

Information Constrained and Finite-Time Distributed Optimisation Algorithms

Bigi Varghese Philip¹, Tansu Alpcan¹, Jiong Jin² and Marimuthu Palaniswami¹

Abstract—This paper studies the delay-accuracy trade-off for an unconstrained quadratic Network Utility Maximization (NUM) problem, which is solved by a distributed, consensus based, constant step-size, gradient-descent algorithm. Information theoretic tools such as entropy power inequality are used to analyse the convergence rate of the algorithm under quantised inter-agent communication. A finite-time distributed algorithm is proposed to solve the problem under synchronised message passing. For a system with N agents, the algorithm reaches any desired accuracy within $2N$ iterations, by adjusting the step-size, α . However, if N is quite large or if the agents are constrained by their memory or computational capacities, asymptotic convergence algorithms are preferred to arrive within a permissible neighbourhood of the optimal solution. The analytical tools and algorithms developed shed light to delay-accuracy trade-off required for many real-time IoT applications, e.g., smart traffic control and smart grid. As an illustrative example, we use our algorithm to implement an intersection management application, where distributed computation and communication capabilities of smart vehicles and road side units increase the efficiency of an intersection.

I. INTRODUCTION

Large-scale deployment of interconnected sensors and/or actuators is an integral part of Internet of Things (IoT). Among the many striking applications of IoT are smart grid and smart transportation [1]. For example, consider smart traffic control in which vehicles and traffic lights collaborate together to implement smooth traffic flow in a city. Such systems, where computing, communication and control technologies are tightly integrated, are broadly categorised as cyber-physical systems (CPS) [2]. CPS requires data from sensor networks to be processed in real-time, as they have to be associated with the control of physical systems. For an efficient operation of such a system, control operation has to be taken by each system agent within fraction of a second. For instance, in a smart traffic control application, if the leading vehicle in a lane has to slow down for some reason, all the vehicles in the

lane should reduce speed as soon as possible.

Due to scalability and delay constraints for real-time applications mentioned earlier, it is preferred to solve the problem in a distributed manner, i.e., without assigning any central node to collect and process the data from each agents. Many problems requiring distributed processing of large data sets can be posed in the framework of convex optimization [3]. For example, resource allocation subject to various constraints related to fairness and efficiency has been formulated as network utility maximization (NUM) problems [4], where agents try to optimise the global objective of maximizing aggregate utility. Further, smart objects should attain global objectives with acceptable latency, using limited local communication, computing and memory capabilities. Thus, distributed IoT algorithms have to be developed to reach at a predetermined neighbourhood of the optimum solution within a limited number of iterations, as required by different applications.

In real-time applications such as smart transportation and smart grid, the underlying communication graph and channel conditions affect the rate of convergence to the optimal solution. It is hence worthwhile studying the effect of graph and channel impairments on algorithm performance.

A. Contributions

In this work, we analyse the performance of a constant step-size based decentralised gradient-descent algorithm under imperfect channel conditions. The algorithm yields a near optimal solution for an unconstrained NUM problem, where agents try to maximise the sum of their quadratic objective functions. In every iteration, each agent, in parallel, performs an averaging and updating step. It first collects the quantised estimates of system variables from its neighbours and does weighted averaging in order to obtain an estimate.

Next, it updates its estimate of the system variables, following a simple constant step-size, gradient-descent on its local objective function. Since the capacity of the wireless communication channel is limited, the agents can only exchange quantised versions of the system variables. Thus, we consider limited bandwidth of channel, analyse its convergence properties using information theoretic tool of conditional differential entropy, for optimum achieving quantisation schemes. In contrast to the previous works, whose results are mainly limited to specific quantisation schemes, our bounds are applicable for a general class of quantisation schemes. Our analysis reveals

¹ Department of Electrical and Electronic Engineering, The University of Melbourne, Parkville, VIC 3010 Australia bigip@student.unimelb.edu.au, {tansu.alpcan, palani}@unimelb.edu.au

² School of Software and Electrical Engineering, Swinburne University of Technology, Hawthorn, VIC 3122 Australia jiongjin@swin.edu.au

Acknowledgement:Dr. Michael Pauley (Department of Electrical and Electronic Engineering, The University of Melbourne, Parkville, VIC 3010 Australia), has helped us in completing a portion of the proofs.

the relationship between the underlying communication graph, quantised channel and the maximum rate of convergence attainable for a consensus-based, constant step-size, gradient-descent algorithm.

The real-time constraints of the problem require us to estimate the number of iterations required by the algorithm to reach within an ϵ proximity of the optimum solution. It is well known that the limit point of the asymptotically converging algorithm reaches the optimum value as the step-size tends to zero [5]. We propose a decentralised method, to attain the limit point of constant step-size based gradient descent algorithm in less than $2N$ iterations, where N is the number of agents in the system. In that sense, our method can be indeed treated as a finite-time distributed optimisation algorithm for unconstrained quadratic NUM problems. There is little emphasis in the literature on distributed algorithms that obtain an optimum solution in a fixed number of time steps.

As an illustrative example, we consider a simple intersection management problem, in which a group of vehicles, along with road side units (RSU), distributively decide on their lane velocities so as to efficiently and safely utilise the intersection.

B. Literature Review

Convergence rate estimates explicitly characterise the accuracy of the generated approximate optimal solutions to the number of iteration required for an asymptotically converging optimisation algorithm. The authors of [3] and [4] have analytically obtained the convergence rate estimates for their corresponding algorithms. Further, convergence of a constant step-size based decentralised gradient descent algorithm under perfect channel conditions has been analysed in [5].

Authors of [6] study the performance degradation of a class of averaging problem, over time-varying topologies, when exchanged data is quantised to their nearest quantisation values. An incremental quantisation method is proposed in [7] for a group of distributed optimisation problem. Analysis of a distributed sub-gradient algorithm operating under a zooming-in technique based quantiser was done in [8].

However, the results developed in these papers are algorithm specific, i.e., they are valid only for a particular set of algorithms implemented using specific quantisation schemes. A universal (quantisation scheme independent) bound on the rate of exponential mean square convergence is obtained by [9]. They analysed the primal-dual algorithm, under quantised message passing between agents and the system (bipartite graph topology). In particular, the information theoretic concept of *entropy power inequality* is used to derive the bound. We extend this work to a totally distributed system implementing consensus based, constant step-size, gradient-descent algorithm.

In the context of smart transportation applications,

authors of [10] suggest a traffic lighting pattern that creates a “green wave”. They advise to delay green lights at each intersection if a sufficiently large convoy is not formed, and conclude that convoys can improve the traffic flow when the large empty areas between the convoys are effectively used by crossing convoys with few interferences. The problem of convoy formation is considered in [11], [12], with the objective of maintaining a fixed distance among agents of a vehicular network. Here all vehicles in a convoy have to adjust to the velocity of the lead vehicle. In order to find the most suited velocity for the convoy, our proposed algorithm can be run on top of such basic convoy forming algorithm.

C. Notation

\mathbf{A} stands for matrix. \mathbf{A}_i and \mathbf{A}^j are the i^{th} column and j^{th} row of \mathbf{A} . $\mathbf{A}_{i,j}$ is the $(i,j)^{\text{th}}$ element of \mathbf{A} . \mathbf{v} and v represent column vectors and scalars, respectively. \mathbf{v}_i is the i^{th} element of \mathbf{v} . \mathbf{v}^i is the estimate of \mathbf{v} by node i . $|X|$ denotes cardinality of X . $h[z|A=a]$ is the conditional differential entropy of variable z , given, variable A takes the value a . We use the terms agents, nodes, vehicles interchangeably in this paper. $\mathbf{1}_n \in \mathbb{R}^N$ is the column vector with all elements 1.

The remaining sections of the paper are organised as follows. Section II gives the underlying communication network model and the optimization problem. Section III discusses the convergence of the proposed distributed algorithm to a near neighbourhood of the optimum solution. Effect of quantised communication on our algorithm is studied in Section IV. An alternate finite-time approach to compute the solution for unconstrained quadratic NUM problem is explained in Section V. Especially, we present an illustrative application of smart traffic intersection management using our algorithm and the simulation details in Section VI. Section VII is dedicated to the result discussion, followed by the concluding remarks in Section VIII.

II. SYSTEM MODEL AND PROBLEM FORMULATION

This section describes the mathematical formulations of underlying communication network and the optimisation problem we are going to solve.

A. Inter-Agent Communication Model

The communication network among agents (here, vehicles in a cluster) can be modelled using an undirected graph, $\mathcal{G} = (\mathcal{N}, \mathcal{E})$, consisting of the set of nodes (vehicles) $\mathcal{N} = \{1, \dots, n\}$ connected by a set of edges \mathcal{E} , where each edge $\{i, j\}$ is an ordered pair of distinct nodes. The graph is assumed to be static and connected. Let \mathcal{N}_i denote the set of neighbours of node i . \mathbf{W} is a doubly stochastic weight matrix. $\mathbf{W}_{i,j} > 0$, iff $\{i, j\} \in \mathcal{E}$, meaning i, j are neighbours; $\mathbf{W}_{i,j} = 0$, otherwise. There are numerous choices for this weight matrix, e.g., the authors of [13] explain a simple method to compute \mathbf{W} in a distributed manner. In more details, if all agents have an estimate of the total number of agents in the system or an upper

limit of N , each agent i could choose its weight vector independently as

$$W_{i,j} = \begin{cases} \frac{1}{N}, & \text{if } j \in \mathcal{N}_i \\ 0, & \text{if } j \notin \mathcal{N}_i, j \neq i \\ 1 - \sum_{k \in \mathcal{N}_i} W_{i,k}, & \text{if } j = i \end{cases}$$

B. The Optimization Problem

In this paper, an unconstrained NUM problem, where a group of agents tries to cooperatively maximise the aggregate of their individual objectives is studied. The system will then jointly compute the solution of the following optimization problem:

$$\text{maximise } F(x) = \sum_{i=1}^N \frac{a_i}{2} x^2 + c_i x \quad (1)$$

where x is the global decision variable (in our example, velocity of the convoy). The parameters, $a_i \leq 0$, $c_i \in \mathbb{R}$, and $N = |\mathcal{N}|$ is the number of agents. The objective function above is concave and the coupling is attained through a global variable x . Thus, the optimization problem (1) can be solved using standard convex optimization techniques. We use the consensus based, constant step-size, gradient-descent algorithm similar to [5] in order to obtain the solution for (1) distributively. The step-size can be chosen based on the delay-accuracy requirement of various applications, as explained later in the paper. The algorithm can be implemented iteratively as,

$$y_i(k) = \sum_{j=1}^N \mathbf{W}_{i,j} x_j(k-1) \quad (2)$$

$$x_i(k) = y_i(k) + \alpha(a_i y_i(k) + c_i) \quad (3)$$

where $x_i(k)$ is the estimate of global variable x , by agent i at the instant k . Let $\mathbf{x}(k)$ denote the column vector whose elements are $x_i(k)$. For ease of analysis, the values of all nodes at time-step k will be aggregated into the value vector, $\mathbf{x}(\mathbf{k}) = [x_1(k) \ x_2(k) \ \dots \ x_N(k)]^T$. After each iteration the state of the system can be described using $\mathbf{x}(\mathbf{k})$,

$$\mathbf{x}(\mathbf{k}) = (I + \alpha \mathbf{Diag}(\mathbf{a})) \mathbf{W} \mathbf{x}(\mathbf{k}-1) + \alpha \mathbf{c} \quad (4)$$

where $\mathbf{Diag}(\mathbf{a})$ is a diagonal matrix with its diagonal entries as elements of vector \mathbf{a} .

III. CONVERGENCE OF THE ALGORITHM

Convergence of the iterative algorithm, implemented using (4) to a near optimal point, for generalized NUM problems of the form: minimise $F(x) = \sum_{i=1}^N f_i(x)$, where ∇f_i is Lipschitz continuous with constant $L_{f_i} > 0$, has been proved in [5]. They define the average objective error as $\bar{r}(k) := f(\bar{x}(k) - f^*)$, where $\bar{x}(k) = \frac{1}{n} \sum x_i(k)$ and $f(x) = \frac{1}{n} F(x)$. Theorem 2 in [5], when applied to our quadratic NUM problem, leads to

Theorem 1: For a quadratic NUM problem of the form (1), if the step-size, $\alpha < \frac{2}{\max_i |a_i|}$, the objective error $\bar{r}(k)$ reduces at a rate of $O(\frac{1}{\alpha k})$ until it converges to an $O(\frac{\alpha}{(1-\beta)})$ - neighbourhood of the optimal solution, where

$\beta < 1$ is the second largest eigenvalue of the weight matrix, \mathbf{W} . Further, from Theorem 1 in [5], for k large enough, $|x_i(k) - \bar{x}(k)| < \frac{\alpha D}{(1-\beta)}$, where D , the upper bound for $\sum \nabla f_i(x_i(k))$, is proved finite.

From Theorem 1, it can be seen that the algorithm attains geometric convergence, and the convergence rate can be adjusted by choosing higher values of α . Now we are to find the point, $\mathbf{x}_{\text{limit}}$, to which the sequence $\mathbf{x}(\mathbf{k})$ converges. Substituting $\mathbf{x}(\mathbf{k}) = \mathbf{x}(\mathbf{k}-1) = \mathbf{x}_{\text{limit}}$ in (4) yields

$$\mathbf{x}_{\text{limit}} = (I + \alpha \mathbf{Diag}(\mathbf{a})) \mathbf{W} \mathbf{x}_{\text{limit}} + \alpha \mathbf{c} \quad (5)$$

$$\implies (\mathbf{I} - \mathbf{W} - \alpha \mathbf{Diag}(\mathbf{a}) \mathbf{W}) \mathbf{x}_{\text{limit}} = \alpha \mathbf{c} \quad (6)$$

$$\implies \mathbf{x}_{\text{limit}} = \alpha (\mathbf{I} - \mathbf{W} - \alpha \mathbf{Diag}(\mathbf{a}) \mathbf{W})^{-1} \mathbf{c} \quad (7)$$

$$= \alpha (\mathbf{I} - \mathbf{D} \mathbf{W})^{-1} \mathbf{c} \quad (8)$$

where $\mathbf{D} = \mathbf{I} + \alpha \mathbf{Diag}(\mathbf{a})$.

As we reduce the step-size, we move more and more close towards the optimum solution. It is evident from our results that the error of the limit point, ϵ_{limit} , decreases with step-size. Thus, we will be able to attain a trade-off between accuracy and the rate of convergence, by varying the design parameter, α . The smaller the constant step-size, the more accurate the solution will be at the expense of a slower convergence rate.

IV. IMPACT OF QUANTISED INTER-AGENT COMMUNICATION

Let us consider the effect of finite channel capacity and quantisation in the above mentioned optimization scenario. Each node, $i \in \mathcal{N}$, communicates $\mathbf{x}_i(k)$ to its neighbour, $j \in \mathcal{N}_i$. A digital channel is assumed between all the nodes. Quantisers, $\mathbb{Q}_{\{i,j\}}^k(\mathbf{x}_i(k)) : \mathbb{R} \rightarrow \mathcal{A}_{\{i,j\}}^k$ is used by agent i to quantise its estimates to be shared with agent j at time k . Each $\mathcal{A}_{\{i,j\}}^k$ is a finite set representing the range of $\mathbb{Q}_{\{i,j\}}^k(\mathbf{x}_i(k))$. The asymptotic average data rate [14] of the link (i, j) , $r_{\{i,j\}} \in \mathbb{R}$ is defined as,

$$r_{\{i,j\}} = \limsup_{k \rightarrow \infty} \frac{1}{k} \sum_{t=0}^{k-1} \log |\mathcal{A}_{\{i,j\}}^t| \quad (9)$$

The quantised version of updates in (3) at each agent i can be written as

$$\mathbf{x}_i(k) = (1 + \alpha a_i) \mathbf{W}_i \mathbb{Q}_{\{*,i\}}^{k-1}(\mathbf{x}(k-1)) + \alpha c_i \quad (10)$$

where, $\mathbb{Q}_{\{*,i\}}^k(\mathbf{x}(\mathbf{k})) = [\mathbb{Q}_{\{1,i\}}^k(x_1(k)) \cdots \mathbb{Q}_{\{j,i\}}^k(x_j(k)) \cdots \mathbb{Q}_{\{N,i\}}^k(x_N(k))]^T$ and \mathbf{W}_i is the i^{th} row of \mathbf{W} .

A. Optimum Achieving Quantisation Scheme

Since our scheme converges to a value in the neighbourhood of the optimum solution, we modify the definition accordingly for optimum achieving quantisation scheme proposed in [9].

Definition 1: The quantisation scheme \mathbb{Q}_a is called an optimum achieving (OA) quantisation scheme if, under \mathbb{Q}_a , the sequence generated by the update rule converges to the same limit, i.e., $\mathbf{x}_{\text{limit}}$ irrespective of the introduction of quantisation. i.e., $\lim_{k \rightarrow \infty} \mathbf{x}(k) = \mathbf{x}_{\text{limit}}$.

Let us assume optimum achieving quantisation scheme is being implemented. We could define $\|e_k\|_2$ as the total distance between values of the system variables at time k and their limit values, i.e.,

$$\|e_k\|_2 = \sqrt{\sum_{i=1}^N (x_i(k) - \mathbf{x}_{\text{limit } i})^2} \quad (11)$$

We assume that the vector \mathbf{x}_0 is randomly drawn according to the probability density function $p_{\mathbf{x}_0}(\mathbf{x})$ with finite entropy. Then, the average error norm square at time k under the quantisation scheme \mathbb{Q} is defined as $E[\|e_k\|_2^2]$. We define error decay exponent (EDE), as in [9], to measure the convergence rate, i.e., $\liminf_{k \rightarrow \infty} \frac{1}{k} \log E[\|e_k\|_2^2]$. Note that EDE is a non-positive quantity, and a more negative EDE implies faster convergence to the limit point. We obtain a lower bound of EDE in the above defined scenario.

Theorem 2: A lower bound of EDE for the constant step size distributed optimization algorithm is

$$\liminf_{k \rightarrow \infty} \frac{1}{k} \log E[\|e_k\|_2^2] \geq \quad (12)$$

$$\frac{2}{N} \left(\left(\sum_{i=1}^N \log(1 + \alpha a_i) \right) + \log|\det(\mathbf{W})| - \sum_{i,j=1, i \neq j}^N r_{i,j} \right)$$

B. Proof of Theorem 2

The following is a proof on lower bound of EDE. The differential entropy power of $z \in \mathbb{R}^N$ conditioned on the event $A = a$, $\mathbb{N}[z|A = a]$ can be defined as $\mathbb{N}[z|A = a] = \frac{1}{2\pi e} e^{\frac{2}{N} h[z|A = a]}$. Given $A = a$, the average conditional entropy power of z , that is, $E_A[\mathbb{N}[z|A = a]]$ can be upper bounded as in [14],

$$E_A[\mathbb{N}[z|A = a]] \leq e^{\frac{1}{N} - 1} E[\|z\|_2^2] \quad (13)$$

Let $D_{i,j}^{k-1} = \{\mathbb{Q}_{\{i,j\}}^n(x_i(n)) = q_{\{i,j\}}(n)\}_{n=0}^{k-1}$, where $q_{\{i,j\}}(n)$ is a possible output of scheme $\mathbb{Q}_{\{i,j\}}$ at time n . Also, let $D_i^{k-1} = \{D_{i,j}^{k-1} | j \in \mathcal{A}_i\}$ and $\mathbf{D}_{k-1} = \{D_1^{k-1}, D_2^{k-1}, \dots, D_n^{k-1}\}$.

Using (13), $E[\|e_k\|_2^2]$ can be lower bounded as:

$$E[\|e_k\|_2^2] \geq e^{1 - \frac{1}{N}} E_A[\mathbb{N}[e_k|D_{k-1}]] \quad (14)$$

$$\geq \frac{e^{1 - \frac{1}{N}}}{2\pi e} e^{\frac{2}{N} E[h[e_k|D_{k-1}]]} \quad (15)$$

(15) follows from Jensen inequality. Now,

$$h[e_k|D_{k-1}] = h[\mathbf{x}_k - \mathbf{x}_{\text{limit}} | D_{k-1}] \quad (16)$$

$$= h[\mathbf{x}_k | D_{k-1}] \quad (17)$$

(17) follows from translation invariance property of differential entropy.

$$h[\mathbf{x}_k | D_{k-1}] = h[(\mathbf{I} + \alpha \mathbf{Diag}(\mathbf{a})) \mathbf{W} \mathbf{x}_{k-1} + \alpha c | D_{k-1}]$$

$$= \sum_{i=1}^N \log(1 + \alpha a_i) + \log|\det(\mathbf{W})| + h[\mathbf{x}_{k-1} | D_{k-1}] \quad (18)$$

$$= h[\mathbf{x}_0 | D_{k-1}] + \sum_{n=0}^{k-1} \left(\left(\sum_{i=1}^M \log(1 + \alpha a_i) \right) + \log|\det(\mathbf{W})| \right) \quad (19)$$

We use translation invariance property of differential

entropy and the fact that $h[\mathbf{Tz}] = \log|\det(\mathbf{T})| + h[\mathbf{z}]$ in (18) to obtain (19). The following lemma from [14] gives a lower bound on $E[h[\mathbf{x}_k | D_{k-1}]]$.

Lemma 1: The average conditional entropy of x_0 given D_{k-1} i.e., $E_D[h[x_0 | D_{k-1}]]$, can be lower bounded as

$$E_D[h[x_0 | D_{k-1}]] \geq h[\mathbf{x}_0] - \sum_{n=0}^{k-1} \left(\sum_{j=1}^N \sum_{i=1, i \neq j}^N \log|\mathcal{A}_{\{i,j\}}^n| \right)$$

Applying Lemma 1 in (19), we have

$$E[h[\mathbf{x}_k | D_{k-1}]] \geq E[h[\mathbf{x}_0]] + \sum_{n=0}^{k-1} \left(\left(\sum_{i=1}^M \log(1 + \alpha a_i) \right) + \log|\det(\mathbf{W})| \right) - \sum_{j=1}^N \sum_{i=1, i \neq j}^N \log|\mathcal{A}_{\{i,j\}}^n| \quad (20)$$

Using Equations (15), (17) and (20), and the assumption that $p_{\mathbf{x}_0}(\mathbf{x})$ has finite entropy, we could obtain lower bound of EDE as,

$$\liminf_{k \rightarrow \infty} \frac{1}{k} \log E[\|e_k\|_2^2] \geq \frac{2}{N} \left(\left(\sum_{i=1}^M \log(1 + \alpha a_i) \right) + \log|\det(\mathbf{W})| - \sum_{i,j=1, i \neq j}^N r_{i,j} \right)$$

□

Remark 1: It should be noted that our theorem does not impose any restriction on the structure of the quantiser used (many previous studies are focused on uniform quantisation), rather it is applicable to any optimum achieving quantisation scheme. In that sense, ours is a universal lower bound.

The lower bound obtained in Theorem 2 is controlled by the number of agents, the slope of objective function, weight matrix and average aggregate data rate under quantisation scheme. Determinant of weight matrix indeed depends on the underlying graph structure.

V. DECENTRALISED SOLUTION IN FINITE NUMBER OF ITERATIONS

This section explores the methods to compute x_{limit} distributively in finite number of iterations. In [13], authors solve a consensus problem of the form $\mathbf{x}(\mathbf{k} + 1) = \mathbf{W}_{N \times N} \mathbf{x}(k)$, where $\mathbf{W}_{N \times N}$ is an $N \times N$ weight matrix, in $N + 1$ iterations. They utilise minimal polynomial of $\mathbf{W}_{N \times N}$ to distributively compute the consensus value. Since our optimisation problem has a similar structure, we could also solve this unconstrained quadratic NUM problem in $2N$ iterations, using a similar approach. Equation (4) can be rewritten as,

$$\mathbf{x}(\mathbf{k}) = \mathbf{B} \mathbf{x}(\mathbf{k} - 1) + \alpha c \quad (21)$$

where $\mathbf{B} = (\mathbf{I} + \alpha \mathbf{Diag}(\mathbf{a})) \mathbf{W}$.

The characteristic polynomial of \mathbf{B} has degree $D < N$. i.e., $\mathbf{B}^D + \alpha_{D-1} \mathbf{B}^{D-1} + \dots + \alpha_1 \mathbf{B} + \alpha_0 \mathbf{I} = 0$. Therefore, for

all $k > 0$, we have

$$\begin{aligned} \mathbf{x}(\mathbf{k} + \mathbf{D}) &= \mathbf{B}\mathbf{x}(k + D - 1) + \alpha\mathbf{c} \\ &= \mathbf{B}^2\mathbf{x}(k + D - 2) + (\mathbf{B} + I)\alpha\mathbf{c} \end{aligned} \quad (22)$$

....

$$\begin{aligned} &= \mathbf{B}^D\mathbf{x}(k) + \left(\sum_{j=1}^{D-1} \mathbf{B}^j + I\right)\alpha\mathbf{c} \\ &= -(\alpha_{D-1}\mathbf{B}^{D-1} + \dots + \alpha_1\mathbf{B} + \alpha_0 I)\mathbf{x}(k) + \left(\sum_{j=1}^{D-1} \mathbf{B}^j + I\right)\alpha\mathbf{c} \\ &= -(\alpha_{D-1}\mathbf{x}(k + D - 1) + \dots + \alpha_1\mathbf{x}(k + 1) + \alpha_0\mathbf{x}(k)) \\ &\quad - \left(\sum_{i=2}^{D-1} \alpha_i \sum_{j=1}^i (\mathbf{B}^j + I) + \alpha_1\alpha\mathbf{c} - \sum_{j=1}^{D-1} (\mathbf{B}^j + I)\right)\alpha\mathbf{c} \\ &= -(\alpha_{D-1}\mathbf{x}(k + D - 1) + \dots + \alpha_1\mathbf{x}(k + 1) + \alpha_0\mathbf{x}(k)) - \hat{\mathbf{c}} \end{aligned} \quad (23)$$

where $\hat{\mathbf{c}} = \left(\sum_{i=2}^{D-1} \alpha_i \sum_{j=1}^i (\mathbf{B}^j + I) + \alpha_1\alpha\mathbf{c} - \sum_{j=1}^{D-1} (\mathbf{B}^j + I)\right)\alpha\mathbf{c}$ is a constant vector. Thus, for all $1 \leq i \leq N$, satisfies a linear difference equation of the form

$$\begin{aligned} \mathbf{x}_i(k + D) &= -(\alpha_{D-1}\mathbf{x}_i(k + D - 1) + \dots \\ &\quad + \alpha_1\mathbf{x}_i(k + 1) + \alpha_0\mathbf{x}_i(k)) - \hat{\mathbf{c}}_i \end{aligned} \quad (24)$$

It can be seen from Section III that $\mathbf{x}_i(\mathbf{k} + \mathbf{D} + 1)$ converges to a limit point, i.e., $k \rightarrow \infty, \mathbf{x}_i(\mathbf{k} + \mathbf{n}) = \mathbf{x}_{\text{limit } i}$, for finite n . If we know $\alpha_{D-1}, \dots, \alpha_1$ and $\hat{\mathbf{c}}_i$, the limit point can be easily found as in [15], from the recurrence relation (24).

$$\begin{aligned} \mathbf{x}_{\text{limit } i} &= -(\alpha_{D-1}\mathbf{x}_{\text{limit } i} + \dots + \alpha_1\mathbf{x}_{\text{limit } i} + \alpha_0\mathbf{x}_{\text{limit } i}) - \hat{\mathbf{c}}_i \\ \implies \mathbf{x}_{\text{limit } i} &= \frac{-\hat{\mathbf{c}}_i}{1 + \sum_{d=0}^{D-1} \alpha_d} \end{aligned} \quad (25)$$

A. Decentralised Calculation of The Minimal Polynomial

We now explain how each agent could compute the unknowns, $\alpha_{D-1}, \dots, \alpha_1$ and $\hat{\mathbf{c}}_i$, in a distributed manner. We assume that each agent knows N (or an upper limit on N). Initially by assuming that $D = N$, the agent has to update and communicate the system variables for $2 * (N + 1)$ iterations of the asymptotic converging algorithm following (4). Thus, providing $N + 1$ equations of the form (24), from which the unknowns can be estimated. Each agent takes N equations of the form (24) and expresses them in a matrix form as

$$\begin{bmatrix} \mathbf{x}_i(N + 1) \\ \vdots \\ \mathbf{x}_i(2N) \end{bmatrix} = \mathbf{M}_{\mathbf{x}_i} \begin{bmatrix} \alpha_{D-1} \\ \vdots \\ \alpha_0 \\ \hat{\mathbf{c}}_i \end{bmatrix} \quad (26)$$

$$\text{where } \mathbf{M}_{\mathbf{x}_i} = - \begin{bmatrix} \mathbf{x}_i(N) & \dots & \mathbf{x}_i(0) & 1 \\ \vdots & & & \\ \mathbf{x}_i(2N - 1) & \dots & \mathbf{x}_i(N) & 1 \end{bmatrix}.$$

Remark 2: It should be noted that agents can stop

iterations when the matrix, $\mathbf{M}_{\mathbf{x}_i}$, loses rank. The rank of $\mathbf{M}_{\mathbf{x}_i}$ will be $D + 1$. Because $D + 1$ linearly independent equations of (26) can be used to solve for $D + 1$ unknowns, a maximum of $N + D + 1$ iterations are required to reach the limit value. An alternative approach, for distributed computation of the minimal polynomial, which requires utmost $(N + 1) \times (N + 2)$ iterations is proposed in [13].

Further, the limit, $N + D - 1$, is independent of the choice of α . Hence, ideally, one can choose α to be arbitrarily small and thereby reach arbitrarily close to the optimum solution in $N + D - 1$ iterations.

B. Finite-Time vs. Asymptotic Convergence: Application Perspective

The permissible error is supposed to be ϵ . We know that the initial values of the system variables lie within a range of δ from the optimum solution. Then, the error estimate for asymptotic convergence algorithm after $2N$ iterations will be $\max_i (|1 - \alpha|a_i|)^{2N} \delta + \|x_{\text{limit}} - x_{\text{opt}}\|$. It is worth pointing out that our finite-time algorithm could attain desired accuracy within $2N$ iterations, by using small enough values of α . If application allows for a delay of $2N$ iterations and the agents have sufficient memory and computing capacities, the finite-time convergence algorithm is always preferred to asymptotic convergence algorithm.

VI. INTERSECTION MANAGEMENT IN SMART TRAFFIC

This section shows the practicality of our proposed algorithm through intersection management as a smart transportation application, where clusters of smart vehicles negotiate for an intersection slot with the aid of Road Side Units (RSUs) and adjust their velocities accordingly. Each vehicle arriving at the intersection may have a desired velocity (e.g., set by user in cruise control). Further, all vehicles are not equally willing to deviate from their preferred velocity (Factors like mass, dynamics of the vehicles, fuel economy etc. decide the willingness of a vehicle to vary its velocity). Therefore, utility (happiness) of vehicle i , at a particular velocity can be modelled using a quadratic utility function of the form $f_i(x) = \frac{a_i}{2}x^2 + c_i x$, where x is the global decision variable (i.e., velocity of the convoy). In our case, we set $a_i \leq 0$ to be $-a$ and c_i to be $a * v$, where a and v are the acceleration and desired velocity of vehicle i . The overall problem of intersection management is decomposed into several sub-problems, using dual decomposition techniques [16]. The master problem (running at RSUs) solves for the constraints and passes a corresponding Lagrangian to each vehicle cluster (sub-problems). Each vehicle cluster then solves an unconstrained quadratic NUM problem and updates the master regarding their updates.

VII. NUMERICAL RESULTS AND DISCUSSION

In order to validate Theorem 2, a sparse network with 5 agents was assumed. Each agent uses 2 bits to quantise its estimate before transmitting to its neighbours. We

calculated a lower bound of EDE to be -9.7564 . It can be seen from Fig. 1 that the EDE is well above the bounds estimated using Theorem 2. For the above

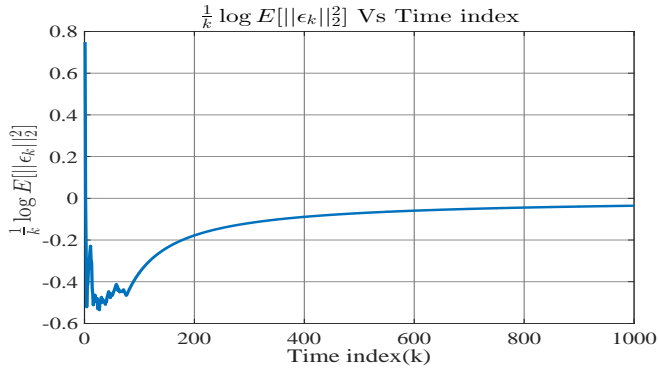


Fig. 1: $\log(\text{Mean Square Error})$ divided by k versus time index under quantisation scheme \mathbb{Q}_a

mentioned sparse network with 5 agents, using the algorithm mentioned in Section V, we were able to compute exactly the limit values of convergence in a decentralised manner in finite number of iterations. In our case, the degree of the characteristic polynomial, D , was found to be 5 for all the agents. D is directly linked to the gradient of the individual objective functions and the structure of the graph. Further, the upper limit on D is the number of agents, N , in the system. Fig. 2

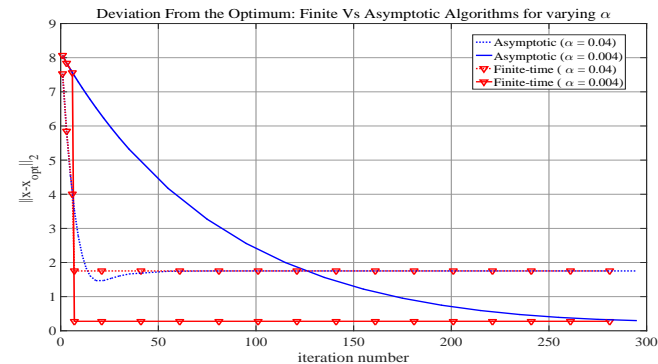


Fig. 2: Comparing convergence of x to near optimal solution in case of Finite-time algorithm and asymptotically converging algorithm for varying step-size α values.

compares the convergence of finite-time and asymptotically converging algorithms, with step-size .04 and .004, to a near optimal solution. Both the algorithms reach a near optimal solution, and the accuracy of the solution reduces with increasing step-size value, α . It has to be noted that the finite-time algorithm could compute the limit point of the corresponding asymptotic algorithm, in $N + D$ iterations, irrespective of the step-size.

VIII. CONCLUSION

In this paper, we analysed an unconstrained quadratic NUM problem in which a group of agents distributively seeks to maximise the sum of their individual utility functions using a consensus based, constant step-size, gradient descent algorithm. First, the effect of quantised inter-agent communication on the performance of the algorithm using the notion of differential entropy power

is analysed. An upper bound on the rate of convergence of the algorithm was established, which depends on the number of agents, the slope of objective function, weight matrix and average aggregate data rate under the quantisation scheme.

Further, a finite-time version of the same algorithm was proposed. It was shown that the algorithm could estimate the global optimum value with desired accuracy in less than $2N$ iterations. Considering the delay-accuracy trade-off of real-time IoT algorithms, for systems with fewer number of agents, it is preferable to switch to the finite-time optimisation scheme. Asymptotic convergence scheme is better if N is large or when the agents are constrained by memory or processing power. A smart traffic intersection management application was briefly included to illustrate our algorithm for practical operation.

REFERENCES

- [1] J. A. Stankovic, "Research directions for the internet of things," *Internet of Things Journal, IEEE*, vol. 1, no. 1, pp. 3–9, 2014.
- [2] K. D. Kim and P. R. Kumar, "Cyber physical systems: A perspective at the centennial," *Proceedings of the IEEE*, vol. 100, no. Special Centennial Issue, pp. 1287–1308, 2012.
- [3] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2011.
- [4] A. Nedic, *Convergence Rate of Distributed Averaging Dynamics and Optimization in Networks*. Now Publishers Inc., 2015.
- [5] K. Yuan, Q. Ling, and W. Yin, "On the convergence of decentralized gradient descent," *SIAM Journal on Optimization*, vol. 26, no. 3, pp. 1835–1854, 2016.
- [6] A. Nedic, A. Olshevsky, A. Ozdaglar, and J. N. Tsitsiklis, "On distributed averaging algorithms and quantization effects," *Automatic Control, IEEE Trans. on*, vol. 54, no. 11, pp. 2506–2517, 2009.
- [7] M. G. Rabbat and R. D. Nowak, "Quantized incremental algorithms for distributed optimization," *IEEE Journal on Selected Areas in Comm.*, vol. 23, no. 4, pp. 798–808, 2005.
- [8] P. Yi and Y. Hong, "Quantized subgradient algorithm and data-rate analysis for distributed optimization," *IEEE Transactions on Control of Network Systems*, vol. 1, no. 4, pp. 380–392, 2014.
- [9] E. Nekouei, G. Nair, and T. Alpcan, "Convergence analysis of quantized primal-dual algorithm in quadratic network utility maximization problems," in *Decision and Control (CDC), 2015 IEEE 54th Annual Conf. on*. IEEE, 2015, pp. 2655–2660.
- [10] S.-B. Cools, C. Gershenson, and B. D'ÁZHooghe, *Self-organizing traffic lights: A realistic simulation*. Springer, 2013, pp. 45–55.
- [11] L. Y. Wang, A. Syed, G. G. Yin, A. Pandya, and H. Zhang, "Control of vehicle platoons for highway safety and efficient utility: Consensus with communications and vehicle dynamics," *Journal of Systems Science and Complexity*, vol. 27, no. 4, pp. 605–631, 2014.
- [12] Q. Jin, G. Wu, K. Boriboonsomsin, and M. Barth, "Platoon-based multi-agent intersection management for connected vehicle," in *16th International IEEE Conf. on Intelligent Transport. Systems*, 2013, pp. 1462–1467.
- [13] S. Sundaram and C. N. Hadjicostis, "Finite-time distributed consensus in graphs with time-invariant topologies," in *American Control Conference, 2007. ACC '07, 2007*, pp. 711–716.
- [14] G. N. Nair and R. J. Evans, "Stabilizability of stochastic linear systems with finite feedback data rates," *SIAM Journal on Control and Optimization*, vol. 43, no. 2, pp. 413–436, 2004.
- [15] G. Everest, *Recurrence sequences*, ser. Mathematical surveys and monographs: v. 104. Providence, RI : American Mathematical Society, c2003., 2003.
- [16] D. P. Palomar and M. Chiang, "A tutorial on decomposition methods for network utility maximization," *IEEE Jnl. on Selected Areas in Comm.*, vol. 24, no. 8, pp. 1439–1451, 2006.



Minerva Access is the Institutional Repository of The University of Melbourne

Author/s:

Philip, BV;Alpcan, T;Jin, J;Palaniswami, M

Title:

Information Constrained and Finite-Time Distributed Optimisation Algorithms

Date:

2017-01-01

Citation:

Philip, B. V., Alpcan, T., Jin, J. & Palaniswami, M. (2017). Information Constrained and Finite-Time Distributed Optimisation Algorithms. 2017 IEEE 56TH ANNUAL CONFERENCE ON DECISION AND CONTROL (CDC), 2018-January, IEEE. <https://doi.org/10.1109/CDC.2017.8264337>.

Persistent Link:

<http://hdl.handle.net/11343/241552>