A Theory of Sufficient Statistics for Teams

Jeffrey Wu[1] and Sanjay Lall[2]

To appear, IEEE Conference on Decision and Control, 2014

Abstract

We introduce a theory of sufficient statistics for team decision problems. Starting from a rigorous definition of team sufficient statistics, we show that they contain all the information needed to make an optimal team decision. We also state general theorems of how to construct, manipulate, and update team sufficient statistics given the previous sufficient statistics and new measurements. Finally, we give the sufficient statistics for the class of partially nested team decision problems, and show that they have intuitive and compelling interpretations.

I Introduction

Team decision problems [2] are a fundamental type of decentralized control problem, where we have multiple players with different information that collaborate to minimize some expected cost. Such a problem consists of random variables $X, Y_1, \dots, Y_n$, where X is called the **state** and Y_i is called the **measurements** for player i. The goal of such problems is to choose functions $\gamma_1, \dots, \gamma_n$ (collectively called a **policy**) to minimize

$$\mathbb{E}\, c(X, \gamma_1(Y_1), \dots, \gamma_n(Y_n))$$

where c is some **cost function**. Team decision problems are harder than single-player problems because the policy variables are *coupled*, so that in general, any decision a player makes depends on the entire policies of the other players.

Although the general team decision problem is NP-hard [3], there are important cases when the problem is readily solvable. These include when $X, Y_1, \dots, Y_n$ are jointly Gaussian and c is convex quadratic [4], or when $X, Y_1, \dots, Y_n$ is finitely discrete, and c can be extended to a convex function. However, the real difficulty with team decision problems is the fact that we usually accumulate measurements over time. As we add measurements, the size of the policies increase, and the problem quickly becomes intractable even if the problem with the initial set of measurements was easy.

The burden of accumulated measurements can be alleviated by *sufficient statistics*. For the above team decision problem, let $(g_1(Y_1), \dots, g_n(Y_n))$ be functions of the measurements for each of the players. We say that $(g_1(Y_1), \dots, g_n(Y_n))$ is **sufficient for optimality** if given any policy $(\gamma_1, \dots, \gamma_n)$, there is another set of functions $\eta_1, \dots, \eta_n$ where

$$\begin{aligned}\mathbb{E}\, c(X, \eta_1(g_1(Y_1)), \dots, \eta_n(g_n(Y_n)))& \\ \leq \mathbb{E}\, c(X, \gamma_1(Y_1), \dots, \gamma_n(Y_n))&\end{aligned}$$

Thus we have converted a search for the optimal $(\gamma_1, \dots, \gamma_n)$ into a search for the optimal $(\eta_1, \dots, \eta_n)$. This conversion can greatly help if $(g_1(Y_1), \dots, g_n(Y_n))$ are smaller than the original measurements, either by taking a smaller number of discrete values, or by inhabiting a lower-dimensional space. The ideal situation is when the sufficient statistics are fixed in size, even as the measurement history grows. For a highly nontrivial example of this, see the discussion of the Gaussian partially nested system after Theorem 19.

Being sufficient for optimality is a necessary, but not a sufficient condition for being a sufficient statistic (which we will define rigorously later). The other important requirement is for a sufficient statistic to be **updatable**, i.e. given a sufficient statistic and some new measurements, we can compute an updated sufficient statistic that is just a function of the previous sufficient statistic and the new measurements. In this way, we do not have to scan through the entire measurement history every time we add new measurements.

Because of their crucial role in reducing the complexity of decentralized control problems, sufficient statistics are a focal point for research in decentralized stochastic control. Early examples include Hsu and Marcus [6] for the one-step delayed sharing pattern, which was extended to k steps of delay by Aicardi, et. al. [7]. More recently, Wu and Lall [8] gave sufficient statistics and a dynamic programming algorithm for a partially nested broadcast system, and Nayyar, Mahajan, and Teneketzis [9] have unified these cases under a common information framework.

The general approach used to derive the decentralized sufficient statistics is to use probabilistic intuition to guess the form of the sufficient statistic, and then check whether the guess is sufficient for optimality and updatable. The checking can be quite difficult, often involving a mess of complicated probabilistic expressions where it

[1] Jeffrey Wu is a senior research engineer at Polaris Wireless, 301 North Whisman Road, Mountain View, CA 94043 USA `jeffwu78@gmail.com`

[2] Sanjay Lall is a professor with the Department of Electrical Engineering, Stanford University, 350 Serra Mall, Stanford, CA 94305 `lall@stanford.edu`

is simultaneously easy to make a mistake while hard to check the work of others.

Another strikingly different approach to decentralized control uses sophisticated algebraic methods developed for linear control. Work that falls under this umbrella includes that of Swigart [10] for partially nested systems with state feedback, an alternate method for the same problem developed by Shah and Parrilo [11], and a solution to the two-player output feedback problem developed by Lessard and Lall [12]. Such approaches involve complicated algebraic expressions, and the role of sufficient statistics is to find meaningful subexpressions that simplify and give structure to the calculations. Finding the right subexpressions can be hard even for modest problems, and interpreting them can be just as challenging.

For the class of team decision problems, the contribution of this work is to replace the *de facto* definition of sufficiency (i.e. being both sufficient for optimality and updatable) with a rigorous one. From this rigorous definition flows a number of theorems, including the desired properties of being sufficient for optimality and updatable, and also theorems about how to explicitly construct and simplify the sufficient statistics. The explicit construction of the team sufficient statistics not only takes the guesswork out the process, but also automatically gives their probabilistic interpretation. Finally, we give the sufficient statistics for the class of partially nested team decision problems, and illustrate the results with a number of examples.

II Notation and Conventions

The support set of a real-valued function $p : \Omega \to \mathbb{R}$ is the set of $\omega \in \Omega$ where $p(\omega) \neq 0$. We denote the support set of p as $\operatorname{supp} p$. We say that p is a (finite) probability distribution if p is nonnegative, $\operatorname{supp} p$ is finite, and

$$\sum_{\omega \in \operatorname{supp} p} p(\omega) = 1$$

The set of all finite probability distributions over a sample space Ω is called a probability simplex, and denoted as $\Delta(\Omega)$.

To avoid the complications of measure theory, all the results in this paper have been proved in the following setting: We have a single underlying finite probability distribution $p \in \Delta(\Omega)$, where Ω is called the sample space, and functions over the sample space are called random variables. Given a random variable X, we use $p_X(x)$ to denote the probability that $X = x$, i.e.

$$p_X(x) = \sum_{\omega \in \operatorname{supp} p : X(\omega) = x} p(\omega)$$

Likewise, we use $p_{X|Y}(x|y)$ to denote the conditional probability of $X = x$ given $Y = y$, and assume it defaults to $p_X(x)$ if $p_Y(y) = 0$.

We use $X \equiv Y$ to denote that X equals Y with probability 1, and use $X \perp\!\!\!\perp Y \mid Z$ to denote that X and Y are conditionally independent given Z.

Although analogous results for continuous probability spaces may involve some nontrivial measure-theoretic concepts, we are confident that all our results have straightforward extensions to continuous case. Thus for the purpose of illustration, we will freely use examples involving Gaussian random variables.

III Team Decisions

Let $X, Y_1, \ldots, Y_n$ be random variables where we call X the state, and Y_i the measurements of player i. We say that $(U_1, \ldots, U_n)$ is a **randomized team decision** if there is a random variable W independent of $(X, Y_1, \ldots, Y_n)$ and functions $f_1, \ldots, f_n$ where

$$U_i = f_i(Y_i, W)$$

The common random variable W allows the players to coordinate their team decision with some randomness. At the same time, the independence of W from $(X, Y_1, \ldots, Y_n)$ ensures that the players do not gain any extra information from W. While allowing for this common randomization does not improve team decision performance (see Theorem 3), it will be very helpful in making a precise, straightforward definition of team sufficient statistics.

In order to be truly useful, however, we will need to extend the notion of randomization to include *hidden randomness*, i.e. cases where we cannot generate a randomized team decision on the current probability space, but *can* generate an equivalent randomized team decision on a separate probability space. To be precise, we say that the tuple of random variables $(U_1, \ldots, U_n)$ is a **(general) team decision** given the state X and measurements $(Y_1, \ldots, Y_n)$ if the joint distribution

$$p_{X, Y_1, \ldots, Y_n, U_1, \ldots, U_n}$$

is a convex combination of distributions of the form

$$p_{X, Y_1, \ldots, Y_n, f_1(Y_1), \ldots, f_n(Y_n)}$$

where $f_1, \ldots, f_n$ are functions. We denote the set of all such team decisions $(U_1, \ldots, U_n)$ as

$$\operatorname{TeamDec}(X|Y_1, \ldots, Y_n)$$

If X is constant, we omit the X and just use the notation

$$\operatorname{TeamDec}(Y_1, \ldots, Y_n)$$

The following theorem states that this definition is equivalent to being able to generate an randomized team decision on a separate probability space. We remark that throughout this paper, we will just sketch an outline of the proof if the details can easily be filled in by the reader. In any case, the full proofs can be found in [1].

Theorem 1. $(U_1,\dots,U_n) \in \mathrm{TeamDec}(X|Y_1,\dots,Y_n)$ iff there is a probability distribution q with associated random variables $\hat{X},\hat{Y}_1,\dots,\hat{Y}_n$ and $\hat{W}$ where

1) $(\hat{X},\hat{Y}_1,\dots,\hat{Y}_n)$ and $\hat{W}$ are independent.
2) $p_{X,Y_1,\dots,Y_n,U_1,\dots,U_n} = q_{\hat{X},\hat{Y}_1,\dots,\hat{Y}_n,f_1(\hat{Y}_1,W),\dots,f_n(\hat{Y}_n,W)}$ for some functions $f_1,\dots,f_n$

Proof. The main idea of the proof is that the coefficients of the convex combination correspond precisely to the probability distribution of the random variable $\hat{W}$.

In the single-player case, the definition for a general team decision reduces to conditional independence.

Theorem 2. $U \in \mathrm{TeamDec}(X|Y)$ iff $U \perp\!\!\!\perp X \mid Y$.

Proof. The idea of the proof is that the conditional distribution of U given Y can always be represented as a convex combination of conditional distributions of the form $p_{f(Y)|Y}$, see Theorem 2.3 of [1]. The rest follows by the elementary properties of conditional independence.

Because the randomization of a general team decision is independent of the state and measurements, allowing for general team decisions does not improve performance over deterministic decisions.

Theorem 3. Let f be a real-valued concave function on the probability simplex $\Delta(\mathcal{X}\times\mathcal{U}_1\times\cdots\times\mathcal{U}_n)$. Then given any general team decision

$$(U_1,\dots,U_n) \in \mathrm{TeamDec}(X|Y_1,\dots,Y_n)$$

where $\mathcal{X},\mathcal{U}_1,\dots,\mathcal{U}_n$ are the codomains of $X,U_1,\dots,U_n$, there are functions $\gamma_1,\dots,\gamma_n$ where

$$f(p_{X,\gamma_1(Y_1),\dots,\gamma_n(Y_n)}) \le f(p_{X,U_1,\dots,U_n})$$

Proof. By definition, $p_{X,Y_1,\dots,Y_n,U_1,\dots,U_n}$ is a convex combination of distributions of the form

$$p_{X,Y_1,\dots,Y_n,\gamma_1(Y_1),\dots,\gamma_n(Y_n)}$$

By marginalizing out $Y_1,\dots,Y_n$, it follows that $p_{X,U_1,\dots,U_n}$ is a convex combination of distributions of the form

$$p_{X,\gamma_1(Y_1),\dots,\gamma_n(Y_n)}$$

Thus by concavity of f, $f(p_{X,U_1,\dots,U_n})$ is greater than or equal to a convex combination of the form

$$f(p_{X,\gamma_1(Y_1),\dots,\gamma_n(Y_n)})$$

and therefore $f(p_{X,U_1,\dots,U_n})$ must be at least as great as one of these values, proving the theorem.

In particular, by defining the *linear* function f where

$$f(p_{X,U_1,\dots,U_n}) = \mathbb{E}\, c(X,U_1,\dots,U_n)$$

for some arbitrary cost function c, Theorem 3 shows that allowing for general team decisions does not reduce the expected cost of a standard team decision problem.

The following workhorse theorem shows when one team decision implies another.

Theorem 4 (Transformation of Team Decisions)**.** Suppose that

1) $(U_1,\dots,U_n) \in \mathrm{TeamDec}(X|Y_1,\dots,Y_n)$
2) $\hat{X},\hat{Y}_1,\dots,\hat{Y}_m$ are functions of $(X,Y_1,\dots,Y_n)$.
3) Each $\hat{U}_i$ is a function of $(\hat{Y}_i,U_{\mathcal{G}(i)})$. Here, $\mathcal{G}(i)$ is a subset of $\{1,\dots,n\}$ such that $Y_{\mathcal{G}(i)}$ is a function of $\hat{Y}_i$.

Then $(\hat{U}_1,\dots,\hat{U}_m) \in \mathrm{TeamDec}(\hat{X}|\hat{Y}_1,\dots,\hat{Y}_m)$.

Proof. By definition, the joint distribution

$$p_{X,Y_1,\dots,Y_n,U_1,\dots,U_n}$$

is a convex combination of distributions of the form

$$p_{X,Y_1,\dots,Y_n,f_1(Y_1),\dots,f_n(Y_n)}$$

By applying to both sides the *linear* transformation that transforms $p_{X,Y_1,\dots,Y_n,U_1,\dots,U_n}$ to $p_{\hat{X},\hat{Y}_1,\dots,\hat{Y}_n,\hat{U}_1,\dots,\hat{U}_n}$, it follows that

$$p_{\hat{X},\hat{Y}_1,\dots,\hat{Y}_n,\hat{U}_1,\dots,\hat{U}_n}$$

is a convex combination of distributions of the form

$$p_{\hat{X},\hat{Y}_1,\dots,\hat{Y}_n,h_1(\hat{Y}_1),\dots,h_n(\hat{Y}_n)}$$

for some functions $h_1,\dots,h_n$.

The next theorem clarifies the role of the state in a team decision. It says that given an existing team decision, the variable X can serve as its state iff X is conditionally independent from the team decision given the measurements.

Theorem 5 (State Decomposition of Team Decisions)**.**

$$\begin{gathered}(U_1,\dots,U_n) \in \mathrm{TeamDec}(X|Y_1,\dots,Y_n)\\ \Updownarrow \\ (U_1,\dots,U_n) \perp\!\!\!\perp X \mid (Y_1,\dots,Y_n)\\ (U_1,\dots,U_n) \in \mathrm{TeamDec}(Y_1,\dots,Y_n)\end{gathered}$$

Proof. The $\Downarrow$ direction follows directly from Theorems 4 and 2. To show the $\Uparrow$ direction, note that the conditional independence implies

$$p_{X,Y_1,\dots,Y_n,U_1,\dots,U_n}(x,y_1,\dots,y_n,\cdot) = p_{X,Y_1,\dots,Y_n}(x,y_1,\dots,y_n)p_{U_1,\dots,U_n|Y_1,\dots,Y_n}(\cdot|y_1,\dots,y_n)$$

But since $(U_1,\dots,U_n) \in \mathrm{TeamDec}(Y_1,\dots,Y_n)$, then it follows that $p_{Y_1,\dots,Y_n,U_1,\dots,U_n}$ is a convex combination of distributions of the form $p_{Y_1,\dots,Y_n,f_1(Y_1),\dots,f_n(Y_n)}$. By conditioning this convex combination on $(Y_1,\dots,Y_n)$ and substituting in the above equation, it follows that $p_{X,Y_1,\dots,Y_n,U_1,\dots,U_n}$ is convex combination of distributions of the form

$$p_{X,Y_1,\dots,Y_n,f_1(Y_1),\dots,f_n(Y_n)}$$

proving the result.

We also show an intuitive rule that says if we make an initial team decision, and then make a subsequent team decision based on the measurements and the previous team decision, then the latter is also a team decision based only on the original measurements.

Theorem 6 (Chain Rule for Team Decisions)**.** If

$$(U_1, \ldots, U_n) \in \mathrm{TeamDec}(X|Y_1, \ldots, Y_n)$$
$$(V_1, \ldots, V_n) \in \mathrm{TeamDec}(X|(Y_1, U_1), \ldots, (Y_n, U_n))$$

then $(V_1, \ldots, V_n) \in \mathrm{TeamDec}(X|Y_1, \ldots, Y_n)$.

Proof. Note that $p_{X,Y_1,\ldots,Y_n,(U_1,V_1),\ldots,(U_n,V_n)}$ is a convex combination of distributions of the form

$$p_{X,Y_1,\ldots,Y_n,(U_1,f_1(Y_1,U_1)),\ldots,(U_n,f_n(Y_n,U_n))} \tag{1}$$

But by Theorem 4,

$$((U_1, f_1(Y_1, U_1)), \ldots, (U_n, f_n(Y_n, U_n))) \in \mathrm{TeamDec}(X|Y_1, \ldots, Y_n)$$

so each distribution in (1) is in turn a convex combination of distributions of the form

$$p_{X,Y_1,\ldots,Y_n,h_1(Y_1),\ldots,h_n(Y_n)}$$

so $((U_1, V_1), \ldots, (U_n, V_n)) \in \mathrm{TeamDec}(X|Y_1, \ldots, Y_n)$. Applying Theorem 4 again yields the result.

A consequence of the chain rule shows how to combine team decisions from two groups of players into a single team decision.

Theorem 7 (Combining Team Decisions)**.** If

$$(U_1, \ldots, U_k) \in \mathrm{TeamDec}(X, Y_{k+1}, \ldots, Y_n|Y_1, \ldots, Y_k)$$
$$(U_{k+1}, \ldots, U_n) \in \mathrm{TeamDec}(X, Y_1, \ldots, Y_k, U_1, \ldots, U_k| Y_{k+1}, \ldots, Y_n)$$

then $(U_1, \ldots, U_n) \in \mathrm{TeamDec}(X|Y_1, \ldots, Y_n)$.

Proof. Applying Theorem 4 to the two team decisions, we have

$$(U_1, \ldots, U_k, Y_{k+1}, \ldots, Y_n) \in \mathrm{TeamDec}(X|Y_1, \ldots, Y_n)$$

and

$$(U_1, \ldots, U_n) \in \mathrm{TeamDec}(X| (Y_1, U_1), \ldots, (Y_n, U_k), Y_{k+1}, \ldots, Y_n)$$

Thus $(U_1, \ldots, U_n) \in \mathrm{TeamDec}(X|Y_1, \ldots, Y_n)$ by Theorem 6.

IV Team Sufficiency

Let X be the state and $(Y_1, \ldots, Y_n)$ be measurements for n players. We say that the derived random variables $(g_1(Y_1), \ldots, g_n(Y_n))$ are a **team sufficient statistic** if every team decision based on the full measurements $(Y_1, \ldots, Y_n)$ is also a team decision on the derived variables $(g_1(Y_1), \ldots, g_n(Y_n))$, i.e.

$$\mathrm{TeamDec}(X|Y_1, \ldots, Y_n) \subseteq \mathrm{TeamDec}(X|g_1(Y_1), \ldots, g_n(Y_n))$$

We denote the set of all such team sufficient statistics as $\mathrm{TeamSuff}(X|Y_1, \ldots, Y_n)$. If X is constant, then we will just express this set as $\mathrm{TeamSuff}(Y_1, \ldots, Y_n)$.

What makes this intuitive-sounding definition work is the careful and very broad definition of team decisions that allow for hidden randomization in its policies. We now show that our definition of team sufficiency implies being sufficient for optimality.

Theorem 8. Let f be a real-valued concave function on the probability simplex $\Delta(\mathcal{X} \times \mathcal{U}_1 \times \cdots \times \mathcal{U}_n)$, and

$$(g_1(Y_1), \ldots, g_n(Y_n)) \in \mathrm{TeamSuff}(X|Y_1, \ldots, Y_n)$$

where $\mathcal{X}, \mathcal{Y}_1, \ldots, \mathcal{Y}_n$ are the codomains of $X, Y_1, \ldots, Y_n$. Then given any functions $\gamma_1 : \mathcal{Y}_1 \to \mathcal{U}_1, \ldots, \gamma_n : \mathcal{Y}_n \to \mathcal{U}_n$, there are functions $\eta_1, \ldots, \eta_n$ where

$$f(p_{X,\eta_1(g_1(Y_1)),\ldots,\eta_n(g_n(Y_n))}) \leq f(p_{X,\gamma_1(Y_1),\ldots,\gamma_n(Y_n)})$$

Proof. By definition of team sufficiency,

$$(\gamma_1(Y_1), \ldots, \gamma_n(Y_n)) \in \mathrm{TeamDec}(X|Y_1, \ldots, Y_n) \subseteq \mathrm{TeamDec}(X|g_1(Y_1), \ldots, g_n(Y_n))$$

The result then follows immediately from Theorem 3.

In particular, by defining the linear function f where

$$f(p_{X,U_1,\ldots,U_n}) = \mathbb{E}\, c(X, U_1, \ldots, U_n)$$

for some arbitrary cost function c, we have shown that team sufficiency implies being sufficient for optimality of a team decision problem.

We now show that basic equivalence which states that a set of reduced measurements is sufficient iff the full measurements can be modeled as a team decision using these reduced measurements.

Theorem 9 (Sufficiency-Decision Equivalence)**.**

$$(g(Y_1), \ldots, g_n(Y_n)) \in \mathrm{TeamSuff}(X|Y_1, \ldots, Y_n)$$
$$\Updownarrow$$
$$(Y_1, \ldots, Y_n) \in \mathrm{TeamDec}(X|g_1(Y_1), \ldots, g_n(Y_n))$$

Proof. Suppose that

$$(g_1(Y_1), \ldots, g_n(Y_n)) \in \mathrm{TeamSuff}(X|Y_1, \ldots, Y_n)$$

Then

$$(Y_1, \ldots, Y_n) \in \mathrm{TeamDec}(X|Y_1, \ldots, Y_n) \subseteq \mathrm{TeamDec}(X|g_1(Y_1), \ldots, g_n(Y_n))$$

Conversely, suppose that

$$(Y_1, \dots, Y_n) \in \mathrm{TeamDec}(X | g_1(Y_1), \dots, g_n(Y_n))$$

Then for any team decision

$$\begin{aligned}(U_1, \dots, U_n) &\in \mathrm{TeamDec}(X | Y_1, \dots, Y_n) \\ &= \mathrm{TeamDec}(X | (g_1(Y_1), Y_1), \dots, \\ &\qquad (g_n(Y_n), Y_n))\end{aligned}$$

we have by Theorem 6,

$$(U_1, \dots, U_n) \in \mathrm{TeamDec}(X | g_1(Y_1), \dots, g_n(Y_n))$$

so $(g_1(Y_1), \dots, g_n(Y_n)) \in \mathrm{TeamSuff}(X | Y_1, \dots, Y_n)$.

The above equivalence is important because it allows us to use theorems about team decisions in order to prove results about team sufficiency. The following theorem is a example of such results. To keep notation simple, in the single-player case, we abbreviate TeamSuff to just Suff.

Theorem 10 (Sufficiency State Decomposition)**.**

$$\begin{aligned}&\mathrm{TeamSuff}(X | Y_1, \dots, Y_n) \\ &\quad = \mathrm{Suff}(X | Y_1, \dots, Y_n) \cap \mathrm{TeamSuff}(Y_1, \dots, Y_n)\end{aligned}$$

Proof. Convert to team decision membership using Theorem 9, apply the state decomposition of Theorem 5, and convert back to a sufficiency membership with Theorem 9.

We now show that in the single-player case, team sufficient is equivalent to being able to generate the posterior distribution—which is the same as the usual definition of sufficient statistic in a Bayesian setting.

Theorem 11 (Posterior Distribution Characterization)**.** $g(Y) \in \mathrm{Suff}(X|Y)$ iff there is a function f where

$$f(g(Y)) \equiv p_{X|Y}(\cdot | Y)$$

One such function f is $f(z) = p_{X|g(Y)}(\cdot | z)$.

Proof. Note that by Theorems 9 and 2, $g(Y) \in \mathrm{Suff}(X|Y)$ iff

$$X \perp\!\!\!\perp Y \mid g(Y)$$

Using fundamental properties of conditional independence, $X \perp\!\!\!\perp Y \mid g(Y)$ implies that

$$p_{X|Y}(\cdot | Y) \equiv p_{X|g(Y)}(\cdot | g(Y))$$

Conversely, if there is some function f where

$$f(g(Y)) \equiv p_{X|Y}(\cdot | Y)$$

then this implies (again using the fundamental properties of conditional independence) that $X \perp\!\!\!\perp Y \mid g(Y)$.

Theorem 11 gives a constructive method to generate single-player sufficient statistics. Section VI will show how to construct team sufficient statistics when there is more than one player.

Example 1 (Gaussian Sufficient Statistics)**.** Suppose (X, Y) is jointly Gaussian with a mean and nonsingular covariance

$$\begin{bmatrix}\mu_X \\ \mu_Y\end{bmatrix}, \quad \begin{bmatrix}\Sigma_X & \Sigma_{XY} \\ \Sigma_{YX} & \Sigma_Y\end{bmatrix}$$

Then the conditional density $p_{X|Y}(\cdot|y)$ is Gaussian with mean and covariance

$$\begin{aligned}\hat{\mu}(y) &= \mu_X + \Sigma_{XY}\Sigma_Y^{-1}(y - \mu_Y) \\ \hat{\Sigma} &= \Sigma_X - \Sigma_{XY}\Sigma_Y^{-1}\Sigma_{YX}\end{aligned}$$

Since the conditional expectation

$$\hat{\mu}(Y) = \mu_X + \Sigma_{XY}\Sigma_Y^{-1}(Y - \mu_Y) = \mathbb{E}(X|Y)$$

can generate $p_{X|Y}(\cdot|Y)$, then it follows that $\mathbb{E}(X|Y) \in \mathrm{Suff}(X|Y)$.

The next result shows how to simplify expressions of single-player sufficiency given a conditional independence.

Theorem 12. If $X \perp\!\!\!\perp Z \mid Y$, then

$$\begin{aligned}\mathrm{Suff}(X|Y, Z) &\supseteq \mathrm{Suff}(X|Y) \\ \mathrm{Suff}(X, Z|Y) &= \mathrm{Suff}(X|Y) \cap \mathrm{Suff}(Z|Y) \\ \mathrm{Suff}(X, Y|Z) &= \mathrm{Suff}(Y|Z)\end{aligned}$$

Proof. Follows from Theorem 11 and the fact that

$$\begin{aligned}p_{X|Y,Z}(\cdot|y, z) &= p_{X|Y}(\cdot|y), \quad \forall y \in \operatorname{supp} p_{Y,Z} \\ p_{X,Z|Y}(x, z|y) &= p_{X|Y}(x|y)p_{Z|Y}(z|y), \quad \forall y \in \operatorname{supp} p_Y \\ p_{X,Y|Z}(x, y|z) &= p_{X|Y}(x|y)p_{Y|Z}(y|z), \quad \forall z \in \operatorname{supp} p_Z\end{aligned}$$

each of which follow from $X \perp\!\!\!\perp Z \mid Y$.

V Updating

The problem of updating involves modifying an existing sufficient statistic under changes of the state or addition of measurements, without having to scan through the full measurement history. We will build to the full updating theorem by some simpler results that are useful in their own right.

The first result shows that sufficiency is preserved if we modify the state with our current sufficient statistic.

Theorem 13 (Left Expansion)**.** If

$$(g_1(Y_1), \dots, g_n(Y_n)) \in \mathrm{TeamSuff}(X | Y_1, \dots, Y_n)$$

then for any function f,

$$\begin{aligned}&(g_1(Y_1), \dots, g_n(Y_n)) \\ &\quad \in \mathrm{TeamSuff}(f(X, g_1(Y_1), \dots, g_n(Y_n)) | Y_1, \dots, Y_n)\end{aligned}$$

Proof. By Theorem 9,

$$(Y_1, \ldots, Y_n) \in \mathrm{TeamDec}(X|g_1(Y_1), \ldots, g_n(Y_n))$$

and therefore by Theorem 4,

$$(Y_1, \ldots, Y_n) \in \mathrm{TeamDec}(f(X, g_1(Y_1), \ldots, g_n(Y_n))| \\ g_1(Y_1), \ldots, g_n(Y_n))$$

and the result follows by Theorem 9.

The next result shows when we can remove parts of the state without affecting the set of team sufficient statistics.

Theorem 14 (Left Contraction)**.** If

$$X' \perp\!\!\!\perp (Y_1, \ldots, Y_n) \mid X$$

then

$$\mathrm{TeamSuff}(X', X|Y_1, \ldots, Y_n) = \mathrm{TeamSuff}(X|Y_1, \ldots, Y_n)$$

Proof. By Theorems 10 and 12,

$$\begin{aligned} &\mathrm{TeamSuff}(X', X|Y_1, \ldots, Y_n) \\ &\quad = \mathrm{Suff}(X', X|Y_1, \ldots, Y_n) \cap \mathrm{TeamSuff}(Y_1, \ldots, Y_n) \\ &\quad = \mathrm{Suff}(X|Y_1, \ldots, Y_n) \cap \mathrm{TeamSuff}(Y_1, \ldots, Y_n) \\ &\quad = \mathrm{TeamSuff}(X|Y_1, \ldots, Y_n) \end{aligned}$$

The third result shows that if we have an existing team sufficient statistic, and construct a new team sufficient statistic using the existing sufficient statistics as measurements, then this new team statistic is actually sufficient with respect to the full measurements.

Theorem 15 (Right Contraction)**.** If

$$(g_1(Y_1), \ldots, g_n(Y_n)) \in \mathrm{TeamSuff}(X|Y_1, \ldots, Y_n)$$

then

$$\mathrm{TeamSuff}(X|g_1(Y_1), \ldots, g_n(Y_n)) \\ \subseteq \mathrm{TeamSuff}(X|Y_1, \ldots, Y_n)$$

Proof. Let

$$(h_1(g_1(Y_1)), \ldots, h_n(g_n(Y_n))) \\ \in \mathrm{TeamSuff}(X|g_1(Y_1), \ldots, g_n(Y_n))$$

Then by definition of team sufficiency,

$$\begin{aligned} &\mathrm{TeamDec}(X|Y_1, \ldots, Y_n) \\ &\subseteq \mathrm{TeamDec}(X|g_1(Y_1), \ldots, g_n(Y_n)) \\ &\subseteq \mathrm{TeamDec}(X|h_1(g_1(Y_1)), \ldots, h_n(g_n(Y_n))) \end{aligned}$$

and the result follows.

We now show the full updating theorem, which shows how to update a team sufficient statistic given very broad conditions about how the new states and measurements evolve.

Theorem 16 (Updating)**.** If

$$\begin{aligned} &(g_1(Y_1), \ldots, g_n(Y_n)) \in \mathrm{TeamSuff}(X|Y_1, \ldots, Y_n) \\ &\quad (X', Y_1', \ldots, Y_n') \perp\!\!\!\perp (Y_1, \ldots, Y_n) \\ &\qquad\qquad \mid (X, g_1(Y_1), \ldots, g_n(Y_n)) \end{aligned}$$

then

$$\mathrm{TeamSuff}(X'|(g_1(Y_1), Y_1'), \ldots, (g_n(Y_n), Y_n')) \\ \subseteq \mathrm{TeamSuff}(X'|(Y_1, Y_1'), \ldots, (Y_1, Y_n'))$$

Proof. By Theorems 13 and 14,

$$\begin{aligned} &(g_1(Y_1), \ldots, g_n(Y_n)) \\ &\quad \subseteq \mathrm{TeamSuff}(X, g_1(Y_1), \ldots, g_n(Y_n)|Y_1, \ldots, Y_n) \\ &\quad = \mathrm{TeamSuff}(X', Y_1', \ldots, Y_n', X, g_1(Y_1), \ldots, g_n(Y_n)| \\ &\qquad Y_1, \ldots, Y_n) \\ &\quad \subseteq \mathrm{TeamSuff}(X', Y_1', \ldots, Y_n'|Y_1, \ldots, Y_n) \end{aligned}$$

By Theorem 9,

$$(Y_1, \ldots, Y_n) \in \mathrm{TeamDec}(X', Y_1', \ldots, Y_n'| \\ g_1(Y_1), \ldots, g_n(Y_n))$$

and therefore by Theorem 4,

$$((Y_1, Y_1'), \ldots, (Y_n, Y_n')) \in \mathrm{TeamDec}(X'| \\ (g_1(Y_1), Y_1'), \ldots, (g_n(Y_n), Y_n'))$$

and the result follows from Theorems 9 and 15.

The conditional independence in the theorem states that the new state and measurements $(X', Y_1', \ldots, Y_n')$ can be modeled as a randomized function of the previous state X and sufficient statistics $(g_1(Y_1), \ldots, g_n(Y_n))$ (which presumably was used to form a team decision at that time). The theorem then states that we can form an updated team sufficient statistic for the full measurement history by simply computing a team sufficient statisic using X' as the state and $(g_1(Y_1), Y_1'), \ldots, (g_n(Y_n), Y_n')$ as the measurements.

VI Elimination Theorems

We now tackle the problem of how to construct team sufficient statistics for more than one player. The main idea is akin to Gaussian elimination, where we choose team sufficient statistics for one group of players, and then choose team sufficient statistics for the remaining players that depend on our choice of sufficient statistics for the first group. The simplest such result is given below.

Theorem 17 (Simple Elimination)**.** If

$$\begin{aligned} (S_{k+1}, \ldots, S_n) &\in \mathrm{TeamSuff}(X, Y_1, \ldots, Y_k|Y_{k+1}, \ldots, Y_n) \\ (S_1, \ldots, S_k) &\in \mathrm{TeamSuff}(X, S_{k+1}, \ldots, S_n|Y_1, \ldots, Y_k) \end{aligned}$$

then $(S_1, \ldots, S_n) \in \mathrm{TeamSuff}(X|Y_1, \ldots, Y_n)$.

Proof. By Theorem 9, the conditions of the theorem can be expressed as

$$\begin{aligned}(Y_{k+1},\dots,Y_n) &\in \mathrm{TeamDec}(X,Y_1,\dots,Y_k|S_{k+1},\dots,S_n)\\ (Y_1,\dots,Y_k) &\in \mathrm{TeamDec}(X,S_{k+1},\dots,S_n|S_1,\dots,S_k)\end{aligned}$$

Since each S_i is just a function of Y_i, it follows by Theorem 7 that

$$(Y_1,\dots,Y_n) \in \mathrm{TeamDec}(X|S_1,\dots,S_n)$$

and the result follows by Theorem 9.

Example 2 (General Sufficient Statistics)**.** By inductively applying Theorem 17, it is easy to show that if

$$S_i \in \mathrm{Suff}(X,Y_1,\dots,Y_{i-1},S_{i+1},\dots,S_n|Y_i)$$

then $(S_1,\dots,S_n) \in \mathrm{TeamSuff}(X|Y_1,\dots,Y_n)$. A special case of these conditions (via Theorem 13) is when

$$S_i \in \mathrm{Suff}(X,Y_1,\dots,Y_{i-1},Y_{i+1},\dots,Y_n|Y_i)$$

Example 3 (Conditionally Independent Measurements)**.** Suppose the measurements $Y_1,\dots,Y_n$ are conditionally independent given the state X. Then by the Theorem 14,

$$\mathrm{Suff}(X,Y_1,\dots,Y_{i-1},Y_{i+1},\dots,Y_n|Y_i) = \mathrm{Suff}(X|Y_i)$$

It follows by Example 2 that if

$$S_i \in \mathrm{Suff}(X|Y_i)$$

then $(S_1,\dots,S_n) \in \mathrm{TeamSuff}(X|Y_1,\dots,Y_n)$.

The simple elimination theorem has trouble if the measurements are nested. For example, consider a two-player system where Y_1 can be derived from Y_2. By Example 2, if

$$\begin{aligned}S_2 &\in \mathrm{Suff}(X,Y_1|Y_2)\\ S_1 &\in \mathrm{Suff}(X,S_2|Y_1)\end{aligned}$$

then $(S_1,S_2) \in \mathrm{TeamSuff}(X|Y_1,Y_2)$. But because Y_2 contains all of Y_1, then this implies that S_2 (and also S_1) contains all of Y_1. Since the full measurements Y_1 usually grows with time, then we conclude that Theorem 17 is too weak to give useful sufficient statistics in this situation.

The next theorem fixes this problem by allowing a player to just keep the *sufficient statistics* of the other players' measurements that it contains.

Theorem 18 (General Elimination)**.** Suppose that

$$\begin{aligned}&((Y_{\mathcal{G}(k+1)},S_{k+1}),\dots,(Y_{\mathcal{G}(n)},S_n))\\ &\qquad \in \mathrm{TeamSuff}(X,Y_1,\dots,Y_k|Y_{k+1},\dots,Y_n)\\ &(S_1,\dots,S_k)\\ &\qquad \in \mathrm{TeamSuff}(X,S_{k+1},\dots,S_n|Y_1,\dots,Y_k)\end{aligned}$$

where each $\mathcal{G}(i)$ is a subset of $\{1,\dots,k\}$ such that $Y_{\mathcal{G}(i)}$ is a function of Y_i. Then

$$\begin{aligned}&(S_1,\dots,S_k,(S_{\mathcal{G}(k+1)},S_{k+1}),\dots,(S_{\mathcal{G}(n)},S_n))\\ &\qquad \in \mathrm{TeamSuff}(X|Y_1,\dots,Y_n)\end{aligned}$$

Proof. By directly applying Theorem 17, we have

$$\begin{aligned}&(Y_1,\dots,Y_k,(Y_{\mathcal{G}(k+1)},S_{k+1}),\dots,(Y_{\mathcal{G}(n)},S_n))\\ &\qquad \in \mathrm{TeamSuff}(X|Y_1,\dots,Y_n)\end{aligned}$$

Moreover, by Theorem 9,

$$(Y_1,\dots,Y_k) \in \mathrm{TeamDec}(X,S_{k+1},\dots,S_n|S_1,\dots,S_k)$$

It follows by Theorem 4 that

$$\begin{aligned}&(Y_1,\dots,Y_k,(Y_{\mathcal{G}(k+1)},S_{k+1}),\dots,(Y_{\mathcal{G}(n)},S_n))\\ &\qquad \in \mathrm{TeamDec}(X|S_1,\dots,S_k,\\ &\qquad\qquad (S_{\mathcal{G}(k+1)},S_{k+1}),\dots,(S_{\mathcal{G}(n)},S_n))\end{aligned}$$

so by Theorems 9 and 15,

$$\begin{aligned}&(S_1,\dots,S_k,(S_{\mathcal{G}(k+1)},S_{k+1}),\dots,(S_{\mathcal{G}(n)},S_n))\\ &\quad \in \mathrm{TeamSuff}(X|Y_1,\dots,Y_k,\\ &\qquad (Y_{\mathcal{G}(k+1)},S_{k+1}),\dots,(Y_{\mathcal{G}(n)},S_n))\\ &\quad \subseteq \mathrm{TeamSuff}(X|Y_1,\dots,Y_n)\end{aligned}$$

Example 4 (Common Measurements)**.** Suppose X is the state and $(Y_0,Y_1),\dots,(Y_0,Y_n)$ are the measurements of the players. To find team sufficient statistics, we create an extra player who only sees the common measurements Y_0. Now obviously,

$$\begin{aligned}&((Y_0,Y_1),\dots,(Y_0,Y_n))\\ &\qquad \in \mathrm{TeamSuff}(X,Y_0|(Y_0,Y_1),\dots,(Y_0,Y_n))\end{aligned}$$

and suppose we choose

$$S_0 \in \mathrm{Suff}(X,Y_1,\dots,Y_n|Y_0)$$

Then by Theorem 18,

$$\begin{aligned}&(S_0,(S_0,Y_1),\dots,(S_0,Y_n))\\ &\qquad \in \mathrm{TeamSuff}(X|Y_0,(Y_0,Y_1),\dots,(Y_0,Y_n))\end{aligned}$$

By removing the extra player (proof: use Theorem 9 to convert to a team decision, use Theorem 4 to remove the player, use Theorem 9 to convert back), we have

$$\begin{aligned}&((S_0,Y_1),\dots,(S_0,Y_n))\\ &\qquad \in \mathrm{TeamSuff}(X|(Y_0,Y_1),\dots,(Y_0,Y_n))\end{aligned}$$

Thus each player i keeps a sufficient statistic of $(X, Y_1,\dots,Y_n)$ given the common measurements Y_0, plus its own private measurements Y_i.

VII Partially Nested Problems

Partially nested team decision problems organize the dependencies between the players as a directed acyclic graph, with the children receiving the measurements of their parents. They were introduced by Ho and Chu [5] in the context of LQG team decision problems. In this section, we define a generalization of a partially nested problem, and find their team sufficient statistics.

In order to define this generalization, we first establish notation for directed acyclic graphs and Bayesian networks. A directed graph $\mathcal{G}$ is a function mapping $\{1,\ldots,n\}$ to subsets of $\{1,\ldots,n\}$. We call the set $\mathcal{G}(i)$ the parents of the node i. We say that (i,j) an an edge of $\mathcal{G}$ if $i \in \mathcal{G}(j)$, and $(i_0,\ldots,i_m)$ is a path of $\mathcal{G}$ of length m if (i_{k-1},i_k) is an edge for each $k \in \{1,\ldots,m\}$.

We say that a node i is an ancestor of j if there is a path of nonzero length from i to j, and let $\mathcal{G}_*(j)$ denote the set of ancestors of j. The function $\mathcal{G}_*$ as a whole is another directed graph, which we call the transitive closure of $\mathcal{G}$. For convenience, we also define the function $\mathcal{G}^*$, where $\mathcal{G}^*(i) = \mathcal{G}_*(i) \cup \{i\}$.

A directed graph $\mathcal{G}$ is acyclic if no node is an ancestor of itself. If $\mathcal{G}$ is acyclic, we also define its transitive reduction $\mathcal{G}_0$, where $\mathcal{G}_0(j)$ is the set of nodes i whose only path to j is the edge (i,j). For this reason, we will call $\mathcal{G}_0(j)$ the **immediate parents** of j. One can show that $\mathcal{G}_0$ is the smallest graph whose transitive closure is $\mathcal{G}_*$.

We say that the tuple $X = (X_1,\ldots,X_n)$ factors according to the directed acyclic graph $\mathcal{G}$ if

$$p_X(x) = \prod_{i=1}^{n} p_{X_i|X_{\mathcal{G}(i)}}(x_i|x_{\mathcal{G}(i)})$$

We now define the desired generalization to a partially nested system. We say that the tuple of random variable pairs $((X_1,Y_1),\ldots,(X_n,Y_n))$ is **partially nested** according the directed acyclic graph $\mathcal{G}$ if

1) $((X_1,Y_1),\ldots,(X_n,Y_n))$ factors according to $\mathcal{G}$.
2) For each ancestor $i \in \mathcal{G}_*(j)$, Y_i is derived from Y_j.

Intuitively speaking, partially nested systems can model situations where if player i affects player j, then j receives all the measurements of i.

The next theorem gives the team sufficient statistics for a partially nested system.

Theorem 19 (Partially Nested Systems)**.**
Let $((X_1,Y_1),\ldots,(X_n,Y_n))$ be partially nested according to $\mathcal{G}$, and suppose that for each node i,

$$S_i \in \mathrm{Suff}(X_{\mathcal{G}^*(i)}|Y_i) \tag{2}$$
$$S_{\mathcal{G}_0(i)} \in \mathrm{Suff}(S_i|Y_{\mathcal{G}_0(i)}) \tag{3}$$

Then

1) $(S_{\mathcal{G}^*(1)},\ldots,S_{\mathcal{G}^*(n)}) \in \mathrm{TeamSuff}(X|Y_1,\ldots,Y_n)$.
2) $((X_1,S_{\mathcal{G}^*(1)}),\ldots,(X_n,S_{\mathcal{G}^*(n)}))$ is partially nested according to $\mathcal{G}_*$.

Proof. The proof is omitted here for space reasons; a detailed line-by-line proof can be found in Theorems 3.9 and 3.10 of [1].

This is a significant theorem that needs a bit of explanation. The S_i variables represent player i's **contribution** to the team sufficient statistic. To form the team sufficient statistic, each player needs to keep track of its own contribution as well as those of its ancestors. Moreover, if we replace the measurements with the team sufficient statistics, the system remains partially nested according to the transitive closure $\mathcal{G}_*$. This last conclusion is crucial for updating: for details see [1].

Equation (2) says that each S_i needs to contain a sufficient statistic of its own state and its ancestors' states given its measurements Y_i. For Gaussian partially nested systems, this corresponds to the conditional expectation $\mathbb{E}(X_{\mathcal{G}^*(i)}|Y_i)$.

Equation (3) considers the immediate parents $\mathcal{G}_0(i)$ of a node i. Each of these parents will need to *split* among them a sufficient statistic of S_i given their joint measurements $Y_{\mathcal{G}_0(i)}$. It does not matter how this splitting is done; all that matters is that their joint contributions $S_{\mathcal{G}_0(i)}$ can generate this sufficient statistic.

In the Gaussian case, this splitting of this sufficient statistic can occur as follows: If we assume that S_i is an affine function of the measurements Y_i (which certainly is the case when i is a leaf node and $S_i = \mathbb{E}(X_{\mathcal{G}^*(i)}|Y_i)$), then the sufficient statistic we need to split is

$$\mathbb{E}(S_i|Y_{\mathcal{G}_0(i)})$$

Since this is just an affine function of $Y_{\mathcal{G}_0(i)}$, then

$$\mathbb{E}(S_i|Y_{\mathcal{G}_0(i)}) = \mu_i + \sum_{j\in\mathcal{G}_0(i)} A_j(Y_j - \mu_j)$$

for some vectors μ_j and matrices A_j. Thus giving each immediate parent j the vector $A_j(Y_j - \mu_j)$ will suitably split the sufficient statistic. It follows by induction that in the Gaussian case, we can construct a team sufficient statistic where

1) The statistics are affine functions of the measurements.
2) The dimensions of the statistics only depend on the size of the state variables X and not the size of the measurement variables Y.

The last point is particularly important because the state variables are often constant in size, while the measurement history grows with time.

The conditions for the team sufficient statistics simplify somewhat if each node has at most one immediate parent.

Corollary 19.1 (Outward Trees)**.** Suppose the graph $\mathcal{G}$ in Theorem 19 satisfies $|\mathcal{G}_0(i)| \leq 1$ for all nodes i, and that

$$S_i \in \begin{cases} \mathrm{Suff}(X_{\mathcal{G}^*(i)}|Y_i), & i \text{ is a leaf node} \\ \bigcap_{j: i \in \mathcal{G}_0(i)} \mathrm{Suff}(S_j|Y_i), & \text{otherwise} \end{cases} \tag{4}$$

Then the conclusions of Theorem 19 still hold.

Proof. We just need to show that $S_i \in \mathrm{Suff}(X_{\mathcal{G}^*(i)}|Y_i)$ for all i. If (i,j) is an edge of $\mathcal{G}_0$ and $S_j \in \mathrm{Suff}(X_{\mathcal{G}^*(j)}|Y_j)$, then

$$\begin{aligned} S_i &\in \mathrm{Suff}(S_j|Y_i) \\ &= \mathrm{Suff}(X_{\mathcal{G}^*(i)}, S_j|Y_i) \\ &\subseteq \mathrm{Suff}(X_{\mathcal{G}^*(i)}|Y_i) \end{aligned}$$

by Theorems 2, 12, and 13, and the result follows by induction.

Example 5 (Gaussian Outward Trees)**.** Suppose $((X_1,Y_1),\ldots,(X_n,Y_n))$ is jointly Gaussian and partially nested according to the graph $\mathcal{G}$ where $|\mathcal{G}_0(i)| \leq 1$ for each i. Let $\mathcal{H}$ denote the graph $\mathcal{G}$ with all its edges reversed. Then one can show by the tower property of conditional expectation that

$$S_i = \mathbb{E}(X_{\mathcal{G}^*(i)\cup\mathcal{H}_*(i)}|Y_i)$$

satisfies the conditions of Corollary 19.1, so that $(S_{\mathcal{G}^*(1)},\ldots,S_{\mathcal{G}^*(n)})$ is a team sufficient statistic. Thus the contribution of node i is the conditional expectation of its own state and those of its ancestors and descendants, given its measurements Y_i.

Example 6 (Fully nested systems)**.** In a fully nested system, the graph $\mathcal{G}$ is triangular, so that $\mathcal{G}(i) = \{1,\ldots,i-1\}$. This means that *any* $((X_1,Y_1),\ldots,(X_n,Y_n))$ factors according to $\mathcal{G}$, and all that is required is that Y_i can be derived from Y_j for each $i < j$. Moreover, the transitive reduction of $\mathcal{G}$ is a simple linear chain, so that $\mathcal{G}_0(i) = \{i-1\}$ for each $i > 1$.

For Gaussian fully nested systems, this means that $(T_1,\ldots,T_n)$ is a team sufficient statistic, where $T_i = (S_1,\ldots,S_i)$, and

$$S_i = \mathbb{E}(X|Y_i)$$

Example 7 (Gaussian Inward Trees)**.** Consider a Gaussian partially nested system where each node has at most one immediate *child* (as opposed to one immediate parent for outward trees). By exploiting the fact that the immediate parents of a node have independent measurements and using Theorem 19, one can show that if we set

$$S_i = \mathbb{E}(X_{\mathcal{G}^*(i)\cup\mathcal{H}_*(i)}|Y_i)$$

where $\mathcal{H}$ is the reversal of $\mathcal{G}$, then $(S_{\mathcal{G}^*(1)},\ldots,S_{\mathcal{G}^*(n)})$ is a team sufficient statistic (see [1] for details). Note that this is the same form of the team sufficient statistic as the outward tree case, even though very different methods were used to derive the sufficient statistics.

References

[1] J. Wu, Sufficient statistics for team decision problems, Ph.D. thesis, Dept. Elect. Eng., Stanford Univ., Stanford, CA, 2013.

[2] J. Marschak, Elements for a theory of teams, Management Science, vol. 1, no. 2, pp. 127-137, 1955.

[3] C. H. Papdimitriou and J. N. Tsitsiklis, On the complexity of designing distributed protocols, Information and Control, vol. 53, no. 3, pp. 211-218, 1982.

[4] R. Radner, Team decision problems, Annals of Mathematical Statistics, vol. 33, no. 3, pp. 857-881, 1962.

[5] Y. C. Ho and K. C. Chu, Team decision theory and information structures in optimal control problems—Part I, IEEE Trans. on Automatic Control, vol. 17, no. 1, pp. 15-22, 1972.

[6] K. Hsu and S. I. Marcus, Decentralized control of finite state Markov processes, IEEE Trans. on Automatic Control, vol. 27, no. 2, pp. 426-431, 1982.

[7] M. Aicardi, F. Davoli, and R. Minciardi, Decentralized optimal control of Markov chains with a common past information set, IEEE Trans. on Automatic Control, vol. 32, no. 11, pp. 1028-1031, 1987.

[8] J. Wu and S. Lall, A dynamic programming algorithm for decentralized Markov decision processes with a broadcast structure, in Proc. IEEE Conf. on Decision and Control, Atlanta, GA, 2010, pp. 6143-6148.

[9] A. Nayyar, A. Mahajan, and D. Teneketzis, Decentralized stochastic control with partial history sharing: A common information approach, IEEE Trans. on Automatic Control, vol. 58, no. 7, pp. 1644-1658, 2013.

[10] J. Swigart, Optimal controller synthesis for decentralized systems, Ph.D. thesis, Dept. Elect. Eng., Stanford Univ., Stanford, CA, 2010.

[11] P. Shah and P. Parrilo, H_2-optimal decentralized control over posets: A state-space solution for state-feedback, in Proc. IEEE Conf. on Decision and Control, Atlanta, GA, 2010, pp. 6722-6727.

[12] L. Lessard and S. Lall, Optimal controller synthesis for the decentralized two-player problem with output feedback, in Proc. American Control Conference, 2012, pp. 6314-6321.