Coding Strategies for a Class of Decentralized Control Problems with Limited Communication

Reza Mirghaderi Sanjay Lall Andrea Goldsmith[1]

IEEE Conference on Decision and Control, p. 4809–4816, 2011

Abstract

We propose a large class of decentralized control problems with non-classical information structure for which a coding strategy is optimal. This class is a generalized version of the hats problem with statistically dependent hat colors where implicit communication via action is allowed. We propose a sufficient condition on the joint distribution of the hat colors which guarantees the optimality of a binary sum coding strategy. We explore the connection between this version of the hats problem and hypercube graph theory, and use that to show that verifying our proposed sufficient condition is computationally tractable.

1 Introduction and Prior Work

The decentralized control problems with non-classical information structures are known to be difficult. One of the oldest problems in this area is Witsenhausen's counterexample [9] which is unsolved. Even the discrete variants of such problems are computationally intractable [8].

The challenge faced by each decision maker under non-classical information is the tension between signaling and individual cost minimization. The distributed nature of the problem leads to an information asymmetry between different decision makers. Furthermore, the non-classical structure of the information implies that the action of the decision maker A is "overheard" by another decision maker B who does not have access to all the information available at A. This introduces an opportunity of communication through the action of the decision maker A. The challenging question here is whether A should use its action to minimize its direct contribution to the cost or to encode information which is potentially useful for player B.

The above tension is particularly restrictive when the communication between different decision makers are limited to the implicit communication through actions. For example, in Witsenhausen's problem, a larger control action by the first player gives him a higher encoding capability but also increases his individual cost directly.

In [7] we presented a class of decentralized control problems with non-classical information for which we answered the above question by showing the priority of information encoding over the individual cost minimization. In particular, we considered a multi-agent distributed estimation problem called the hats problem [2], where we further assumed that the estimate given by each player is heard by all the other players. For this problem we proposed a binary sum coding strategy and proved its optimality. Under this coding strategy, one player, called the encoding agent, has to forgo his individually optimal estimation in order to embed information in his declared estimate. Having overheard this estimate, the other players can use the encoded information to make error-free estimations.

Under the setup considered in [7], the variable to be estimated at the encoding agent is independent of the information observed by that agent. In fact we showed that the individually optimal estimate for the encoding agent in [7] is a constant value independent of his observations. The question addressed in this paper is what happens if the observations at each player are correlated with the variable to be estimated by that player. As this correlation can potentially increase the performance of estimation by each player, one would expect more incentives for individual cost minimization rather than information encoding. However, we show that for a large class of such correlated hats problems with communication, the coding strategy considered in [7] is still optimal.

We propose a sufficient condition for the optimality of the binary sum coding strategy over the hats problems with dependent state variables. We will use the hypercube graph representation of the hats problem to show that verifying our sufficient condition is computationally tractable. We will also introduce specific instances of the hats problems for which our coding strategy is optimal. In particular, we will see scenarios in which, under the optimal strategy, the encoding agent has to make erroneous estimation although he perfectly observes his own variable to be estimated.

The rest of the paper is organized as follows: In Section 2 we present the problem formulation and show the

[1]R. Mirghaderi and A. Goldsmith are members of the Department of Electrical Engineering, Stanford University, 350 Serra Mall, Stanford, CA, USA, Email: `rezam@stanford.edu`, `andrea@wsl.stanford.edu`. S. Lall is with the Department of Electrical Engineering and the Department of Aeronautics and Astronautics, Stanford University, Stanford, CA 94305, USA, Email: `lall@stanford.edu`. This work is supported by the NSF Center for the Science of Information under Award CCF-0939370.

optimality of deterministic strategies for any hats problem. In Section 3, we will review the binary sum coding strategy and provide a sufficient condition for the optimality of this strategy. We characterize the problems which satisfy the sufficient condition and present different examples. The complexity of verifying the validity of the sufficient condition for an instance of the hats problem is discussed in Section 4, where we use a hypercube graph representation of the hats problem. Section 5 concludes the paper.

Notation. In the remainder of the paper we use bold letters (e.g. $\mathbf{x}$, $\mathbf{u}$, $\mathbf{I}$) to represent the random variables and the corresponding regular letters (e.g. x, u, I) to represent their realizations. We use subscripts to denote particular elements of a vector. The indicator function for event A is denoted by $[A]$. We denote $\mathcal{B}^n$ to be the n-fold Cartesian product of the set $\{0,1\}$. That is $\mathcal{B}^n = \{0,1\} \times \cdots \times \{0,1\}$ n-times, with the interpretation that $\mathcal{B}^0 = \emptyset$.

2 The Hats Problem

An n-player hats problem with hats probability distribution $\mathcal{P}$ over $\mathcal{B}^n$ (simply referred to as the hats problem $\mathcal{P}$) can be described as follows. Player i is assigned a binary state variable (hat color) $\mathbf{x}_i$. The state vector $(\mathbf{x}_1, \ldots, \mathbf{x}_n)$ is distributed according to the distribution $\mathcal{P}$. Player i observes state variables $\mathbf{x}_{i+1}, ..., \mathbf{x}_n$. The action (estimate) of player i is a binary variable $\mathbf{u}_i$ and is observed by all other players. The cost of player i's action is $[\mathbf{u}_i \neq \mathbf{x}_i]$. The objective is to find a decision strategy that minimizes the expected total cost

$$\mathbb{E}\Big(\sum_{i=1}^{n} [\mathbf{x}_i \neq \mathbf{u}_i]\Big).$$

A *randomized decision strategy*, denoted by the pair (ρ, γ), consists of a permutation ρ and a randomized decision rule γ. The permutation $\rho = \big(\rho_1, \cdots, \rho_n\big)$ is an order in which the players make decisions. Thus $\rho_k = i$ implies that player i makes its decision at the k^{th} step. A player making a decision at the k^{th} step knows the previous actions $\mathbf{u}_{\rho_1}, \cdots, \mathbf{u}_{\rho_{k-1}}$ as well as the state of players $\mathbf{x}_{\rho_k+1}, \cdots, \mathbf{x}_n$. Let $\mathbf{I}_k^{\rho}$ be the information vector available to a player making a decision at the k^{th} step under the decision strategy (ρ, γ). Then, we have

$$\mathbf{I}_k^{\rho} = \big(\mathbf{u}_{\rho_1}, \cdots \mathbf{u}_{\rho_{k-1}}, \mathbf{x}_{\rho_k+1} \cdots, \mathbf{x}_n\big)$$

For example if $n = 3$ and $\rho = (2, 1, 3)$, then
Step 1: $\mathbf{u}_2$ is chosen based on $\mathbf{I}_1^{\rho} = (\mathbf{x}_3)$,
Step 2: $\mathbf{u}_1$ is chosen based on $\mathbf{I}_2^{\rho} = (\mathbf{u}_2, \mathbf{x}_2, \mathbf{x}_3)$,
Step 3: $\mathbf{u}_3$ is chosen based on $\mathbf{I}_3^{\rho} = (\mathbf{u}_1, \mathbf{u}_2)$.

The second component of a randomized decision strategy is the decision rule $\gamma = \big(\gamma_1, \cdots, \gamma_n\big)$ which maps the information available to a player making a decision at the k^{th} step to the probability of action $\mathbf{u}_{\rho_k} = 0$. That is,

$$\gamma_k : \mathcal{B}^{k-1+n-\rho_k} \rightarrow [0,1],$$

and

$$\mathbf{u}_{\rho_k} = \begin{cases} 0 & \text{with probability } \gamma_k\big(\mathbf{I}_k^{\rho}\big) \\ 1 & \text{with probability } 1 - \gamma_k\big(\mathbf{I}_k^{\rho}\big) \end{cases}.$$

Given a strategy (ρ, γ), the cost function is given as

$$J(\rho, \gamma) = \mathbb{E}\Big(\sum_{k=1}^{n} [\mathbf{u}_{\rho_k} \neq \mathbf{x}_{\rho_k}]\Big),$$

where the expectation is with respect to the randomness in the *state vector* $(\mathbf{x}_1, \ldots, \mathbf{x}_n)$ and the uncertainty in the decision rule. Note that a deterministic decision strategy is a degenerate version of a randomized strategy when, for any k, the function γ_k maps any information vector to either 0 or 1. We denote a *deterministic strategy* by the pair (ρ, μ) where ρ is the decision order and $\mu = \big(\mu_1, \cdots, \mu_n\big)$ is the set of decision functions such that

$$\mu_k : \mathcal{B}^{k-1+n-\rho_k} \rightarrow \mathcal{B},$$

and the action $\mathbf{u}_{\rho_k} = \mu_k\big(\mathbf{I}_k^{\rho}\big)$. In the following theorem we will show that for any instance of the hats problem, there exists a deterministic strategy that achieves the minimum cost.

Theorem 1. *For any hats problem $\mathcal{P}$, there exists a deterministic strategy that achieves its optimal cost.*

Proof. Suppose the theorem does not hold. Denoting the set of optimal strategies by Λ_{opt}, this implies that all the strategies in Λ_{opt} should have a non-deterministic component in their decision rule. Let's define for any $(\rho, \gamma) \in \Lambda_{\text{opt}}$, the set $\mathcal{A}_{(\rho,\gamma)}$ of steps at which the decision rule is non-deterministic. For any $i \in \mathcal{A}_{(\rho,\gamma)}$, there exists an information vector realization I_i^{ρ} such that

$$\gamma_i(I_i^{\rho}) \in (0, 1).$$

Among all the optimal strategies consider the one for which the first randomized decision making happens the latest. That is, consider the strategy (ρ^*, γ^*) and the k^{th} step such that

$$(\rho^*, \gamma^*) = \text{argmax}_{(\rho,\gamma) \in \Lambda_{\text{opt}}} \min \mathcal{A}_{(\rho,\gamma)}, \tag{1}$$

and

$$k = \min \mathcal{A}_{(\rho^*,\gamma^*)}, \tag{2}$$

We can write the expected cost corresponding to the policy (ρ^*, γ^*) as

$$J(\rho^*, \gamma^*) = \sum_{x \in \{0,1\}^n} \mathcal{P}(x) \mathbb{E}\left(J(\rho^*, \gamma^*)|x\right), \quad (3)$$

where x represents a realization of the state vector (hat color configurations) and $\mathbb{E}\left(J(\rho^*, \gamma^*)|x\right)$ is the expected cost (with respect to the randomness in the strategy) when the state vector is x. Let $\mathbf{u}_{\rho_i^*}(x)$ and $\mathbf{I}_i^{\rho^*}(x)$ denote the action variable and the information vector of the player making a decision at the i^{th} step, respectively, when the state vector is x. Denoting the total cost incurred from step i to the last step by $J_i(\rho^*, \gamma^*)$, we can write for any $x \in \{0,1\}^n$,

$$\begin{aligned}
\mathbb{E}\left(J(\rho^*, \gamma^*)|x\right) &= \sum_{i=1}^{k-1} [\mathbf{u}_{\rho_i^*}(x) \neq x_{\rho_i^*}] \\
&+ \gamma_k^*(\mathbf{I}_k^{\rho^*}(x)) \Big([x_{\rho_k^*} \neq 0] + \mathbb{E}\left(J_{k+1}(\rho^*, \gamma^*)|x, \mathbf{u}_{\rho_k^*} = 0\right) \Big) \\
&+ (1 - \gamma_k^*(\mathbf{I}_k^{\rho^*}(x))) \Big([x_{\rho_k^*} \neq 1] + \mathbb{E}\left(J_{k+1}(\rho^*, \gamma^*)|x, \mathbf{u}_{\rho_k^*} = 1\right) \Big) \\
&= \beta(x) + \gamma_k^*(\mathbf{I}_k^{\rho^*}(x)) \alpha(x),
\end{aligned}$$

where

$$\begin{aligned}
\beta(x) &= \sum_{i=1}^{k-1} [\mathbf{u}_{\rho_i^*}(x) \neq x_{\rho_i^*}] \\
&+ [x_{\rho_k^*} \neq 1] + \mathbb{E}\left(J_{k+1}(\rho^*, \gamma^*)|x, \mathbf{u}_{\rho_k^*} = 1\right)
\end{aligned}$$

and

$$\begin{aligned}
\alpha(x) &= [x_{\rho_k^*} \neq 0] + \mathbb{E}\left(J_{k+1}(\rho^*, \gamma^*)|x, \mathbf{u}_{\rho_k^*} = 0\right) \\
&- [x_{\rho_k^*} \neq 1] - \mathbb{E}\left(J_{k+1}(\rho^*, \gamma^*)|x, \mathbf{u}_{\rho_k^*} = 1\right).
\end{aligned}$$

Note that given the assumption that decisions at steps $1, ..., k-1$ are all deterministic, $\beta(x)$ and $\alpha(x)$ do not depend on the probability distribution of the decision rule at the k^{th} step. Also note that due to the same assumption of deterministic decisions at steps $1, ..., k$, given the state x, there is no uncertainty in determining $\mathbf{I}_k^{\rho^*}(x)$. Using the above facts and derivations, we can write the total cost as

$$\begin{aligned}
J(\rho^*, \gamma^*) &= \sum_{x \in \{0,1\}^n} \mathcal{P}(x) \beta(x) \quad (4) \\
&+ \sum_{\omega \in \mathcal{B}^{k-1+n-\rho_k^*}} \gamma_k^*(\omega) \sum_{\mathbf{x}: \mathbf{I}_k^{\rho^*}(x) = \omega} \mathcal{P}(x) \alpha(x).
\end{aligned}$$

Now let's define a new strategy (ρ', γ'), where $\rho' = \rho^*$ and the only difference between γ' and γ^* is in the decision rule at step k, where

$$\gamma_k'(\omega) = \begin{cases} 0 & \text{if} \quad \sum_{x: \mathbf{I}_k^{\rho^*}(x) = \omega} \mathcal{P}(x) \alpha(x) \geq 0 \\ 1 & \text{otherwise} \end{cases},$$

for all $\omega \in \mathcal{B}^{k-1+n-\rho_k'}$. Note that in the above strategy, players $1, .., k$ have deterministic decision rules. Also using (4), we can easily check that

$$J(\rho', \gamma') \leq J(\rho^*, \gamma^*),$$

which contradicts our assumption that (ρ^*, γ^*) is the optimal decision strategy for which the randomization happens the latest (assumptions stated in equations 1 and 2). ∎

Given the above result, we will only focus on deterministic strategies (ρ, μ) in the rest of the paper.

3 Binary Sum Coding Strategy

We showed in the previous section that one can always find an optimal deterministic strategy for the hats problem. A simple yet useful class of deterministic strategies for the hats problem is the *binary sum coding strategy* introduced in [7].

Definition 2. *The binary sum coding strategy (BSCS) for an n-player hats problem, consists of the decision order $\rho^{code} = (1, 2, \cdots, n)$ and the decision rule $\mu^{code} = (\mu_1^{code}, \cdots, \mu_n^{code})$, where*

$$\mu_k^{code}(\mathbf{u}_1, \cdots, \mathbf{u}_{k-1}, \mathbf{x}_{k+1}, \cdots, \mathbf{x}_k) = \mathbf{u}_1 \oplus \cdots \oplus \mathbf{u}_{k-1} \oplus \mathbf{x}_{k+1} \oplus \cdots \oplus \mathbf{x}_n.$$

It is easy to check (and was proved in [7]) that BSCS results in correct estimates for all players except possibly player 1. We proved in [7] that BSCS is optimal for the hats problems with i.i.d. Bernoulli state variables. In this section, we show the optimality of BSCS over a broader class of distributions. This includes a large subclass of independent distributions, Markov distributions and the distributions with star-shaped dependency graphs. The proofs of all the results presented in this section are provided in the appendix.

Definition 3. *Given n, the **even parity set** S^{even} and the **odd parity set** S^{odd} are defined as*

$$\begin{aligned}
S^{even} &= \{x \in \mathcal{B}^n \mid x_1 \oplus x_2 \oplus \cdots \oplus x_n = 0\}, \\
S^{odd} &= \mathcal{B}^n - S^{even}.
\end{aligned}$$

Definition 4. *For any set $R \subseteq \mathcal{B}^n$ we define the internal Hamming distance of the set R (denoted by $d(R)$) as*

$$d(R) = \min_{\substack{a,b \in R, \\ a \neq b}} \sum_{i=1}^{n} [a_i \neq b_i].$$

Definition 5. *A hats problem $\mathcal{P}$ is called **strongly even** if*

$$\mathcal{P}(S^{even}) \geq \mathcal{P}(R),$$

for any $R \subset \mathcal{B}^n$ such that

$$d(R) \geq 2.$$

As will be shown in the next section, any set R with the above internal Hamming distance property is an independent set in the hypercube graph corresponding to the hats problem.

Theorem 6. *Given a hats problem $\mathcal{P}$, the binary sum coding strategy $(\rho^{code}, \mu^{code})$ is optimal if $\mathcal{P}$ is strongly even.*

The sufficient condition provided in the above theorem introduces a class of BSCS-optimal hats problems (distributions), namely strongly even problems. In the following, we characterize strongly even problems and show some examples.

Proposition 7. *Given a set Γ of strongly even n-player hats problems (distributions), $\mathcal{P}$ is strongly even if*

$$\mathcal{P} \in Conv(\Gamma),$$

where $Conv(\Gamma)$ is the convex hull of all the distributions in Γ.

Proposition 8. *Given n, the hats problem $\mathcal{P}$ is strongly even if for all $x \in \mathcal{B}^n$*

$$\mathcal{P}(x) = \prod_{i=1}^{n} \mathcal{P}_i(x_i),$$

where for $1 \leq i \leq n$, $\mathcal{P}_i$ is the probability distribution of the i^{th} state variable and

$$\mathcal{P}_i(0) \geq \frac{1}{2}.$$

Proposition 9. *Given n, the hats problem $\mathcal{P}$ is strongly even if n is even and the following conditions hold*

- *$\mathcal{P}$ satisfies the Markov property*

  $$\mathcal{P}(x_i|x_{i-1}, ..., x_1) = \mathcal{P}(x_i|x_{i-1}),$$

  for all $x \in \mathcal{B}^n$ and for $i = 2, ..., n$

- *$\mathcal{P}(\mathbf{x}_i = 0|\mathbf{x}_{i-1} = 0) = \mathcal{P}(\mathbf{x}_i = 1|\mathbf{x}_{i-1} = 1) = q$, for $i = 2, ..., n$ and for a fixed $q \in [\frac{1}{2}, 1]$*

- *$1 - q \leq \mathcal{P}(\mathbf{x}_1 = 0) \leq q$.*

Proposition 10. *Given n, the hats problem $\mathcal{P}$ is strongly even if n is even and the following conditions hold*

- *Given $\mathbf{x}_1$, all the other state variables are conditionally independent*

  $$\mathcal{P}(\mathbf{x}_2, ..., \mathbf{x}_n|\mathbf{x}_1) = \prod_{i=2}^{n} \mathcal{P}(\mathbf{x}_i|\mathbf{x}_1),$$

- *For $i = 2, ..., n$,*

  $$\mathcal{P}(\mathbf{x}_i = 0|\mathbf{x}_1 = 1) \leq \frac{1}{2} \leq \mathcal{P}(\mathbf{x}_i = 0|\mathbf{x}_1 = 0).$$

A special example covered by Proposition 10 is when $\mathbf{x}_1 = \mathbf{x}_2$ and the states $\mathbf{x}_3, ..., \mathbf{x}_n$ are independent of $\mathbf{x}_1$ and are i.i.d. with symmetric Bernoulli distributions. The proposition implies that although the first player can perfectly estimate his hat color through his observation of the second hat, he should make a lossy guess by declaring the sum of the other hats as his own estimate.

Note that while Theorem 6 shows the optimality of BSCS over a potentially large class of hat problems, BSCS (or any other linear coding strategy) is not always optimal. As an example, consider a 3-player hat problem $\mathcal{P}$ such that

$$\mathcal{P}(0, 0, 0) = \mathcal{P}(1, 1, 1) = 0.5.$$

Under this scenario, it is optimal for the first player to estimate 0 when the other hats are all 0 and estimate 1, when other hats are all 1, which means the action of the first player does not follow the binary sum of the other hats.

4 BSCS Optimality and Hypercubes

In this section we will show that verifying the strongly even condition for $\mathcal{P}$ stated in Definition 5 is computationally tractable in terms of the size of the hats problem. Here by the size of the hats problem we mean the degree of freedom for the distribution on state vectors which is 2^n.[1]

Theorem 11. *Given n, the problem of determining whether $\mathcal{P}$ is strongly even is solvable in polynomial time in terms of the size of the hats problem. That is, there exits a constant $\zeta > 0$, such that the strongly evenness of $\mathcal{P}$ can be verified in*

$$O(2^{\zeta n}).$$

We will prove the theorem by showing that the above problem can be translated to a well-studied problem in graph theory called the *maximum weight independent set (MWIS)* problem on hypercube graphs. Let's first define a hypercube graph [6].

[1] Or more precisely $2^n - 1$ since the probabilities should add up to 1.

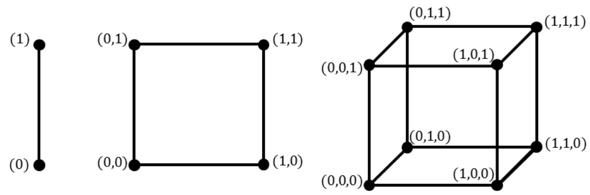

Figure 1: n-dimensional Hypercubes for $n = 1, 2, 3$

Definition 12. *Given n, an n-dimensional hypercube graph is defined as an undirected graph $Q = (V_Q, E_Q)$, where $V_Q = \mathcal{B}^n$ is the set of vertices and E_Q is the set of edges. Every two nodes $x, x' \in V_Q$ are connected via $(x, x') \in E_Q$ iff x and x' only differ in one component i.e. $d(\{x, x'\}) = 1$.*

Figure 1 illustrates the hypercube graphs for $n = 1, 2, 3$.

Definition 13. *An independent set of a graph $G = (V, E)$ is a subset $R \subseteq V$ of its vertices such that no two vertices in R are connected in G.*

Given the above two definitions we can observe that $R \subseteq V_Q$ is an independent set of the hypercube Q iff

$$d(R) \geq 2. \quad (5)$$

In particular, one can easily show that S^{even} and S^{odd} are both independent sets of Q.

Definition 14. *Given a graph $G = (V, E)$ and a weight function $w : V \mapsto \mathbb{R}^+$, the weighted independence number $\alpha_w(G)$ is defined as*

$$\alpha_w(G) = \max_{R \subseteq V : R \text{ is independent}} \sum_{x \in R} w(x).$$

Proposition 15. *A hats problem $\mathcal{P}$ is strongly even iff*

$$\alpha_{\mathcal{P}}(Q) = \mathcal{P}(S^{even}),$$

where $\alpha_{\mathcal{P}}(Q)$ is the weighted independence number of the hypercube graph Q given the weighting function which assigns to each vertex the probability of its corresponding vector under the distribution $\mathcal{P}$.

Proof. Follows from Definition 5, inequality 5, Definition 14 and the fact that S^{even} is an independent set of Q. ∎

Lemma 16. *Q is a bipartite graph. That is, V_Q can be partitioned into two independent sets.*

Proof. Partition V_Q into S^{even} and S^{odd}. ∎

The problem of finding the weighted independence number is $NP-$complete for general graphs [4]. However, it can be solved in polynomial time (as a function of the size of the vertex set) for bipartite graphs. This is a well understood result and can be shown using different techniques (see for example [1] and [5]). In the following, we adopt a recent exposition provided in [3] which shows that the problem of finding the weighted independence number of a bipartite graph is equivalent to finding the minimum cut (maximum flow) in an induced $st-$graph which is solvable in polynomial time [1].

Definition 17. *Let $G = (V, E)$ be a vertex-weighted bipartite graph $G = (V, E)$ with independent sets S and $T = V - S$ and vertex weighting function $w : V \mapsto \mathbb{R}^+$. We add two new vertices $\{s, t\}$ and define the induced $st-$graph for G as an edge-weighted directed graph $G^{st} = (V^{st}, E^{st})$ with edge weighting function $c : V^{st} \times V^{st} \mapsto \mathbb{R}^+$, such that*

$$\begin{aligned} V^{st} &= V \cup \{s\} \cup \{t\}, \\ E^{st} &= \{(x, x') : (x, x') \in E, x \in S, x' \in T\} \\ &\cup \{(s, x) : x \in S\} \\ &\cup \{(x, t) : x \in T\}, \end{aligned}$$

and

$$c(x, x') = \begin{cases} 0 & \text{if } (x, x') \notin E^{st} \\ w(x') & \text{if } x = s \text{ and } x' \in S \\ w(x) & \text{if } x \in T \text{ and } x' = t \\ \infty & \text{if } (x, x') \in E. \end{cases}.$$

We also define a $st-$cut as a partition of V^{st} into two sets V_s and V_t where $s \in V_s$ and $t \in V_t$. The capacity of the cut is given by

$$C(V_s, V_t) = \sum_{x \in V_s, x' \in V_t} c(x, x')$$

Figure 2 illustrates sample $st-$cuts with finite capacity on an induced graph for a 3-dimensional hypercube whose vertices are weighted according to the distribution $\mathcal{P}$.

Lemma 18. *Let $G^{st} = (V^{st}, E^{st})$ be an induced $st-$graph of a bipartite graph $G = (V, E)$ with vertex weighting function $w : V \mapsto \mathbb{R}^+$. The weighted independence number of G is related to the minimum capacity cut of G^{st} as follows*

$$\alpha_w(G) = \sum_{x \in V} w(x) - \min_{V_s} C(V_s, V_t).$$

Proof. Let S and $T = V - S$ denote the two independent partitions of V. From any independent set $R \subseteq V$,

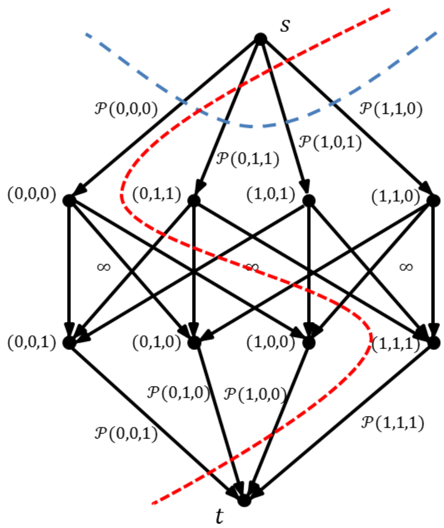

Figure 2: Induced graph and sample finite capacity cuts for a 3-dimensional hypercube with vertex probability (weight) distribution $\mathcal{P}$

one can obtain a unique finite capacity cut (V_s, V_t) in G^{st} such that

$$\begin{aligned} V_s &= \{s\} \cup (S \cap R) \cup (T - R), \\ V_t &= \{t\} \cup (S - R) \cup (T \cap R). \end{aligned}$$

Note that the edges from V_s to V_t in the above cut only consist of the ones connected to the nodes outside R from s or the ones originated from outside R and connected to t. Therefore, using the definition of edge weighting in an induced graph, we have

$$\sum_{v \in R} w(v) = \sum_{x \in V} w(x) - C(V_s, V_t). \tag{6}$$

Conversely, given a finite capacity cut (V_s, V_t), we know that no edge in the original graph G appears between V_s and V_t. Therefore, since the graph is bipartite, we can construct an independent set as follows

$$R = (V_s \cap S) \cup (V_t \cap T).$$

Similarly, one can check that equation 6 is true in this case as well. The proof of the lemma follows from the one to one mapping between the finite capacity cut sets and independent sets and the validity of equation 6. ■

Proof of Theorem 11 From Proposition 15, checking strongly evenness of $\mathcal{P}$ is equivalent to solving maximum weight independent set problem on the hypercube graph Q with weighting distribution $\mathcal{P}$. Therefore, since Q is bipartite (Lemma 16), Lemma 18 implies that verifying strongly evenness is equivalent to the minimum cut problem on the hypercube's induced graph whose vertex cardinality is $|V^{st}| = O(2^N)$. It is well known that the minimum cut problem is equivalent to the maximum flow problem which is solvable in polynomial time in terms of the vertex size of the graph (for e.g. see [1]). ■

5 Discussion and Conclusion

This paper presents optimal coding strategies for a class of distributed estimation problems with implicit communication among the different decision makers. We considered a variant of the hats problem with communication [7] in which the colors of the hats are statistically dependent. While the dependence of the hat colors in general motivates individual cost control rather than encoding information through the actions, we showed that coding is optimal over a large class of hats problems (hat color distributions). We presented a sufficient condition for the optimality of our proposed coding strategy and used the connection between the hats problem and the hypercube graph theory to show that verifying our sufficient condition is computationally tractable.

A Proof of Theorem 6

The proof follows from the similar steps as in [7]. We first define a perfect set for a deterministic strategy as follows.

Definition 19. *For the n-player hats problem and for any deterministic strategy (ρ, μ), the* perfect set *associated with that strategy (denoted by $S(\rho, \mu)$) is given by*

$$S(\rho, \mu) = \left\{ x \in \mathcal{B}^n \mid u_{\rho_k} = x_{\rho_k} \quad \text{for all } k = 1, 2, \cdots, n \right\},$$

where $u_{\rho_k} = \mu_k(I_k^\rho)$ and $I_k^\rho = (u_{\rho_1}, \cdots u_{\rho_{k-1}}, x_{\rho_k+1} \cdots x_n)$.

In other words, the perfect set associated with any strategy is a set of all possible initial hat configurations such that all resulting actions (estimates) match the states (hat colors).

Lemma 20. *Given n, the even parity set S^{even} is the perfect set associated with the binary sum coding strategy*

Proof. See the proof of Lemma 4 in [7]. ■

Lemma 21. *Given any deterministic decision strategy (ρ, μ) and a hats problem $\mathcal{P}$, we have $J(\rho, \mu) \geq 1 - \mathcal{P}(S(\rho, \mu))$ with equality if the strategy is the binary sum coding strategy.*

Proof. See the proof of Lemma 7 in [7]. ■

Lemma 22. *For any strategy (ρ, μ), the internal Hamming distance associated with the perfect set $S(\rho, \mu)$ is at least 2. That is,*

$$d(S(\rho, \mu)) \geq 2.$$

Proof. See the proof of Lemma 10 in [7]. ■

Proof of Theorem 6. Given the result of Theorem 1, it is sufficient to show the optimality over all deterministic strategies. For any deterministic strategy (ρ, μ), we can write

$$\begin{aligned} J(\rho, \mu) &\geq 1 - \mathcal{P}(S(\rho, \mu)) \\ &\geq 1 - \mathcal{P}(S^{even}) \\ &= J(\rho^{code}, \mu^{code}), \end{aligned}$$

where the first inequality follows from Lemma 21, the second inequality follows from the combination of Lemma 22 and the assumption that $\mathcal{P}$ is strongly even and the third inequality follows from Lemma 20. ■

B Proofs of Propositions 7-10

The proof of Proposition 7 simply follows from the linearity of the condition stated in Definition 5 for strongly evenness.

Before we prove Propositions 8-10, let's first present a simple lemma which will be used in the proofs frequently. Given a set $R \subseteq \mathcal{B}^n$ and a binary sequence $(b_1, ..., b_l) \in \mathcal{B}^l$ for $l < n$ we define the set $R_{(b_1,...,b_l)} \subseteq \mathcal{B}^{n-l}$ as follows

$$R_{(b_1,...,b_l)} = \{x \in \mathcal{B}^{n-l} : (b_1, ..., b_l, x) \in R\}.$$

Lemma 23. *Consider a set $R \subseteq \mathcal{B}^n$ which satisfies $d(R) \geq 2$. The following holds for any $l < n$*

1. *Given any $(b_1, ..., b_l) \in \mathcal{B}^l$,*
$$d(R_{(b_1,...,b_l)}) \geq 2$$
2. *Given $(b_1, ..., b_l), (b'_1, ..., b'_l) \in \mathcal{B}^l$, if $d(\{(b_1, ..., b_l), (b'_1, ..., b'_l)\}) = 1$*
$$R_{(b_1,...,b_l)} \cap R_{(b'_1,...,b'_l)} = \emptyset$$

Proof. The proof of the first part follows from the fact that $d(R_{(b_1,...,b_l)}) \leq d(R)$. To prove the second part, note that any common element between $R_{(b_1,...,b_l)}$ and $R_{(b'_1,...,b'_l)}$ leads to two vectors in R with a Hamming distance equal to 1. ■

The proofs presented in the following include inductive steps which involve iterations over different values of n. To distinguish between parity sets for different values of n, we will use subscript n (e.g. S_n^{even}, S_n^{odd}).

Proof of Proposition 8 The case of $n = 1$ is trivial. Let's assume the validity of the proposition for values of n up to k and consider the case of $n = k + 1$. Let's consider the distribution $\mathcal{P}'$ over $x \in \mathcal{B}^k$ as follows

$$\mathcal{P}'(x) = \prod_{i=1}^{k} \mathcal{P}_{i+1}(x_i).$$

Note that given the definition of the even parity set we have

$$\mathcal{P}(S_{k+1}^{even}) = \mathcal{P}_1(0)\mathcal{P}'(S_k^{even}) + \mathcal{P}_1(1)(1 - \mathcal{P}'(S_k^{even})). \tag{7}$$

For any $R \subseteq \mathcal{B}^{k+1}$, if R satisfies $d(R) \geq 2$, we can write

$$\begin{aligned} \mathcal{P}(R) &= \mathcal{P}_1(0)\mathcal{P}'(R_{(0)}) + (1 - \mathcal{P}_1(0))\mathcal{P}'(R_{(1)}) \\ &\leq \mathcal{P}_1(0)\mathcal{P}'(R_{(0)}) + (1 - \mathcal{P}_1(0))(1 - \mathcal{P}'(R_{(0)})) \\ &= (2\mathcal{P}_1(0) - 1)\mathcal{P}'(R_{(0)}) + 1 - \mathcal{P}_1(0) \\ &\leq (2\mathcal{P}_1(0) - 1)\mathcal{P}'(S_k^{even}) + 1 - \mathcal{P}_1(0) \\ &= \mathcal{P}(S_{k+1}^{even}), \end{aligned}$$

where the first inequality is a consequence of Lemma 23.2, the second inequality follows from Lemma 23.1, the induction assumption and the fact that $2\mathcal{P}_1(0) - 1 \geq 0$, and finally the last equality follows from equation 7. ■

Proof of Proposition 9 The validity of the proposition for $n = 2$ can be verified by inspection. Assuming the proposition holds for even values of n up to $2k$, we consider the case of $n = 2(k + 1)$. Note that given the Markov property, we have

$$\mathcal{P}(\mathbf{x}_3, ..., \mathbf{x}_{2(k+1)}|\mathbf{x}_1, \mathbf{x}_2) = \mathcal{P}(\mathbf{x}_3, ..., \mathbf{x}_{2(k+1)}|\mathbf{x}_2).$$

Let's define for $b = 0, 1$, the distribution $\mathcal{P}'_b$ over $\mathcal{B}^{2k}$ as

$$\mathcal{P}'_b(v_1, ..., v_{2k}) = \mathcal{P}(\mathbf{x}_3 = v_1, ..., \mathbf{x}_{2(k+1)} = v_{2k}|\mathbf{x}_2 = b).$$

Note that $\mathcal{P}'_b$ satisfies the conditions stated in the proposition and therefore, according to the induction assumption, for $b = 0, 1$, $\mathcal{P}'_b$ is strongly even. For any $R \subseteq \mathcal{B}^{2(k+1)}$, if R satisfies $d(R) \geq 2$, we can write

$$\begin{aligned} \mathcal{P}(R) &= \sum_{x_1, x_2} \mathcal{P}(x_1, x_2)\mathcal{P}(R_{(x_1,x_2)}|x_2) \\ &= \mathcal{P}(\mathbf{x}_1\mathbf{x}_2 = 00)\mathcal{P}'_0(R_{(0,0)}) + \mathcal{P}(\mathbf{x}_1\mathbf{x}_2 = 10)\mathcal{P}'_0(R_{(1,0)}) \\ &+ \mathcal{P}(\mathbf{x}_1\mathbf{x}_2 = 11)\mathcal{P}'_1(R_{(1,1)}) + \mathcal{P}(\mathbf{x}_1\mathbf{x}_2 = 01)\mathcal{P}'_1(R_{(0,1)}). \end{aligned}$$

According to Lemma 23.2, $R_{(0,b)} \cap R_{(1,b)} = \emptyset$ for $b = 0, 1$. Therefore using $\mathcal{P}'_b(R_{(1,b)}) \leq 1 - \mathcal{P}'_b(R_{(0,b)})$, we can upper bound the above quantity by

$$\begin{aligned} \mathcal{P}(R) &\leq (\mathcal{P}(\mathbf{x}_1\mathbf{x}_2 = 00) - \mathcal{P}(\mathbf{x}_1\mathbf{x}_2 = 10))\,\mathcal{P}'_0(R_{(0,0)}) \\ &+ (\mathcal{P}(\mathbf{x}_1\mathbf{x}_2 = 11) - \mathcal{P}(\mathbf{x}_1\mathbf{x}_2 = 01))\,\mathcal{P}'_1(R_{(1,1)}) \\ &+ \mathcal{P}(\mathbf{x}_1\mathbf{x}_2 = 10) + \mathcal{P}(\mathbf{x}_1\mathbf{x}_2 = 01). \end{aligned}$$

It can be easily verified that the conditions on $\mathcal{P}(x_1)$ and $\mathcal{P}(x_2|x_1)$ stated in the proposition guarantee that

$$\begin{array}{l}\mathcal{P}(\mathbf{x}_1\mathbf{x}_2 = 00) - \mathcal{P}(\mathbf{x}_1\mathbf{x}_2 = 10) \geq 0 \\ \mathcal{P}(\mathbf{x}_1\mathbf{x}_2 = 11) - \mathcal{P}(\mathbf{x}_1\mathbf{x}_2 = 01) \geq 0 \qquad .\end{array}$$

Therefore, we can extend the upper bound on $\mathcal{P}(R)$ as follows

$$\begin{aligned}\mathcal{P}(R) &\leq (\mathcal{P}(\mathbf{x}_1\mathbf{x}_2 = 00) - \mathcal{P}(\mathbf{x}_1\mathbf{x}_2 = 10))\,\mathcal{P}'_0(S^{even}_{2k}) \\ &+ (\mathcal{P}(\mathbf{x}_1\mathbf{x}_2 = 11) - \mathcal{P}(\mathbf{x}_1\mathbf{x}_2 = 01))\,\mathcal{P}'_1(S^{even}_{2k}) \\ &+ \mathcal{P}(\mathbf{x}_1\mathbf{x}_2 = 10) + \mathcal{P}(\mathbf{x}_1\mathbf{x}_2 = 01) \\ &= \sum_{x_1 \oplus x_2 = 0} \mathcal{P}(x_1, x_2)\mathcal{P}'_{x_2}(S^{even}_{2k}) \\ &+ \sum_{x_1 \oplus x_2 = 1} \mathcal{P}(x_1, x_2)\mathcal{P}'_{x_2}(\mathcal{B}^{2k} - S^{even}_{2k}) \\ &= \mathcal{P}(S^{even}_{2(k+1)}),\end{aligned}$$

where in the first inequality, we have used the strongly evenness of $\mathcal{P}'_b$ for $b \in \mathcal{B}$ (induction assumption) which implies that for any $(b_1, b_2) \in \mathcal{B}^2$, $\mathcal{P}'_b(R_{(b_1,b_2)}) \leq \mathcal{P}'_b(S^{even}_{2k})$. ■

Remark 24. *Using the inductive step of the above proof, one can show that if $\mathcal{P}'$ and $\mathcal{P}''$ are strongly even over $\mathcal{B}^{n-2}$, then $\mathcal{P}$ is strongly even over $\mathcal{B}^n$ if*

- *For all $(x_1, x_3, ..., x_n) \in \mathcal{B}^{n-1}$,*

$$\begin{array}{l}\mathcal{P}(x_1, 0, x_3, ..., x_n) = \mathcal{P}(x_1, \mathbf{x}_2 = 0)\mathcal{P}'(x_3, ..., x_n), \\ \mathcal{P}(x_1, 1, x_3, ..., x_n) = \mathcal{P}(x_1, \mathbf{x}_2 = 0)\mathcal{P}''(x_3, ..., x_n).\end{array}$$

- *For $b \in \mathcal{B}$*

$$\mathcal{P}(\mathbf{x}_1 = b, \mathbf{x}_2 = b) \geq \mathcal{P}(\mathbf{x}_1 = 1 - b, \mathbf{x}_2 = b).$$

Proof of Proposition 10 Assuming $n = 2k$, let's define for $b = 0, 1$, the distribution $\mathcal{P}'_b$ over $\mathcal{B}^{2k-1}$ as

$$\mathcal{P}'_b(v_1, ..., v_{2k-1}) = \prod_{i=2}^{2k} \mathcal{P}(\mathbf{x}_i = v_{i-1}|\mathbf{x}_1 = b).$$

Given the condition on $\mathcal{P}(\mathbf{x}_i = 0|\mathbf{x}_1 = 0)$ stated in the proposition and using the result of Proposition 8, we can conclude that $\mathcal{P}'_0$ is strongly even. Now let's define the inverting function $\mathrm{Inv} : \mathcal{B}^{2k-1} \mapsto \mathcal{B}^{2k-1}$ as $\mathrm{Inv}(v) = (1 \oplus v_1, ..., 1 \oplus v_{2k-1})$ and consider the distribution $\mathcal{P}^{\mathrm{Inv}}$ such that

$$\mathcal{P}^{Inv}(v) = \mathcal{P}'_1(\mathrm{Inv}(v)).$$

Given the condition on $\mathcal{P}(\mathbf{x}_i = 0|\mathbf{x}_1 = 1)$ stated in the proposition and using the result of Proposition 8, we can conclude that $\mathcal{P}^{\mathrm{Inv}}$ is strongly even. This implies that for any $R \subset \mathcal{B}^{2k-1}$ with $d(R) \geq 2$

$$\mathcal{P}^{\mathrm{Inv}}(S^{even}_{2k-1}) \geq \mathcal{P}^{\mathrm{Inv}}(R).$$

One can easily verify that $S^{odd}_{2k-1} = \mathrm{Inv}(S^{even}_{2k-1})$ and therefore

$$\mathcal{P}^{\mathrm{Inv}}(S^{even}_{2k-1}) = \mathcal{P}'_1(S^{odd}_{2k-1}),$$

which further leads to

$$\mathcal{P}'_1(S^{odd}_{2k-1}) \geq \mathcal{P}'_1(R), \tag{8}$$

for any $R \subset \mathcal{B}^{2k-1}$ which satisfies $d(R) \geq 2$. Note that the sets S^{even}_{2k}, S^{even}_{2k-1} and S^{odd}_{2k-1} are related by the following equation

$$\mathcal{P}(S^{even}_{2k}) = \mathcal{P}(\mathbf{x}_1 = 0)\mathcal{P}'_0(S^{even}_{2k-1}) + \mathcal{P}(\mathbf{x}_1 = 1)\mathcal{P}'_1(S^{odd}_{2k-1}). \tag{9}$$

Using the above derivations, we can now write for any $S \subset \mathcal{B}^{2k}$ which satisfies $d(R) \geq 2$

$$\begin{aligned}\mathcal{P}(R) &= \mathcal{P}(\mathbf{x}_1 = 0)\mathcal{P}'_0(R_{(0)}) + \mathcal{P}(\mathbf{x}_1 = 1)\mathcal{P}'_1(R_{(1)}) \\ &\leq \mathcal{P}(\mathbf{x}_1 = 0)\mathcal{P}'_0(S^{even}_{2k-1}) + \mathcal{P}(\mathbf{x}_1 = 1)\mathcal{P}'_1(S^{odd}_{2k-1}) \\ &= \mathcal{P}(S^{even}_{2k}).\end{aligned}$$

The first inequality in the above derivation is a consequence of the strongly evenness of $\mathcal{P}'_1$ and equation 8. Also the last line is copied from equation 9. ■

References

[1] R.K. Ahuja, T.L. Magnanti, and J.B. Orlin. *Network Flows*. Prentice Hall, 1993.

[2] S. Butler, M. T. Hajiaghayi, R. D. Kleinberg, and T. Leighton. Hat guessing games. *SIAM Review*, 51(2):399–413, 2009.

[3] U. Faiglea and G. Frahling. A combinatorial algorithm for weighted stable sets in bipartite graphs. *Discrete Applied Mathematics*, 154:1380–1391, 2006.

[4] M. R. Garey and D. S. Johnson. *Computers and Intractability : A guide to the theory of NP-completeness*. W. H. Freeman, 1979.

[5] M. Grotschel, L. Lovasz, and A. Schrijver. Polynomial algorithms for perfect graphs. *Annals of Discrete Mathematics*, 21:325–356, 1984.

[6] J. P. Hayes and H. Wu. A survey of the theory of hypercube graphs. *Computers & Mathematics with Applications*, 15:277–289, 1988.

[7] R. Mirghaderi, S. Adlakha, S. Lall, and A. Goldsmith. Hat guessing games and the use of coding for decentralized control. *Proceedings of the 49th IEEE Conference on Decision and Control*, pages 979–985, 2010.

[8] J. N. Tsitsiklis and M. Athans. On the complexity of decentralized decision making and detection problems. *IEEE Transactions on Automatic Control*, (5), 1985.

[9] H. S. Witsenhausen. A counterexample in stochastic optimum control. *SIAM Journal on Control*, 6(1), 1968.