

An Information-Theoretic Analysis of Distributed Resource Allocation

Tansu Alpcan and Subhrakanti Dey

Abstract—Solving a resource allocation problem in a distributed way requires communication between the system and its users. This information exchange is, however, limited by communication constraints, delays, and distortions in most practical problems. This paper presents a quantitative analysis of information (flow) in a well-known distributed resource allocation algorithm using concepts from Shannon information theory. For this purpose, an entropy-based measure is adopted to quantify information which is defined as uncertainty reduction. Then, information flow in a certain class of iterative algorithms is studied. The relationships between the rate and total amount of information exchanged, and convergence of the algorithm are investigated under certain assumptions. The concepts introduced and the obtained results are illustrated using numerical examples.

I. INTRODUCTION

Distributed resource allocation problems are encountered in a wide variety of systems ranging from electrical power grid to communication networks [4], [13], [24], [27]. Their common characteristics include heterogeneous preferences of individual users sharing the limited resources and the desirability –sometimes necessity– of a decentralized solution which involves information exchange between the users and the system. Iterative, e.g. gradient, algorithms have been used in the literature extensively to address such problems [3], [14], [16], [17], [27], [29].

Although most of the iterative resource allocation schemes in the literature involve communication between the users and the system, the information aspect of the problem is often treated only implicitly. As a rare example, the effect of using single-bit marks for signaling link price information is investigated in [7] in the specific context of congestion control. While recent works on networked control with information constraints [11], [12], [15], [18], [20]–[22], [25], [28] study the interplay between stability and information, they do not focus on explicitly quantifying information flow in the context of distributed resource allocation using information-theoretic concepts.

This paper presents an initial set of results on quantifying information explicitly in iterative (gradient) algorithms for distributed resource allocation using concepts from information theory. Iterative methods which converge to a unique solution, especially the well-known primal (gradient)

algorithm [17], [27], are analyzed to illustrate the concepts introduced. Building upon the earlier results on the role information plays in optimization [1], [2], an information measure based on the concept of uncertainty reduction is defined to quantify the information gain, and hence information flow, in the iterative algorithms analyzed. Specifically, *entropy power* [26] which is closely related to differential entropy is utilized as a measure of uncertainty.

The results presented in this paper constitute an initial step toward answering challenging questions such as (1) what are the information-related implications of iterative algorithms? (2) how to quantify information flow and the effects of information limitations? (3) how to design optimal communication systems for distributed resource allocation? A better quantitative understanding of information will help answering these questions through a novel synthesis of distributed optimization, learning, and information theory.

The main contributions of the paper are:

- 1) Establishing the relationships between the convergence of a certain class of iterative algorithms for distributed resource allocation and the information gain at each iteration.
- 2) A novel use of information-theoretic concepts such as entropy power and Shannon information (defined as uncertainty reduction) in the context of distributed optimization.
- 3) Quantitative analysis of information flow in (primal) gradient algorithms under *communication limitations*.
- 4) Investigation of the trade-off between the total amount of information (data) exchanged until convergence of a distributed algorithm and the information rate (at each iteration) under certain simplifying assumptions.

The rest of the paper is organized as follows. The next section provides an overview of a well-known distributed resource allocation problem and its iterative solution. Section III introduces basic information-theoretic concepts within the context of the problem considered. A set of results on the relationship between the convergence properties of the iterative algorithm and their information implications are presented under the assumption of perfect communication (with infinite bandwidth) between the system and users. Subsequently, Section IV quantifies the information flow in the system and explores the trade-off between total amount of data exchanged until the algorithm converges versus the communication limitations. The paper ends with the concluding remarks and a discussion on the future work in Section V.

This work was supported by an Interdisciplinary Seed Funding at The University of Melbourne.

T. Alpcan is with the Department of Electrical and Electronic Engineering, the University of Melbourne, Australia. tansu.alpcan@unimelb.edu.au S. Dey is with the Uppsala University, Sweden. Subhra.Dey@signal.uu.se He was with the University of Melbourne during part of this work.

II. A DISTRIBUTED AND ITERATIVE ALGORITHM FOR ALLOCATION OF A SINGLE RESOURCE

The problem of allocating a limited resource to independent users arises in a variety of systems. A representative example is distributing a fixed divisible quantity C such as bandwidth, water, or energy to a set of N users, \mathcal{A} . Let the scalar $x_i \geq 0$ denote the share each user $i \in \mathcal{A}$ receives from the resource, C . Then, the resource constraint is formulated as

$$\sum_{i=1}^N x_i \leq C. \quad (1)$$

Assuming that every user has a possibly different preference for the resource, the preference of user i is captured by the utility function $U_i(x_i) : \mathbb{R}^+ \rightarrow \mathbb{R}$, which is assumed to be continuous, twice-differentiable, and concave for analytical tractability. Concavity property of the utility function is motivated by the fact the marginal “utility” decreases for a large variety of resources. In some settings, even the actual utility may decrease capturing the situation where having too much of the resource has a negative effect on the user, e.g. having too many pieces of pie results in stomach pain, if the resource shared is a big pie. Based on the preferences of users, U , a commonly used objective is to maximize the aggregate utility of the users under the resource constraint (1), which leads to the optimization problem:

$$\max_x V(x) = \sum_i U_i(x_i) \text{ subject to } \sum_i x_i \leq C, \quad (2)$$

where $x = [x_1, \dots, x_N]$.

The “system problem” in (2) is a “convex” optimization problem due to concavity of $U_i \forall i$. Hence, it can be solved using standard methods in a centralized manner [10] yielding the unique solution (x^*, λ^*) , if the system knows the private preferences of all the users.

The Lagrangian function associated with (2) is

$$\mathcal{L}(x) = \sum_i U_i(x_i) + \lambda \left(C - \sum_i x_i \right),$$

where $\lambda > 0$ is a scalar Lagrange multiplier.

In many resource allocation problems there is no way of directly asking users of their preferences and a decentralized solution to (2) is more desirable or even necessary. For this purpose, the objective function $V(x)$ is decomposed [9] to obtain N convex user optimization problems:

$$\max_{x_i} U_i(x_i). \quad (3)$$

If each user solves (3) independently, then the resource constraint (1) may be violated, which results in a lose-lose situation for participants called “tragedy of commons” or “the price of anarchy” [23]. Therefore, the users should be informed of the resource limitations and be given an incentive to act accordingly. A standard way of addressing this issue is to modify the user problem by introducing a signaling or “pricing” term leading to

$$\max_{x_i} U_i(x_i) - p x_i, \quad (4)$$

where p acts as an incentive signal effectively imposing the resource constraint. It was shown in [5], [6] that a linear signaling term is sufficient. Next, the signal p is chosen such that the global system problem (2) is aligned with the N local user problems (4). Due to convexity of user problems, the first-order necessary and sufficient condition $dU_i/dx_i - p = 0$ indicates that simply choosing $p = \lambda$ results in the desired alignment. Hence, the problem reduces to finding $\lambda^* = p^*$ in a distributed way.

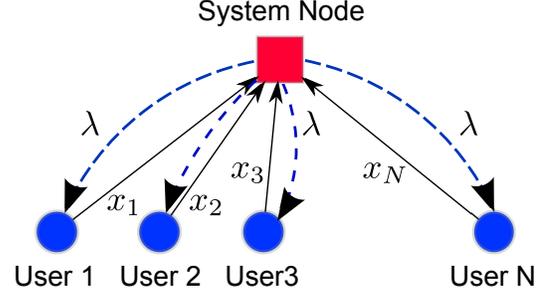


Fig. 1. A visualization of the distributed system in Algorithm 1. The communication between the system node and users may be imperfect in many problems.

A solution to the joint user and system problem is provided by the iterative distributed optimization scheme summarized in the Algorithm 1, which is visualized in Figure 1 and represented by the mapping $T : \mathbb{R}^{N+1} \rightarrow \mathbb{R}^{N+1}$,

$$\text{Users: } x_i(t+1) = f_i^u(x_i(t), \lambda(t)), \quad i = 1, \dots, N,$$

$$\text{System: } \lambda(t+1) = f^\lambda(x(t), \lambda(t)), \quad (5)$$

$$\iff z(t+1) = T(z(t)), \quad t = 0, 1, \dots$$

where the variable $z := [x, \lambda]$ is defined for notational convenience.

Algorithm 1 Iterative Distributed Optimization Algorithm

- 1: **repeat**
 - 2: *System*: observe user actions, $x_i \forall i$ or $\sum_i x_i$.
 - 3: *System*: update the signal $\lambda (= p)$ taking into account $\sum_i x_i$ and C .
 - 4: **for** *User* $i \in \mathcal{A}$ **do**
 - 5: update x_i by solving the user problem (4).
 - 6: **end for**
 - 7: **until** $z = (x, \lambda)$ converges to $z^* = (x^*, \lambda^*)$, the unique solution of the system problem (2).
-

The iterative approach of Algorithm 1 (5) is applicable to a broad class of problems. For example, one or multiple users can collaboratively act as a virtual “system node” if there is no physical node to handle this task. An underlying assumption here is that although users have their own preferences, they do not try to cheat the system by misrepresenting them. The research field of mechanism design investigates that problem [19], which is outside the scope of this paper.

III. INFORMATION GAIN IN ITERATIVE ALGORITHMS

The first step towards analysis of Algorithm 1 using an information-theoretic approach is to define the vector

z^* , which is the (unique) solution of (5), as the unknown quantity to be “learned”. For most problems, z belongs to a broad compact set $\mathcal{Z} := [0, \Psi]^{N+1}$, where Ψ is chosen sufficiently large such that $x_i \leq \Psi \forall i$ and $\lambda \leq \Psi$. The maximum uncertainty of z , i.e. complete lack of information, is captured by choosing $\phi_0(z)$, the joint probability density function (pdf) on \mathcal{Z} , as uniform. The next definition presents information-theoretic measures relevant to the problem.

Definition 1. Let $\phi(z)$ be the uniform joint probability density function (pdf) on $\mathcal{Z} := [0, \Psi]^{N+1}$ that captures the uncertainty of the vector $z \in \mathcal{Z}$ due to not knowing its actual value, $z^* \in \mathcal{Z}$, which is the (unique) solution of (5). Then, the following hold:

- 1) A variant of entropy power [26], $h^e(z) : \mathcal{Z} \rightarrow [0, \infty)$, defined as

$$h^e(z) := e^{\frac{1}{N+1} h(z)}$$

quantifies the uncertainty of $z \in \mathcal{Z}$, where

$$h(z) = - \int_{\mathcal{Z}} \phi(z) \log(\phi(z)) dz$$

is the differential entropy.

- 2) The maximum uncertainty of $z \in \mathcal{Z}$ is

$$h_0^e(z) = \Psi.$$

Here, entropy power [26] is used to measure the uncertainty of z instead of differential entropy because it has a bounded range whereas differential entropy may diverge to negative infinity in some cases.

Next, the information gain on $z(t)$ after each iteration is defined as a reduction in its uncertainty using Definition 1.

Definition 2. The information gain at iteration t of Algorithm 1 (and mapping (5)) is defined as the difference between the uncertainty (entropy power) before and after the iteration,

$$I(t+1) := h^e(z(t)) - h^e(z(t+1)).$$

The information metric in Definition 2 is related to yet different from standard mutual information $I(X; Y) = h(X) - h(X|Y)$ between two random continuous variables X, Y with a joint density function. In this case, the result of the iteration at time t implicitly plays the role of the hidden variable Y , which leads to a slightly different interpretation of information outside the context of communication.

The following straightforward yet important result establishes a relationship between the geometric convergence of the Algorithm 1 and the information obtained at each iteration of the mapping (5).

Theorem 3. Let the mapping $z(t+1) = T(z(t))$ in (5) be a contraction

$$\|T(z_1) - T(z_2)\|_{\infty} \leq \alpha_t \|z_1 - z_2\|_{\infty}, \forall z_1, z_2 \in \mathcal{Z},$$

where $\alpha \in [0, 1)$ and $\|\cdot\|_{\infty}$ denotes the maximum (infinity) norm. Define the infinite sequence

$$\{\alpha_t := \|z(t+1) - z^*\|_{\infty} / \|z(t) - z^*\|_{\infty}, t = 1, 2, \dots\}$$

Then, the following hold:

- 1) The mapping T has a unique fixed point $z^* \in \mathcal{Z}$.
- 2) The iteration $z(t+1) = T(z(t))$ converges to z^* geometrically from any starting point $z(0) \in \mathcal{Z}$,

$$\|z(t) - z^*\|_{\infty} \leq \alpha^t \|z(0) - z^*\|_{\infty}, \forall t \geq 0.$$

- 3) The entropy power $h^e(z(t))$ is bounded by

$$0 \leq h^e(z(t)) \leq h_0^e e^{-t(N+1) \ln(1/\alpha)},$$

and converges to zero

$$\lim_{t \rightarrow \infty} h^e(z(t)) = 0.$$

- 4) The information gain converges to zero,

$$\lim_{t \rightarrow \infty} I(t) = 0.$$

Proof. [Overview]

Parts 1) and 2) directly result from Proposition 1.1 in [8, Chap 3, pp 182].

Part 3) follows from 2) and Definition 1. The differential entropy of z is bounded above by

$$\begin{aligned} h(z(t)) &\leq \min \left[\sum_{i=1}^N \ln(\alpha^t 2 |x(0) - x_i^*|) \right. \\ &\quad \left. + \ln(2\alpha^t |\lambda(0) - \lambda^*|), (N+1) \ln(\Psi) \right] \\ &\leq -t(N+1) \ln(1/\alpha) + (N+1) \ln(\Psi) \\ &\leq h_0 - t(N+1) \ln(1/\alpha). \end{aligned}$$

Part 4) immediately follows from Part 3) and Definition 2. \square

Numerical Example 1 [Information Gain]:

Consider a gradient algorithm as a special case of (7), with $C = 100$, $\mu = 0.01$, and $U_i(x_i) = \alpha_i \log(x_i)$ such that $\sum_i \alpha_i = 100$. The following evolution of the Lagrangian multiplier

$$\lambda(t+1) = \lambda(t) + \mu \left(\frac{100}{\lambda(t)} - C \right). \quad (6)$$

admits the unique solution $\lambda^* = 2$. Let $\lambda(0) = 4$ which results in the initial (maximum) uncertainty $h_0^e = 4$. The evolution of uncertainty of λ and the information gain for each iteration of (6) are depicted in Figures 2 and 3, respectively. It can be observed in Figure 3 that $\lim_{t \rightarrow \infty} I(\lambda(t)) = 0$.

IV. INFORMATION FLOW IN GRADIENT ALGORITHMS

In Section III, Theorem 3 presents an information-theoretic method for analyzing the outcome of (5). However, it was assumed that all the variables involved, e.g. z , are of infinite precision and the information exchange between users and the system occurs over infinite bandwidth in Algorithm 1. Next, the information flow in the algorithm and its effect on the outcome are investigated by relaxing these commonly used assumptions.

A well-known special case of the Algorithm 1 is when a gradient search is performed on the system problem while

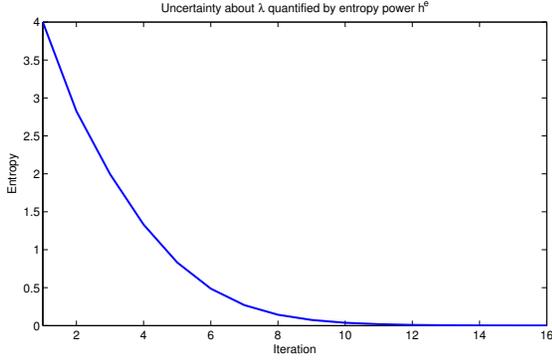


Fig. 2. The evolution of the uncertainty of λ in (6) quantified by power entropy, $h^e(\lambda(t))$, in Definition 1.

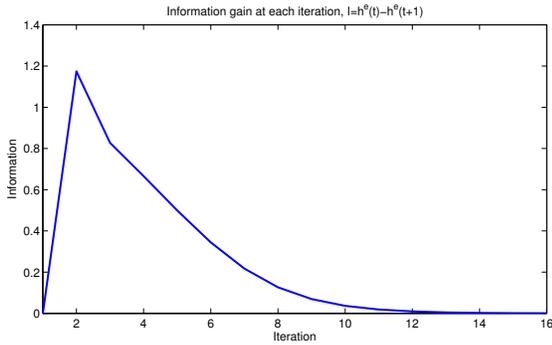


Fig. 3. The evolution of information I obtained on λ in each iteration of (6) quantified by the information metric in Definition 2.

the users adopt a best response scheme. Hence, the mapping (5) becomes the *primal algorithm* [17], [27]:

$$\begin{aligned} \text{Users: } x_i(t+1) &= \arg \max_{x_i} U_i(x_i) - \lambda(t)x_i, \quad \forall i, \\ \text{System: } \lambda(t+1) &= \lambda(t) + \mu (\sum_i x_i(\lambda(t)) - C), \end{aligned} \quad (7)$$

where μ is the step-size constant. Under the strict concavity conditions imposed on U , each user problem above admits a unique solution corresponding to a static mapping from $\lambda(t)$ to $x_i(t+1)$. Imposing a sufficient condition on μ , see e.g. [8], the primal algorithm (7) becomes a contraction and the results in Theorem 3 apply.

In real life digital systems, the communication channels from the users to the system and vice versa (see Figure 1) have limited bandwidth and are imperfect. Such information flow limitations affect user problems as well as the system problem in (7). The following assumptions summarize various relevant aspects:

- **A1** The computations at the nodes are of (practically) infinite precision.
- **A2** The vector x and scalar λ are quantized for transmission between the nodes.
- **A3** The communication channel may distort (x, λ) due to delays and errors.

The system problem in (7) provides a good starting point for the information flow analysis. Let $x_i^q(t) := x_i(t) + \varepsilon_i(t)$ be the solution of user problems in (7) communicated to the

system node under the Assumptions A1-A3, where $\varepsilon_i(t)$ represents the distortion due to quantization and communication errors. Then, the distorted information x_i^q is an input to the system problem leading to

$$\begin{aligned} \lambda(t+1) &= \lambda(t) + \mu s(t) \\ &= \lambda(t) + \mu (g(t) + \bar{\varepsilon}(t)), \end{aligned} \quad (8)$$

where $\bar{\varepsilon}(t) := \sum_i \varepsilon_i(t)$ and $g(t) := \sum_i x_i(\lambda(t)) - C$.

In the case of infinite bandwidth, i.e. $\bar{\varepsilon} = 0$, a sufficient condition for (8) converge to λ^* is $\mu \leq 2/K$, where K is the Lipschitz constant of the gradient $\sum_i x_i(\lambda(t)) - C$ [8]. On the other hand, under Assumptions A2-A3, the aggregate quantization and communication error $\bar{\varepsilon}$ does have an impact on the convergence rate of (8) as it distorts the original gradient $g(t)$.

In many problems where communication costs are not prohibitive, it is desirable to design information flow in the system such that the gradient algorithm (7) deterministically and asymptotically converges to the solution. The following theorem presents a result on the relationship between deterministic convergence of (8) and the amount of information required.

Theorem 4. *Let λ^* be the solution to (7). Define a region $R(r) := |\lambda - \lambda^*| < r$ around it, where $r \in \mathbb{R}^+$ is a positive constant. An upper-bound on the quantization and communication errors outside $R(r)$ is defined as*

$$|\bar{\varepsilon}| \leq \theta |g(t)|,$$

where $g(t) = \sum_i x_i(\lambda(t)) - C$ and $\lambda(t) \notin R(r)$. Consider a communication system (quantization level) design and choose r such that there exists a $\theta \in (0, 1)$.

If the constant step-size μ in (8) satisfies

$$\mu \leq \frac{2}{K}(1 - \theta),$$

where K is the Lipschitz constant of the gradient $g(t)$, then there exists a finite L such that for all $t > L$, $\lambda(t) \in R(r)$, i.e. the system dynamics (8) converge near the solution in finite time.

Proof. [Overview] The proof is based on the Propositions 2.1 and 2.3 in [8, pp. 203-206]. The sufficient conditions of those propositions simplify here to

$$|s(t)| \geq K_1 |g(t)| \quad \text{and} \quad s(t)g(t) \geq K_2 |g(t)|^2,$$

for positive K_1 and K_2 with $s(t)$ given in (8). Since outside the region $R(r)$ it is assumed that

$$|s(t)| \geq (1 - \theta) |g(t)|,$$

there exists a positive K_1 and a $K_2 \leq (1 - \theta)$. Thus, the restriction on the step size $0 < \mu < 2K_2/K$ satisfies the sufficient condition on μ for asymptotic convergence. Hence, the scalar $\lambda(t)$ enters the region $R(r)$ in finite time. \square

Remark 1. The sufficient conditions in Theorem 4, especially the ones on quantization and communication errors are quite stringent in order to maintain the deterministic convergence

of the algorithm. These conditions can be relaxed in certain problems where probabilistic and/or slower convergence is acceptable.

Theorem 4 indicates an interesting relationship between the information flow, convergence region, and gradient step-size affecting the convergence speed. Next, a bound on the total amount of information necessary for the trajectory $\lambda(t)$ to enter the region $R(r)$ is derived under certain simplifying assumptions.

Proposition 5. Assume that (i) the user information x_i are uniformly quantized on the interval $[0, x_{max}]$ to Q_x levels (ii) the errors ε in (8) are only due to quantization (iii) the users receive the value $\lambda(t)$ with infinite precision, i.e. the return channel has infinite bandwidth (iv) the sufficient conditions in Theorem 4 hold.

Then, for a given starting point $\lambda(0)$ and $r > 0$, an upper-bound on the total amount of user information needed for the gradient algorithm (8) to enter the region $R(r)$ is

$$I_{total}(\lambda(0), r, U) = L(Q_x, \lambda(0), r) \cdot N \cdot \ln(Q_x), \quad (9)$$

where L is the number of steps for entering the region $R(r)$ and N the number of users.

Choosing a finer quantization or increasing quantization levels, Q_x , clearly decreases the errors ε_i , $\bar{\varepsilon}$, and hence θ . A smaller θ allows for a larger step size and decreases the number of steps, L . Thus, there is a clear trade-off between the terms L and $\ln(Q_x)$ in (9).

An interesting and important question that immediately follows from the discussion above is: *what is the best information flow structure, more specifically the uniform quantization scheme that effectively determines the bit rate, for the gradient algorithm (8) to converge with minimum amount of total information exchange?* This complex question is analyzed numerically in the next example.

Numerical Example 2 [Gradient Algorithm]

A special case is of the gradient algorithm (7) is simulated for $N = 10$ users with utility functions $U_i(x_i) = \alpha_i \log(x_i)$, where α_i are randomly generated user-specific preference parameters in the interval $a_i \in [1, 10]$, $\forall i$. Each $x_i \in [0, 100]$ is uniformly quantized to $Q_x = 2^v$ levels with different number of bits in each simulation $v \in \{6, 7, \dots, 19, 20\}$. Then, the respective upper-bound on the (worst-case) aggregate quantization error $\bar{\varepsilon} = \sum_i \varepsilon_i$ is $w = N * 100 / Q_x$. The initial value in (8) is $\lambda(0) = 1.8$ for all cases and $\lambda^* = 0.65$.

Figure 4 shows the total information I_{total} sent from users to the system as defined in (9) until convergence (in the sense of entering the interval $[\lambda^* - 0.01, \lambda^* + 0.01]$) for different bit rates based on uniform quantization of x_i . The reverse channel is assumed to be of infinite bandwidth here for simplicity. Figure 5 depicts the total number of iterations until convergence. As expected, sending more precise data (more information) does not decrease the number of steps after a lower bound is reached due to the limitation on the step size. As Figures 4 and 5 indicate, there is an optimal uniform quantization scheme, specifically 12 bit uniform

quantization of x_i per iteration, that minimizes the aggregate data flow by balancing the number of iterations needed for convergence and the amount of information transmitted at each iteration.

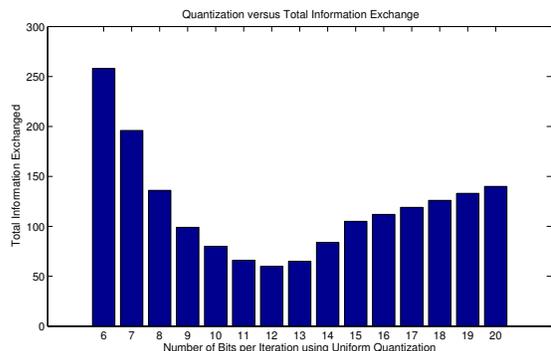


Fig. 4. Total information I_{total} flow from users to the system as defined in (9) for different number of bits, $v = \log_2(Q_x)$, per iteration in the case of uniform quantization.

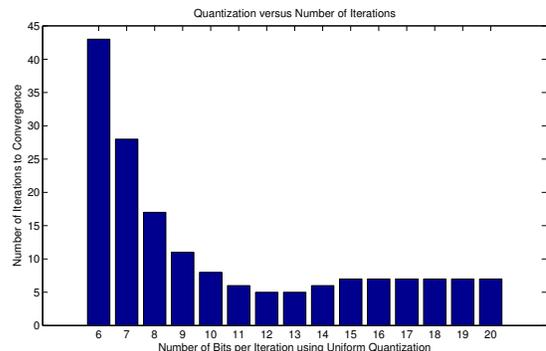


Fig. 5. The number of iterations in (9) for different number of bits, $v = \log_2(Q_x)$, per iteration in the case of uniform quantization.

V. CONCLUSION

An initial set of results is presented toward a better quantitative understanding of the role information plays in distributed resource allocation. Firstly, the information gain in a certain class of iterative algorithms for distributed allocation of a single resource is quantified by establishing the relationships between convergence and information under a perfect communication assumption. Information-theoretic concepts such as entropy (power) and information as uncertainty reduction are used for this purpose. Secondly, the information flow in (primal) gradient algorithms under communication limitations between the system and the users is analyzed. Specifically, the trade-off between the total amount of information (data) exchanged until convergence of the distributed algorithm and the information rate (at each iteration) is investigated under certain simplifying assumptions. The theoretical analysis is supported by illustrative numerical examples.

This paper differs from recent studies on networked control with information constraints, which study the interplay between stability and information, by its explicit focus on quantifying information flow in distributed optimization algorithms using information-theoretic measures. There remains a substantial amount of future work in the presented research direction such as quantitative analysis of Bayesian learning schemes for distributed optimization from an information perspective, applications to N -person strategic games, and extensions to mechanism design.

VI. ACKNOWLEDGMENTS

The authors thank Girish Nair for helpful comments and discussions.

REFERENCES

- [1] T. Alpcan, "A framework for optimization under limited information," *Journal of Global Optimization*, pp. 1–26, 2012.
- [2] —, "A Risk-Based Approach to Optimisation under Limited Information," in *Proc. of the 20th Intl. Symp. on Mathematical Theory of Networks and Systems (MTNS)*, July 2012.
- [3] T. Alpcan and T. Başar, "A Utility-Based Congestion Control Scheme for Internet-Style Networks with Delay," *IEEE Trans. on Networking*, vol. 13, no. 6, pp. 1261–1274, December 2005.
- [4] —, *Network Security: A Decision and Game Theoretic Approach*. Cambridge University Press, 2011. [Online]. Available: <http://www.tansu.alpcan.org/book.php>
- [5] T. Alpcan and L. Pavel, "Nash Equilibrium Design and Optimization," in *Proc. of Intl. Conf. on Game Theory for Networks (GameNets 2009)*, Istanbul, Turkey, May 2009.
- [6] T. Alpcan, L. Pavel, and N. Stefanovic, "A Control Theoretic Approach to Noncooperative Game Design," in *Proc. of 48th IEEE Conf. on Decision and Control*, Shanghai, China, December 2009.
- [7] L. Andrew, S. Hanly, S. Chan, and T. Cui, "Adaptive deterministic packet marking," *Communications Letters, IEEE*, vol. 10, no. 11, pp. 790–792, 2006.
- [8] D. Bertsekas and J. N. Tsitsiklis, *Parallel and Distributed Computation: Numerical Methods*. Upper Saddle River, NJ: Prentice Hall, 1989.
- [9] S. Boyd, L. Xiao, A. Mutapcic, and J. Mattingley, "Notes on decomposition methods," 2007.
- [10] S. Boyd and L. Vandenberghe, *Convex Optimization*. New York, NY, USA: Cambridge University Press, 2004.
- [11] J. S. Freudenberg, R. H. Middleton, and J. H. Braslavsky, "Minimum variance control over a Gaussian communication channel," *IEEE Trans. Automat. Contr.*, vol. 56, pp. 1751–1765, 2011.
- [12] J. S. Freudenberg, R. H. Middleton, and V. Solo, "Stabilization and disturbance attenuation over a Gaussian communication channel," *IEEE Trans. Automat. Contr.*, vol. 55, pp. 795–799, 2010.
- [13] Z. Han, D. Niyato, W. Saad, T. Başar, and A. Hjørungnes, *Game Theory in Wireless and Communication Networks: Theory, Models, and Applications*, ser. Game Theory in Wireless and Communication Networks: Theory, Models, and Applications. Cambridge University Press, 2011.
- [14] J. Huang, R. Berry, and M. Honig, "Distributed Interference Compensation for Wireless Networks," *IEEE Journal on Selected Areas in Communications*, vol. 24, no. 5, pp. 1074–1084, May 2006.
- [15] H. Ishii, K. Okano, and S. Hara, "Achievable sensitivity bounds for MIMO control systems via an information theoretic approach," *Syst. Contr. Lett.*, vol. 60, pp. 111–118, 2011.
- [16] R. Johari and J. N. Tsitsiklis, "Efficiency of scalar-parameterized mechanisms," *Operations Research*, vol. 57, no. 4, pp. 823–839, July 2009. [Online]. Available: <http://www.stanford.edu/~rjohari/pubs/char.pdf>
- [17] F. Kelly, A. Maulloo, and D. Tan, "Rate control in communication networks: Shadow prices, proportional fairness and stability," *Journal of the Operational Research Society*, vol. 49, pp. 237–252, 1998.
- [18] N. C. Martins and M. A. Dahleh, "Feedback Control in the Presence of Noisy Channels: Bode-like Fundamental Limitations of Performance," *IEEE Trans. Automat. Contr.*, vol. 53, pp. 1604–15, 2008.
- [19] E. Maskin, "Nash Equilibrium and Welfare Optimality," *Review of Economic Studies*, vol. 66, no. 1, pp. 23 – 38, December 2003.
- [20] A. S. Matveev and A. V. Savkin, "An analogue of Shannon information theory for detection and stabilization via noisy discrete communication channels," *SIAM Jour. Contr. Optim.*, vol. 46, no. 4, pp. 1323–67, 2007.
- [21] G. N. Nair and R. J. Evans, "Stabilizability of stochastic linear systems with finite feedback data rates," *SIAM Jour. Contr. Optim.*, vol. 43, no. 2, pp. 413–36, July 2004.
- [22] G. Nair, F. Fagnani, S. Zampieri, and R. Evans, "Feedback Control Under Data Rate Constraints: An Overview," *Proceedings of the IEEE*, vol. 95, no. 1, pp. 108–137, January 2007.
- [23] S. Robinson, "The Price of Anarchy," *SIAM News*, vol. 37, no. 5, pp. 1–4, June 2004. [Online]. Available: <http://www.siam.org/pdf/news/232.pdf>
- [24] W. Saad, Z. Han, H. V. Poor, and T. Başar, "Game-Theoretic Methods for the Smart Grid: An Overview of Microgrid Systems, Demand-Side Management, and Smart Grid Communications," *IEEE Signal Process. Mag.*, vol. 29, no. 5, pp. 86–105, 2012. [Online]. Available: <http://dx.doi.org/10.1109/MSP.2012.2186410>
- [25] A. Sahai and S. Mitter, "The necessity and sufficiency of anytime capacity for stabilization of a linear system over a noisy communication link part 1: scalar systems," *IEEE Trans. Info. Theory*, vol. 52, no. 8, pp. 3369–95, 2006.
- [26] C. E. Shannon, "A Mathematical Theory of Communication," *The Bell System Technical Journal*, vol. 27, pp. 379–423, 623–656, july, october 1948. [Online]. Available: <http://cm.bell-labs.com/cm/ms/what/shannonday/shannon1948.pdf>
- [27] R. Srikant, *The Mathematics of Internet Congestion Control*, ser. Systems & Control: Foundations & Applications. Boston, MA: Birkhauser, 2004.
- [28] S. Tatikonda, A. Sahai, and S. Mitter, "Stochastic linear control over a communication channel," *IEEE Trans. Automat. Contr.*, vol. 49, no. 9, pp. 1549–61, Sep. 2004.
- [29] S. Yang and B. Hajek, "VCG-Kelly mechanisms for allocation of divisible goods: Adapting VCG mechanisms to one-dimensional signals," *IEEE JSAC*, vol. 25, no. 6, pp. 1237–1243, August 2007.