A gossip-like distributed optimization algorithm for reactive power flow control

Saverio Bolognani and Sandro Zampieri

Department of Information Engineering,
University of Padova, Padova, Italy
e-mail: {saverio.bolognani, zampi}@dei.unipd.it

April 14, 2011

Abstract

We considered the problem of minimizing reactive power flows in a smart microgrid. First we modeled this problem as a linearly constrained quadratic optimization, in which the decision variables are the amount of reactive power that compensators inject into the network. Then, we designed a distributed algorithm in which agents are clustered into overlapping subsets, according to a given communication graph that allows them to coordinate and to exchange information. At each time, one subset is triggered, and agents belonging to it update their states in order to minimize the reactive power flows on the grid. We showed that, by sensing the network at their points of connection, agents can perform this minimization with just the data that they can gather from the other agents belonging to the subset. We characterized convergence of this algorithm in term of conditions on the subsets and on the randomized triggering sequence. Moreover, we studied the rate of convergence, obtaining also a convenient upper bound. We finally analyzed the rate of convergence for some specific topologies of the grid and for some choices of the agents communication topologies.

1 Introduction

Distributed optimization has historically been intended as the problem of dispatching part of a large scale optimization algorithm to different computational units, e.g. [10, 1].

More recently, distributed optimization has been applied to complex, large-scale systems. Large-scale systems consist in a large, sometimes unknown and time-varying, number of agents. These agents can usually communicate, they interact with an underlying physical system by sensing and actuating it, and none of them have a complete knowledge of the system state and parameters.

When dealing with distributed optimization in this context, we face the problem of designing an algorithm for the agents that allows them to drive the system in a state that minimizes some global cost function while enforcing some feasibility constraints (see for example [7] and references therein). The behavior of each agent must necessarily depend only on local data and measurements, and on the information that each node can gather from the nodes that are in its communication neighborhood.

A notable "success story" in this sense is the application of distributed optimization to the Internet: since the work of [6], large-scale data networks have been probably the preferred testbed for these algorithms.

In this work, we focus on the problem of optimal reactive power compensation in power distribution networks. This application is part of the extremely important framework of ancillary services in smart-grids ([8], [5]), which can be considered among the most interesting and intriguing testbeds at the moment.

The problem belongs to the class of single-commodity network flow problems. We show in Section 2 that optimal reactive power compensation can be modeled as a quadratic optimization problem with a linear constraint, in which the decision variables are the amounts of reactive power that are injected by the agents, connected to different nodes of the power grid.

Because of a coupling constraint and because the cost function is not separable, coordination among agents is required. In Section 3 we show how it is possible to reduce the need for global coordination

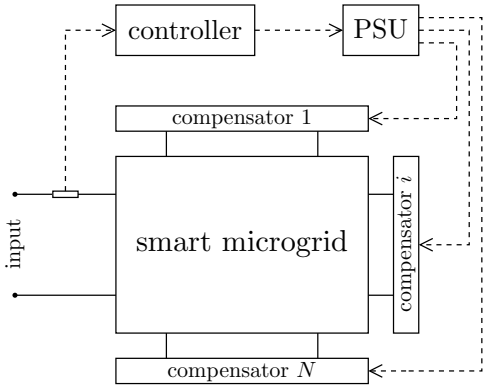

Figure 1: The controller structure of [9].

by clustering the agents into groups and by defining some optimization subproblems that have to be executed iteratively.

In Section 4 we characterize the convergence of the algorithm by giving conditions on the clustering choices and on the randomized execution of the subproblems. In Section 5 we analyze the speed of convergence of this algorithm, giving a convenient bound for it and comparing different clustering choices for some network topologies (both analytically and numerically).

Finally, in Section 6, we derive some conclusions and we list some possible future direction of investigations, motivated by the results of this work.

2 Problem description

We define a *smart microgrid* as a portion of the electrical power distribution network that connects to the transmission grid in one point and that is managed autonomously from the rest of the network.

In particular, ancillary services are taken care by some microgrid controllers, whose objective is to operate the microgrid in an optimal way while satisfying some constraint on how the microgrid interfaces with the rest of the network. Among them, we focus on the problem of optimal reactive power compensation. This problem has been described in greater detail in [2], and will be quickly reviewed in this section.

Both residential and industrial users belonging to the microgrid may require a sinusoidal current which is not in phase with voltage. A convenient description for that consists in saying that they demand reactive power together with active power, associated with out-of-phase and in-phase components of the current, respectively.

Reactive power is not a "real" physical power, meaning that there is no energy conversion involved nor fuel costs to produce it. Like active power flows, reactive power flows contribute to power losses on the lines, cause voltage drop, and may lead to grid instability. It is therefore preferable to minimize reactive power flows by producing it as close as possible to the users that need it.

One possible approach has been proposed in [9], and is sketched in Figure 1.

It consists in a centralized controller that measures the reactive power flow at the input port of the microgrid, i.e. where the microgrid connects with the main grid. According to this measurement, the controller produces a reference for the amount of reactive power that has to be produced inside the microgrid. This reference has to be split by a power sharing unit (PSU) among some devices connected to the network that can produce a commanded amount of reactive power (*compensators*), in a way that minimizes reactive power flows inside the microgrid. In this paper we focus on the optimization problem faced by the PSU, therefore assuming that the total amount of reactive power to be produced is fixed and given.

Let us introduce a mathematical model for this problem. Let the electrical connections in the microgrid be described by a tree of $\bar{N}$ agents. Each agent injects a certain amount of reactive power into the network. N of them (the compensators) can be commanded to inject a given amount of reactive power, while the other nodes (users) inject (or are supplied with, if negative) a fixed and unknown amount (see Figure 2). Flows f_i on the tree edges are oriented outbound from the tree root and indexed as the child node they point to. Reactive power obeys regular flow conservation equations.

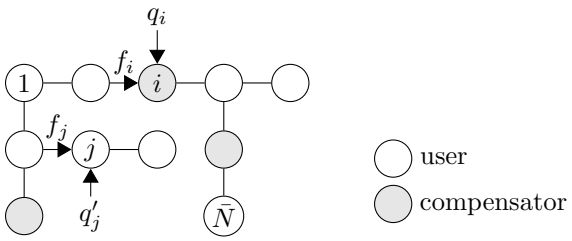

Figure 2: Direction of the flows and naming conventions in the tree of users and compensators.

As power losses are a quadratic function of the reactive power flowing on a line, the optimization problem of having minimal power losses corresponds to the cost function

$$F(f_2, \dots, f_N) = \sum_{i=2}^{N} k_i f_i^2 \tag{1}$$

where k_i is the resistance of the edge i (which goes linearly with the length of the line).

Define q as the vector of all the amounts of reactive power injected by the compensators, and by q' those injected by the users that cannot be commanded. From the conservation of reactive power constraint, any flow f_i can be expressed as the sum of some of the reactive power injected into the network by those agents. In other words we can determine the matrices K and K' with entries in $\{0, 1\}$ such that

$$f = Kq + K'q'$$

This, together with the global flow conservation law $\mathbf{1}^T q + \mathbf{1}^T q' = 0$, allow us to rewrite the optimization problem (1) as

$$\begin{aligned} \min_{q} \quad & F(q) = q^T \frac{M}{2} q + m^T q \\ \text{subject to} \quad & \mathbf{1}^T q = c, \end{aligned} \tag{2}$$

where

$$\begin{aligned} M &= 2K^T \operatorname{diag}(k_2, \dots, k_N) K, \\ m^T &= 2{q'}^T {K'}^T \operatorname{diag}(k_2, \dots, k_N) K, \\ c &= -\mathbf{1}^T q'. \end{aligned}$$

The problem of minimizing reactive power flows inside a microgrid has therefore been expressed as a quadratic, linearly constrained, optimization problem, whose analytic solution is:

$$q^* = -M^{-1} m + \frac{\mathbf{1}^T M^{-1} m + c}{\mathbf{1}^T M^{-1} \mathbf{1}} M^{-1} \mathbf{1}. \tag{3}$$

The size of this problem (i.e. the number of compensators) can be very large, as the electronic interface of any distributed generator (wind turbines, combined heat and power generators, micro hydroelectric, solar panels) can also produce reactive power at no additional cost. Each of these units is capable of:

- sensing the electric network at its point of connection to the grid;
- performing some amount of computation and data processing;
- communicating with other agents, according to some communication graph that may or may not coincide with the electric network;
- actuating the system, by injecting a certain amount of reactive power.

The agents may have a partial knowledge of the problem parameters M and m (which depend on the electrical network topology and on the reactive power demand), while no agent knows them entirely.

For these reasons, in the next section we will explore the possibility of solving (2) via a *distributed algorithm.*

3 Optimization problem decomposition

Let the agents be divided into ℓ possibly overlapping sets $\mathcal{C}_1, \dots, \mathcal{C}_\ell$, with $\bigcup_{i=1}^{\ell} \mathcal{C}_i = \{1, \dots, N\}$. This family of subsets can be interpreted as a hypergraph defined over the node set $\{1, \dots, N\}$.

Nodes belonging to the same set are able to communicate each other (they form a clique in the communication graph), and they are therefore capable of coordinating and sharing measurements and local knowledge of the problem parameters M and m. We assume that, by using this information, nodes belonging to the same set are capable of driving their state in a new feasible state that minimizes $F(q)$, solving the optimization *subproblem* in which all nodes that are not in $\mathcal{C}_i$ keep their state constant:

$$\begin{aligned} \arg\min_{\Delta q} \quad & F(q + \Delta q) \\ \text{subject to} \quad & \Delta q \in \mathcal{S}_i, \end{aligned}$$

where

$$\mathcal{S}_i := \left\{ q \in \mathbb{R}^N \ : \ \sum_{j \in \mathcal{C}_i} q_j = 0 \ , q_j = 0 \ \forall j \notin \mathcal{C}_i \right\}.$$

One possible way in which nodes in $\mathcal{C}_i$ can solve this optimization subproblem is described hereafter.

Let $\mathcal{A}, \mathcal{B} \subseteq \{1, \dots, N\}$ be two nonempty set of indices. Define $M_{\mathcal{A}\mathcal{B}}$ as the submatrix of M obtained by selecting the rows indexed by $\mathcal{A}$ and the columns indexed by $\mathcal{B}$. Let the same definition hold for a vector, i.e. $v_{\mathcal{A}}$ is the subvector of v obtained by selecting the elements indexed by $\mathcal{A}$. Let moreover $\bar{\mathcal{C}}_i$ be the set complement $\{1, \dots, N\} \backslash \mathcal{C}_i$.

The optimization problem faced by the nodes in $\mathcal{C}_i$ can then be rewritten as

$$\begin{aligned} \min_{q_{\mathcal{C}_i}} \quad & q_{\mathcal{C}_i}^T \frac{M_{\mathcal{C}_i \mathcal{C}_i}}{2} q_{\mathcal{C}_i} + \left(q_{\bar{\mathcal{C}}_i} M_{\bar{\mathcal{C}}_i \mathcal{C}_i} + m_{\mathcal{C}_i}^T \right) q_{\mathcal{C}_i} \\ \text{subject to} \quad & \mathbf{1}^T q_{\mathcal{C}_i} = c - \mathbf{1}^T q_{\bar{\mathcal{C}}_i}. \end{aligned} \tag{4}$$

It is easy to see that agents in $\mathcal{C}_i$ can reach the optimal solution by adding to $q_{\mathcal{C}_i}$ the increment $\Delta q_{\mathcal{C}_i}$ which is equal to

$$\Delta q_{\mathcal{C}_i} = -M_{\mathcal{C}_i \mathcal{C}_i}^{-1} \nabla F_{\mathcal{C}_i} + \frac{\mathbf{1}^T M_{\mathcal{C}_i \mathcal{C}_i}^{-1} \nabla F_{\mathcal{C}_i}}{\mathbf{1}^T M_{\mathcal{C}_i \mathcal{C}_i}^{-1} \mathbf{1}} M_{\mathcal{C}_i \mathcal{C}_i}^{-1} \mathbf{1},$$

where $M_{\mathcal{C}_i \mathcal{C}_i}^{-1}$ is the inverse of submatrix $M_{\mathcal{C}_i \mathcal{C}_i}$ and

$$\nabla F_{\mathcal{C}_i} = M_{\mathcal{C}_i \mathcal{C}_i} q_{\mathcal{C}_i} + M_{\mathcal{C}_i \bar{\mathcal{C}}_i} q_{\bar{\mathcal{C}}_i} + m_{\mathcal{C}_i}$$

is the subvector of the gradient of $F(q)$ corresponding to the agents belonging to $\mathcal{C}_i$.

It has been shown in [2], that an estimate of the gradient $\nabla F(q) = Mq + m$ can be obtained by sensing the network when in the state q. More precisely, it has been shown that, under a certain assumption on the impedance of the lines, the steady state voltage measurement $u \in \mathbb{R}^N$ approximates $\nabla F(q)$ up to a common additive term, namely

$$u \approx \nabla F(q) + \alpha \mathbf{1},$$

with α unknown.

Nodes in $\mathcal{C}_i$ can therefore solve their corresponding optimization subproblem by performing the update

$$\Delta q_{\mathcal{C}_i} = -M_{\mathcal{C}_i \mathcal{C}_i}^{-1} u_{\mathcal{C}_i} + \frac{\mathbf{1}^T M_{\mathcal{C}_i \mathcal{C}_i}^{-1} u_{\mathcal{C}_i}}{\mathbf{1}^T M_{\mathcal{C}_i \mathcal{C}_i}^{-1} \mathbf{1}} M_{\mathcal{C}_i \mathcal{C}_i}^{-1} \mathbf{1},$$

as the uncertain term $\alpha \mathbf{1}$ get canceled in the expression.

The data that any node k in $\mathcal{C}_i$ has to know are then its corresponding row in $M_{\mathcal{C}_i \mathcal{C}_i}^{-1}$, the voltage measurements $u_{\mathcal{C}_i}$ from the other nodes in $\mathcal{C}_i$, and its own state q_k.

Notice that the update law requires only local information or information that can be gathered inside the subset $\mathcal{C}_i$ (which is a clique in the communication graph). This is possible because we are able to estimate the gradient, that otherwise would depend on the whole system state, from some local measurements. Moreover, the elements of $M_{\mathcal{C}_i \mathcal{C}_i}$ depends only on the length of the electric paths that connect nodes in $\mathcal{C}_i$, and therefore we can assume that this information is available and shared among the nodes of the same cluster.

The proposed optimization algorithm will therefore consists of the following, repeated steps:

i) a set $\mathcal{C}_i$ is chosen according to a sequence of symbols $\sigma(t) \in \{1, \dots, \ell\}$;

ii) agents in $\mathcal{C}_i$ sense the network and obtain, directly or via some filtering, an estimate of the gradient;

iii) they determine a feasible update step that minimizes the given cost function, possibly by coordinating their actions and communicating;

iv) they actuate the system by updating their state (the injected reactive power).

The iterated algorithm will then results in the following discrete time system for q

$$\begin{aligned} q(t+1) = T_{\sigma(t)}[q(t)] \quad := \quad & \arg\min_{q} \quad F(q) \\ & \text{subject to} \quad q - q(t) \in \mathcal{S}_{\sigma(t)}, \end{aligned} \tag{5}$$

with initial conditions $q(0)$ such that $\mathbf{1}^T q(0) = c$.

The following notation will be useful in the rest of the paper. Define the $N \times N$ matrices

$$\Omega_i = I_{\mathcal{C}_i} - \frac{1}{|\mathcal{C}_i|} \mathbf{1}_{\mathcal{C}_i} \mathbf{1}_{\mathcal{C}_i}^T$$

where $|\mathcal{C}_i|$ is the cardinality of the set $\mathcal{C}_i$, $I_{\mathcal{C}_i}$ is the diagonal matrix having diagonal entries 1 in positions belonging to $\mathcal{C}_i$ and zero elsewhere and $\mathbf{1}_{\mathcal{C}_i}$ is the column vector having entries 1 in positions belonging to $\mathcal{C}_i$ and zero elsewhere. Notice that

$$\Omega_i = \frac{1}{2|\mathcal{C}_i|} \sum_{h,k \in \mathcal{C}_i} (e_h - e_k)(e_h - e_k)^T$$

where e_i is the column vector having entry 1 in position i and zero elsewhere. Observe that $\mathcal{S}_i = \operatorname{Im} \Omega_i$.

4 Convergence results

To study the convergence of the proposed algorithm and its speed, we introduce the auxiliary variable $x = q - q^*$, where q^* is given in (3). By substitution, it can be shown that the optimization problem (2) is equivalent to

$$\begin{aligned} \min_{x} \quad & V(x) = x^T \frac{M}{2} x \\ \text{subject to} \quad & \mathbf{1}^T x = 0, \end{aligned} \tag{6}$$

and that the subproblems described in the previous section are equivalent to the subproblems

$$\begin{aligned} \min_{\Delta x} \quad & V(x + \Delta x) \\ \text{subject to} \quad & \mathbf{1}^T \Delta x = 0, \\ & \Delta x \in \operatorname{Im} \Omega_i. \end{aligned} \tag{7}$$

In this notation, it is possible to explicitly express the solution of the individual subproblems as a linear function of the starting point $x(t)$:

$$x(t+1) = F_i x(t), \quad F_i = I - (\Omega_i M \Omega_i)^\sharp M, \tag{8}$$

where $\sharp$ means pseudoinverse.

The discrete time system (5) in the x coordinates results then to be a linear time varying system of the form

$$x(t+1) = F_{\sigma(t)} x(t). \tag{9}$$

It is easy to verify from the properties of the pseudoinverse that $\ker(\Omega_i M \Omega_i)^\sharp = \ker \Omega_i$ and $\operatorname{Im}(\Omega_i M \Omega_i)^\sharp = \operatorname{Im} \Omega_i$. The matrices F_i are projection operators, i.e. $F_i^2 = F_i$, and they are orthogonal projections with respect to the inner product $\langle \cdot, \cdot \rangle_M$, defined as $\langle x, y \rangle_M := x^T M y$. In other words, $\langle F_i x, F_i x - x \rangle_M = x^T M (F_i x - x) = 0$. Moreover, they are self-adjoint matrices with respect to the inner product $\langle \cdot, \cdot \rangle_M$, i.e. $F_i^T M = M F_i$.

The following results characterize the uniqueness of the equilibrium for all maps $T_i[x] = F_i x$.

Lemma 1. *Consider the family of linear transformations* $\{F_i\}$ *as described in* (8). $\bar{x} = 0$ *is the only point in* $\ker \mathbf{1}^T$ *that is invariant for all* F_i*'s if and only if*

$$\mathrm{Im}[\Omega_1 \dots \Omega_\ell] = \ker \mathbf{1}^T.$$

Proof. Let us prove the reverse implication first. If $\mathrm{Im}[\Omega_1 \dots \Omega_\ell] = \ker \mathbf{1}^T$, then we can express $\bar{x}$ as

$$\bar{x} = \sum_i \Omega_i y_i.$$

Moreover, as $F_i \bar{x} = \bar{x}$ for all i, then $M\bar{x} \in \ker \Omega_i$. We therefore have

$$\bar{x}^T M \bar{x} = \sum_i y_i^T \Omega_i M \bar{x} = 0,$$

and so, since M is positive definite, $\bar{x} = 0$.

Suppose conversely that

$$\ker \begin{bmatrix} \mathbf{1}^T \\ I - F_1^T \\ \vdots \\ I - F_\ell^T \end{bmatrix} = 0$$

and take any $\bar{x} \in \ker \mathbf{1}^T$. Then there exist a scalar α and vectors y_i such that

$$\begin{aligned} M\bar{x} &= \alpha \mathbf{1} + \sum_i (I - F_i)^T y_i \\ &= \alpha \mathbf{1} + \sum_i^T M(\Omega_i M \Omega_i)^\sharp y_i \end{aligned}$$

Then

$$\bar{x} = M^{-1}\alpha \mathbf{1} + \sum_i^T (\Omega_i M \Omega_i)^\sharp y_i$$

Now observe that $\mathbf{1}^T \bar{x} = \alpha \mathbf{1}^T M^{-1} \mathbf{1}$, which, since M^{-1} is positive definite, implies that $\alpha = 0$ and so $\bar{x} \in \mathrm{Im}[\Omega_1 \dots \Omega_\ell]$. The converse inclusion is trivial. □

The condition $\mathrm{Im}[\Omega_1 \dots \Omega_\ell] = \ker \mathbf{1}^T$ is then a necessary condition for the convergence of the algorithm. Notice that this condition is equivalent to the fact that $L + \mathbf{1}\mathbf{1}^T$ is positive definite, where

$$\begin{aligned} L &:= [\Omega_1 \dots \Omega_\ell] \,\mathrm{diag}\{2|\mathcal{C}_1|I, \dots, 2|\mathcal{C}_\ell|I\} [\Omega_1 \dots \Omega_\ell]^T \\ &= \sum_i 2|\mathcal{C}_i| \Omega_i = \sum_{i=1}^{\ell} \sum_{h,k \in \mathcal{C}_i} (e_h - e_k)(e_h - e_k)^T \\ &= \sum_{h,k=1}^{N} (e_h - e_k)(e_h - e_k)^T \sum_{i=1}^{\ell} \delta_{\mathcal{C}_i}(h) \delta_{\mathcal{C}_i}(k), \end{aligned}$$

and where the symbol $\delta_{\mathcal{C}_i}(\cdot)$ means the characteristic function of the set $\mathcal{C}_i$, namely a function of the nodes that is 1 when the node belongs to $\mathcal{C}_i$ and is zero otherwise. The matrix L can be interpreted an the Laplacian matrix of a weighted graph $\mathcal{G}$ having nodes $\{1, \dots, N\}$ and weights on the edge h, k equal to the number of the sets $\mathcal{C}_i$ which contains both h and k. As we did before, we can interpret the family of sets $\{\mathcal{C}_1, \dots, \mathcal{C}_\ell\}$ as an hypergraph $\mathcal{H}$. It is quite easy to see that the hypergraph $\mathcal{H}$ with edges $\mathcal{C}_i$ is connected if and only if $\mathcal{G}$ is a connected graph. From these arguments we can state the following proposition.

Proposition 2. *The condition* $\mathrm{Im}[\Omega_1 \dots \Omega_\ell] = \ker \mathbf{1}^T$ *holds if and only if* $\mathcal{H}$ *is a connected hypergraph.*

We characterize now the convergence of the algorithm under the following assumption on the sequence $\sigma(t)$.

Assumption 3. *The sequence* $\sigma(t)$ *is a sequence of independently, uniformly distributed symbols in* $\{1, \dots, \ell\}$.

Let us review the main tool that we need for this, namely the formalism of *set-valued* maps and the Invariance Principle for these maps.

A set-valued map $T : X \rightrightarrows X$ associates to an element of X a subset of X. T is non-empty if $T(x) \neq \emptyset$ for all $x \in X$. An evolution of the dynamical system determined by a non-empty set-valued map T is a sequence $\{x_t\}_{t\in\mathbb{Z}_{\geq 0}}$ with the property that $x_{t+1} \in T(x_t)$ for all $t \in \mathbb{Z}_{\geq 0}$. A set W is strongly positively invariant for T if $T(w) \subset W$ for all $w \in W$. The following theorem holds.

Theorem 4 (Th. 4.5 in [3])**.** *Let (X, d) be a metric space. Given a collection of maps $T_1, \dots, T_\ell$, define the set-valued map $T : X \rightrightarrows X$ by $T(x) = \{T_1(x), \dots, T_\ell(x)\}$. Given a stochastic process $\sigma : \mathbb{Z}_{\geq 0} \to \{1, \dots, \ell\}$, consider an evolution $\{x_n\}_{n\in\mathbb{Z}_{\geq 0}}$ of T satisfying*

$$x_{n+1} = T_{\sigma(n)}(x_n).$$

Assume that

i) there exists a compact set $W \subseteq X$ that is strongly positively invariant for T;

ii) there exists a function $U : W \to \mathbb{R}$ such that $U(w') < U(w)$, for all $w \in W$ and $w' \in T(w)\backslash\{w\}$;

iii) the maps T_i, for $i \in \{1, \dots, \ell\}$, and U are continuous on W; and

iv) there exists $p \in]0, 1[$ and $h \in \mathbb{N}$ such that, for all $i \in \{1, \dots, \ell\}$ and $n \in \mathbb{Z}_{\geq 0}$

$$\mathbb{P}[\sigma(n + h) = i|\sigma(n), \dots, \sigma(1)] \geq p.$$

If $x_0 \in W$, then there exists $c \in \mathbb{R}$ such that almost surely the evolution $\{x_n\}_{n\in\mathbb{Z}_{\geq 0}}$ approaches the set

$$(J_1 \cap \dots \cap J_\ell) \cap U^{-1}(c),$$

where $J_i = \{w \in W|T_i(w) = w\}$ is the set of fixed points of T_i in W, $i \in \{1, \dots, \ell\}$.

We can then state the following.

Theorem 5. *Consider the discrete time system* (9)*, under Assumption 3. If* $\mathrm{Im}[\Omega_1 \dots \Omega_\ell] = \ker \mathbf{1}^T$*, then*

$$x(t) \to 0 \quad as \quad t \to \infty \quad almost\ surely$$

for all $x(0) \in \mathbb{R}^N$.

Proof. Consider the maps $F_i(x) = F_i x$ and the set-valued map $T(x) = \{F_1(x), \dots, F_\ell(x)\}$. Let W be the compact set $V^{-1}(x(0))$. V is strongly positively invariant for T as $V(F_i x) \leq V(x)$ for all x, i (as $F_i x$ solves the optimization subproblems initialized in x). As F_i's are orthogonal projection matrices for the norm $\langle\cdot,\cdot\rangle_M$, $V(F_i x) = V(x)$ implies $F_i x = x$ (as $V(x) = \|x\|_M^2/2$), and then $F_i x \neq x$ implies $V(F_i x) < V(x)$. Moreover, because of Assumption 3, for all n, i we have

$$\mathbb{P}[\sigma(n + 1) = i|\sigma(n), \dots, \sigma(1)] = \mathbb{P}[\sigma(n + 1) = i] = \frac{1}{\ell} > 0.$$

Theorem 4 then applies. Because of Lemma 1, the intersection of the fixed points of the maps F_i reduces to $x = 0$, and therefore $x(t) \to 0$ almost surely as $t \to \infty$. □

5 Rate of convergence

Consider the performance metric

$$R = \sup_{x(0)\in\ker \mathbf{1}^T} \limsup v(t)^{1/t}$$

where $v(t) = \mathbb{E}\left[V(x(t))\right]$. R describes the exponential rate of convergence to zero of $v(t)$ and so also the exponential rate of convergence of $q(t)$ to the optimal solution q^*. Using (8), we have

$$\begin{aligned} v(t) &= \frac{1}{2}\mathbb{E}\left[x(t)^T M x(t)\right] \\ &= \frac{1}{2}\mathbb{E}\left[x(t)^T \Omega M \Omega x(t)\right] \\ &= \frac{1}{2}\mathbb{E}\left[x(t-1)^T F_{\sigma(t-1)}^T \Omega M \Omega F_{\sigma(t-1)} x(t-1)\right] \\ &= \frac{1}{2}x(0)^T \mathbb{E}\left[F_{\sigma(0)}^T \cdots F_{\sigma(t-1)}^T \Omega M \Omega F_{\sigma(t-1)} \cdots F_{\sigma(0)}\right] x(0). \end{aligned}$$

Let us then define

$$\Delta(t) = \mathbb{E}\left[F_{\sigma(0)}^T \cdots F_{\sigma(t-1)}^T \Omega M \Omega F_{\sigma(t-1)} \cdots F_{\sigma(0)}\right].$$

Via Assumption 3, we can derive the following linear system:

$$\begin{aligned} \Delta(t+1) &= \mathbb{E}\left[F^T \Delta F\right] = \mathcal{L}(\Delta(t)), \quad \Delta(0) = \Omega M \Omega \\ \Xi(t) &= \Omega \Delta(t) \Omega, \end{aligned} \tag{10}$$

and express the expected cost function as

$$\mathbb{E}\left[V(x(t))\right] = v(t) = \frac{1}{2} x(0)^T \Xi(t) x(0).$$

Let denote by $\mathbf{F}$ the $N^2 \times N^2$ matrix associated with the linear transformation $\mathcal{L}$:

$$\operatorname{vec}\left(\Delta(t+1)\right) = \mathbf{F} \operatorname{vec}\left(\Delta(t)\right),$$

where $\operatorname{vec}(\cdot)$ is the operation of vectorization. We then have

$$\mathbf{F} = \mathbb{E}\left[F^T \otimes F^T\right],$$

which is self-adjoint with respect to the inner product $\langle \cdot, \cdot \rangle_{M^{-1} \otimes M^{-1}}$. Therefore $\mathbf{F}$ has real eigenvalues. We can define the function

$$\lambda_{\mathcal{L}}(i) : \{1, \dots, N^2\} \to \mathbb{R}$$

that returns the i-th eigenvalue of $\mathbf{F}$. We assume that the function is monotonically non increasing, i.e. $\lambda_{\mathcal{L}}(i) \geq \lambda_{\mathcal{L}}(i+1)$ for all i. We can represent this map as an N^2-dimensional ordered vector (in decreasing order, with possible repetitions) $\lambda_{\mathcal{L}} = [\lambda_{\mathcal{L}}(1) \cdots \lambda_{\mathcal{L}}(N^2)]^T$. Let moreover $\Delta_{\mathcal{L}}(i)$ be an eigenvector associated with the eigenvalue $\lambda_{\mathcal{L}}(i)$.

By decomposing $\Omega M \Omega$ into $\sum_i \alpha_i \Delta_{\mathcal{L}}(i)$ we can then express the convergence rate R as

$$R = \max\left\{|\lambda_{\mathcal{L}}(i)| \mid \alpha_i \neq 0,\ \Omega \Delta_{\mathcal{L}}(i) \Omega \neq 0\right\}. \tag{11}$$

The following proposition relates the convergence result of Theorem 5 with the approach of this section, showing how the same conditions for convergence also guarantee asymptotic stability of the dynamics of (10).

Proposition 6. *Let* $\operatorname{Im}\left[\Omega_1 \cdots \Omega_\ell\right] = \ker \mathbf{1}^T$. *Then* $R < 1$.

Proof. Let us define $\mathcal{L}_i$ as the linear transformation $\mathcal{L}_i(\Delta) = F_i^T \Delta F_i$. The N^2 eigenvalues of $\mathcal{L}_i$ are $\lambda(F_i^T \otimes F_i^T)$, and therefore belong to the set $\{0, 1\}$. As $F_i^T \otimes F_i^T$ is self-adjoint, it follows that $\|F_i^T \otimes F_i^T\|_2 \leq 1$. By using the fact that $\mathcal{L}(\Delta)$ is a convex combination of the elements of $\{\mathcal{L}_i(\Delta)\}$, we have

$$\max\{|\lambda_{\mathcal{L}}(i)|\} \leq \|\mathbb{E}\left[F^T \otimes F^T\right]\| \leq 1.$$

Let us then consider $\lambda_{\mathcal{L}}(i)$ such that $|\lambda_{\mathcal{L}}(i)| = 1$, and let $x = \operatorname{vec}(\Delta_{\mathcal{L}}(i))$ be the corresponding eigenvector of $\mathbf{F}$. We have

$$\|x\| = \left\|\mathbb{E}\left[F^T \otimes F^T\right] x\right\| \leq \mathbb{E}\left[\left\|F^T \otimes F^T x\right\|\right] \leq \|x\|.$$

We therefore must have

$$\left\|F_i^T \otimes F_i^T x\right\| = \|x\| \quad \forall i.$$

$F_i^T \otimes F_i^T$ has only 0 and 1 eigenvalues, and eigenvectors $v_h^{(i)} \otimes v_k^{(i)}$, where $v_{h,k}^{(i)}$ are right eigenvectors of F_i. Therefore we must have

$$(F_i^T \otimes F_i^T) x = x \quad \forall i$$

and then

$$x = v_h \otimes v_k, \quad \Omega_i^T v_h = \Omega_i^T v_k = 0 \quad \forall i.$$

As

$$\operatorname{Im}\left[\Omega_1 \cdots \Omega_\ell\right] = \ker \mathbf{1}^T \quad \Rightarrow \quad \bigcap_i \ker \Omega_i^T = \operatorname{Im} \mathbf{1},$$

we have $v_h = v_k = \mathbf{1}$, and therefore the only eigenvector of $\mathcal{L}$ corresponding to an eigenvalue of norm 1 is $\Delta_{\mathcal{L}}(1) = \mathbf{1}\mathbf{1}^T$. As it is not observable ($\Omega \mathbf{1}\mathbf{1}^T \Omega = 0$), we conclude that $R < 1$. □

Computing R as defined in (11) is in general not simple. In the following, we will derive an upper bound for R that can be computed from $\bar{F} = \mathbb{E}[F]$. We first state a few technical lemmas.

Lemma 7. *Let $P, Q \in \mathbb{R}^{N\times N}$ and $P \geq Q$. Then $\mathcal{L}^k(P) \geq \mathcal{L}^k(Q)$ for all $k \in \mathbb{Z}_{\geq 0}$.*

Proof. From the definition of $\mathcal{L}$, we have

$$\begin{aligned} x^T\left[\mathcal{L}(P) - \mathcal{L}(Q)\right]x &= x^T\left[\mathbb{E}\left[F^TPF\right] - \mathbb{E}\left[F^TQF\right]\right]x \\ &= \mathbb{E}\left[x^TF^T(P-Q)Fx\right] \\ &\geq 0. \end{aligned}$$

By iterating these steps k times we then obtain $\mathcal{L}^k(M) \geq \mathcal{L}^k(N)$. □

Lemma 8. *The following holds for all Δ:*

$$\Omega\mathcal{L}^t(\Omega\Delta\Omega)\Omega = \Omega\mathcal{L}^t(\Delta)\Omega.$$

Proof. Proof is by induction. The statement is true for $t = 0$, as $\Omega^2 = \Omega$. Suppose it is true up to t. We then have

$$\begin{aligned} \Omega\mathcal{L}^{t+1}(\Delta)\Omega &= \Omega\mathcal{L}(\mathcal{L}^t(\Delta))\Omega \\ &= \Omega\mathcal{L}(\Omega\mathcal{L}^t(\Delta)\Omega)\Omega \\ &= \Omega\mathcal{L}(\Omega\mathcal{L}^t(\Omega\Delta\Omega)\Omega)\Omega \\ &= \Omega\mathcal{L}^{t+1}(\Omega\Delta\Omega)\Omega. \end{aligned}$$

□

Lemma 9. *Let $\bar{F} = \mathbb{E}[F]$. If $\operatorname{Im}\left[\Omega_1 \cdots \Omega_\ell\right] = \ker \mathbf{1}^T$, then all the eigenvalues of $\bar{F}$ have absolute value not larger than 1, and its only eigenvalue on the unitary circle is $\lambda = 1$, with associated left eigenvector $\mathbf{1}$ and right eigenvector $M^{-1}\mathbf{1}$.*

Proof. The fact that all eigenvalues lie inside or on the unit circle follows from the fact that $\bar{F}$ is the convex combination of matrices F_i that satisfies $\|F_i\|_M \leq 1$ for all i's. Consider then an eigenvector x such that $\|x\| = \|\bar{F}x\|$. We have

$$\|\bar{F}x\| \leq \mathbb{E}\left[\|F_ix\|\right] \leq \|x\|,$$

and therefore $\|F_ix\| = \|x\|$ for all i's. As F_i are projection matrices, it means that $F_ix = x$ and then $Mx \in \ker \Omega_i^T, \forall i$. Similarly to what we have done in the proof of Proposition 6, using the fact that $\operatorname{Im}\left[\Omega_1 \cdots \Omega_\ell\right] = \ker \mathbf{1}^T$, we necessarily have $x = M^{-1}\mathbf{1}$. By inspection we can verify that the left eigenvector corresponding to the same eigenvalue is $\mathbf{1}$. □

We can then state the following result.

Theorem 10. *Consider the linear system* (10) *and the rate of convergence R defined in* (11)*. Define*

$$\beta = \max\{|\lambda| \mid \lambda \in \lambda(\bar{F}), \lambda \neq 1\}$$

where $\bar{F} = \mathbb{E}[F]$. Then $R \leq \beta$.

Proof. Let us first prove that $\Omega\mathcal{L}(\Omega M\Omega)\Omega \leq \beta\Omega M\Omega$. Indeed, we have

$$\begin{aligned} x^T\Omega\mathcal{L}(\Omega M\Omega)\Omega x &= \mathbb{E}\left[x^T\Omega F^T\Omega M\Omega F\Omega x\right] \\ &= \mathbb{E}\left[x^T\Omega F^TMF\Omega x\right] \\ &= x^T\Omega M^{1/2}\mathbb{E}\left[M^{1/2}FM^{-1/2}\right]M^{1/2}\Omega x, \end{aligned}$$

where we use the fact that $\Omega F\Omega = F\Omega$ and that $F_i^TMF_i = MF_i$.

$\mathbb{E}\left[M^{1/2}FM^{-1/2}\right] = M^{1/2}\bar{F}M^{-1/2}$ is symmetric and, by Lemma 9, it has only one eigenvalue on the unit circle (precisely in 1), with eigenvector $M^{-1/2}\mathbf{1}$. As $M^{1/2}\Omega x \perp M^{-1/2}\mathbf{1}$ for all x, we have

$$x^T\Omega\mathcal{L}(\Omega M\Omega)\Omega x \leq \beta\Omega M\Omega,$$

with $\beta = \max\{|\lambda| \mid \lambda \in \lambda(\bar{F}), \lambda \neq 1\}$.

From this result, using Lemmas 7 and 8, we can conclude

$$\begin{aligned}\Omega\mathcal{L}^t(\Omega M\Omega)\Omega &= \Omega\mathcal{L}^{t-1}\left(\mathcal{L}(\Omega M\Omega)\right)\Omega \\ &= \Omega\mathcal{L}^{t-1}\left(\Omega\mathcal{L}(\Omega M\Omega)\Omega\right)\Omega \\ &\leq \Omega\mathcal{L}^{t-1}\left(\beta\Omega M\Omega\right)\Omega \\ &= \beta\Omega\mathcal{L}^{t-1}\left(\Omega M\Omega\right)\Omega \\ &\leq \cdots \leq \beta^t\Omega M\Omega,\end{aligned}$$

and therefore $R \leq \beta$. □

Before analyzing the tightness of the bound for some specific cases, we state a result that allows us to compute R when the spectra of $\mathcal{L}$ and $\bar{F}$ are available (analytically or numerically).

Let define $\mathcal{O}$ as the non observable space for the system (10):

$$\mathcal{O} = \left\{\Delta \in \mathbb{R}^{N\times N} \mid \Omega\Delta\Omega = 0\right\}.$$

We can then introduce the rate

$$R_{\mathcal{O}} = \max\left\{|\lambda_{\mathcal{L}}(i) \mid \Delta_{\mathcal{L}}(i) \notin \mathcal{O}\right\}. \tag{12}$$

The following proposition holds.

Proposition 11. *Let R and $R_{\mathcal{O}}$ be defined by* (11) *and* (12) *respectively. Then*

$$R = R_{\mathcal{O}}.$$

Proof. For any eigenvector $\Delta_{\mathcal{L}}(i)$, there exists $\gamma > 0$ such that $\Delta_{\mathcal{L}}(i) \leq \gamma M$. We then have $\Omega\mathcal{L}^t(\Delta_{\mathcal{L}}(i))\Omega \leq \gamma\Omega\mathcal{L}^t(M)\Omega$, and therefore $\lambda_{\mathcal{L}}(i)^t\Omega\Delta_{\mathcal{L}}(i)\Omega \leq \gamma\Omega\mathcal{L}^t(M)\Omega$. If $\Delta_{\mathcal{L}}(i) \notin \mathcal{O}$, then we must have $\lambda_{\mathcal{L}}(i) \leq R$, therefore $R_{\mathcal{O}} \leq R$. As of course $R_{\mathcal{O}} \geq R$, we conclude that $R = R_{\mathcal{O}}$. □

Remembering that $\lambda_{\mathcal{L}} \in \mathbb{R}^{N^2}$ and $\lambda_{\bar{F}} \in \mathbb{R}^N$ are the ordered vector of *possibly repeated* eigenvalues of $\mathcal{L}$ and $\bar{F}$, we can state the following result, illustrated also in Figure 3.

Theorem 12. *The elements of the vector*

$$\lambda'_{\bar{F}} = [\lambda_{\bar{F}}(2), \ldots, \lambda_{\bar{F}}(N)]$$

appear twice in the vector

$$\lambda'_{\mathcal{L}} = [\lambda_{\mathcal{L}}(2), \ldots, \lambda_{\mathcal{L}}(N^2)],$$

and so R is the largest element in absolute value of the remaining ones in $\lambda'_{\mathcal{L}}$.

Proof. Via Lemma 8 it is possible to show that $\mathcal{O}$ is an invariant set:

$$\Omega\mathcal{L}(\Delta)\Omega = \Omega\mathcal{L}(\Omega\Delta\Omega)\Omega = 0 \quad \forall\Delta \in \mathcal{O}.$$

As the dimension of $\mathcal{O}$ is $2N-1$ (the dimension of the kernel of $\Omega \otimes \Omega$), there must exist $2N-1$ eigenvectors of $\mathcal{L}$ in $\mathcal{O}$. These eigenvectors can be constructed from the eigenvectors of $\bar{F}^T$. Indeed, consider N linearly independent vectors $v_1, \ldots, v_N$ such that $\bar{F}^T v_i = \mu_i v_i$ with $1 = \mu_1 \geq \cdots \geq \mu_N$. We have for all i

$$\begin{aligned}\mathcal{L}(\mathbf{1}v_i^T) &= \mathbb{E}\left[F^T\mathbf{1}v_i^T F\right] = \mathbf{1}v_i^T\mathbb{E}\left[F\right] = \mu_i\mathbf{1}v_i^T \\ \mathcal{L}(v_i\mathbf{1}^T) &= \mathbb{E}\left[F^T v_i\mathbf{1}^T F\right] = \mathbb{E}\left[F\right]v_i\mathbf{1}^T = \mu_i v_i\mathbf{1}^T.\end{aligned}$$

For all these eigenvectors we have $\Omega v_i\mathbf{1}^T\Omega = \Omega\mathbf{1}v_i^T\Omega = 0$. We therefore constructed a basis of $2N-1$ linearly independent eigenvectors of $\mathcal{O}$. One of them, $\Delta_{\mathcal{L}}(1) = \mathbf{1}\mathbf{1}^T$, corresponds to the eigenvalue $\lambda_{\mathcal{L}}(1) = 1$. The remaining $2(N-1)$ correspond to the eigenvalues $\lambda_{\bar{F}}(2), \ldots, \lambda_{\bar{F}}(N)$, taken twice. According to Proposition 11, R is then the largest among the eigenvalues left when removing (twice) $[\lambda_{\bar{F}}(2), \ldots, \lambda_{\bar{F}}(N)]$ from $[\lambda_{\mathcal{L}}(2), \ldots, \lambda_{\mathcal{L}}(N^2)]$. □

In the following, we will analyze the rate of convergence R and the bound β for different networks and different decomposition choices. In some of them (namely for the 1-dimensional case) it is possible to compute them analytically, gaining some insight on how they scale with the number of nodes. For a more general case, the rate of convergence has been studied numerically and compared with simulations.

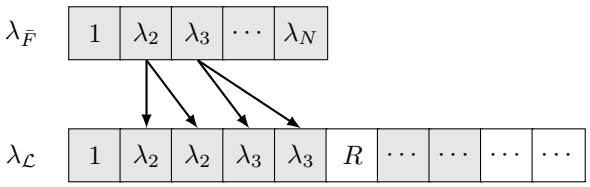

Figure 3: Representation of the eigenvalues of $\mathcal{L}$ and of $\bar{F}$, according to Theorem 12.

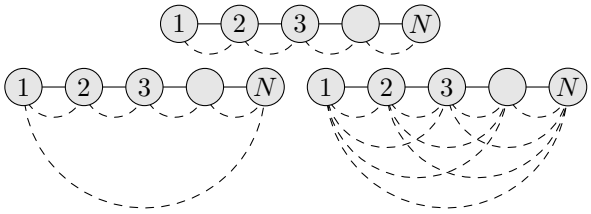

Figure 4: Three possible clustering choices (*1-step*, *circle*, *complete*) illustrated via their corresponding hypergraphs $\mathcal{H}$ (in dashed line). Edges of $\mathcal{H}$ connect nodes that are allowed to update their state together (each edge corresponds to a different subproblem). The graph in continuous line describes the physical system.

5.1 1-dimensional case

Consider the specific case of a 1-dimensional graph, i.e. an electrical network consisting in one single line with compensators equally distributed at unitary distances along the line[1]. Loads (passive agents) can be connected everywhere in this line, as their location and their demands are uninfluential on the matrix M, and therefore on the speed of convergence of the optimization algorithm.

We consider three different decompositions of the optimization problem, corresponding to different clustering of the nodes into subsets. In all of them we assume that compensators are allowed to update their state in pairs. As proposed in Section 4, this can be conveniently described by an hypergraph $\mathcal{H}$ (a graph in this case) where an edge connecting node i with node j corresponds to the optimization subproblem in which only the states q_i and q_j are updated.

We will consider the three following graphs $\mathcal{H}$ (see Figure 4), corresponding to different clustering choices:

- edges of $\mathcal{H}$ connect compensators which are adjacent in the electric line (*1-step*);
- edges of $\mathcal{H}$ connect compensators which are adjacent in the electric line and moreover the first agent is connected with the last agent (*circle*);
- edges of $\mathcal{H}$ connect any pair of compensators (*complete*).

The Hessian M for the 1-dimensional electric network takes the form

$$M = M_0 - \begin{bmatrix} 0 & 1 & 2 & \cdots & N-1 \\ 1 & 1 & 2 & & N-1 \\ 2 & 2 & 2 & & N-1 \\ \vdots & & & & \vdots \\ N-1 & N-1 & N-1 & \cdots & N-1 \end{bmatrix},$$

where $M_0 = m_0 \mathbf{1}\mathbf{1}^T$ and therefore it can be safely ignored, as $x^T M_0 x = 0$ for all $x \in \ker \mathbf{1}^T$.

For the 1-step case, the i-th element of $\{F_i\}_{i=1}^{N-1}$ corresponds to the subproblem in which node i and

[1]The hypothesis of equally spaced compensators on the line allows easier comparison between strategies and simplifies the analytical results, but it not critical. Indeed, some of the rates and bounds are independent from the distance between compensators. These cases will be pointed out later, when they appear in the text.

node $i+1$ are allowed to update their state:

$$F_i = \begin{bmatrix} 1 & & & & & & & \\ & \ddots & & & & & & \\ & & 1 & & & & & \\ -1 & \cdots & -1 & 0 & & & & \\ 1 & \cdots & \cdots & \cdots & 1 & & & \\ & & & & & \ddots & & \\ & & & & & & 1 & \end{bmatrix}.$$

It is possible to analytically compute $\bar{F}$, which results to be a lower triangular matrix with elements on the diagonal (and eigenvalues):

$$\lambda_{\bar{F}} = \left[1, \underbrace{1-\frac{1}{N-1}, \ldots, 1-\frac{1}{N-1}}_{N-1}\right]^T.$$

In this specific case it is also possible to compute the matrix $\mathbf{F}$ associated with the linear transformation $\mathcal{L}$, which results to be an upper triangular $N^2 \times N^2$ matrix whose elements on the diagonal are

$$\lambda_{\mathcal{L}} = \left[1, \underbrace{1-\frac{1}{N-1}, \ldots, 1-\frac{1}{N-1}}_{3(N-1)}, \underbrace{1-\frac{2}{N-1}, \ldots, 1-\frac{2}{N-1}}_{(N-1)(N-2)}\right]^T.$$

From the analysis of the previous section, we can then state that

$$\beta_{\text{1step}} = R_{\text{1step}} = 1 - \frac{1}{N-1}.$$

Interestingly, both β_{1step} and R_{1step} do not depend on the length of the electric paths between adjacent compensators.

Consider now the case in which the graph $\mathcal{H}$ is a circle, i.e. every node can communicate with its closest neighbors on the line and in addition an edge connects the first and the last node.

In this case the set of matrices $\{F_i\}_{i=1}^N$ includes the F_i's of the previous case (1-step), together with

$$F_N = \begin{bmatrix} 0 & \frac{1}{N-1}-1 & \frac{2}{N-1}-1 & \cdots & 0 \\ & 1 & & & \\ & & \ddots & & \\ & & & 1 & \\ 1 & 1-\frac{1}{N-1} & \cdots & 1-\frac{N-2}{N-1} & 1 \end{bmatrix}.$$

Also in this case, by exploiting the block-triangular structure of the resulting $\bar{F}$, it is possible to list its eigenvalues:

$$\lambda_{\bar{F}} = \left[1, \underbrace{1-\frac{1}{N}, \ldots, 1-\frac{1}{N}}_{N-2}, 1-\frac{2}{N}\right]^T,$$

and therefore obtain

$$\beta_{\text{circle}} = 1 - \frac{1}{N}.$$

β_{circle} too does not depend on the length of the paths between compensators.

In this case, however, $\lambda_{\mathcal{L}}$ (and therefore R) cannot be easily expressed analytically as we did for the *1step* case. We therefore computed R_{circle} numerically, together with the both the bound β_{complete} and the exact rate R_{complete} for the third clustering choice, in which every couple of nodes is allowed to communicate.

In Table 1 it is possible to compare the rate of convergence of these different clustering (or decomposition) choices for different values of N, and to realize how the bound is tight. The tightness of the

	$N = 10$		$N = 100$		$N = 500$
	$1-\beta$	$1-R$	$1-\beta$	$1-R$	$1-\beta$
1-step	0.11111	0.11111	0.010101	0.010101	20.040×10^{-4}
circle	0.10000	0.10572	0.010000	0.010051	20.000×10^{-4}
complete	0.03129	0.05196	0.000427	0.000560	0.204×10^{-4}

Table 1: Exact convergence rate R and bound β for different network lengths N and different communication topologies.

bound justifies our choice of including in the table also the a larger network ($N = 500$), for which the problem of computing the exact convergence rate R results to be numerically intractable.

It is worth noticing that the well studied problem of randomized gossip algorithms for average consensus can be casted into the framework of this paper by choosing $M = I$. These results are therefore quite interesting in the fact that they contrast with the phenomena generally observed in gossip consensus algorithms (e.g. [4]), in which long-distance communication, by decreasing the diameter of the graph, tends to be extremely beneficial for the rate of convergence.

5.2 General case

The convergence rate and its bound based on $\bar{F}$ have also been computed for a more general case. We considered a tree of height 6, with 33 nodes and an average of 2.4 children for every node that is not a leaf.

We implemented two clustering choices: in the first one, only nodes that are neighbors on the tree can communicate (*1-step*); in the second one, every pair of node is allowed to communicate (*complete*).

We obtained (numerically) the following values for the convergence rates and the bound β:

$$\beta_{\text{1step}} = 0.9688, \qquad R_{\text{1step}} = 0.9688,$$
$$\beta_{\text{complete}} = 0.9967, \qquad R_{\text{complete}} = 0.9937.$$

In the upper part of Figure 5 we plotted the signal $\bar{v}(t)$ corresponding to the cost function averaged over 100 realizations, for the two strategies. In the lower part, instead, we plotted the function $\bar{v}(t)^{1/t}$, together with the computed rates of convergence (and bounds).

Even in this case, we can see how adding long distance links (i.e. enabling communication between agents with are connected to distant points of the distribution network) seems to be detrimental for the convergence speed of the algorithm. On the contrary, it looks like the optimal strategy consists in choosing a clustering hypergraph which resembles (or is the same in the case of clusters of two nodes) the graph describing the physical interconnection of the electric network.

6 Conclusion

The randomized algorithm proposed in this work seems to be an effective way to tackle the problem of optimal reactive power flows, as it requires local knowledge of the problem structure and of the system state at the agent level, and it exploits physical features of the system to reduce the need for communication (via the gradient estimation from the voltage measurements).

The main degree of freedom in the algorithm implementation consists in the choice of the clusters of cooperating nodes. The analysis on the convergence rate that has been carried out in the paper allows comparison between different choices, producing some interesting observations that will be the subject of future investigation. In particular, it seems that clustering nodes that are close in the power network is beneficial for the speed of convergence of this algorithm, and therefore the design problem of building a communication and coordination graph among nodes seems to be tightly coupled with the structure of the physical system.

Moreover, this work can be considered a valuable starting point for the design of a dynamic optimization algorithm, to tackle the more realistic problem in which reactive power demands are time-varying,

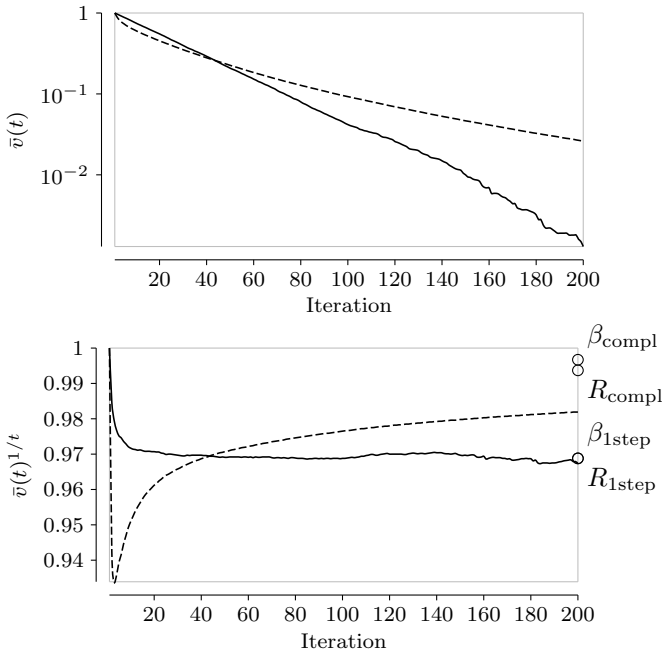

Figure 5: Algorithm behavior when applied to the general case described in Section 5.2. The continuous line refers to the *1-step* communication strategy, while the dashed line refers to the *complete* one.

compensators are subject to operating limits, and the estimation of the gradient of the cost function consists in an appropriate filtering of the voltage measurement signals.

References

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

[2] Saverio Bolognani and Sandro Zampieri. Distributed quasi-newton method and its application to the optimal reactive power flow problem. In *Proc. NECSYS'10*, Annecy, France, 2010.

[3] Francesco Bullo, Ruggero Carli, and Paolo Frasca. Gossip coverage control for robotic networks: Dynamical systems on the the space of partitions. *SIAM Journal on Control and Optimization*, 2010. Submitted.

[4] Fabio Fagnani and Sandro Zampieri. Randomized consensus algorithms over large scale networks. *IEEE J. on Selected Areas in Comm.*, 26(4):634 – 649, May 2008.

[5] Ali Ipakchi and Farrokh Albuyeh. Grid of the future. are we ready to transition to a smart grid? *IEEE Power Energy Mag.*, 7(2):52 –62, mar-apr 2009.

[6] F. P. Kelly, A. K. Maulloo, and D. K. H Tan. Rate control for communication networks: Shadow prices, proportional fairness and stability. *J. Oper. Res. Soc.*, 49:237–252, 1998.

[7] M. Rotkowitz and S. Lall. A characterization of convex problems in decentralized control. *IEEE Transactions on Automatic Control*, 51(2):274–286, February 2006.

[8] Enrique Santacana, Gary Rackliffe, Le Tang, and Xiaoming Feng. Getting smart. with a clearer vision of the intelligent grid, control emerges from chaos. *IEEE Power Energy Mag.*, 8(2):41–48, March 2010.

[9] Elisabetta Tedeschi, Paolo Tenti, and Paolo Mattavelli. Synergistic control and cooperative operation of distributed harmonic and reactive compensators. In *IEEE PESC'08*, 2008.

[10] John N. Tsitsiklis, Dimitri P. Bertsekas, and Michael Athans. Distributed asynchronous deterministic and stochastic gradient optimization algorithms. *IEEE Trans. Automat. Contr.*, 31(9):803–812, September 1986.