\widehat{U}_C^{-1} P_{12}^T P_{11} P_{21}^T L_E^{-T})^{SS} (L_E D_E)^{-1} \tag{13}$$

where $P_{21}P_{21}^T = L_E D_E L_E^T$ from Lemma 3, and $P_{12}^T P_{12} = \widehat{U}_E \widehat{D}_E \widehat{L}_E$ from Lemma 17.

Proof. The proof here parallels the proof of Theorem 7. The only difference, as noted above, is that the sparse antisymmetric factorization is performed on $P_{12}^T P_{12}$ here instead, and $P_{21}P_{21}^T$ has the standard LDL^T factorization. ■

7 Spatial Skyline Example

To better illustrate the factorization methods described above, we provide an example of the horizontal skyline structure. As noted in Section 6, this structure should be viewed a distributed information structure since some decision makers may have less information than previous decision makers.

To highlight this fact, let us consider a type of soldier-general problem. We have two decision makers: a soldier who is on the front lines receiving observations in real time, and a general who receives observations from the soldier but with a fixed communication delay. For simplicity we will assume that they take turns making decisions, though any sequence of decisions can be handled by our framework. The result of such an information pattern is a SS structure, which might look something like Figure 2(a). In other words, using our notation from Section 6, we represent the controller subspace $\mathcal{SS}$ by the set of matrices with the sparsity structure in Figure 2(a).

We consider the system (1), when the system matrices are time-invariant, so that $A = \operatorname{diag}(A_0, A_0, \ldots)$, and similarly for R, Q, B. From Lemma 16, we know that we need a sparse antisymmetric factorization for $P_{12}^T P_{12}$ in order to solve (9). For the structure in Figure 2(a), the corresponding set $\mathcal{SA}^{(\mathcal{J},Q)}$ can be represented by the set of sparse antisymmetric matrices which have structure of Figure 2(b).

Notice that if we group the above matrix into 2x2 blocks, that this courser block structure is lower triangular with the blocks on the diagonal being upper triangular. This hints at the recursion needed to perform the factorization. Namely, we must first perform a UDL factorization of the form

$$P_{12}^T P_{12} = \begin{bmatrix} I_2 & U_{12} & U_{13} \\ & I_2 & U_{23} \\ & & I_2 \end{bmatrix} \begin{bmatrix} D_1 & & \\ & D_2 & \\ & & D_3 \end{bmatrix} \begin{bmatrix} I_2 & & \\ L_{21} & I_2 & \\ L_{31} & L_{32} & I_2 \end{bmatrix}$$

where we use I_2 to denote the identity matrix for the courser 2x2 blocks. With this factorization, we now perform an LDU factorization on the diagonal elements themselves, so $D_i = U_i^T D_{ii} U_i$, where D_{ii} is diagonal and

$$U_i = \begin{bmatrix} I_1 & U_{i,12} \\ & I_1 \end{bmatrix} \quad (14)$$

where I_1 represents the identity matrix on the finer scale.

With these two factorizations, it is immediately apparent that we can express $P_{12}^T P_{12}$ as $L^T D L$, where D is diagonal and

$$L = \begin{bmatrix} U_1 & & \\ U_2 L_{21} & U_2 & \\ U_3 L_{31} & U_3 L_{32} & U_3 \end{bmatrix}$$

It is clear that $L \in \mathcal{SA}^{(\mathcal{J},Q)}$, so we have found the factorization required to solve this problem.

Let us now look at these factorizations in detail. For the course UDL factorization, we define the matrices

$$M_0 = \begin{bmatrix} 0 & A_0 \\ 0 & A_0^2 \end{bmatrix} \qquad L_0 = \begin{bmatrix} I_1 & \\ A_0 & I_1 \end{bmatrix}$$

and let $M = \operatorname{diag}(M_0, M_0, \ldots)$ and $L = \operatorname{diag}(L_0, L_0, \ldots)$. Similar to I_2, we define Z_2 as the 2x2 block analog of Z. Then, we can write

$$P_{12}^T P_{12} = R + L^T Z_2^T (I - Z_2 M)^{-T} Q (I - Z_2 M)^{-1} Z_2 L$$

With this construction, we can use the result from Lemma 2 to write $P_{12}^T P_{12} = L_C^T D_C L_C$, where

$$L_C = I + K(I - Z_2 M)^{-1} Z_2 L$$
$$K = (R + L^T Z_2^T P Z_2 L)^{-1} L^T Z_2^T P Z_2 M$$

and $D_C \in \mathbb{R}^{mN \times mN}$ is a block diagonal matrix given by

$$D_C = R + L^T Z_2^T P Z_2 L$$

Here, P satisfies the Riccati recursion

$$P = Q + M^T Z_2^T P Z_2 M - M^T Z_2^T P Z_2 L K$$

Since $P = \operatorname{diag}(P_0, P_1, \ldots)$ and $P_i = \operatorname{diag}(Q_0, P_{ii})$, for the second factorization of each diagonal element, we have

$$D_i = R + L_0^T P_i L_0 = U_i^T \operatorname{diag}(N_1, N_2) U_i$$

with U_i given in (14) and

$$U_{i,12} = (R_0 + B_0^T Q_0 B_0 + B_0^T A_0 P_{ii} A_0 B_0)^{-1} B_0^T A_0^T P_{ii} B_0$$

and

$$N_1 = R_0 + B_0^T Q_0 B_0 + B_0^T A_0 P_{ii} A_0 B_0$$
$$N_2 = R_0 + B_0^T P_{ii} B_0 - B_0^T P_{ii} A_0 B_0 U_{i,12}$$

We end our example here, having now performed the factorization needed to solve (9). While we chose a rather simple example, it highlights one of the important aspects of these types of problems; namely, diagonal elements in the factorization which are a result of two Riccati recursions. In the classical case, these diagonal elements satisfy the single Riccati recursion (10). However, we see a need to perform two Riccati recursions here: one on a course partitioning of the system matrices, and another on the finer scale. These simultaneous Riccati recursions are something we are not aware of in the literature to date, and might have implications in other control problems.

8 Conclusion

We have considered a general feedback system with two nonclassical information feedback structures. Using the quadratic invariance property of each information structure, we first expressed each system as a convex optimization problem. These optimization problems could then be solved via a spectral factorization approach, whereby we could preserve the structure of the optimality condition while simultaneously solving for the optimal policy. In the classical information case, these factorizations took the form of upper and lower triangular matrices. In the temporal skyline case, we showed the need for a special sparse antisymmetric factorization for the estimator side of the equation. Conversely, for the spatial skyline case, the sparse antisymmetric factorization was required for the controller side. We also provided an example to illustrate this factorization method. Using these factorizations, we were then able to partition the optimality condition and solve for the optimal policy independently of the Lagrange multipliers.

References

[1] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.

[2] G. Casalino, F. Davoli, R. Minciardi, P. P. Puliafito, and R. Zoppoli. Partially nested information structures with a common past. *IEEE Transactions on Automatic Control*, 29(9):846–850, 1984.

[3] G. Golub and C. Van Loan. *Matrix Computations*. The Johns Hopkins University Press, 1996.

[4] 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.

[5] B.P. Molinari. The stabilizing solution of the discrete algebraic riccati equation. *IEEE Transactions on Automatic Control*, pages 396–399, 1975.

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

[7] R. Radner. Team decision problems. *Annals of mathematical statistics*, 33:857–881, 1962.

[8] M. Rotkowitz and S. Lall. Decentralized control information structures preserved under feedback. *Proceedings of the IEEE Conference on Decision and Control*, pages 569–575, 2002.

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

[10] H. S. Witsenhausen. Separation of estimation and control for discrete time systems. *Proceedings of the IEEE*, 59(11):1557–1566, 1971.

[11] T. Yoshikawa. Dynamic programming approach to decentralized stochastic control problems. *IEEE Transactions on Automatic Control*, pages 796–797, 1975.