52nd IEEE Conference on Decision and Control
December 10-13, 2013. Florence, Italy

A Total Variation based Approach for Robust Consensus in Distributed Networks

Walid Ben-Ameur, Pascal Bianchi and Jérémie Jakubowicz

***Abstract*— Consider a connected network of agents endowed with local cost functions representing private objectives. Agents seek to find an agreement on some minimizer of the aggregate cost, by means of repeated communications between neighbors. This paper investigates the case where some agents are unreliable in the sense that they permanently inject some false value in the network. We introduce a new relaxation of the initial optimization problem. We show that the relaxed problem is equivalent to the initial one under some regularity conditions which are characterized. We propose two iterative distributed algorithms for finding minimizers of the relaxed problem. When all agents are reliable, these algorithms converge to the sought consensus provided that the above regularity conditions are satisfied. In the presence of misbehaving agents, we show in simple scenario that our algorithms converge to a solution which remains in the vicinity of the sought consensus. Unlike standard distributed algorithms, our approach turns out to be unsensitive to large perturbations. Numerical experiments complete our theoretical results.**

I. Introduction

Consensus algorithms designate a class of distributed methods allowing a set of connected agents / nodes to find an agreement on a some global parameter value [1]. The latter parameter is often defined as a minimizer of a global objective function defined as the sum of some local regret functions held by the agents [2], [3], [4]. As we shall see below, an important special case is obtained when the aim is to compute the average over the network of some local values held by the agents. The latter scenario will be refered to as the *average consensus* case. It has been well-studied in the literature [1], [5], [6]. The most widespread approach to achieve average consensus is through iteration of linear operations mimicking the behaviour of heat equation [6]: at each round, nodes average the values in their neighborhood (including themselves). Similarly, in the more general framework of distributed optimization, many algorithms have been proposed: some of them are based on distributed (sub)gradient approaches [3], [4], [7], [8] while others use splitting methods such as the Alternating Direction Method of Multipliers (ADMM) [9]. Under certain hypotheses, such approaches can be shown to converge to a state where each node in the network eventually has the same value – the sought parameter.

However, most of this works share a common view of the network: all agents show good will. They do not, for instance, deliberately introduce some false value inside the network, or refuse to update their value. There are a few recent work raising the problem of misbehaving agents in the gossip process [10], [11]. In such scenarios, standard consensus algorithms not only fail, but can be driven arbitrarily far away from the sought consensus [12]. A first approach to increase consensus robustness in unreliable networks is to detect misbehaving agents, identify them and finally exclude them from the network. Of course, cleaning the network beforehand is certainly beneficial whenever feasible, however misbehaving agents are not necessarily detectable and even if they are, may be detectable only by using involved and computationaly expensive algorithms [13]. An alternative is to design simple algorithms that naturally show good robustness properties. This paper follows this perspective.

Contributions. *i)* We introduce a new relaxation of standard distributed optimization problems over a network. The latter is based on the introduction of a Total-Variation based penalization term which penalizes the configurations when the network is far away from consensus.

ii) We provide verifiable sufficient regularity conditions under which the minimizers of the relaxed problem coincide with the sought minimizers of the initial optimization problem.

iii) We provide two iterative distributed algorithms for computing the minimizers of the relaxed problem. The first one is based on a natural subgradient method. The second one uses ADMM. Both of them use only limited communications between nodes at each iteration. When the above regularity conditions hold and when no misbehaving agent is present, these algorithms drive the network to the sought consensus.

iv) As a sanity check for the robustness of our algorithms, we analyze the convergence of our algorithms in the presence of stubborn agents that permanently introduce some false value in the network. We prove that unlike traditional approaches, our algorithms ensure that the estimates cannot be driven arbitrarily far away from the sought consensus.

The paper is organized as follows. Section II introduces the problem and the notations. Section III provides preliminary theoretical materials. Section IV introduces the relaxed problem and characterizes the minimizers. Algorithms are proposed in Section V. In Section VI, we analyze the convergence of our algorithm in a scenario where subborn agents are present. Section VII presents the numerical results.

W. Benameur and J. Jakubowicz are with Institut Télécom and CNRS / SAMOVAR (UMR 5157) – Télécom SudParis, 9 rue Charles Fourier, 91000 Evry, France [firstname].[lastname]@it-sudparis.eu

P. Bianchi is with Institut Télécom - CNRS/LTCI (UMR 5141) - Télécom Paristech, 37-39 rue Dareau, 75014 Paris, France [firstname].[lastname]@telecom-paristech.fr

II. Framework and notations

A. The Problem

Consider a network of agents represented by an undirected graph $G = (V, E)$ where V is a finite set of agents and $\{v, w\}$ belongs to E if and only if agent v and agent w are able to communicate. We investigate the following optimization problem:

$$\inf_{x\in\mathbb{R}} \sum_{v\in V} F_v(x) \tag{1}$$

where $F_v : \mathbb{R} \to (-\infty, +\infty]$ is a function which can be interpreted as the regret of agent v when the network lies in a state x. Merely for notational convenience, this paper is restricted to the case where parameter x is real. Generalization to the case where x belongs to an arbitrary Euclidean space is however straightforward. We assume the following.

Assumption 1:
i) For any $v \in V$, F_v is a proper closed convex function.
ii) The infimum of (1) is finite and is attained for some point $x \in \mathbb{R}$.

Example 1: We shall pay a special attention to the following particular case, which we shall refer to as the *Average Consensus (AC)* case:

$$\text{(AC)} \quad F_v(x) = \frac{1}{2}(x - x_0(v))^2 \tag{2}$$

where $x_0(v)$ represents some initial value held by agent v. In that case, problem (1) is equivalent to the distributed computation of the average $\overline{x}_0 = (1/|V|)\sum_v x_0(v)$ where $|V|$ is the cardinal of V.

Each agent v is supposed to hold some value $x_n(v)$ at each time $n \in \mathbb{N}$. The aim of this paper is to introduce and analyze distributed algorithms which, under some assumptions, drive all sequences $(x_n(v))_{v\in V}$ to a common minimizer of (1) as n tends to infinity. Moreover, the proposed algorithms should be robust to the presence of misbehaving agents. By robust, we mean that the final estimate of regular (well-behaved) agents should remain in an acceptable vincinity of the sought consensus even in the case when other agents permanently introduce some false value in the network.

B. Network Model

Throughout this paper, we assume a synchronous network where a global clock allows the agents to communicate with each other at each clock tick. In this paper, we refer to a *distributed algorithm* as an iteration of the form:

$$x_{n+1}(v) = h_{n,v}\left((x_k(v), x_k(w) : w \sim v, 0 \le k \le n)\right) \tag{3}$$

for some specified functions $h_{n,v}$. Nevertheless, we shall sometimes assume that some subset $S \subset V$ of agents do not follow the specified update rule (3). Such agents will be called *irregular*. An irregular agent $v \in S$ is called *stubborn* if for any $n \ge 0$,

$$x_n(v) = x_0(v)\,. \tag{4}$$

We denote by S the set of irregular agents and by $R = V\backslash S$ the set of *regular* agents.

C. Notations

We denote by ι_A the indicator function of a set A: $\iota_A(x) = 0$ if $x \in A$ and $\iota_A(x) = +\infty$ otherwise. We denote by $d(v)$ the degree of a vertex v *i.e.*, the number of neighbors in G. Sometimes we are going to need an (arbitrary) orientation to each edge. $\vec{E}$ denotes whatever compatible set of directed edges, in the sense that $(v, w) \in \vec{E}$ implies that $\{v, w\} \in E$ and $(w, v) \notin \vec{E}$; reciprocally $\{v, w\} \in E$ implies either $(v, w) \in \vec{E}$ or $(w, v) \in \vec{E}$. Of course, our results will not depend on the particular orientation we choose.

For a given set A, vector space $\mathbb{R}^A$ denotes the set of functions $A \to \mathbb{R}$, it is endowed with its standard vector space structure and scalar product $\langle f, g\rangle_A = \sum_{v\in A} f(v)g(v)$. Subscript A will be omitted when no confusion can occur. 0_A stands for the constant function $v \in A \mapsto 0$ and 1_A stands for the constant function $v \in A \mapsto 1$. The set $\mathcal{C}$ of functions which are proportional to 1_V is called the *consensus subspace*. The cardinal of a set A is denoted $|A|$. The average of $x \in \mathbb{R}^A$ is denoted $\bar{x} = \frac{1}{|A|}\sum_{v\in A} x(v)$. Notation grad accounts for the linear operator:

$$\begin{aligned} \mathrm{grad} : \mathbb{R}^V \quad &\to \quad \mathbb{R}^{\vec{E}} \\ \big(v \mapsto f(v)\big) \quad &\mapsto \quad \big((v, w) \mapsto f(w) - f(v)\big)\,. \end{aligned}$$

Notation div accounts for the operator:

$$\begin{aligned} \mathrm{div} : \mathbb{R}^{\vec{E}} \quad &\to \quad \mathbb{R}^V \\ \xi \quad &\mapsto \quad \Big(v \mapsto \sum_{(v,w)\in\vec{E}} \xi(v, w) - \sum_{(w,v)\in\vec{E}} \xi(w, v)\Big)\,. \end{aligned}$$

III. Total Variation Basics

Let us denote by $\mathbb{R}^V_0$ the set $\{x \in \mathbb{R}^V : \langle x, 1_V\rangle = 0\}$ of zero-mean functions over V. It is straightforward to check that function $x \in \mathbb{R}^V_0 \mapsto \sum_{e\in\vec{E}} |\,\mathrm{grad}\, x|(e)$ is a semi-norm on $\mathbb{R}^V_0$ and a norm when G is connected. It is denoted $\|\cdot\|_{\mathrm{TV}}$ throughout the paper. Although operator grad depends on the orientation chosen for $\vec{E}$, note that $\|\cdot\|_{\mathrm{TV}}$ does not.

A. Dual space of $(\mathbb{R}^V_0, \|\cdot\|_{\mathrm{TV}})$

The dual space $(\mathbb{R}^V_0)^*$ identified with $\mathbb{R}^V_0$ using the standard scalar product is equipped with the dual norm:

$$\|u\|_* = \max_{\|x\|_{\mathrm{TV}}\le 1} \langle x, u\rangle\,. \tag{5}$$

We introduce the unit ball:

$$B_* = \{u : \|u\|_* \le 1\}\,.$$

Another characterization of the dual norm is the following. For a vector field $\xi \in \mathbb{R}^{\vec{E}}$, we denote by $\|\xi\|_\infty = \max\{|\xi(e)|; e \in \vec{E}\}$. The following proposition provides a characterization of the dual norm. The proof is omitted due to the lack of space.

Proposition 1: If G is a connected graph, the following equality holds true:

$$\|u\|_* = \inf\{\|\xi\|_\infty : u = \mathrm{div}\,\xi\}\,. \tag{6}$$

The following property is a consequence of a general fact about subdifferentials of support functions:

Proposition 2: If $\partial\|x\|_{\rm TV}$ denotes the subdifferential of norm $\|\cdot\|_{\rm TV}$ at point x, one has:

$$\partial\|x\|_{\rm TV} = \{u \in \mathbb{R}_0^V : \|u\|_* \leq 1, \langle u, x\rangle = \|x\|_{\rm TV}\}$$

For instance, $\partial\|0\|_{\rm TV} = B_*$.

Proof: Let u be such that $\|u\|_* \leq 1$ and $\langle u, x\rangle = \|x\|_{\rm TV}$. Then, $\langle u, y-x\rangle + \|x\|_{\rm TV} = \langle u, y\rangle \leq \|u\|_*\|y\|_{\rm TV} = \|y\|_{\rm TV}$ which means that $u \in \partial\|x\|_{\rm TV}$. Conversely, assume $u \in \partial\|x\|_{\rm TV}$ and x_u is s.t. $\|x_u\|_{\rm TV} = 1$ and $\langle u, x_u\rangle = \|u\|_*$. Define $y_u = \|x\|_{\rm TV}x_u$; one has: $\|y_u\|_{\rm TV} - \|x\|_{\rm TV} \geq \langle u, y_u - x\rangle$, which gives $0 \geq \|u\|_*\|x\|_{\rm TV} - \langle u, x\rangle$. By inequality $\|u\|_*\|x\|_{\rm TV} - \langle u, x\rangle \geq 0$, one has $\langle u, x\rangle = \|u\|_*\|x\|_{\rm TV}$. Moreover, as $u \in \partial\|x\|_{\rm TV}$ $\|2x\|_{\rm TV} - \|x\|_{\rm TV} \geq \langle u, x\rangle = \|u\|_*\|x\|_{\rm TV}$. Hence $\|u\|_* = 1$. ∎

B. Co-area Formula

First remark that $\|\cdot\|_{\rm TV}$ can be extended into a semi-norm on $\mathbb{R}^V$ using the same definition:

$$\|x\|_{\rm TV} = \sum_{e\in\vec{E}} |\operatorname{grad} x|(e) \,.$$

Using this definition, one has $\|x + c1_V\|_{\rm TV} = \|x\|_{\rm TV}$ for any $c \in \mathbb{R}$ and any $x \in \mathbb{R}^V$. The perimeter $\operatorname{Per}(S)$ of a subset $S \subset V$ is defined as

$$\operatorname{Per}(S) = \|1_S\|_{\rm TV} \,.$$

Proposition 3: For a function $x \in \mathbb{R}^V$, we denote by $\{x \geq \lambda\} = \{v \in V : x(v) \geq \lambda\}$ the upper-level set associated with level λ. The following equality holds true:

$$\|x\|_{\rm TV} = \int_{-\infty}^{+\infty} \operatorname{Per}(\{x \geq \lambda\})\mathrm{d}\lambda \,.$$

Proof: For each edge $\{v, w\} \in E$, define the interval $I_e = [x(v)\wedge x(w), x(v)\vee x(w)] \subset \mathbb{R}$. Now, it is easy to check that $\operatorname{Per}(\{x \geq \lambda\}) = \sum_{e\in E} 1_{I_e}(\lambda)$. Hence, $\int_{-\infty}^{+\infty}\operatorname{Per}(\{x \geq \lambda\})\mathrm{d}\lambda = \sum_{e\in E}\int_{-\infty}^{+\infty} 1_{I_e}(\lambda)\mathrm{d}\lambda = \sum_{e\in E}|I_e|$ where $|I|$ denotes the length $b-a$ of interval $I = [a, b]$. The rightmost term is equal to $\|x\|_{\rm TV}$, which completes the proof. ∎

A consequence of this formula is the following useful result, which can be seen as an extension of the immediate formula $\|u\|_* = \max_{x\in\mathbb{R}^V}\langle u, x\rangle/\|x\|_{\rm TV}$.

Proposition 4: Assume u is in $(\mathbb{R}_0^V, \|\cdot\|_*)$. Then, using the canonical embedding $\mathbb{R}_0^V \subset \mathbb{R}^V$ and the standard inner product $\langle\cdot,\cdot\rangle$ over $\mathbb{R}^V$, the following equality holds true:

$$\|u\|_* = \max_{\emptyset\subsetneq S\subset V}\frac{\langle u, 1_S\rangle}{\|1_S\|_{\rm TV}} \,.$$

Proof: Since $\mathbb{R}_0^V$ is finite dimensional, there exists $x_u \in \mathbb{R}_0^V$ with $\|x_u\|_{\rm TV} = 1$ such that $\|u\|_* = \langle u, x_u\rangle$. Since $\langle u, 1_V\rangle = 0$ one has $\langle u, \tilde{x}_u\rangle = \langle u, x_u\rangle$ with $\tilde{x}_u = x_u - (\min_v x_u(v))1_V$. Now, let us consider subsets of V having the form $S_\mu = \{\tilde{x}_u \geq \mu\}$ for $\mu \in \mathbb{R}$. Notice that $S_\mu = V$ for $\mu \leq 0$ and $S_\mu = \emptyset$ for $\mu > M$ with $M > 0$ large enough. Hence, the following integral is well defined: $\int_{-\infty}^{+\infty}\langle u, 1_{S_\mu}\rangle\mathrm{d}\mu$. And, $\int_{-\infty}^{+\infty}\langle u, 1_{S_\mu}\rangle\mathrm{d}\mu = \int_0^M\langle u, 1_{S_\mu}\rangle\mathrm{d}\mu = \langle u, \int_0^M 1_{S_\mu}\mathrm{d}\mu\rangle$. where $\int_0^M 1_{S_\mu}\mathrm{d}\mu$ denotes function $v \mapsto \int_0^M 1_{S_\mu}(v)\mathrm{d}\mu$. Moreover $\forall v \in V$, $\int_0^M 1_{S_\mu}(v)\mathrm{d}\mu = \int_0^{+\infty} 1\{\mu \leq \tilde{x}_u(v)\}\mathrm{d}\mu = \tilde{x}_u(v)$. Hence $\int_{-\infty}^{+\infty}\langle u, 1_{S_\mu}\rangle\mathrm{d}\mu = \langle u, \tilde{x}_u\rangle$. By definition of $\|u\|_*$ and the fact that (i) $\langle u, 1_V\rangle = 0$, (ii) $\|x\|_{\rm TV} = \|x + c1_V\|_{\rm TV}$, one has $\langle u, 1_{S_\mu}\rangle \leq \|u\|_*\|1_{S_\mu}\|_{\rm TV}$. Hence, function $\mu \mapsto \langle u, 1_{S_\mu}\rangle - \|u\|_*\|1_{S_\mu}\|_{\rm TV}$ is nonpositive. Integrating and using the co-area formula, one gets, for almost every $\mu \in \mathbb{R}$, $\langle u, 1_{S_\mu}\rangle = \|u\|_*\|1_{S_\mu}\|_{\rm TV}$. *A fortiori*, the set of such μ is not empty, which concludes the proof. ∎

Remark 1: As will be made clear in Section IV, our algorithms shall converge to the sought consensus under some technical conditions which involve the dual norm. In order to verify these conditions, it is therefore important to have in practice an efficient algorithm for the computation of the dual norm. In that perspective, Proposition 4 helps. In an extended version of this paper, we shall introduce a strongly polynomial-time combinatorial algorithm to compute the dual norm of a vector (the description of this algorithm is omitted in this paper due to the lack of space).

IV. Variational Framework

Consider replacing problem (1) with:

$$\min_{x\in\mathbb{R}^V}\sum_{v\in V} F_v(x(v)) + G(x) \tag{7}$$

where $G : \mathbb{R}^V \to \mathbb{R}$ is a convex regularization term which penalizes the functions $x \in \mathbb{R}^V$ that are away from the consensus space $\mathcal{C}$. There are several ways to choose G. The most immediate one is $G = \iota_{\mathcal{C}}$ as the indicator function of $\mathcal{C}$. In that case, problem (7) is equivalent to problem (1). From an intuitive point of view, setting $G = \iota_{\mathcal{C}}$ means that consensus must be achieved *at any price*. However, in the presence of irregular agents, it is sometimes beneficial to break the diktat of consensus, in order to allow regular agents to possibly disagree with irregular ones. Of course, for $G \neq \iota_{\mathcal{C}}$, it can no longer be expected that the minimizers of (7) coincide in all generality with those of (1). This can be seen as the price to pay for an increased robustness. Nevertheless, we propose a way to select G such that the minimizers of (7) coincide with those of (1) *at least for a certain class of functions* $(F_v)_{v\in V}$.

A. Definition

In the sequel, we consider the following optimization problem:

$$\min_{x\in\mathbb{R}^V}\sum_{v\in V} F_v(x(v)) + \lambda\,\|x\|_{\rm TV} \,. \tag{8}$$

Here, $\lambda > 0$ is a parameter to be specified. Intuitively, when λ is large enough, the regularity term is dominant and the minimizer of (8) is forced to the consensus subspace. In the AC problem, functions F_v are given by (2) and the problem (8) reduces to:

$$\min_{x\in\mathbb{R}^V}\frac{1}{2}\|x - x_0\|_2^2 + \lambda\,\|x\|_{\rm TV} \,. \tag{9}$$

In the context of image processing, the particular objective function (9) is referred to as the ROF (Rudin-Osher-Fatemi) energy [14]. We will refer to the general objective function in (8) as a regularized energy, and to the minimizers of (8) as regularized minimizers.

Our aim is threefolds: *i)* to prove that the minimizers of (1) coincide with the regularized minimizers at least for a specified class of functions F_v; *ii)* to propose distributed algorithms to find regularized minimizers, *iii)* to analyze the limit points of the algorithms in the presence of irregular (stubborn) agents.

B. Regularized minimizers

Define function $F : \mathbb{R}^V \to \mathbb{R}$ by $F(x) = \sum_v F_v(x(v))$. For any $x^\star \in \mathbb{R}$, one has:

$$\partial F(x^\star 1_V) = \left\{ u \in \mathbb{R}^V : \forall v \in V,\, u(v) \in \partial F_v(x^\star) \right\} .$$

When all F_v's are differentiable, note that $\partial F(x^\star 1_V)$ is a singleton $\{(F_v'(x^\star))_{v\in V}\}$. Recall that B_* is the unit ball associated with the dual norm.

Theorem 1: Among the following statements, 1) 2) 3) are equivalent and imply 4).

1) $x^\star 1_V$ is a minimizer of (8) ;
2) $\partial F(x^\star 1_V) \cap \lambda B_*$ is nonempty ;
3) There exists $u \in \partial F(x^\star 1_V)$ such that $\sum_{v\in V} u(v) = 0$ and for all $A \subset V$,
$$\sum_{v\in A} u(v) \le \lambda \operatorname{Per}(A) .$$
4) $x^\star$ is a minimizer of (1).

Proof: [1) $\Leftrightarrow$ 2)] Note that $x^\star 1_V$ is a minimizer of $F + \lambda\|\cdot\|_{\mathrm{TV}}$ iff $0 \in \partial F(x^\star 1_V) + \lambda\partial\|x^\star 1_V\|_{\mathrm{TV}}$. From Proposition 1, $\partial\|x^\star 1_V\|_{\mathrm{TV}} = B_*$. Therefore, 1) holds iff there exists $u \in \partial F(x^\star 1_V)$ such that $0 \in u + \lambda B_*$. Otherwise stated, there exists $u \in \partial F(x^\star 1_V)$ such that $u \in \lambda B_*$. [2) $\Leftrightarrow$ 3)] is a consequence of Proposition 4. [3) $\Rightarrow$ 4)] As $\sum_v \partial F_v = \partial(\sum_v F_v)$, condition $\sum_{v\in V} u(v) = 0$ implies that $0 \in \partial(\sum_v F_v)(x^\star)$. Thus, $x^\star$ is a minimizer of $\sum_v F_v$. ■

We now review the consequences of Theorem 1 regarding the Average Consensus problem described earlier in the introduction.

C. Average Consensus Problem

Proposition 5: The following statements are equivalent.

1) $\overline{x}_0 1_V$ is the unique minimizer of (9) ;
2) $x_0 - \overline{x}_0\, 1_V \in \lambda B_*$;
3) For all $A \subset V$,
$$\left| \frac{\sum_{v\in A} x_0(v)}{|A|} - \bar{x}_0 \right| \le \lambda \frac{\operatorname{Per}(A)}{|A|}$$

Proposition 5 is a consequence of Theorem 1 (the proof is left to the reader). It quantifies how much a local average can fluctuate around $\bar{x}_0$ in order to preserve the sought equilibrium at $\bar{x}_0 1$: the larger the ratio $\operatorname{Per}(A)/|A|$ the more it can fluctuate safely inside A. The heuristic behind this argument is that the ratio $\operatorname{Per}(A)/|A|$ measures how well a given region A is connected to the rest of the network, since $\operatorname{Per}(A)$ is the number of edges connecting S to its complementary set $V\backslash A$ and $|A|$ is a measure of its size. A large ratio $\operatorname{Per}(A)/|A|$ amounts to say that relatively to its size, A has a lot of connections to its outside. Being well connected, a subset A is more able to make its member agree on $\bar{x}_0$ using its outside neighbors in $V\backslash A$; while a little connected part A should already agree right from the start in order to hope a local consensus.

For λ large enough, the critical value being $\|x_0 - \bar{x}_0 1\|_*$, the minimizer of (9) is the sought equilibrium point. In other terms, there is a whole range of values for λ, namely the interval $[\|x_0 - \bar{x}_0\|_*, +\infty)$ for which the minimizer is *exactly* $\bar{x}_0 1$.

Notice that if the data x_0 belong to a known bounded interval, then it is easy to compute an upper bound of $\|x_0 - \bar{x}_0\|_*$ and to select λ above this bound. Checking this condition requires the computation of $\|x_0 - \bar{x}_0\|_*$. As far as computation of the dual norm is concerned, we refer to Remark 1.

V. Proposed Algorithms

A. Subgradient Algorithm

Note that function $\lambda\|x\|_{\mathrm{TV}}$ is non-differentiable. Perhaps the most simple and natural approach is to use the subgradient algorithm associated with problem (8). This naturally yields the following distributed algorithm, where each node v holds an estimate $x_n(v)$ of the minimizer at time n and combine it with the ones received from its neighbors.

Algorithm 1:

$$x_{n+1}(v) = x_n(v) + \gamma_n \big[g_n(v) + \lambda \sum_{w\sim v} \operatorname{sign}\left(x_n(w) - x_n(v)\right) \big]$$

where $g_n(v) \in -\partial F_v(x_n(v))$, typically $g_n(v) = x_0(v) - x_n(v)$ in the (AC) case.

Standard convex optimization arguments can be used to prove that function x_n converges to a minimizer of (7) under the hypothesis of decreasing step size. The arguments being standard, the proof is omitted and we refer to [15], [16] or references therein.

B. Alternating Direction of Multipliers (ADMM)

The subgradient method is known to be rather slow in terms of convergence rate. Many alternatives do exist in order to speed up the convergence [17], [15]. Among these solutions, we propose an approach which can be seen as a special case of the ADMM [9], [17].

Let us denote by $(V, \overset{\rightleftharpoons}{E})$ the directed graph such that: $(v,w) \in \overset{\rightleftharpoons}{E}$ iff $\{v,w\} \in E$. Each edge $\{v,w\} \in E$ yields two edges in $\overset{\rightleftharpoons}{E}$. Problem (8) is equivalent to

$$\min_{(x,z)} \sum_{v\in V} F_v(x(v)) + \lambda \sum_{\{v,w\}\in E} |z(v,w) - z(w,v)|$$

where the minimum is taken w.r.t. $(x, z) \in \mathbb{R}^V \times \mathbb{R}^{\overleftrightarrow{E}}$ such that $z(v,w) = x(w)$ for any $(v,w) \in \overleftrightarrow{E}$. The augmented Lagrangian writes:

$$\mathcal{L}(x,z;\eta) = \sum_{v\in V} F_v(x(v)) + \sum_{\{v,w\}\in E} \lambda|z(v,w) - z(v,w)| + \sum_{(v,w)\in\overleftrightarrow{E}} T_\rho\left(\eta(v,w), z(v,w) - x(w)\right)$$

where we set $T_\rho(\alpha,\beta) = \alpha\beta + \frac{\rho}{2}\beta^2$. The ADMM consists in generating three sequences $(x_n, z_n, \eta_n)_{n\geq 0}$ recursively defined by

$$\begin{aligned} x_{n+1} &= \arg\min_{x\in\mathbb{R}^V} \mathcal{L}(x, z_n; \eta_n) \\ z_{n+1} &= \arg\min_{z\in\mathbb{R}^{\overleftrightarrow{E}}} \mathcal{L}(x_{n+1}, z; \eta_n) \\ \eta_{n+1}(v,w) &= \eta_n(v,w) + \rho\left(z_{n+1}(v,w) - x_{n+1}(w)\right) \end{aligned}$$

for all $(v,w) \in \overleftrightarrow{E}$. Making the above update equation explicit is tedious but straightforward. After some algebra, it can be shown that the update equation in x_n is actually given by Algorithm 2 below.

Denote by $\text{proj}_{[-\omega,\omega]}(x)$ the projection of x onto $[-\omega,\omega]$ and by $\text{prox}_{f,\rho}(x) = \arg\min_y f(y) + \frac{\rho}{2}(y-x)^2$ the proximal operator associated with a real function f.

Algorithm 2:

At each time n, agent $v \in R$ receives $(x_n(w) : w \sim v)$ and makes the following updates:

$$\mu_{n+1}(w,v) = \text{proj}_{[-2\lambda/\rho, 2\lambda/\rho]}\left(\mu_n(w,v) + x_n(w) - x_n(v)\right)$$

for any $w \sim v$ in its neighborhood. Next,

$$x_{n+1}(v) = \text{prox}_{F_v,\,\rho d(v)}\left(x_n(v) + \tfrac{3}{2}\,\tilde{\mu}_{n+1}(v) - \tfrac{1}{2}\,\tilde{\mu}_n(v)\right)$$

where we set $\tilde{\mu}_n(v) = \frac{1}{d(v)}\sum_{w\sim v}\mu_n(w,v)$.

Here, each regular agent v not only maintains an estimate $x_n(v)$ of the minimizer, but also holds in its memory one scalar $\mu_n(w,v)$ for any of its neighbors $w \sim v$. Note however that values $\mu_n(w,v)$ are purely private in the sense that they are not exchanged by agents. Agents only exchange their estimates $x_n(v)$. In the (AC) case, operator $\text{prox}_{F_v,\rho}$ has a simple expression. In that case, the update equation in x_n simplifies to:

$$x_{n+1}(v) = \frac{x_0(v) + \rho d(v)\left(x_n(v) + \frac{3}{2}\,\tilde{\mu}_{n+1}(v) - \frac{1}{2}\,\tilde{\mu}_n(v)\right)}{1 + \rho d(v)}.$$

As Algorithm 2 can be seen as a special case of a standard ADMM, the following result follows directly from [17].

Theorem 2: Assume that $R = V$ (all agents are regular). Under Assumption 1, sequence x_n converges to the minimizers of (8).

VI. Stubborn Agents

One of the claims of this paper is that the above algorithms are attractive in order to provide robustness against misbehaving agents. This claim is motivated by the example below. Assume that some agents, called *stubborn*, never change their state. The rationale behind this model is twofold: either these agents are malfunctioning, or they might want to deliberately pollute or influence the network.

For ease of interpretation, we will focus on the average consensus case. We represent the state vector as $x_n = (x_n^R, x_n^S)$ where x_n^R (resp. x_n^S) is the restriction of x_n to regular agents (resp. stubborn agents). By definition, $x_n^S = x_0^S$ for any n. As a straightforward extension of Theorem 2, it is not difficult to see that in the presence of stubborn agents, the sequence x_n^R generated either by Algorithms 1 or 2 converges to the minimizers of the following *perturbed* optimization problem

$$\min_{x\in\mathbb{R}^R} \frac{1}{2}\|x - x_0^R\|_2^2 + \lambda\,\|x\|_{\text{TV}} + \lambda \sum_{\substack{v\in R, w\in S \\ v\sim w}} |x(v) - x_0(w)|\,. \tag{10}$$

where, here, $\|x\|_{\text{TV}}$ is to be understood as the total variation of a function $x \in \mathbb{R}^R$ on the subgraph $G(R)$ *i.e.*, the restriction of G to the set of regular agents. To ease the reading and with no risk of ambiguity, we still keep the same notations $\|\,.\,\|_{\text{TV}}$ and $\|\,.\,\|_*$ to designate the TV and the dual norms associated $G(R)$.

A simple compact analytical expression of the minimizers of (10) seems unfortunately out of reach in the general case. In the sequel, we make the following assumption.

Assumption 2: Any stubborn agent is connected to all regular agents. In addition, there exist $a \in \mathbb{R}$ such that $x_0(s) = a$ for any $s \in S$.

One might think of Assumption 2 as a worst-case situation in the sense that each stubborn agent directly disturbs *all* regular agents and, intuitively, has therefore the widest possible impact.

Let $\overline{x}_0^R = \frac{1}{|R|}\sum_{v\in R} x_0^R(v)$.

Theorem 3: Assume that $\lambda \geq \|x_0^R - \overline{x}_0^R 1_R\|_*$ and let

$$x^* = \begin{cases} a & \text{if } |\overline{x}_0^R - a| \leq \lambda|S| \\ \overline{x}_0^R + \lambda|S| & \text{if } \overline{x}_0^R + \lambda|S| < a \\ \overline{x}_0^R - \lambda|S| & \text{if } \overline{x}_0^R - \lambda|S| > a \end{cases}.$$

Then, under Assumption 2, $x^* 1_R$ is the unique minimizer of Problem (10).

The proof of Theorem 3 will be provided in an extended version of this paper. Observe that even if the common value a of the *stubborn* coalition is very far from $\overline{x}_0^R$, we will reach a consensus within a distance $\lambda|S|$ from $\overline{x}_0^R$. The same conclusion holds when a is already close to $\overline{x}_0^R$ (we still reach a consensus within a distance $\lambda|S|$ from $\overline{x}_0^R$). The quantity $\lambda|S|$ can be interpreted as the robustness level of our algorithms in the sense of a maximum error margin. Therefore the proposed algorithm is unlike more standard gossip algorithms which can be driven arbitrarily far away from the sought consensus [12]. Note that a small λ reduces the error margin. The tradeoff is related to the fact that the selection of a small λ also reduces the set of functions x_0^R satisfy the regularity condition $\|x_0^R - \overline{x}_0^R 1_R\|_* \leq \lambda$.

VII. Numerical Experiments

The underlying network is the complete graph with $N = 100$ agents. The initial data is represented by circles in Figure 1. There is one stubborn agent corresponding to index 1 in all experiments. The average value of regular agents is 0.13. We represent iteration 1, 5 and 10 on this data (see Figure 1). Consensus is numerically attained at iteration 375. In order to measure how far regular agents are from

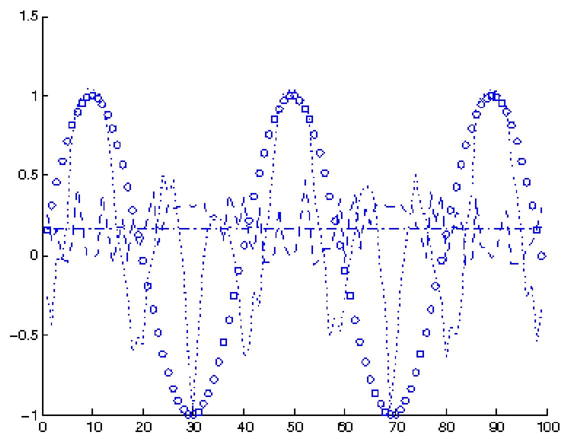

Fig. 1. Iteration 5 in dotted line, iteration 10 in dashed line, iteration 375 in dashdot.

consensus, we represent the evolution of $\log \|J^{\perp} x_n^R\|$ with n where $J^{\perp}$ is the projector onto the space of constant functions on $G(R)$ (see Figure 2). Consensus is achieved at high rate. Oscillatory patterns one can observe in Figure 2 are typical of subgradient methods that are bound to oscillate about the objective at the rate given by the decreasing size of the steps. To illustrate the (in)sensitivity to outliers of our Algorithm table I represents the average attained over regular agents for several value of the stubborn agent. First it is striking that our algorithm is not perturbed by huge values of the stubborn agent. Second, as predicted by Theorem 3, consensus value is separated by $\lambda = .04$ in our experiment from the "true" average value for $\boldsymbol{x}^0$ over the regular agents.

Stubborn Agent Value	0	-10	-10^9	10	10^9
Consensus Attained	0.09	0.09	0.09	0.17	0.17

TABLE I

Compare to "true" average 0.13. Note that this is in perfect accordance with Theorem 3 since $\lambda = .04$ in this experiment.

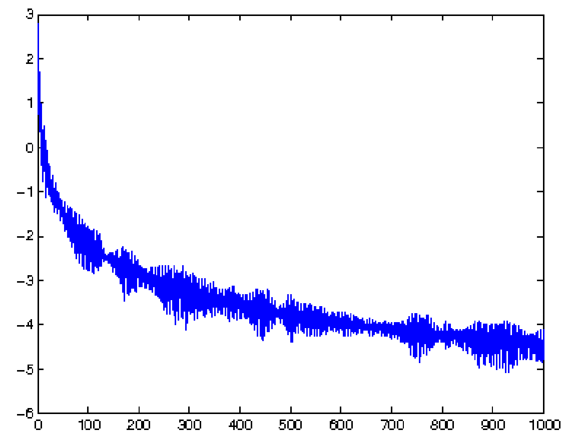

Fig. 2. $n \mapsto \log \|J^{\perp} x_n^R\|$

References

[1] D. P. Bertsekas and J. N. Tsitsiklis, *Parallel and Distributed Computation: Numerical Methods*, Athena Scientific, 1997.

[2] C. Lopes and A.H. Sayed, "Distributed processing over adaptive networks," in *Adaptive Sensor Array Processing Workshop*, June 2006, pp. 1–5.

[3] S.S. Ram, A. Nedic, and V. Veeravalli, "Distributed stochastic subgradient projection algorithms for convex optimization," *Journal of Optimization Theory and Applications*, vol. 147, pp. 516–545, 2010.

[4] P. Bianchi and J. Jakubowicz, "Convergence of a multi-agent projected stochastic gradient algorithm for non-convex optimization," *IEEE Trans. on Automatic Control*, vol. 58, no. 2, 2013.

[5] D. Kempe, A. Dobra, and J. Gehrke, "Gossip-based computation of aggregate information," in *Foundations of Computer Science, Annual IEEE Symposium on*. 2003, IEEE Computer Society.

[6] S. Boyd, A. Ghosh, B. Prabhakar, and D. Shah, "Randomized Gossip Algorithms," *IEEE Transactions on Inform. Theory*, vol. 52, no. 6, pp. 2508–2530, 2006.

[7] Dusan Jakovetic, Joao Xavier, and Jose MF Moura, "Fast distributed gradient methods," *arXiv preprint arXiv:1112.2972*, 2011.

[8] John C Duchi, Alekh Agarwal, and Martin J Wainwright, "Dual averaging for distributed optimization: convergence analysis and network scaling," *Automatic Control, IEEE Transactions on*, vol. 57, no. 3, pp. 592–606, 2012.

[9] I.D. Schizas, A. Ribeiro, and G.B. Giannakis, "Consensus in ad hoc WSNs with noisy links-Part I: Distributed estimation of deterministic signals," *IEEE Transactions on Signal Processing*, vol. 56, no. 1, pp. 350–364, 2008.

[10] C. Eksin and A. Ribeiro, "Distributed network optimization with heuristic rational agents," *IEEE Trans. Signal Process,(submitted)*, 2011.

[11] D. Acemoglu, G. Como, F. Fagnani, and A. Ozdaglar, "Opinion fluctuations and persistent disagreement in social networks," in *IEEE CDC 2011*, 2011, pp. 2347–2352.

[12] P. Benameur, P. Bianchi, and J. Jakubowicz, "Robust average consensus using total variation gossip algorithm," in *ValueTools' 12: 6th International Conference on Performance Evaluation Methodologies and Tools*, 2012.

[13] Fabio Pasqualetti, Antonio Bicchi, and Francesco Bullo, "Consensus computation in unreliable networks: A system theoretic approach," *Automatic Control, IEEE Transactions on*, vol. 57, no. 1, pp. 90–104, 2012.

[14] Leonid I Rudin, Stanley Osher, and Emad Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D: Nonlinear Phenomena*, vol. 60, no. 1, pp. 259–268, 1992.

[15] Francis Bach, Rodolphe Jenatton, Julien Mairal, and Guillaume Obozinski, "Optimization with sparsity-inducing penalties," *arXiv preprint arXiv:1108.0775*, 2011.

[16] S. Boyd, L. Xiao, and A. Mutapcic, "Subgradient methods," Tech. Rep., October 2003, Lecture Notes.

[17] Patrick L. Combettes and Jean christophe Pesquet, *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*, chapter Proximal Splitting Methods in Signal Processing, pp. 185–222, Springer, 2011.