ON A FUNDAMENTAL PREDICTABILITY-DISTORTION TRADE-OFF IN DECENTRALIZED DECISIONAL NETWORKS

Lorenzo Miretti, Paul de Kerret, and David Gesbert

EURECOM, Sophia Antipolis, France

ABSTRACT

Decentralized decisional networks are composed of agents that take coordinated decisions on the basis of individual noisy information about the system state, i.e. under a so called distributed state information configuration. In such scenarios, the application of algorithms directly derived from classical centralized optimization often incurs severe performance degradation. This is because the agents fail to predict each other's decision due to local state information noise, hence they do not coordinate properly. On the other hand coordination can be easily enhanced by letting one agent reduce its dependency with respect to its state information, thus making its decision more predictable but less adapted to the system state. This observation naturally leads to formulating a fundamental trade-off, coined here predictability-distortion trade off. The goal of this paper is to formulate this trade-off and propose a framework to explore it, based on a concept of *quantization under predictability constraint*.

Index Terms— quantization, coordination, multi-agent systems, team decision problems, Lloyd algorithm

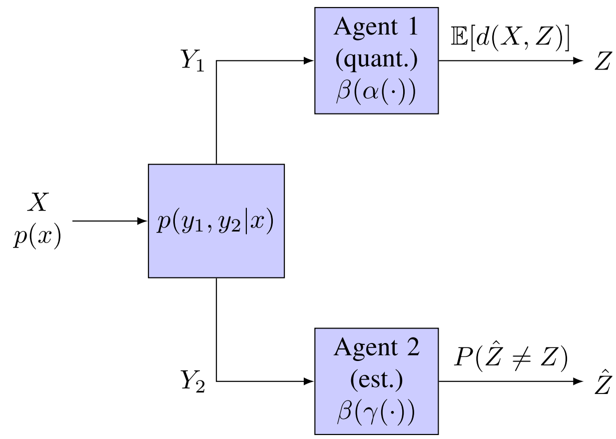


Fig. 1. System model

1. INTRODUCTION

Many types of networks benefit from the cooperating behavior of its agents. Examples include car networks, energy networks, interference limited radio networks [1], etc. Unfortunately, agents are often limited in the quality of the system state information over which they seek to adapt their decision so as to maximize system performance. More importantly, noisy state information (with possibly unequal noise levels) at the agents prevents them from reliably predict each other decision hence coordinate. For instance, a reliable distributed resource allocation scheme in an interference limited mobile network is difficult to obtain when the available channel state information is very noisy at some of the radio devices.

In this paper we consider a system where two agents wish to cooperatively coordinate their decisions based on possibly imperfect observations of the system state X, a so called *distributed state information* configuration. When the state information is perfectly shared among all agents, perfect coordination is achieved by simply running at each agent an instance of a centralized decision algorithm. We call in fact such a configuration a *logically centralized* configuration. However, when this condition is not met, the situation becomes significantly more complex as each agent needs to take its decision as a function of its beliefs over the actions of the other agents [2]. This fall into the framework of so-called *team decision problems* [3], which have remained vastly unsolved despite significant efforts (see [4] for an overview related to wireless communications). An alternative approach to alleviate the complexity of such problems is to consider systems with *hierarchical* state information, in which one agent has access to the state information of the other agent, or equivalently, to the decision taken at the other agent. In that setting, the coordination task is alleviated and efficient low complexity heuristics are known [4].

In this work, we focus on enforcing hierarchical state information through quantization. Indeed, it is clear that quantizing the information (or the decision) makes it easier to successfully guess it at the other node. However, it is intuitive that there is a trade-off between the information loss, or distortion, introduced by the quantizer, and the predictability, which is here measured in terms of the probability that the other agent guesses the right value.

The main goal of this paper is precisely to investigate the

L. Miretti, P. de Kerret, and D. Gesbert are supported by the European Research Council under the European Union's Horizon 2020 research and innovation program (Agreement no. 670896).

fundamental interplay between distortion and predictability, and to discuss algorithms achieving performances as close as possible to the fundamental limits.

2. SYSTEM MODEL AND PROBLEM STATEMENT

Let us consider a source symbol $X \in \mathcal{X}$, and two correlated observations $Y_1 \in \mathcal{Y}_1, Y_2 \in \mathcal{Y}_2$ distributed according to $p_{XY_1Y_2}$ available at two different agents. We let then $Z \in \mathcal{Z}$ be a quantized representation of X obtained from the observation Y_1 at Agent 1, where $\mathcal{Z} \subset \mathcal{X}$ is a finite codebook of cardinality $|\mathcal{Z}| = M$. Furthermore, we let $\hat{Z} \in \mathcal{Z}$ be an estimate of Z obtained from Y_2 at Agent 2.

We wish to design the quantizer at Agent 1 and the estimator at Agent 2 such that Z can be reliably estimated, while minimizing an expected distortion. More precisely, we look at the problem

$$\begin{aligned} &\text{minimize} \quad D = \mathbb{E}[d(X,Z)] \\ &\text{subject to} \quad P_e = P(\hat{Z} \neq Z) \leq \epsilon, \end{aligned} \tag{1}$$

where $d : \mathcal{X} \times \mathcal{X} \to \mathbb{R}^+$ is an arbitrary distortion measure, and where $\epsilon \in [0,1]$ defines a given estimation reliability tolerance. The minimization is carried over the set of tuples (α, β, γ), where $\alpha : \mathcal{Y}_1 \to \mathcal{I} = \{1, \ldots, M\}$ is an encoder that maps $\mathcal{Y}_1$ into indexes, $\beta : \mathcal{I} \to \mathcal{Z}$ is a decoder that bijectively maps the indexes into the reconstruction codewords $z_i = \beta(i)$, and where $\gamma : \mathcal{Y}_2 \to \mathcal{I}$ is an estimator that maps Y_2 into estimated indexes. We assume w.l.o.g. $z_i \neq z_j$, for all $i \neq j$, so that, in order to estimate Z, Agent 2 needs only to produce an estimate $\hat{i} = \gamma(y_2)$ of $i = \alpha(y_1)$ and then apply $\hat{z}_i = \beta(\hat{i})$.

Let us denote with $D^\star(\epsilon)$ the optimum of Problem (1) for a given predictability ϵ. The function $D^\star(\epsilon)$ is non-increasing, but in general non-convex. However, any point in the convex envelope of $D^\star(\epsilon)$ can be achieved by allowing time-sharing between two solutions (α, β, γ) and $(\alpha, \beta, \gamma)'$. Hence, the goal of this work is to provide a characterization of the *distortion-predictability function*

$$D(\epsilon) = \mathrm{Conv}\{D^\star(\epsilon)\}. \tag{2}$$

A conventional technique for characterizing the convex envelope (2) of the curve $D^\star(\epsilon)$ is to minimize the Lagrangian functional

$$L_\lambda(\alpha, \beta, \gamma) := \mathbb{E}[d(X, \beta(\alpha(Y_1))] + \lambda P(\gamma(Y_2) \neq \alpha(Y_1)) \tag{3}$$

for a fixed Lagrangian multiplier $\lambda \in [0, \infty)$. In fact, for problems of the type of (1), the optimal $L^\star(\lambda)$ identifies a line of slope λ supporting the convex envelope of $D^\star(\epsilon)$. Hence, by sweeping λ from zero to infinity, and given that we can compute the minimum of $L_\lambda(\alpha, \beta, \gamma)$, it is possible to completely characterize $D(\epsilon)$.

For the aforementioned reasons, in the reminder of this paper, we focus on the following unconstrained problem

$$\underset{(\alpha,\beta,\gamma)}{\text{minimize}} \quad L_\lambda(\alpha, \beta, \gamma), \quad \text{for } \lambda \geq 0. \tag{4}$$

3. CHARACTERIZATION OF THE DISTORTION-PREDICTABILITY FUNCTION

In this section we propose an algorithm for solving (4), hereafter referred to as the *Predictability Constrained Quantizer* (PCQ), that guarantees convergence to a local optimum. The proposed algorithm is a modification of the celebrated Lloyd algorithm [5], in a similar spirit to [6], i.e. based on an alternating minimization procedure.

3.1. Necessary conditions for optimality

We start by recalling a necessary conditions for the optimality of the decoder β, known as the *centroid* condition, developed for standard vector quantization [7] and readily extensible to the problem considered in here.

Lemma 1. *Let (α, γ) be any tuple of encoders and estimators. Define the set $\mathcal{I}_{\text{eff}} := \{i \in \mathcal{I} \,|\, P(\alpha(y_1) = i) > 0\}$. Then, the optimal decoder $\beta^\star(i) = z_i^\star$ that minimizes the Lagrangian* (3) *is given by*

$$\forall i \in \mathcal{I}_{\text{eff}}, \quad z_i^\star \in \arg\min_{z \in \mathcal{X}} \mathbb{E}[d(X,z) \,|\, \alpha(Y_1) = i].$$

The points $z_i^\star$ are called the centroids of the quantization regions $\{y_1 \in \mathcal{Y}_1 \,|\, \alpha(y_1) = i\}$.

Proof. Since the constraint set of problem (1) does not depend on β, and so does the term $\lambda P(\gamma(Y_2) \neq \alpha(Y_1))$ of (3), the proof follows directly from the classical result in [7] for unconstrained minimization. ☐

Note that for $i \in \mathcal{I} \setminus \mathcal{I}_{\text{eff}}$, the centroids are not well-defined, but they can be selected arbitrarily without impacting the cost function. Furthermore, if $d(x,z)$ is convex in z, the centroids can be easily computed by standard convex optimization techniques. As a classical example, if the distortion measure is the squared-error $d(x,z) = \|x - z\|_2^2$ over an Euclidean space, then the centroids can be computed as

$$z_i^\star = \mathbb{E}[X | \alpha(Y_1) = i]. \tag{5}$$

Next, we state a necessary condition for the optimality of the encoder α.

Lemma 2. *Let (β, γ) be any tuple of decoders and estimators. Then, the optimal encoder $\alpha^\star$ that minimizes the Lagrangian* (3) *is given by*

$$\begin{aligned} \alpha^\star(y_1) \in \arg\min_i \Big\{ &\mathbb{E}[d(X, z_i) | Y_1 = y_1] \\ &+ \lambda P(\gamma(Y_2) \neq i | Y_1 = y_1) \Big\}. \end{aligned} \tag{6}$$

Proof. We lower bound the objective $L_\lambda(\alpha,\beta,\gamma)$ as

$$\begin{aligned}L_\lambda(\alpha,\beta,\gamma) &= \int_{\mathcal{Y}_1}\Big\{\mathbb{E}[d(X,\beta(\alpha(y_1))|Y_1=y_1]\\ &\qquad + \lambda P(\gamma(Y_2)\neq\alpha(y_1)|Y_1=y_1)\Big\}dP_{Y_1}(y_1)\\ &\geq \int_{\mathcal{Y}_1}\min_i\Big\{\mathbb{E}[d(X,z_i)|Y_1=y_1]\\ &\qquad + \lambda P(\gamma(Y_2)\neq i|Y_1=y_1)\Big\}dP_{Y_1}(y_1),\end{aligned}$$

which is achieved by a mapping α that minimizes the integrand almost everywhere, i.e. by (6). □

As an example, for the squared-error distortion we obtain

$$\alpha^\star(y_1)\in\arg\min_i\Big\{|\mathbb{E}[X|Y_1=y_1]-z_i|^2 + \lambda P(\gamma(Y_2)\neq i|Y_1=y_1)\Big\}.$$

Finally, we state a necessary condition for the optimality of the estimator γ.

Lemma 3. *Let (α,β) be any tuple of encoder and decoder. Then, the optimal estimator $\gamma^\star$ that minimizes the Lagrangian* (3) *is given by*

$$\gamma^\star(y_2)\in\arg\max_i P(\alpha(Y_1)=i|Y_2=y_2),$$

which corresponds to the maximum a posteriori (MAP) estimator.

Proof. Similar to the proof of Lemma 2, hence omitted. □

We conclude this part by noticing that the above necessary conditions can be equivalently stated by designing the quantizer according to a modified distortion measure $\tilde{d}:\mathcal{Y}_1\times\mathcal{X}\to\mathbb{R}^+$ operating directly on the noisy source Y_1, given by [8]

$$\tilde{d}(y_1,z) := \mathbb{E}[d(X,z)|Y_1=y_1], \tag{7}$$

since $D=\mathbb{E}[d(X,Z)]=\mathbb{E}[\mathbb{E}[d(X,Z)|Y_1]]=\mathbb{E}[\tilde{d}(Y_1,Z)]$.

3.2. The PCQ algorithm

We now describe the proposed PCQ algorithm for solving Problem (4), based on the necessary conditions for optimality developed in the previous section.

Starting from an arbitrary initial tuple $(\alpha^{(0)},\beta^{(0)},\gamma^{(0)})$, and fixing a predictability weight $\lambda\in[0,\infty)$, we form a sequence $\{(\alpha^{(k)},\beta^{(k)},\gamma^{(k)})\}_{k=0}^\infty$ defined by the following alternating minimization procedure:

$$\begin{aligned}\alpha^{(k+1)} &\in \arg\min_\alpha L_\lambda(\alpha,\beta^{(k)},\gamma^{(k)})\\ \beta^{(k+1)} &\in \arg\min_\beta L_\lambda(\alpha^{(k+1)},\beta,\gamma^{(k)})\\ \gamma^{(k+1)} &\in \arg\min_\gamma L_\lambda(\alpha^{(k+1)},\beta^{(k+1)},\gamma).\end{aligned}$$

Every step of iteration $k\geq 1$ is obtained by applying the optimality conditions in Lemma 1, Lemma 2, and Lemma 3, and the modified distortion measure (7). In case of multiple minima, the algorithm picks one solution at random. By definition, the sequence $L_\lambda^{(k)} := L_\lambda(\alpha^{(k)},\beta^{(k)},\gamma^{(k)})$ satisfies $0\leq L_\lambda^{(k+1)}\leq L_\lambda^{(k)}$, hence the convergence of the algorithm to a local minimum is guaranteed. In practice, the algorithm is executed until a given stopping criterion $\left(L_\lambda^{(k+1)}-L_\lambda^{(k)}\right)/L_\lambda^{(k+1)}\leq\delta$ is met. The algorithm is summarized in Figure 2.

It is easy to see that for $\lambda=0$ the PCQ algorithm reduces to the classical unconstrained quantizer of [7, 8]. For $\lambda\to\infty$ instead, where the distortion D becomes asymptotically negligible, the Lagrangian is minimized by a perfectly predictable quantizer of rate zero, where all the probability mass is put on a single quantization point $z\in\arg\min_z\mathbb{E}[X]$.

As for the classical Lloyd algorithm, it is important to underline that the PCQ algorithm does not specify the codebook cardinality M. Indeed, the larger M, the more precise is the characterization of $D(\epsilon)$, as the optimization problem (1) is defined for any M, and it should be in principle optimized. Clearly, for the extreme point $D(1)$ the optimal number of codewords is often unbounded. In practice, we set M as high as possible with respect to the computation capabilities. However, for the region of interest of most applications, i.e. for ϵ relatively small, the PCQ algorithm usually puts zero probability mass in most of the quantization regions, resulting in a low-rate quantizer.

Finally, since the PCQ algorithm cannot guarantee global optimality, the trade-off region $D(\epsilon)$ is characterized only in terms of an achievable upper bound.

3.3. Statistical inference of the PCQ algorithm

In some applications the distribution $p_{XY_1Y_2}$ might be unknown, or the update rules in Figure 2 might be too difficult to be treated analytically. However, given the availability of a training sequence $\{x_1^N,y_1^N,y_2^N\}$ generated according to the joint distribution $\prod_{n=1}^N p_{XY_1Y_2}(x,y_1,y_2)$, it is still possible to approximately infer the PCQ.

Similarly to [7], we can consider the minimization of the functional

$$\hat{L}_\lambda(\alpha,\beta,\gamma) := \frac{1}{N}\sum_{n=1}^N\Big\{\tilde{d}\left(y_{1n},\beta(\alpha(y_{1n}))\right) + \lambda\mathbb{1}[\gamma(y_{2n})\neq\alpha(y_{1n})]\Big\}, \tag{8}$$

where $\mathbb{1}[\cdot]$ denotes the indicator function. This 'empirical average' functional corresponds to the 'expected' functional (3) with respect to the sample distribution of the training data. By the strong law of large number, we have that $\hat{L}_\lambda\to L_\lambda$ for $N\to\infty$ almost surely. It is possible to minimize (8) by simply substituting all the statistical operators $\mathbb{E}[\cdot]$ and $P(\cdot)$

Algorithm 1: PCQ

Set $(\alpha^{(0)}, \beta^{(0)}, \gamma^{(0)})$, $\lambda \geq 0$, $M \gg 1$, $\delta > 0$, $L_\lambda^{(0)} = \infty$

while $(L_\lambda^{(k+1)} - L_\lambda^{(k)})/L_\lambda^{(k+1)} > \delta$ **do**

$\quad \alpha(y_1) = \arg\min_i \{\tilde{d}(y_1, \beta(i)) + \lambda P(\gamma(Y_2) \neq i | Y_1 = y_1)\}$;

$\quad \beta(i) = \arg\min_{z \in \mathcal{X}} \mathbb{E}[\tilde{d}(Y_1, z) \,|\, \alpha(Y_1) = i]$;

$\quad \gamma(y_2) = \arg\max_i P(\alpha(Y_1) = i | Y_2 = y_2)$;

end

Fig. 2. The proposed PCQ algorithm for finding a local minimum of $L_\lambda(\alpha, \beta, \gamma)$.

in Figure 2 with their respective empirical operators, i.e. their expression with respect to the sample distribution of the training sequence $\{y_1^N, y_2^N\}$. Note that (α, γ) are defined over the finite alphabets $\hat{\mathcal{Y}}_1, \hat{\mathcal{Y}}_2$ composed by the occurrences of Y_1, Y_2 in the training data. The resulting scheme is indeed a clustering algorithm over the training set.

The method described above has some important limitations. Firstly, it cannot be applied to continuous alphabets. In fact, due to the particular structure of the sample distribution derived from continuous alphabets, it can be easily seen that, $\forall \lambda$, the γ update step degenerates to $\gamma^{(k+1)}(y_{2n}) = \alpha^{(k)}(y_{1n})$ almost surely, leading to extremely suboptimal solutions. Secondly, it requires the knowledge of $\tilde{d}$, i.e. at least the knowledge of the marginal p_{XY_1}. Similarly to [8], if this information is not available, it is not possible to adapt the previous scheme in a straight-forward way.

A possible heuristic to address the first problem is to consider instead a discretized version (Y_{1Q}, Y_{2Q}) of (Y_1, Y_2), obtained by applying some quantizer of properly tuned (finite) rate. For the second problem, one can obtain approximations of $\tilde{d}$ from samples $\{x_1^N, y_1^N\}$ by applying density estimation techniques [9] (see [10] for the squared-error distortion).

4. EXAMPLE

We consider Y_1 and Y_2 to be Gaussian noisy measurements of a Gaussian source $X \sim \mathcal{N}(0, 1)$. More specifically, we let

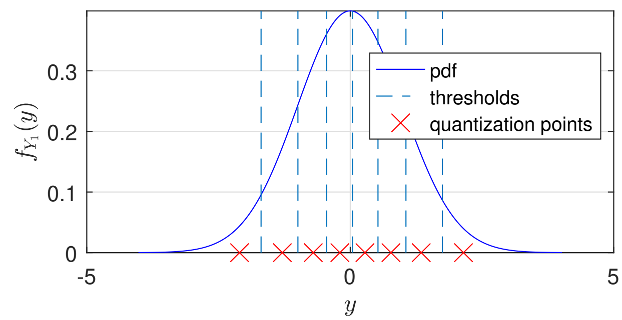
$$Y_1 = X + N_1, \quad N_1 \sim \mathcal{N}(0, \mathrm{SNR}_1^{-1}),$$
$$Y_2 = X + N_2, \quad N_2 \sim \mathcal{N}(0, \mathrm{SNR}_2^{-1}),$$

for a given pair of signal-to-noise ratios (SNR), and where N_1, N_2, and X are independent. As distortion measure, we choose the squared-error distortion $d(x, z) = (x - z)^2$.

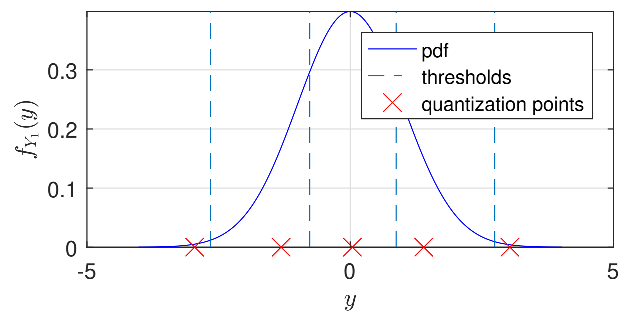
Under these assumptions, it is easy to see that

$$\tilde{d}(y_1, z) = (\kappa y_1 - z)^2 + c,$$

where $\kappa = \sigma_{XY_1}\sigma_{Y_1}^{-2}$ is the minimum mean-square error (MMSE) estimator of X given Y_1, and where c is a constant



(a) $\lambda = 0$, $M = 8$, $\mathrm{SNR}_1 = 30\mathrm{dB}$, $\mathrm{SNR}_2 = 20\mathrm{dB}$.



(b) $\lambda = 10$, $M = 8$, $\mathrm{SNR}_1 = 30\mathrm{dB}$, $\mathrm{SNR}_2 = 20\mathrm{dB}$

Fig. 3. Illustration of the behaviour of the quantizer for 2 different choices of the predictability weight λ. In Figure (a), for $\lambda = 0$, the solution is similar to the classical M-levels Lloyd quantizer. In Figure (b), for $\lambda = 10$, we see that the PCQ algorithm produces fewer quantization regions containing more probability mass.

that depends only on the conditional distribution $p_{X|Y_1}$ and hence it can be neglected through the optimization steps of the PCQ algorithms (it is indeed the MMSE value itself, i.e. the distortion floor for every quantizer on Y_1). The decoder update rule is then given by

$$\beta(i) = \kappa \mathbb{E}[Y_1 | \alpha(Y_1) = i],$$

which is evaluated analytically in terms of Q-functions. The encoder update is obtained numerically by discretizing the alphabet $\mathcal{Y}_1$, by evaluating analytically the augmented distances $\tilde{d}(y_1, \beta(i)) + \lambda P(\gamma(Y_2) \neq i | Y_1 = y_1)$ for every i and y_1, and by solving the optimization problem through exhaustive search over $i = 1, \ldots, M$. The estimator update rule is obtained in a similar way. The PCQ algorithm is run several times with random initialization $(\alpha^{(0)}, \beta^{(0)}, \gamma^{(0)})$, and for several λ. An achievable $D(\epsilon)$ curve is obtained by taking the convex hull of the obtained (D, ϵ) points.

Figure 3 illustrates the effect of the choice of λ on the output of the PCQ algorithm. Note that we only plot the thresholds and the quantization points of the quantizer $\beta(\alpha(\cdot))$ applied at agent 1, since γ simply corresponds to a MAP estimator.

Figure 4 shows the achievable $D(\epsilon)$ curve obtained via the PCQ algorithm, normalized by the Minimum Mean Squared Estimation error (MMSE) of X given Y_1. We also show the performance of a heuristic approach based on the Lloyd algorithm for noisy sources [8], and where different degrees of predictability are enforced by simply varying the number M of quantization points. We denote this second approach as *Predictable Lloyd Quantizer* (PLQ).

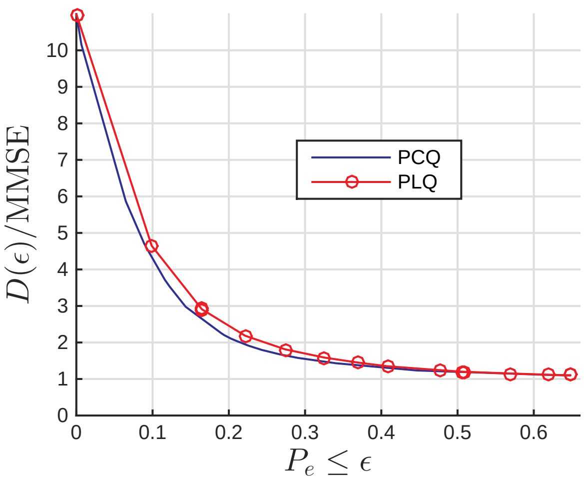


Fig. 4. Achievable distortion-predictability function for the considered jointly Gaussian example with squared-error distortion. $M = 32$, $\mathrm{SNR}_1 = 30\mathrm{dB}$, $\mathrm{SNR}_2 = 20\mathrm{dB}$.

5. DISCUSSION

In this preliminary study we describe a method for exploring the trade-off between *predictability* and *distortion* in a setting composed by a couple of distributed measurements Y_1, Y_2 of a random system state X. As a natural next step, we are currently investigating the benefits of applying the quantizer developed in here to obtain suboptimal solutions to team decision problems. Specifically, by limiting the optimization space to the class of decision functions derived by *hierarchical* information structure, a close-to-optimal operating point can be obtained by varying the parameter λ of the proposed PCQ algorithm, applied to Y_1 as a pre-processing step to enhance coordination. Note that a similar idea has been successfully applied in [11] for the problem of distributed precoding in cooperative wireless networks.

6. REFERENCES

[1] D. Gesbert, S. Hanly, H. Huang, S. Shamai Shitz, O. Simeone, and W. Yu, “Multi-cell mimo cooperative networks: A new look at interference,” *IEEE Journal on Selected Areas in Communications*, vol. 28, no. 9, pp. 1380–1408, December 2010.

[2] J. C. Harsanyi, “Games with incomplete information played by Bayesian players,” *Management Science*, 1967.

[3] R. Radner, “Team decision problems,” *The Annals of Mathematical Statistics*, 1962.

[4] D. Gesbert and P. de Kerret, “Team methods for device cooperation in wireless networks,” in *Cooperative and Graph Signal Processing*, pp. 469–487. Elsevier, 2018.

[5] S. Lloyd, “Least squares quantization in PCM,” Unpublished Bell Laboratories Technical Note. Published in the March 1982 special issue on quantization of the *IEEE Transactions on Information Theory*, 1957.

[6] P. A. Chou, T. Lookabaugh, and R. M. Gray, “Entropy-constrained vector quantization,” *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 1989.

[7] Y. Linde, A. Buzo, and R. Gray, “An algorithm for vector quantizer design,” *IEEE Transactions on communications*, vol. 28, no. 1, pp. 84–95, 1980.

[8] R. M. Ephraim, Y.and Gray, “A unified approach for encoding clean and noisy sources by means of waveform and autoregressive model vector quantization,” *IEEE Transactions on Information Theory*, 1988.

[9] B. W. Silverman, *Density Estimation for Statistics and Data Analysis*, Chapman and Hall, 1986.

[10] Y. A. Ghassabeh and F. Rudzicz, “Noisy source vector quantization using kernel regression,” *IEEE Transactions on Communications*, vol. 62, no. 11, pp. 3825–3834, 2014.

[11] A. N. Bazco, L. Miretti, P. de Kerret, D. Gesbert, and N. Gresset, “Achieving vanishing rate loss in decentralized network MIMO,” in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*. 2019, IEEE.