Proceedings of the
47th IEEE Conference on Decision and Control
Cancun, Mexico, Dec. 9-11, 2008

On Information Structures, Convexity, and Linear Optimality

Michael Rotkowitz
Department of Electrical and Electronic Engineering
The University of Melbourne
Parkville VIC 3010, Australia
Email: mcrotk@unimelb.edu.au

Abstract

In 1968, Witsenhausen introduced his celebrated counterexample, which illustrated that when an information pattern is nonclassical, the controllers which optimize an expected quadratic cost may be nonlinear. For the special invited session commemorating the fortieth anniversary of the counterexample, we address one of the four follow-up questions listed in his original paper; namely, whether there is a relation between the convexity of finding the optimal affine controller, and whether that controller is in fact optimal. In particular, we discuss the connections between partially nested structures, for which linear controllers are known to be optimal, and quadratically invariant structures, for which optimal linear control is known to be convex.

1 Introduction

A classical information pattern assumes that, at every time step, the controller can access not only information from that time, but from all preceding times as well. When this holds, and when the dynamics are linear, the cost quadratic, and the noise Gaussian, optimal controllers are linear.

The Witsenhausen Counterexample [10] showed that when a nonclassical information pattern exists, then affine controllers may be suboptimal for the LQG cost. To put this another way, the class of affine functions is not always complete when the information pattern is not classical.

The conclusions of his paper were as follows:

9. Conclusions. (i) *Further study of linear, Gaussian, quadratic control problems with general information patterns appears to be required.*
(ii) *The existence of an optimum and the question of completeness of the class of affine designs must be examined as a function of the information pattern.*
(iii) *It would be interesting if a relation could be found between the appearance of several local minima over the affine class and lack of completeness of this class.*
(iv) *Algorithms for approaching an optimal solution need to be developed. Because of the occurrence of local minima, this appears to be a most difficult task.*

(i) is obviously open-ended, and his work indeed motivated and catalyzed forty years of research into decentralized control. (ii) was largely answered by Ho and Chu in 1972 [2] with the introduction of partially nested structures, for which linear controllers are indeed optimal. These will be discussed at length later. (iv) is fairly open-ended as well, and there has certainly been a progression of nonlinear optimization theory and tools, but there have also been specific efforts to approach the optimal solution for the Witsenhausen counterexample, or to use it as a testbed for new optimization tools. In particular, the best known achievable cost (for benchmark values of the parameters) was driven to a new low in [3] and was further improved upon in [4].

In this paper, we address the third question.

2 Preliminaries

We review the Witsenhausen counterexample [10], partially nested structures [2], and quadratic invariance [7, 8].

2.1 Witsenhausen counterexample

We review the Witsenhausen counterexample, blending his original notation with variable names commonly used in modern control frameworks.

Given noise

$$w = \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}$$

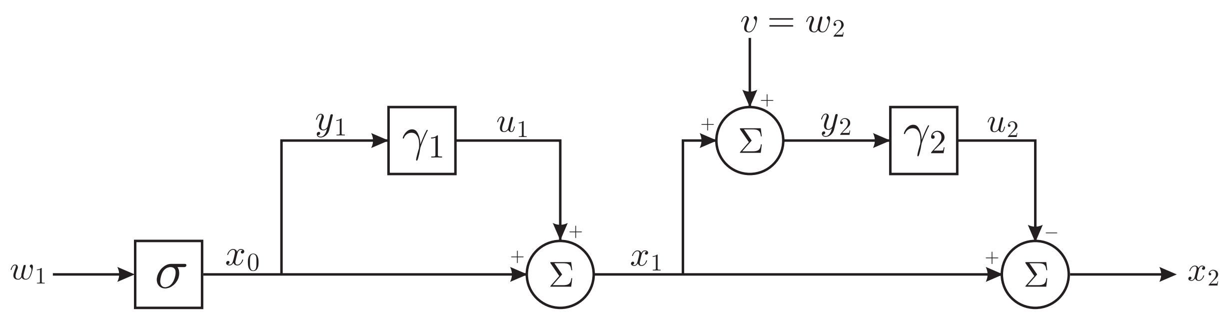

Figure 1: The Witsenhausen Counterexample

and control laws

$$u_1 = \gamma_1(y_1) \qquad u_2 = \gamma_2(y_2)$$

the system then evolves as follows, as indicated in Figure 1

$$\begin{aligned} x_0 &= \sigma w_1 & v &= w_2 \\ y_1 &= x_0 & y_2 &= x_1 + v \\ u_1 &= \gamma_1(y_1) & u_2 &= \gamma_2(y_2) \\ x_1 &= x_0 + u_1 & x_2 &= x_1 - u_2 \end{aligned}$$

and we wish to keep the following variables small

$$\begin{aligned} z_1 &= k\ u_1 \\ z_2 &= x_2. \end{aligned}$$

The noise was normally distributed $w \sim \mathcal{N}(0, I)$ and we seeked γ_1, γ_2 to minimize $\mathbb{E}\ \|z\|_2^2$.

This differs from a standard LQG problem only in that the second controller cannot access the information that the first controller can access, and this subtle difference is shown to cause the optimal controller to no longer be linear, as well as to cause the problem of finding the optimal linear controller to no longer be convex. The obvious follow-up questions of, for which information structures are linear controllers optimal, and, for which information structures is the optimal control problem convex, were largely answered by the introduction of partially nested structures and quadratically invariant structures, respectively. These are the topics of the next two subsections.

2.2 Partially Nested

We now review the framework and the main results of [2]. We only alter notation from the original when it is necessary to avoid doubly-defined variables later. We also add a little notation where it will ease later discussions, and try to clarify when we do.

Problem setup. Suppose we have N team members, and let $\mathcal{I} := \{1, \ldots, N\}$.

We assume that we have an underlying random variable $\xi \in \mathbb{R}^n$ on probability space $(\mathbb{R}^n,\ \mathcal{F},\ P)$ and that $\xi \sim \mathcal{N}(0, X)$, for some $X > 0$.

Each team member i can access the measurement $\tilde{y}_i$ where

$$\tilde{y}_i = H_i \xi + \sum_j D_{ij} u_j \quad \forall\ i \in \mathcal{I} \tag{1}$$

and H_i and D_{ij} are matrices of appropriate dimension.

While not explicitly defined as such, H_i' and D_{ij}' are used to indicate the components of H_i and D_{ij} which correspond to a new measurement, while the rest can correspond to measurements of other team members passed on to member i. To help make this distinction as we go forward, we will refer to this new measurement as y_i, which will also be commensurate with our notation in the next subsection.

For example, suppose that the measurement for the first team member is random

$$\tilde{y}_1 = y_1 = H_1 \xi = H_1' \xi$$

and that the second team member can access this measurement, as well as a new measurement affected by the first member's control input:

$$\tilde{y}_2 = \begin{bmatrix} \tilde{y}_1 \\ y_2 \end{bmatrix} = \underbrace{\begin{bmatrix} H_1 \\ H_2' \end{bmatrix}}_{H_2} \xi + \underbrace{\begin{bmatrix} 0 \\ D_{21}' \end{bmatrix}}_{D_{21}} u_1 + \underbrace{\begin{bmatrix} 0 \\ 0 \end{bmatrix}}_{D_{22}} u_2.$$

D_{ij}' thus represents if and how control input j can affect the new measurement i, and D_{ij} thus represents if and how control input j can affect the overall information accessible to member i. The sparsity structure of these matrices then defines the information structure.

They further assume that

$$D_{ij} \neq 0 \implies D_{ji} = 0 \quad \forall\ i, j \in \mathcal{I} \tag{2}$$

where it seems to be presumed that each team member decision corresponds to a particular time, so that this would follow from causality. This is not always desired if the team members correspond to, for example, different subsystems which may affect one another; we will revisit this assumption later. In fact, what they actually require is more restrictive than (2), since they want to ensure that if control action u_j affects the information $\tilde{y}_i$, then u_i cannot affect the information $\tilde{y}_j$. For instance, the following,

$$D = \begin{bmatrix} 0 & 0 & a \\ b & 0 & 0 \\ 0 & c & 0 \end{bmatrix}$$

while not violating (2), would violate the actual assumptions of their paper for any $a, b, c \neq 0$ due to the cycle that it creates. To more formally state the actual assumptions of the partially nested result, they require that

$$\nexists\ r_1, \ldots, r_\kappa \in \mathcal{I} \quad \text{such that} \quad D_{r_1 r_2}, \ldots, D_{r_{\kappa-1} r_\kappa}, D_{r_\kappa r_1} \neq 0 \tag{3}$$

for any postive integer κ. This includes the formerly implicit assumption that

$$D_{ii} = 0 \quad \forall\ i \in \mathcal{I} \tag{4}$$

which similarly was sensible for the types of problems envisioned, but is not always desired. We will revisit this as well.

The control input of each team member is given as $u_i = \gamma_i(\tilde{y}_i)$. Let Γ_i be the set of all Borel-measurable functions $\gamma_i : \mathbb{R}^{q_i} \to \mathbb{R}^{k_i}$, and let $\Gamma = \{\ [\gamma_1, \ldots, \gamma_N] \mid \gamma_i \in \Gamma_i\ \forall\ i \in \mathcal{I}\}$. The objective is then to find $\gamma \in \Gamma$ to minimize

$$J \ = \ \mathbb{E}\big(\frac{1}{2}u^T Q u + u^T V \xi + u^T c\big) \tag{5}$$

for some $Q \in \mathbb{R}^{k \times k},\ Q > 0,\ S \in \mathbb{R}^{k \times n},\ c \in \mathbb{R}^k$.

Main PN result. Let $\mathcal{Z}_i \subset \mathcal{F}$ be the sub-σ-algebra induced by fixing $\gamma_i \in \Gamma_i$.

They say that j is a ***precedent*** of i, and denote $j \prec i$, iff $D_{ij} \neq 0$ or there exist $r_1, \ldots, r_\kappa \in \mathcal{I}$ such that $D_{ir_1}, D_{r_1 r_2}, \ldots, D_{r_{\kappa-1} r_\kappa}, D_{r_\kappa j} \neq 0$. This can be interpreted as saying that j is a precedent of i iff the jth control input u_j can affect the ith measurement $\tilde{y}_i$.

An information structure is then defined to be ***partially nested (PN)*** if

$$j \prec i \implies \mathcal{Z}_j \subset \mathcal{Z}_i \quad \forall\ i, j \in \mathcal{I},\ \forall\ \gamma \in \Gamma \tag{6}$$

This translates to saying that an information structure is partially nested if whenever the jth control input u_j can affect the ith measurement $\tilde{y}_i$, then the jth measurement $\tilde{y}_j$ can be deduced from $\tilde{y}_i$. In others words, whenever someone can affect what you see, you can see everything that they can.

Their main result is then that if the information structure is partially nested, then the optimal control exists, is unique, and is linear.

The key to their proof was showing that when the information structure is partially nested, the problem is equivalent to what is known as a static team problem, which is known to have a unique linear optimum [5]. In this notation, a static team problem is simply one for which $D = 0$, while also assuming that all H_i are fat and full rank.

2.3 Quadratically Invariant

General problem setup. We now introduce the framework in which quadratic invariance has been developed.

Suppose that we have linear spaces $\mathcal{U}, \mathcal{W}, \mathcal{Y}, \mathcal{Z}$, and a generalized plant $P \in \mathcal{L}(\mathcal{W} \times \mathcal{U}, \mathcal{Z} \times \mathcal{Y})$ partitioned as

$$P = \begin{bmatrix} P_{11} & P_{12} \\ P_{21} & G \end{bmatrix}$$

so that $P_{11} : \mathcal{W} \to \mathcal{Z}$, $P_{12} : \mathcal{U} \to \mathcal{Z}$, $P_{21} : \mathcal{W} \to \mathcal{Y}$ and $G : \mathcal{U} \to \mathcal{Y}$. Suppose $K \in \mathcal{L}(\mathcal{Y}, \mathcal{U})$. If $I - P_{22}K$ is invertible, define $f(P, K) \in \mathcal{L}(\mathcal{W}, \mathcal{Z})$ by

$$f(P, K) = P_{11} + P_{12}K(I - GK)^{-1}P_{21}$$

The map $f(P, K)$ is called the (lower) ***linear fractional transformation*** (LFT) of P and K, and is also referred to as the ***closed-loop map***. This represents the map from w to z with the interconnection depicted in Figure 2. Note the abbreviation $G := P_{22}$ made to simplify later discussion.

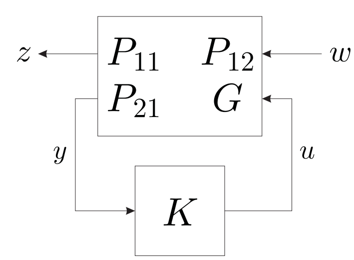

Figure 2: Linear-fractional, or two-input two-output framework.

Given linear spaces $\mathcal{U}, \mathcal{W}, \mathcal{Y}, \mathcal{Z}$, generalized plant $P \in \mathcal{L}(\mathcal{W} \times \mathcal{U}, \mathcal{Z} \times \mathcal{Y})$, and a subspace $S \subset \mathcal{L}(\mathcal{Y}, \mathcal{U})$, the objective is to solve the following problem:

$$\begin{aligned} \text{minimize} \quad & \|f(P, K)\| \\ \text{subject to} \quad & K \in S \end{aligned} \tag{7}$$

Here $\|\cdot\|$ is an arbitrary system norm on $\mathcal{L}(\mathcal{W}, \mathcal{Z})$, which can be chosen to correspond to an expected quadratic cost if desired, and S is a subspace of admissible controllers, which can be used to encapsulate the information structure of a decentralized / nonclassical problem.

Main QI result. S is said to be ***quadratically invariant (QI)*** with respect to G iff

$$KGK \in S \quad \forall\ K \in S. \tag{8}$$

The main result is then that if the information constraint is quadratically invariant, then the optimal control problem can be cast as a convex optimization problem.

Similar results have been obtained in this framework for different spaces. In [7], the linear spaces considered were arbitrary Banach spaces and thus the additional constraint that $(I - GK)$ be invertible was required. In [8], extended spaces (L_{2e}, ℓ_{2e}) were considered and thus the additional constraint that K stabilize P. For now, we will proceed assuming that the spaces are real vectors.

This setup is extremely general, and in many senses contains the previous framework as a special case. We now show how it is used for considering multiple subsystems or team members, each with access to different information.

Sparsity constraints. Suppose that there are n_y separate measurements, and n_u separate controller actions. Often, we will have $n_y = n_u = N$ as each measurement and control action corresponds to a given subsystem, but that does not have to be the case. We partition the sensor measurements and control actions as

$$y = \begin{bmatrix} y_1^T & \dots & y_{n_y}^T \end{bmatrix}^T \qquad u = \begin{bmatrix} u_1^T & \dots & u_{n_u}^T \end{bmatrix}^T$$

and then further partition G and K as

$$G = \begin{bmatrix} G_{11} & \dots & G_{1n_u} \\ \vdots & & \vdots \\ G_{n_y1} & \dots & G_{n_yn_u} \end{bmatrix} \qquad K = \begin{bmatrix} K_{11} & \dots & K_{1n_y} \\ \vdots & & \vdots \\ K_{n_u1} & \dots & K_{n_un_y} \end{bmatrix}$$

so that G_{ij} maps the jth control input u_j to the ith measurement y_i, and so that K_{kl} maps the lth measurement y_l to the kth control input u_k.

Now, if one wants to consider the case where subsystem controllers can access the measurements from some subsystems but not from others, this information constraint manifests itself as a sparsity constraint on the overall controller K to be designed.

We define a bit of notation that will allow this case to be dealt with systematically. Given a binary matrix $A^{\text{bin}} \in \{0,1\}^{m\times n}$, define the subspace

$$\begin{aligned} &\text{Sparse}(A^{\text{bin}}) \\ &\quad = \left\{ B \in \mathbb{R}^{m\times n} \mid B_{ij} = 0 \quad \forall\ i,j \ \text{ s.t. } \ A^{\text{bin}}_{ij} = 0 \right\} \end{aligned}$$

and conversely, given $B \in \mathbb{R}^{m\times n}$, let $A^{\text{bin}} = \text{Pattern}(B)$ be the binary matrix given by

$$A^{\text{bin}}_{ij} = \begin{cases} 0 & \text{if } B_{ij} = 0 \\ 1 & \text{otherwise} \end{cases}$$

In words, $\text{Sparse}(\cdot)$ gives the set of all matrices with the given sparsity pattern (represented with a binary matrix), and $\text{Pattern}(\cdot)$ gives the sparsity pattern (in the form of a binary matrix) of a given matrix.

Returning to the problem of interest, given an information structure, we can construct an associated binary matrix K^{bin}, where $K^{\text{bin}}_{kl} = 0$ indicates that controller k cannot access measurement l, and the problem of finding the optimal controller with that information structure becomes problem (7) with $S = \text{Sparse}(K^{\text{bin}})$. If we further let $G^{\text{bin}} = \text{Pattern}(G)$ give the sparsity structure of the plant, it was shown in [7, 8] that S is quadratically invariant under G iff $\forall\ i,l = 1,\dots,n_y$, and $\forall\ j,k = 1,\dots,n_u$,

$$K^{\text{bin}}_{ki}\ G^{\text{bin}}_{ij}\ K^{\text{bin}}_{jl}\ (1 - K^{\text{bin}}_{kl}) = 0. \tag{9}$$

Thus we can systematically test, with these $n_y^2 n_u^2$ equations, whether the information structure is quadratically invariant, and thus, whether the optimal (linear) control problem is convex.

3 The Relationship

We will now revisit the partially nested setup and results with an eye towards this framework. We will mostly focus on the information structure, but also note that the objective of a problem given in the form of [7] matches up with the objective function of (5) if the 2−norm is chosen and if

$$Q = 2P_{12}^T P_{12}, \qquad V = 2P_{12}^T P_{11} X^{-\frac{1}{2}}, \qquad c = 0.$$

Conversely, given a problem in the form of [2], this can be made to match up with the framework (7) if we consider the 2−norm and choose

$$P_{11} = \frac{1}{\sqrt{2}} Q^{-\frac{1}{2}} V X^{\frac{1}{2}}, \qquad P_{12} = \frac{1}{\sqrt{2}} Q^{\frac{1}{2}}, \tag{10}$$

$$P_{21} = H' X^{\frac{1}{2}}, \qquad G = D' \tag{11}$$

as well as $z_{\text{des}} = -\frac{1}{\sqrt{2}} Q^{-\frac{1}{2}} c$, where the objective is adjusted to keep $z - z_{\text{des}}$ as small as possible. While [7, 8] do not explicitly consider $z_{\rm des} \neq 0$, and this would of course effect the optimal controller, it does not affect the convexity of the optimal control problem, and the main results regarding quadratic invariance thus still hold.

Let us now construct a binary matrix $K^{\rm bin}$ to represent the information structure in the PN framework

$$K^{\rm bin}_{ki} = \begin{cases} 1 & \text{if } \gamma_k \text{ can access } y_i \\ 0 & \text{otherwise} \end{cases}$$

and a binary matrix $D^{\rm bin}$ to represent which measurements depend on which inputs

$$D^{\rm bin}_{ij} = \begin{cases} 0 & \text{if } D'_{ij} = 0 \\ 1 & \text{otherwise} \end{cases}$$

In addition to the implicit assumptions mentioned in Section 2.2, note also that the framework from that section also leads to the implicit assumption that $K^{\rm bin}_{ii} = 1 \ \forall \ i$.

We now state our main theorem, which shows that, at least for the framework where partially nested is explicitly defined, it is actually equivalent to quadratic invariance.

Theorem 1. *The information structure is partially nested iff* $\forall \ i,j,k,l \in \mathcal{I}$,

$$K^{bin}_{ki} \ D^{bin}_{ij} \ K^{bin}_{jl} \ (1 - K^{bin}_{kl}) = 0. \qquad (12)$$

Proof. Suppose first the the information structure is partially nested. Then for any k, i such that $K^{\rm bin}_{ki} = 1$, γ_k can access y_i. Thus what γ_k sees is affected by u_j for all j such that $K^{\rm bin}_{ki} = 1, D^{\rm bin}_{ij} = 1$, and so γ_k must have access to whatever γ_j has access to for all i,j,k such that $K^{\rm bin}_{ki} = 1, D^{\rm bin}_{ij} = 1$. It then follows that γ_k must have access to y_l for all i,j,k,l such that $K^{\rm bin}_{ki} = 1, D^{\rm bin}_{ij} = 1, K^{\rm bin}_{jl} = 1$, and thus we must have $K^{\rm bin}_{kl} = 1$ for all i,j,k,l such that $K^{\rm bin}_{ki} = 1, D^{\rm bin}_{ij} = 1, K^{\rm bin}_{jl} = 1$, and (12) follows.

Now suppose that the information structure is not partially nested. We will show by induction that (12) must fail for some i,j,k,l. Consider the statement P_κ: "If there exists a path of length κ such that $D^{\rm bin}_{i_1 i_2}, \ldots, D^{\rm bin}_{i_\kappa j} = 1$, $K^{\rm bin}_{jl} = 1$, and $K^{\rm bin}_{i_1 l} = 0$, then $\exists \ i,j,k,l$ such that (12) fails." If we can show that P_κ holds for all integers $k \geq 1$, we will be finished.

For $k = 1$, we have $K^{\rm bin}_{i_1 i_1} \ D^{\rm bin}_{i_1 j} \ K^{\rm bin}_{jl} \ (1 - K^{\rm bin}_{i_1 l}) = 1$, and P_1 holds. Now assume that $P_{\kappa-1}$ is true, consider a path of length κ, $D^{\rm bin}_{i_1 i_2} \cdots D^{\rm bin}_{i_\kappa j} = 1$, and consider the value $D^{\rm bin}_{i_2 i_3} \cdots D^{\rm bin}_{i_\kappa j} K^{\rm bin}_{jl} (1 - K^{\rm bin}_{i_2 l})$. If this is equal to 1, then P_κ follows from $P_{\kappa-1}$. If it is equal to 0, then $K^{\rm bin}_{i_2 l} = 1$, and thus $K^{\rm bin}_{i_1 i_1} D^{\rm bin}_{i_1 i_2} K^{\rm bin}_{i_2 l} (1 - K^{\rm bin}_{i_1 l}) = 1$, and P_κ follows. ∎

We have thus shown that for problems where partially nested is defined, it is equivalent to quadratic invariance! Thus the problems for which the optimal linear control problem is convex, that is, for which there are not multiple local minima over the linear class, are the same as those for which this class is complete.

4 Remaining Questions

4.1 LQG

With the exception of the constant c addressed earlier, the framework of quadratic invariance is much more broad, and we now discuss what can be said about some of the quadratically invariant problems for which partially nested is not defined.

First, the constraint that $K^{\rm bin}_{ii} \ \forall \ i$, implicit in PN but not in QI, is not actually restrictive. For any controller that we label γ_i, we can (re)label its information or a part of its information as y_i. Should we choose not to do that, we could still salvage the rest of the PN results after redefining *precedent* such that l is a precedent of k ($l \prec k$) iff there exist $i_1, \ldots, i_\kappa \in \mathcal{I}$ and $j_1, \ldots, j_\kappa \in \mathcal{I}$ such that $K^{\rm bin}_{k i_1}, D^{\rm bin}_{i_1 j_1}, K^{\rm bin}_{j_1 i_2}, D^{\rm bin}_{i_2 j_2}, \ldots, D^{\rm bin}_{i_\kappa j_\kappa}, K^{\rm bin}_{j_\kappa l} = 1$ which could then as before be interpreted as saying that l is a precedent of k iff the lth control input u_l can affect the kth measurement $\tilde{y}_k$.

Let us now turn our attention to the class of problems with cycles that PN does not address; that is, problems for which (3) is violated. The proofs of the main results in [2] indeed break down in the presence of such cycles. However, if such a cycle exists, and if the problem is QI, then through iterative application of (9) we can show that each controller in the cycle must be able to access all of the information in the cycle. We can then define

$$u_{\tilde{\kappa}} = \begin{bmatrix} u^T_{r_1} & \cdots & u^T_{r_\kappa} \end{bmatrix}^T \qquad y_{\tilde{\kappa}} = \begin{bmatrix} y^T_{r_1} & \cdots & y^T_{r_\kappa} \end{bmatrix}^T$$

and would then simply have $K^{\rm bin}_{\tilde{\kappa}\tilde{\kappa}} = 1$ and $K^{\rm bin}_{\tilde{\kappa} j} = 1$ for any j such that $K^{\rm bin}_{r_i j} = 1$ for some $i \in 1, \ldots, \kappa$; that is, the controller $\gamma_{\tilde{\kappa}}$ giving the new block variable $u_{\tilde{\kappa}}$ must be able to access all of the information which had been available to any members of the cycle. We would then of course also have $D_{\tilde{\kappa}\tilde{\kappa}} \neq 0$; that is, $D^{\rm bin}_{\tilde{\kappa}\tilde{\kappa}} = 1$, and so we see that the problem of extending the PN/QI relationship to cycles reduces to whether we can relax (4).

The proof of the main result of [2] indeed still breaks down when $D_{ii} \neq 0$ for some i. However once the problem is reduced just to these internal cycles, more standard centralized control techniques could be used, such as finding the optimal map $\tilde{\gamma}_i(\cdot)$ from the external inputs $H_i \xi + \sum_{j \prec i, j \neq i} D_{ij} u_j$ to u_i (which would be affine) and then recovering $\gamma_i(\cdot) = \tilde{\gamma}_i(I + D_{ii} \tilde{\gamma}_i)^{-1}$.

Formalizing the above argument, to show whether all problems with quadratically invariant sparsity constraints over real vector spaces have linear optimal controllers, even those for which the concept of partially nested is undefined, could be a topic of future work. Other quadratically invariant problems which could be studied to see whether there is a link to some generalization of partially nested structures or other proofs of linear optimality include control over other spaces besides real vectors, as well as other constraints besides sparsity constraints.

Lastly, we note that our main results regarding PN and QI depend only upon the structure of the plant and the information structure, and hold regardless of the parameters of the cost. When more specific costs are considered, problems can be found which violate these conditions but which have linear optimal controllers or for which finding optimal controllers is tractable. For instance, [1] shows that the Witsenhausen counterexample itself, with certain values of its constants, will have linear optimal controllers.

4.2 Other Cost Criteria

The relationship between information structure and linear optimality is not well-understood for other cost criteria besides LQG, but it seems that this connection between convexity and linear optimality is lost when one considers other criteria. Two results in particular make this point. In [9], the ℓ_1-"norm" is considered (that is, the cost induced by the $\infty-$norm), and an example was presented for which linear controllers are suboptimal, even though the information structure was centralized. Centralized problems are trivially PN/QI, and so this clearly represents a breakdown of the connection between convexity and linear optimality. In [6], the same problem setup as the Witsenhausen counterexample is considered, but for the minimization of the cost induced by the 2-norm, sometimes called uniformly optimal control, rather than the LQG criterion. It is shown that linear controllers are then optimal. As the Witsenhausen counterexample is of course not PN/QI, this represents the opposite type of breakdown of the connection between convexity and linear optimality.

5 Conclusion

We have shown that partially nested and quadratically invariant are actually equivalent conditions for problems where they are both well-defined. The former is a condition for which optimal controllers are linear, and the latter is a condition for which the optimal linear control problem is convex. Such a relationship was surmised after the Witsenhausen counterexample elucidated the difficulties of optimal decentralized control, both in terms of nonlinearity and nonconvexity, and we have shown for the fortieth anniversary that he had an incredible amount of foresight.

References

[1] R. Bansal and T. Basar. Stochastic teams with nonclassical information revisited: When is an affine law optimal? *IEEE Transactions on Automatic Control*, 32(6):554–559, June 1987.

[2] 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, January 1972.

[3] J.T. Lee, E. Lau, and Y-C. Ho. The Witsenhausen counterexample: A hierarchical search approach for nonconvex optimization problems. *IEEE Transactions on Automatic Control*, 46(3):382–397, March 2001.

[4] N. Li. *Learning approaches to the Witsenhausen counterexample from a view of potential games*. B.S. thesis, Zhejiang University, 2007.

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

[6] M. Rotkowitz. Linear controllers are uniformly optimal for the Witsenhausen counterexample. In *Proc. IEEE Conference on Decision and Control*, pages 553–558, 2006.

[7] M. Rotkowitz and S. Lall. Decentralized control information structures preserved under feedback. In *Proc. IEEE Conference on Decision and Control*, pages 569–575, 2002.

[8] M. Rotkowitz and S. Lall. A characterization of convex problems in decentralized control. *IEEE Transactions on Automatic Control*, 51(2):274–286, 2006.

[9] A. A. Stoorvogel. Nonlinear l_1 optimal controllers for linear systems. *IEEE Transactions on Automatic Control*, 40(4):694–696, 1995.

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