53rd IEEE Conference on Decision and Control
December 15-17, 2014. Los Angeles, California, USA

Sparse Solutions to the Average Consensus Problem via l_1-norm Regularization of the Fastest Mixing Markov-Chain Problem

Giorgio Gnecco, Rita Morisi and Alberto Bemporad

Abstract— **In the "consensus problem" on multi-agent systems, in which the states of the agents are "opinions", the agents aim at reaching a common opinion (or "consensus state") through local exchange of information. An important design problem is to choose the degree of interconnection of the subsystems so as to achieve a good trade-off between a small number of interconnections and a fast convergence to the consensus state, which is the average of the initial opinions under mild conditions. This paper addresses this problem through l_1-norm regularized versions of the well-known fastest mixing Markov-chain problem, which are investigated theoretically. In particular, it is shown that such versions can be interpreted as "robust" forms of the fastest mixing Markov-chain problem. Theoretical results useful to guide the choice of the regularization parameters are also provided, together with a numerical example.**

I. Introduction

Many dynamical systems (e.g., wireless sensor networks, robotic teams, social networks) can be decomposed into a large number of subsystems (or "agents"), whose interactions are local and can be modeled by weighted edges in a "communication" graph, in which the vertices are the subsystems. Control problems on such multi-agent systems enjoy properties related to the structure of the communication graph, described, e.g., by weighted/unweighted adjacency and graph-Laplacian matrices [1], [2]. A paradigmatic example of such control problems is the "consensus problem" [3], in which the states of the subsystems are "opinions", and the agents aim at reaching a common opinion (or "consensus state") through local exchange of information, without any form of centralization. A typical example is distributed estimation in wireless sensor networks [4]. Under mild conditions, one can prove that the consensus state is the average of the initial opinions, and the problem is called the "average consensus problem" [3]. In both problems, the variables to be chosen are the weights to be assigned to the edges of the communication graph. Such weights define a weighted adjacency matrix and, in the case of undirected communication graphs, also a weighted Laplacian matrix, whose spectral properties (i.e., properties expressed in terms of the eigenvalues/eigenvectors of such matrices) determine the rate of convergence to the consensus state [3], [5]. Interestingly, in the undirected case, determining the weights that optimize such spectral properties can be formulated as a convex optimization problem [5] (specifically, as a semi-definite program (SDP)), which is known as the Fastest Mixing Markov-Chain (FMMC) problem and can be solved efficiently, e.g., using interior-point methods. In several cases, one may be interested to find a suitable compromise between the desired spectral properties of the graph and the number of non-zero weights of the edges, thus obtaining sparse solutions to the consensus/average consensus problems with satisfactory rate of convergence. This is motivated, e.g., in the case of a high cost of communication associated with each edge. In this context, some recent results towards such sparse solutions are provided in [6], that investigates sparse graphs with certain symmetries, for which one has closed-form expressions for the Laplacian eigenvalues. However, the results are restricted to specific kinds of graphs. Another recent result in this direction is provided by [7], which shows that every Laplacian matrix associated with a symmetric graph can be "well-approximated" by the Laplacian matrix of a sparse subgraph, thus keeping the desired spectral properties but with a small number of edges. In principle, such a property could be exploited to find a good sparse solution to the average consensus problem, starting from a dense graph with good spectral properties, then sparsifying it, trying to preserve such spectral properties.

The present work focuses on the average consensus problem and, in the case of undirected graphs, proposes a different trade-off between good spectral properties of the communication graph and its sparsity, using an approach based on an l_1-norm regularized version of the FMMC problem. This is motivated by the fact that, due to geometrical properties of the l_1 norm [8], the introduction of such a regularization term in the objective of a convex optimization problem often enforces the sparsity of an optimal solution of the regularized version of that problem. We also consider another variation of FMMC in which, besides the introduction of the l_1-norm regularization term, the weights of some edges are fixed. Then, we provide both a theoretical analysis of the two variations of FMMC and a numerical example modeling a wireless sensor network, comparing their solutions with the one obtained by solving FMMC. Related approaches were proposed in [9, Section 7.2] and in [10]; the novel contributions of the present work are highlighted in Section VI.

The paper is organized as follows. Section II summarizes the FMMC problem and introduces its equivalent formulation. Then, Section III presents two modifications of such a formulation (both obtained adding an l_1-norm regularization term to enforce sparsity, and fixing also some weights in the second one), which are investigated in Section IV from a theoretical point of view. Section V shows their application

Giogio Gnecco, Rita Morisi and Alberto Bemporad are with the IMT - Institute for Advanced Studies, Piazza S. Ponziano 6, 55100 Lucca, Italy (e-mails: giorgio.gnecco, rita.morisi, alberto.bemporad@imtlucca.it).

978-1-4673-6090-6/14/$31.00 ©2014 IEEE

to the design of a wireless sensor network. Finally, Section VI discusses possible extensions of the work.

II. The fastest mixing Markov-chain problem

The consensus problem consists in determining the strengths of the interconnections among the subsystems of a multi-agent system, so that their states converge to a common state, subject to given constraints on the admissible connections. In the simplest case, the subsystems are linear, their states $x_i \in \mathbb{R}$ are scalar-valued, and the evolution of each subsystem i is determined by the discrete-time dynamics

$$x_i(t+1) = \sum_{i=1}^{n} P_{ij} x_j(t)\,, t = 0, 1, \ldots\,, \tag{1}$$

where $P \in \mathbb{R}^{n \times n}$ is a matrix of interconnections with non-negative entries, satisfying the conditions $P\underline{1}_n = \underline{1}_n$ (here, $\underline{1}_n \in \mathbb{R}^n$ denotes a column vector of dimension n whose components are all equal to 1) and

$$P_{ij} = 0, \quad \text{if } i \neq j \text{ and } (i,j) \notin \mathcal{E}\,, \tag{2}$$

where $\mathcal{E}$ is a given set of admissible interconnections. In a design phase, the elements of the matrix P can be chosen arbitrarily, provided that the conditions above on P are satisfied.

The non-negativity assumption on P, together with the condition $P\underline{1}_n = \underline{1}_n$, implies that 1 is the eigenvalue of the matrix P with maximum absolute value (this is proved, e.g., by an application of Gersghorin's theorem), and that the state $x_i(t+1)$ at time $t+1$ is a convex combination of the states $x_j(t)$ at time t. Note that the diagonal elements of P may be different from 0, so, when $P_{ii} > 0$, this means that $x_i(t+1)$ is influenced by $x_i(t)$.

It is well known (see, e.g., [3]) that, when the eigenvalue 1 has algebraic multiplicity equal to 1, and all the other eigenvalues of P have absolute value smaller than 1, the states of the subsystems converge to the same "consensus" state x_c, when $t \to \infty$:

$$x_i(t) \overset{t\to\infty}{\longrightarrow} x_c := \sum_{j=1}^{n} \alpha_j x_j(0) \text{ for all } i \in \{1, \ldots, n\}\,, \tag{3}$$

where the α_j's are suitable non-negative constants such that

$$\sum_{j=1}^{n} \alpha_j = 1\,. \tag{4}$$

In the particular case in which the matrix P is symmetric, one can show [3] that

$$\alpha_j = \frac{1}{n}\,, \forall j \in \{1, \ldots, n\}\,, \tag{5}$$

and the consensus state is simply the average of the initial states (in such case, the problem is called "average consensus problem"). In the following, we will focus on such a situation, therefore assuming $P = P^T$.

A particularly important aspect of the average consensus problem is the rate of convergence to the average consensus state, which is related to the second-largest eigenvalue modulus of P [3]:

$$\mu(P) := \max_{j=2,\ldots,n} |\lambda_j\{P\}|\,, \tag{6}$$

where the eigenvalues $\lambda_j\{P\}$, $j = 1, \ldots, n$, have been ordered with their multiplicity in a nonincreasing order (i.e., $1 = \lambda_1\{P\} \geq \lambda_2\{P\} \geq \ldots \lambda_j(P) \geq \ldots \geq \lambda_n\{P\} > -1$). In particular, the smaller $\mu(P)$, the faster the convergence to the consensus state.

In addition, a related quantity is the mixing time [5]

$$\tau(P) := \frac{1}{\log\left(\frac{1}{\mu(P)}\right)}\,, \tag{7}$$

which is an asymptotic measure of the number of steps required for reducing by the factor e (i.e., the Euler's number) a suitable distance (the "total variation distance") between the global state vector and the vector whose components are equal to the average consensus state.

Since the symmetric matrix P has non-negative elements and satisfies $P\underline{1}_n = \underline{1}_n$, its generic element P_{ij} can be interpreted as a "transition probability" from the vertex i to the vertex j of a graph (including self-loops), whose vertices are the subsystems. This interpretation is useful since also the rate of convergence of the Markov chain with transition probabilities P_{ij} to its stationary distribution depends on $\mu(P)$ (and again, the smaller the second-largest eigenvalue modulus $\mu(P)$, the faster the convergence to the stationary distribution [5]). So, the problem of determining the coefficients P_{ij} that minimize $\mu(P)$ subject to a given topology of the graph is called the "Fastest Mixing Markov-Chain" problem (Problem FMMC, in the following). In this context, the elements P_{ij} with $i \neq j$ are interpreted as weights of edges in the graph between the two different vertices i and j, whereas P_{ii} is the weight of a self-loop edge. In [5], the problem is formulated as

$$\textbf{Problem FMMC (first formulation)}:$$

$$\begin{array}{ll} \text{minimize} & \mu(P) \\ \text{subject to} & P\underline{1}_n = \underline{1}_n,\ P = P^T\,, \\ & P_{ij} \geq 0,\ \forall i,j \in \{1, \ldots, n\}\,, \\ & P_{ij} = 0,\ \text{if } (i,j) \notin \mathcal{E}\,. \end{array} \tag{8}$$

Interestingly, this is a convex optimization problem, since

$$\mu(P) = \lambda_{\max}\left\{P - \frac{1}{n}\underline{1}_n\underline{1}_n^T\right\} \tag{9}$$

(see [5] for a proof). Moreover, it can also be written as a semi-definite program [5, Section 2.3].

A. An equivalent formulation of Problem FMMC

Now, we introduce an equivalent version of Problem FMMC, using a notation suitable for its sparse extensions presented in Section III and for their theoretical investigation in Section IV.

In the following, we denote by $\underline{w} \in \mathbb{R}^m$ the column vector of weights associated with the m edges joining different vertices, and by $\underline{w}_{\text{sl}} \in \mathbb{R}^n$ the column vector of weights associated with the n self-loop edges. Hence, we can represent the weighted adjacency matrix P as a linear function $P(\underline{w}, \underline{w}_{\text{sl}})$ of such weights. For instance, for $n = 3$ and $m = n(n-1)/2$ (the case of a complete graph), one obtains the symmetric matrix

$$P(\underline{w}, \underline{w}_{\text{sl}}) = \begin{bmatrix} w_{\text{sl},1} & w_1 & w_2 \\ w_1 & w_{\text{sl},2} & w_3 \\ w_2 & w_3 & w_{\text{sl},3} \end{bmatrix} . \quad (10)$$

Moreover, setting

$$\underline{w}_{\text{sl}} := \underline{1}_n - M\underline{w} , \quad (11)$$

and introducing the vertex-edge incidence matrix $M \in \mathbb{R}^{n\times m}$, whose elements are defined as follows:

$$M_{ij} = \begin{cases} 1 , & \text{if the vertex } i \text{ is an endpoint of} \\ & \text{the (non-self-loop) edge } j , \\ 0 , & \text{otherwise} , \end{cases} \quad (12)$$

the constraints

$$P_{ij} \geq 0 \text{ for any } i, j \in \{1, \ldots, n\} \text{ and } P\underline{1}_n = \underline{1}_n \quad (13)$$

are equivalent to

$$w_i \geq 0 \text{ for any } i \in \{1, \ldots, m\} \text{ and } M\underline{w} \leq \underline{1}_n . \quad (14)$$

Using (11), the matrix P becomes an affine function $P(\underline{w})$ of the weight vector $\underline{w}$, and the second-largest eigenvalue modulus of P becomes a convex function - denoted by $\mu(\underline{w})$ - of the weight vector $\underline{w}$, since convexity is preserved by affine mappings [11, Section 3.2]. With the notations just introduced, Problem (8) can be compactly rewritten as

$$\textbf{Problem FMMC (second formulation)} :$$

$$\begin{array}{ll} \text{minimize}_{\underline{w}\in\mathbb{R}^m} & \mu(\underline{w}) \\ \text{subject to} & \underline{w} \geq \underline{0}_m, \\ & M\underline{w} \leqslant \underline{1}_n . \end{array} \quad (15)$$

III. Two sparse variations of the fastest mixing Markov-chain problem

We now consider the following sparse variations of Problem FMMC.

A. Problem FMMC with an l_1-norm regularization term

In order to find a good compromise between sparsity of $\underline{w}$ and a small value of the second-largest eigenvalue modulus of the weighted adjacency matrix $P(\underline{w})$, we consider, for any $\eta > 0$, the following variation of Problem FMMC, in which an l_1-regularization term with regularization parameter η is added to the objective (here, $\|\underline{w}\|_1 := \sum_{i=1}^m |w_i|$):

$$\textbf{Problem FMMC-}\boldsymbol{l_1(\eta)}:$$

$$\begin{array}{ll} \text{minimize}_{\underline{w}\in\mathbb{R}^m} & (\mu(\underline{w}) + \eta\|\underline{w}\|_1) \\ \text{subject to} & \underline{w} \geq \underline{0}_m, \\ & M\underline{w} \leqslant \underline{1}_n . \end{array} \quad (16)$$

The main motivation for the inclusion of $\eta\|\underline{w}\|_1$ in Problem FMMC-$l_1(\eta)$ is that, due to geometrical properties of the l_1 norm, adding such a term to a convex optimization problem often induces sparsity of a resulting optimal solution $\underline{w}^\circ(\eta)$ [8], that is many components of $\underline{w}^\circ(\eta)$ will be usually equal to 0.

Remark 1: In general, the l_0 "pseudo-norm"

$$\|\underline{w}\|_0 := \text{number of non-zero components of } \underline{w} , \quad (17)$$

would be a more natural way to enforce sparsity than the l_1 norm. However, it is a nonconvex function, which would make the resulting optimization problem difficult to solve. Nevertheless, as pointed out in several references (e.g., [12]), in optimization problems where sparsity is desired, a common approach is to relax the l_0 pseudo-norm by the l_1 norm, which is a convex function, thus making the optimization problem a tractable one.

B. Problem FMMC with fixed edges and an l_1-norm regularization term

An interesting variation of Problem FMMC-$l_1(\eta)$ consists in fixing some components of the weight vector $\underline{w}$. This is motivated, e.g., when one is interested in imposing some additional structure on the topology of the graph resulting from the optimization of the weight vector (e.g., enforcing the presence of given subgraphs, such as trees connecting important "backbone" vertices). Without loss of generality, in the following we assume (up to a permutation of the indices) that the fixed weights are the first m_{fixed} ones (where $1 \leq m_{\text{fixed}} < m$), whereas the last $m_{\text{free}} : m - m_{\text{fixed}}$ weights are not fixed. We then decompose the column vector $\underline{w}$ as

$$\underline{w} = \text{col}(\underline{w}_{\text{fixed}}, \underline{w}_{\text{free}}) \quad (18)$$

and the vertex-edge incidence matrix M as

$$M = [M_{\text{fixed}}|M_{\text{free}}] , \quad (19)$$

and we express μ as a function $\mu(\underline{w}_{\text{free}})$ of the unfixed weights only. Then, for a given choice of the weight vector $\underline{w}_{\text{fixed}}$, we consider the following optimization problem:

$$\textbf{Problem FMMC}_{\textbf{constr}}\textbf{-}\boldsymbol{l_1(\eta)}:$$

$$\begin{array}{ll} \text{minimize}_{\underline{w}_{\text{free}}\in\mathbb{R}^{m_{\text{free}}}} & (\mu(\underline{w}_{\text{free}}) + \eta\|\underline{w}_{\text{free}}\|_1) \\ \text{subject to} & \underline{w}_{\text{free}} \geq \underline{0}_{m_{\text{free}}}, \\ & M_{\text{free}}\underline{w}_{\text{free}} \leqslant \underline{1}_n - M_{\text{fixed}}\underline{w}_{\text{fixed}} . \end{array} \quad (20)$$

Problem FMMC$_{\text{constr}}$-$l_1(\eta)$ has a form that is similar to the one of Problem FMMC-$l_1(\eta)$. Of course, we always assume in the following that the fixed weights have been chosen in such a way that the feasible set

$$\{\underline{w} \in \mathbb{R}^{m_{\text{free}}} : \underline{w}_{\text{free}} \geq \underline{0}_{\text{free}}, M_{\text{free}}\underline{w}_{\text{free}} \leqslant \underline{1}_n - M_{\text{fixed}}\underline{w}_{\text{fixed}}\} \quad (21)$$

of Problem FMMC$_{\text{constr}}$-$l_1(\eta)$ is nonempty.

In the next section we provide some theoretical results about the optimal solutions of Problems FMMC-$l_1(\eta)$ and FMMC$_{\text{constr}}$-$l_1(\eta)$.

IV. Theoretical results

We first consider the case of Problem FMMC-$l_1(\eta)$; extensions of the results to Problem FMMC$_{\text{constr}}$-$l_1(\eta)$ are considered later in this section.

a) Effect of the regularization parameter.

Solving Problem FMMC-$l_1(\eta)$ involves finding a good compromise between the minimization of the term $\mu(\underline{w})$ and the one of $\|\underline{w}\|_1$. Next Proposition 1 shows that the regularization parameter η has opposite effects on the two terms $\mu(\underline{w})$ and $\|\underline{w}\|_1$, when evaluated at an optimal solution, whose existence can be proved by an application of Weierstrass theorem, for every value of $\eta > 0$.

Proposition 1: Let $0 < \eta_1 < \eta_2$, and $\underline{w}^\circ(\eta_1), \underline{w}^\circ(\eta_2)$ be optimal solutions to Problem FMMC-$l_1(\eta_1)$ and Problem FMMC-$l_1(\eta_2)$, respectively. Then,
i) $\mu(\underline{w}^\circ(\eta_1)) \leq \mu(\underline{w}^\circ(\eta_2))\,,$
ii) $\|\underline{w}^\circ(\eta_1)\|_1 \geq \|\underline{w}^\circ(\eta_2)\|_2\,.$

b) Conditions under which $\underline{w} = \underline{0}_m$ is an optimal solution to Problem FMMC-$l_1(\eta)$.

Under some conditions on η, $\underline{w} = \underline{0}_m$ is an optimal (trivial) solution to Problem FMMC-$l_1(\eta)$.

Proposition 2: Let $\eta \geq 2$. Then $\underline{w} = \underline{0}_m$ is an optimal solution to Problem FMMC-$l_1(\eta)$. If $\eta > 2$, then $\underline{w} = \underline{0}$ is its unique optimal solution.

The following example demonstrates that the bound shown in Proposition 2 is tight, at least if one does not impose further restrictions on the class of graphs to be considered.

Example 1: Let $n = 2$ and $m = 1$. Then, the matrix $P(\underline{w})$ has the expression

$$P(\underline{w}) = \begin{bmatrix} 1 - w_1 & w_1 \\ w_1 & 1 - w_1 \end{bmatrix}\,, \tag{22}$$

whose eigenvalues are 1 and $1 - 2w_1$. Hence, on the set $[0, 1]$ of admissible solutions to Problem FMMC-$l_1(\eta)$, its objective is

$$1 - 2w_1 + \eta w_1\,, \tag{23}$$

and $w_1 = 0$ is, respectively, the unique optimal solution to Problem FMMC-$l_1(\eta)$ for $\eta > 2$, one of its (infinite) optimal solutions for $\eta = 2$, and a suboptimal solution for $0 < \eta < 2$.

c) Choice of the regularization parameter and reoptimization.

The above theoretical results justify the following practical rule for choosing the regularization parameter η:

- given a positive integer N and a maximal acceptable increase ε for the second-largest eigenvalue modulus of P with respect to its optimal value μ°_{FMMC} in Problem FMMC, solve Problem FMMC-$l_1(\eta)$ in correspondence of N values $\eta^{(j)}$ for η such that
  - $0 < \eta^{(1)} < \eta^{(2)} < \ldots < \eta^{(N)} < 2$ (to avoid the trivial optimal solution $\underline{w}^\circ = \underline{0}_m$), and
  - $\mu(\underline{w}^\circ(\eta^{(j)})) \leq \mu^\circ_{\text{FMMC}} + \varepsilon$ $(j = 1, \ldots, N)$;
- choose $j^\circ \in \{1, \ldots, N\}$ that maximizes the sparsity

$$\begin{aligned} s(\underline{w}^\circ(\eta^{(j)})) &:= 1 - \|\underline{w}^\circ(\eta^{(j)})\|_0/m \\ &= \text{ fraction of zero elements of } \underline{w}^\circ(\eta^{(j)})\,; \end{aligned}$$

- perform a "reoptimization step" solving Problem FMMC on the graph obtained deleting all the edges i for which $w^\circ_i(\eta^{(j^\circ)}) = 0$, obtaining another weight vector $\underline{w}^\circ_{\text{reopt}}$. Of course, $\mu(\underline{w}^\circ_{\text{reopt}}) \leq \mu(\underline{w}^\circ(\eta^{(j^\circ)}))$ (due to the optimality of $\underline{w}^\circ_{\text{reopt}}$ on Problem FMMC on the new graph, and the feasibility of $\underline{w}^\circ(\eta^{(j^\circ)})$ for such a problem), and $s(\underline{w}^\circ_{\text{reopt}}) \geq s(\underline{w}^\circ(\eta^{(j^\circ)}))$ by construction.

Finally, a possible way to choose the tolerance parameter ε (which has to be in any case smaller than $1 - \mu^\circ_{\text{FMMC}}$, again to avoid trivial optimal solutions) consists in expressing it in terms of the maximal allowable ratio ρ between the mixing time $\tau(P)$ and its optimal value $\tau^\circ_{FMMC} := \frac{1}{\log\left(\frac{1}{\mu^\circ_{\text{FMMC}}}\right)}$, which is obtained when solving Problem FMMC, i.e., one sets

$$\varepsilon = (\mu^\circ_{\text{FMMC}})^{\frac{1}{\rho}} - \mu^\circ_{\text{FMMC}}\,. \tag{24}$$

d) Interpretation of Problem FMMC-$l_1(\eta)$ as a robust Problem FMMC.

Problem FMMC-$l_1(\eta)$ has also the following interpretation. Let us suppose that, for any given "nominal" choice of the weights w_i $(i = 1, \ldots, m)$, one has an "uncertainty" Δw_i such that $|\Delta w_i| \leq \delta |w_i|$, for some fixed $\delta > 0$. Then, an application of Gersghorin's theorem and Weyl's inequalities in matrix-perturbation theory shows that the second-largest eigenvalue modulus $\mu(\underline{w} + \Delta\underline{w})$ is bounded from above as

$$\mu(\underline{w} + \Delta\underline{w}) \leq \mu(\underline{w}) + 2\delta\|\underline{w}\|_1\,. \tag{25}$$

Then, an optimal "robust" choice of the nominal weight vector $\underline{w}$ is obtained minimizing the objective $\mu(\underline{w}) + 2\delta\|\underline{w}\|_1$ on the set of admissible weight vectors $\underline{w}$, i.e., solving a robust version of Problem FMMC which takes into account the uncertainty of the weights, replacing the objective $\mu(\underline{w})$ with $\mu(\underline{w}) + 2\delta\|\underline{w}\|_1$. However, this is equivalent to solving Problem FMMC-$l_1((\eta)$ with the choice $\eta = 2\delta$.

e) Extension to Problem FMMC$_{\text{constr}}$-$l_1(\eta)$.

Apart from the tightness of the bound on the minimal value of the regularization parameter η for which $\underline{w}_{\text{free}} = \underline{0}_{m_{\text{free}}}$ is an optimal solution, all the results above can be extended to Problem FMMC$_{\text{constr}}$-$l_1(\eta)$. In particular, Propositions 1 and 2 can be extended to Problem FMMC$_{\text{constr}}$-$l_1(\eta)$, with $\underline{w}$ replaced by $\underline{w}_{\text{free}}$.

f) Formulation through semi-definite programming (SDP).

Likewise Problem FMMC, Problems FMMC-$l_1(\eta)$ and FMMC$_{\text{constr}}$-$l_1(\eta)$ can be formulated as semi-definite programs, allowing the use of interior-point methods for finding their optimal solutions. More precisely, if one expresses the edges $e_1, \ldots, e_m$ in terms of their endpoints as (i, j), and considers the set

$$\begin{aligned} \mathcal{E} := &\{(i,j) : i \neq j, i, j \in \{1, \ldots, n\}, \text{ and} \\ &\exists k \in \{1, \ldots, n\} \text{ such that } M_{ik} = M_{jk} = 1\}\,, \end{aligned} \tag{26}$$

one obtains the following alternative formulation of Problem FMMC-$l_1(\eta)$.

Problem FMMC-$l_1(\eta)$ (SDP formulation) :

$$\begin{aligned} &\text{minimize}_{s\in\mathbb{R},P\in\mathbb{R}^{n\times n}} && \left(s+\frac{\eta}{2}\sum_{i\neq j,i,j=1}^{n} P_{ij}\right) \\ &\text{subject to} && -sI \preceq P-\frac{1}{n}\underline{1}_n\underline{1}_n^T \preceq sI, \\ & && P\underline{1}_n=\underline{1}_n,\ P=P^T, \\ & && P_{ij}\geq 0,\ \forall i,j\in\{1,\ldots,n\}, \\ & && P_{ij}=0,\ \text{if } (i,j)\notin\mathcal{E}. \end{aligned} \tag{27}$$

Moreover, by introducing the subset $\mathcal{E}_{\text{fixed}} \subseteq \mathcal{E}$ of edges (i,j) associated with fixed weights $P_{ij,\text{fixed}} = P_{ji,\text{fixed}}$, one obtains the following alternative formulation of Problem FMMC$_{\text{constr}}$-$l_1(\eta)$.

Problem FMMC$_{\text{constr}}$-$l_1(\eta)$ (SDP formulation) :

$$\begin{aligned} &\text{minimize}_{s\in\mathbb{R},P\in\mathbb{R}^{n\times n}} && \left(s+\frac{\eta}{2}\sum_{i\neq j,i,j=1}^{n} P_{ij}\right) \\ &\text{subject to} && -sI \preceq P-\frac{1}{n}\underline{1}_n\underline{1}_n^T \preceq sI, \\ & && P\underline{1}_n=\underline{1}_n,\ P=P^T, \\ & && P_{ij}\geq 0,\ \forall i,j\in\{1,\ldots,n\}, \\ & && P_{ij}=0,\ \text{if } (i,j)\notin\mathcal{E}, \\ & && P_{ij}=P_{ij,\text{fixed}},\ \text{if } (i,j)\in\mathcal{E}_{\text{fixed}}. \end{aligned} \tag{28}$$

Of course, the fixed weights P_{ij} can be removed from the objective of the optimization problem above, without changing its optimal solution.

We solve both Problems FMMC-$l_1(\eta)$ and FMMC$_{\text{constr}}$-$l_1(\eta)$ through a modified version of the MATLAB function `fmmc.m` in the CVX package (`http://cvxr.com/cvx/download/`), which solves the SDP formulation of Problem FMMC presented in [5] and [9].

V. NUMERICAL RESULTS

In this section, we provide some numerical results about the optimal solutions of Problems FMMC-$l_1(\eta)$ and FMMC$_{\text{constr}}$-$l_1(\eta)$, and a comparison with the one of Problem FMMC. As a test example, we consider a vertex-edge incidence matrix M corresponding to a model of a wireless sensor network with 50 vertices and 200 edges, generated in a similar way as the one in [9, Section 5.1].

The first two plots in Figure 1, which refers to the behavior of an optimal solution $\underline{w}^\circ(\eta)$ with respect to η, confirm the statement of Proposition 1 about the opposite monotonic dependence on η of $\mu(\underline{w}^\circ(\eta))$ and $\|\underline{w}^\circ(\eta))\|_1$. The bottom plot shows its sparsity $s(\underline{w}^\circ(\eta))$ as a function of η, which in this particular case is not a monotonic function of η. However, the plots also show that $\underline{w}^\circ(\eta)$ is more sparse than the optimal solution of Problem FMMC (for which one has $s(\underline{w}^\circ_{FMMC}) = 0.41$), for all the considered values of η. So, they highlight the possibility of choosing a value of the parameter η for which the second-largest eigenvalue modulus $\mu(\underline{w}^\circ(\eta))$ is not much larger than its minimum possible value μ°_{FMMC}, and that, at the same time, provides a satisfactory sparsity of $\underline{w}^\circ(\eta)$.

In order to find such a parameter, following the procedure illustrated in Section IV c), we chose $\rho = 1.5$, associated with the tolerance $\varepsilon = 0.027$, as $\mu^\circ_{\text{FMMC}} = 0.9165$ (see formula (24)). We also considered $N = 20$ values $\eta^{(1)}, \ldots, \eta^{(N)}$ for the regularization parameter η (uniformly spaced in the interval $[2\cdot 10^{-5}, 5\cdot 10^{-3}]$, see Figure 1), obtaining $j^\circ = 5$ and $\eta^{(j^\circ)} = 1.1\cdot 10^{-3}$ as the optimal regularization parameter. For this value, we obtained $\mu(\underline{w}^\circ(\eta^{(j^\circ)})) = 0.9186$, $\|\underline{w}^\circ(\eta^{(j^\circ)})\|_1 = 17.45$, and $s(\underline{w}^\circ(\eta^{(j^\circ)})) = 0.545$. Compared with the optimal solution $\underline{w}^\circ_{FMMC}$ of Problem FMMC (for which $\mu(\underline{w}^\circ_{FMMC}) = 0.9165$, $\|\underline{w}^\circ_{FMMC}\|_1 = 23.71$, and $s(\underline{w}^\circ_{FMMC}) = 0.41$), the increase of the second-largest eigenvalue modulus, the decrease of the l_1 norm of the weight vector, and the increase of its sparsity were, respectively, about $0.2\%, 26\%$, and 25%. In terms of the mixing time (7), we obtained an increase of about 3% with respect to the value associated with $\underline{w}^\circ_{FMMC}$.

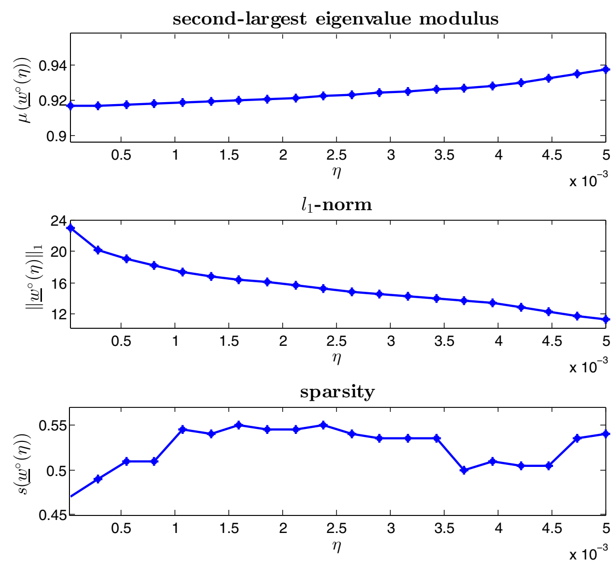

Fig. 1. Dependence on η of the second-largest eigenvalue modulus of the weighted adjacency matrix $P(\underline{w})$ evaluated at an optimal solution $\underline{w}^\circ(\eta)$ of Problem FMMC-$l_1(\eta)$, the l_1-norm of the weight vector $\underline{w}^\circ(\eta)$, and its sparsity.

Figure 2 shows: the original graph associated with the given vertex-edge incidence matrix M (top-left panel); its subgraph obtained by keeping only the edges associated with non-zero weights in the optimal solution $\underline{w}^\circ_{FMMC}$ to Problem FMMC (top-right panel); the one obtained by keeping only the edges associated with the non-zero weights of $\underline{w}^\circ(\eta^{(j^\circ)})$ (middle-left panel); a comparison of the two subgraphs (middle-right panel), obtained merging such subgraphs and highlighting in blue the non-zero-weighted edges appearing in both graphs and in green (resp., red) the non-zero weighted edges of the optimal solution to Problem FMMC (resp., Problem FMMC-$l_1(\eta^{(j^\circ)})$) that are associated with zero weights in the optimal solution to Problem FMMC-$l_1(\eta^{(j^\circ)})$ (resp., Problem FMMC). In particular, starting from the original 200 edges joining different vertices, the optimal solution of Problem FMMC keeps 118 edges, while the optimal solution to Problem FMMC-$l_1(\eta^{(j^\circ)})$ keeps only 91 edges. The percentage of edge reduction when moving from

$\underline{w}^\circ_{FMMC}$ to $\underline{w}^\circ(\eta^{(j^\circ)})$ is therefore about 23%.

As described in Section IV c), after finding the parameter $\eta^{(j^\circ)}$, an additional improvement may be obtained performing a "reoptimization step", solving Problem FMMC on the sparser subgraph obtained deleting the edges associated with zero weights in the optimal solution $\underline{w}^\circ(\eta^{(j^\circ)})$ to Problem FMMC-$l_1(\eta^{(j^\circ)})$. This step is illustrated in the bottom-left and bottom-right panels of Fig. 2, which shows in red the edges deleted by the reoptimization step. In this way, a new weight vector $\underline{w}^\circ_{\text{reopt}}$ is obtained with $\mu(\underline{w}^\circ_{\text{reopt}}) \leq \mu(\underline{w}^\circ(\eta^{(j^\circ)}))$ and $s(\underline{w}^\circ_{\text{reopt}}) \geq s(\underline{w}^\circ(\eta^{(j^\circ)}))$. So, compared with $\underline{w}^\circ_{FMMC}$, the sparsity of the weight vector $\underline{w}^\circ_{\text{reopt}}$ either remains the same or even increases, whereas the second-largest eigenvalue modulus either remains the same or even decreases. Indeed, after the reoptimization step, we obtained $\mu(\underline{w}^\circ_{\text{reopt}}) = 0.9169$ and $s(\underline{w}^\circ_{\text{reopt}}) = 0.56$.

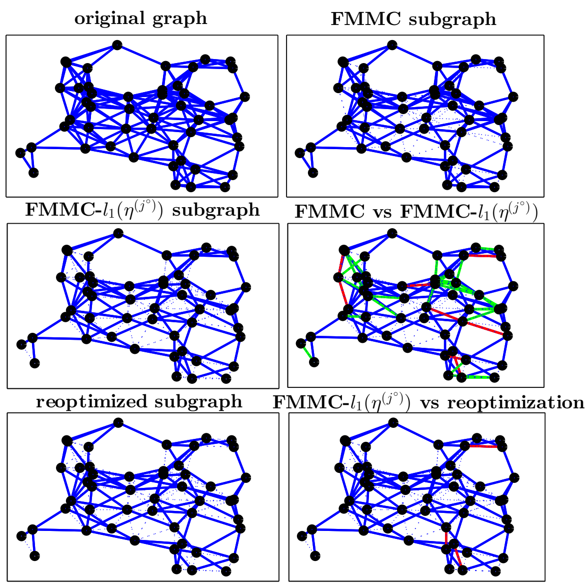

Fig. 2. A comparison of the subgraphs associated with non-zero weights in the optimal solutions to Problems FMMC and FMMC-$l_1(\eta^{(j^\circ)})$.

VI. Conclusions

We have presented some theoretical and numerical results about two sparse variations of the fastest mixing Markov-chain problem. The variations of Problem FMMC presented in Section III are similar to one already proposed and investigated numerically in [9, Section 7.2], with the only difference that the l_1-norm term in that reference appears inside an additional constraint instead than in the objective. However, up to our knowledge, their theoretical analysis presented in this paper includes novel contributions in Sections IV b), d), e). Such sections contain theoretical results that are specific to the l_1-regularized Problem FMMC and were not derived in [9]. In particular, to the best of our knowledge, the interpretation of Problem FMMC-$l_1(\eta)$ as a robust version of the fastest mixing Markov-chain problem is novel, together with the theoretical results shown in Section IV that can be proved using Gershgorin's theorem. Instead, Sections IV a) and c) provide results common also to l_1-norm regularizations of other convex optimization problems (and stated here for completeness, and for their applicability to Problems FMMC-$l_1(\eta)$ and FMMC$_{\text{constr}}$-$l_1(\eta)$), whereas Section IV f) provides semi-definite programming formulations similar to the one presented in [9, Section 7.2], which are useful for solving Problems FMMC-$l_1(\eta)$ and FMMC$_{\text{constr}}$-$l_1(\eta)$ numerically. We also mention that, for the average consensus problem in the presence of disturbances, a similar graph-sparsification optimization problem was also recently considered in [10], and solved in a distributed way through the Alternating Direction Method of Multipliers (ADMM) [13].

Among possible future developments, we mention the use of other sparsity-enforcing regularization terms (such as the group LASSO [14] and the sparse group LASSO [15]), an investigation of theoretical bounds on the degree of sub-optimality of the obtained solution with respect to the one achieved by using the l_0 pseudo-norm, and an extension of the theoretical analysis to nonlinear consensus problems.

References

[1] M. Mesbahi and M. Egerstedt, *Graph Theoretic Methods in Multiagent Networks*. Princeton University Press, 2010.

[2] Y.-Y. Liu, J.-J. Slotine, and A.-L. Barabási, "Control of complex networks," *Nature*, vol. 473, pp. 167–173, 2011.

[3] E. Lovisari and S. Zampieri, "Performance metrics in the average consensus problem: A tutorial," *Annual Reviews in Control*, vol. 36, pp. 26–41, 2012.

[4] R. Olfati-Saber, J. A. Fax, and R. M. Murray, "Consensus and cooperation in networked multi-agent systems," in *Proceedings of the IEEE*, vol. 95, no. 1, 2007, pp. 215–233.

[5] S. Boyd, P. Diaconis, and L. Xiao, "Fastest mixing Markov chain on a graph," *SIAM Review*, vol. 46, pp. 667–689, 2004.

[6] R. Carli, F. Fagnani, A. Speranzon, and S. Zampieri, "Communication constraints in the average consensus problem," *Automatica*, vol. 44, pp. 671–684, 2008.

[7] J. Batson, A. Spielman, N. Srivastava, and S.-H. Teng, "Spectral sparsification of graphs: Theory and algorithms," *Communications of the ACM*, vol. 56, pp. 87–94, 2013.

[8] R. Tibshirani, "Regression shrinkage and selection via the LASSO," *Journal of the Royal Statistical Society, Series B*, vol. 58, pp. 267–288, 1994.

[9] L. Xiao and S. Boyd, "Fast linear iterations for distributed averaging," *Systems and Control Letters*, vol. 53, pp. 65–78, 2004.

[10] N. Dhingra, F. Lin, M. Fardad, and M. R. Jovanović, "On identifying sparse representations of consensus networks," in *Proceedings of the 3rd IFAC Workshop on Distributed Estimation and Control in Networked System, volume 3, part 1*, 2012, pp. 305–310.

[11] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.

[12] E. J. Candès, M. B. Wakin, and S. P. Boyd, "Enhancing sparsity by reweighted l_1 minimization," *Journal of Fourier Analysis and Applications*, vol. 14, pp. 877–905, 2008.

[13] 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, pp. 1–122, 2011.

[14] Y. E. and L. Y., "Model selection and estimation in regression with grouped variables," *Journal of the Royal Statistical Society, Series B*, vol. 68, pp. 49–67, 2007.

[15] N. Simon, J. Friedman, T. Hastie, and R. Tibshirani, "A sparse group LASSO," *Journal of Computational and Graphical Statistics*, vol. 22, pp. 231–245, 2013.