2018 Annual American Control Conference (ACC)
June 27–29, 2018. Wisconsin Center, Milwaukee, USA

Distributed proximal augmented Lagrangian method for nonsmooth composite optimization

Sepideh Hassan-Moghaddam and Mihailo R. Jovanović

***Abstract*— We study a class of nonsmooth composite optimization problems in which the convex objective function is given by a sum of differentiable and nondifferentiable terms. By introducing auxiliary variables in nondifferentiable terms, we provide an equivalent consensus-based characterization that is convenient for distributed implementation. The Moreau envelope associated with the nonsmooth part of the objective function is used to bring the optimization problem into a continuously differentiable form that serves as a basis for the development of a primal-descent dual-ascent gradient flow method. This algorithm exploits separability of the objective function and is well-suited for in-network optimization. We prove global asymptotic stability of the proposed algorithm and solve the problem of growing undirected consensus networks in a distributed manner to demonstrate its effectiveness.**

***Index Terms*— Proximal augmented Lagrangian method, alternating direction method of multipliers, consensus, distributed feedback design, large-scale systems, optimization, primal-dual method, sparsity-promoting optimal control.**

I. Introduction

We study a class of nonsmooth composite optimization problems in which the convex objective function is a sum of differentiable and nondifferentiable functions. Among other applications, these problems emerge in machine learning, compressive sensing, and control. Recently, regularization has been used as a promising tool for enhancing utility of standard optimal control techniques. In this approach, commonly used performance measures are augmented with regularization functions that are supposed to promote some desired structural features in the distributed controller, e.g., sparsity. Such an approach has received significant attention in recent years [1]–[7], but computing optimal solutions in large-scale problems still remains a challenge.

Distributed control techniques are critically important in the design of large-scale systems. In these systems, conventional control strategies that rely on centralized computation and implementation are often prohibitively expensive. For example, finding the optimal controller requires computation of the solution to the algebraic Riccati equations which is often infeasible because of high computational requirements. This necessitates the development of theory and techniques that utilize distributed computing architectures to cope with large problem sizes.

Generic descent methods cannot be used in the nonsmooth composite optimization problems due to the presence of a nondifferentiable component in the objective function. Moreover, these standard methods are not well-suited for distributed implementation. An alternative approach is to separate the smooth and nonsmooth parts of the objective function and use the alternating direction method of multipliers (ADMM). In [8], we exploit separability of the objective function and utilize an ADMM-based consensus algorithm to solve the regularized optimal control problem in a distributed manner over multiple processors. Even though the optimal control problem is in general non-convex, recent results can be utilized to show convergence to a local minimum [9]. However, in an update step of the ADMM algorithm, all the processors halt to compute the weighted average (the gathering step) [10].

Herein, we build on recent work [11] in which the structure of proximal operators associated with nonsmooth regularizers was exploited to bring the augmented Lagrangian into a continuously differentiable form. Such an approach is suitable for developing an algorithm based on primal-descent dual-ascent gradient method. We use the Arrow-Hurwicz-Uzawa gradient flow dynamics [12] and propose an algorithm that can be implemented in a fully distributed manner over multiple processors. This increases the computational efficiency and reduces the overall computation time. By exploiting convexity of the smooth part of the objective function, we show asymptotic convergence of our algorithm.

The point of departure of our work from [11] is that we study a more general form of consensus optimization problems in which the optimization variable is a matrix and develop a fully distributed algorithm. Furthermore, while most existing primal-dual techniques for nonsmooth distributed optimization employ subgradient flow methods [13]–[15], our approach yields a gradient flow dynamics with a continuous right-hand side even for nonsmooth problems.

The rest of the paper is structured as follows. In Section II, we formulate the nonsmooth composite optimization problem, discuss a motivating example, and provide background on proximal operators and the consensus-based ADMM algorithm. In Section III, by exploiting the structure of proximal operators, we introduce the proximal augmented Lagrangian. In Section IV, we use the Arrow-Hurwicz-Uzawa method to develop the gradient flow dynamics that are well-suited for distributed computations and prove global asymptotic stability. In Section V, we discuss distributed implementation, in Section VI, we provide examples, and, in Section VII, we offer concluding remarks.

Financial support from the National Science Foundation under award ECCS-1739210 is gratefully acknowledged.

S. Hassan-Moghaddam and M. R. Jovanović are with the Ming Hsieh Department of Electrical Engineering, University of Southern California, Los Angeles, CA 90089. E-mails: hassanmo@usc.edu, mihailo@usc.edu.

II. Problem formulation

We consider a composite convex optimization problem,

$$\underset{x}{\text{minimize}} \quad \sum_{i=1}^{n} f_i(x) \; + \; g(x) \tag{1}$$

where $x \in \mathbb{R}^m$ is the optimization variable, the functions f_i are continuously differentiable, and the function g is possibly nondifferentiable. This problem can be brought into a standard consensus form by introducing n local variables x_i and a global variable z,

$$\begin{array}{ll} \underset{x_i, \, z}{\text{minimize}} & \displaystyle\sum_{i=1}^{n} f_i(x_i) \; + \; g(z) \\ \text{subject to} & x_i - z = 0, \quad i = 1, \ldots, n. \end{array} \tag{2}$$

Even though this reformulation increases the number of optimization variables, it facilitates distributed computations by bringing the objective function into a separable form. Clearly, the solutions to (1) and (2) coincide but, in contrast to (1), optimization problem (2) is convenient for distributed implementation. Solving (2) on a single processor is not necessarily more computationally efficient than solving the original problem via a centralized algorithm. However, optimization problem (2) can be split into n separate subproblems over n different processors. In such a setup, each processor solves an optimization problem that involves a local objective function f_i of a single local variable x_i. This is advantageous for large-scale systems where centralized implementation is prohibitively complex and cannot be afforded.

A. Motivating application

Problem (1) arises in feedback design when a performance metric, e.g., the $\mathcal{H}_2$ norm, is augmented with a regularization function to promote structural features in the optimal controller. Herein, we discuss the problem of growing undirected consensus networks and show that the objective function is separable; thereby, it completely fits into the framework (2).

We consider the controlled undirected network,

$$\begin{aligned} \dot{\psi} \; &= \; -L_p \, \psi \; + \; d \; + \; u \\ \zeta \; &= \; \begin{bmatrix} Q^{1/2} \\ 0 \end{bmatrix} \psi \; + \; \begin{bmatrix} 0 \\ R^{1/2} \end{bmatrix} u \end{aligned}$$

where d and u are the disturbance and control inputs, ψ is the state, and ζ is the performance output. The dynamic matrix L_p is the Laplacian matrix of the plant network and symmetric matrices $Q \succeq 0$ and $R \succ 0$ specify the state and control weights in the performance output. For memoryless control laws,

$$u \; = \; -L_x \, \psi$$

where L_x is the Laplacian matrix of the controller graph, the closed-loop system is given by

$$\begin{aligned} \dot{\psi} \; &= \; -(L_p \, + \, L_x) \, \psi \; + \; d \\ \zeta \; &= \; \begin{bmatrix} Q^{1/2} \\ -R^{1/2} L_x \end{bmatrix} \psi. \end{aligned} \tag{3}$$

In the absence of exogenous disturbances, the network converges to the average of the initial node values $\bar{\psi} = (1/n) \sum_i \psi_i(0)$ if and only if it is connected [16]. Let $Q := I - (1/n) \mathbb{1}\mathbb{1}^T$ penalizes the deviation of individual node values from average. The objective is to minimize the mean square deviation from the network average by adding a few additional edges, specified by the graph Laplacian L_x of a controller network. If E is the incidence matrix of the controller graph, L_x can be written as

$$L_x \; = \; E \, \text{diag} \, (x) \, E^T$$

where $\text{diag}\,(x)$ is a diagonal matrix containing the optimization variable $x \in \mathbb{R}^m$ (i.e., the vector of the edge weights in the controller graph). Regularization terms may be used to promote sparsity of the controller network or to impose some additional constraints on the edge weights. The matrix L_x that optimizes the closed-loop performance and has certain structural properties can be obtained by solving the regularized optimal control problem

$$\underset{x}{\text{minimize}} \quad f(x) \; + \; g(x). \tag{4}$$

Here, f is the function that quantifies the closed-loop performance, i.e. the $\mathcal{H}_2$ norm, and g is the regularization function that is introduced to promote certain structural properties of L_x. For example, when it is desired to design L_x with a specified pattern of zero elements, g is an indicator function of the set that characterizes this pattern [17]. When it is desired to promote sparsity of L_x, the ℓ_1 norm $g(x) = \gamma \sum_i |x_i|$ can be used as a sparsity-enhancing regularizer, where γ is the positive parameter that characterizes emphasis on sparsity [2].

Next, we exploit the square-additive property of the $\mathcal{H}_2$ norm to provide an equivalent representation that is convenient for large-scale and distributed optimization. As shown in [6], up to an additive constant, the square of the $\mathcal{H}_2$ norm (from d to ζ) is determined by

$$\begin{aligned} f(x) \; = \; & \text{trace} \left((E \, \text{diag} \, (x) \, E^T \; + \; L_p)^\dagger (I \; + \; L_p \, R \, L_p) \right) + \\ & \text{diag} \left(E^T R \, E \right)^T x \end{aligned}$$

where the pseudo-inverse of the closed-loop graph Laplacian is given by

$$(E \text{diag}(x) E^T + L_p)^\dagger = (E \text{diag}(x) E^T + (1/n) \, \mathbb{1}\mathbb{1}^T + L_p)^{-1}.$$

It is easy to show that $f(x)$ can be written as

$$f(x) \; = \; \sum_{i=1}^{n} f_i(x)$$

where

$$\begin{aligned} f_i(x) \; &= \; \xi_i^T \left(E \, \text{diag} \, (x) \, E^T + (1/n) \, \mathbb{1}\mathbb{1}^T + L_p \right)^{-1} \xi_i \; + \\ & \qquad (1/n) \, \text{diag} \left(E^T R \, E \right)^T x. \end{aligned} \tag{5}$$

Here, $\xi_i = (I \, + \, L_p \, R \, L_p)^{1/2} \, \mathrm{e}_i$ is the ith column of the square root of the matrix $(I \, + \, L_p \, R \, L_p)$. Moreover, it can be shown that the gradient of $f_i(x)$ is given by

$$\nabla f_i(x) \;=\; (1/n)\,\mathrm{diag}\left(E^T R\, E\right) \,-\, \nu_i(x) \circ \nu_i(x) \quad (6)$$

where $\circ$ is the elementwise multiplication and

$$\nu_i(x) \;=\; E^T \left(E \,\mathrm{diag}\,(x)\, E^T + (1/n)\, \mathbb{1}\mathbb{1}^T + L_p\right)^{-1} \xi_i. \quad (7)$$

In what follows, we provide essential background on the proximal operators that we utilize for the latter developments.

B. Background

1) Proximal operators: The proximal operator of the function g is given by

$$\mathbf{prox}_{\mu g}(v) \;:=\; \underset{z}{\operatorname{argmin}}\; g(z) \,+\, \frac{1}{2\mu}\,\|z \,-\, v\|^2,$$

and the Moreau envelope determines the corresponding value function,

$$M_{\mu g}(v) \;:=\; g(\mathbf{prox}_{\mu g}(v)) \,+\, \frac{1}{2\mu}\,\|\mathbf{prox}_{\mu g}(v) \,-\, v\|_2^2. \quad (8)$$

Irrespective of differentiability of g, Moreau envelope is a continuously differentiable function and its gradient is given by [18],

$$\nabla M_{\mu g}(v) \;=\; \frac{1}{\mu}\,(v \,-\, \mathbf{prox}_{\mu g}(v)). \quad (9)$$

The above defined functions play essential role in our subsequent developments.

2) Alternating Direction Method of Multipliers (ADMM): We next demonstrate that a standard consensus algorithm based on the ADMM [10] can be used to solve the problem (2). This algorithm is well-suited for distributed implementation in which each processor solves an optimization problem. More details can be found in [8].

The augmented Lagrangian associated with (2) is given by

$$\mathcal{L}_{\mu_i}(x_i, z; \lambda) \;:=\; g(z) \,+\, \sum_{i\,=\,1}^{n} \left(f_i(x_i) \,+\, \langle \lambda_i, x_i - z\rangle \,+\, \frac{1}{2\,\mu_i}\,\|x_i - z\|_2^2\right) \quad (10)$$

where λ_i's are the Lagrange multipliers and μ_i's are positive parameters. The distributed ADMM algorithm consists of the following iterative steps,

$$\begin{aligned} x_i^{k+1} \;&=\; \underset{x_i}{\operatorname{argmin}}\; f_i(x_i) \,+\, \frac{1}{2\,\mu_i}\,\|x_i \,-\, u_i^k\|_2^2 \\ z^{k+1} \;&=\; \underset{z}{\operatorname{argmin}}\; g(z) \,+\, \sum_{i=1}^{n} \frac{1}{2\,\mu_i}\,\|z \,-\, v_i^k\|_2^2 \\ \lambda_i^{k+1} \;&=\; \lambda_i^k \,+\, \frac{1}{\mu_i}\left(x_i^{k+1} \,-\, z^{k+1}\right) \end{aligned}$$

where

$$\begin{aligned} u_i^k \;&:=\; z^k \,-\, \mu_i\, \lambda_i^k \\ v_i^k \;&:=\; x_i^{k+1} \,+\, \mu_i\, \lambda_i^k. \end{aligned}$$

The x_i-minimization step can be done via distributed computation by spreading subproblems to n different processors. On the other hand, the update of z amounts to the evaluation of the proximal operator of the function g,

$$z^{k+1} \;=\; \mathbf{prox}_{\hat{\mu} g}\left(\hat{\mu}\, \sum_{i\,=\,1}^{n} \frac{1}{\mu_i}\, v_i^k\right).$$

where $\hat{\mu} := (\sum_i 1/\mu_i)^{-1}$. Thus, the update of z requires gathering each x_i^{k+1} and the associated Lagrange multipliers λ_i^k in order to form v_i^k.

The above presented consensus algorithm is standard (e.g., see [10]). We have previously used this algorithm for distributed design of structured feedback gains in [8]. Recently, convergence of this algorithm was established even for problems with non-convex objective functions f_i [9].

III. Proximal augmented Lagrangian

The interconnection graph between the nodes in the consensus-based formulation in (2) is given by a star graph. Each node has access to the internal node of the star graph with the state z. This topology yields the z-update in ADMM that requires gathering the states of all subsystems. Recently, an algorithm based on the proximal augmented Lagrangian for solving (2) was developed in [11]. To avoid the above described computational requirement in the z-update of the ADMM algorithm, we propose a primal-dual algorithm based on the proximal augmented Lagrangian that can be implemented in a fully distributed manner.

Problem (2) can be equivalently written as,

$$\begin{array}{ll} \underset{x_i,\, z}{\text{minimize}} & \displaystyle\sum_{i\,=\,1}^{n} f_i(x_i) \,+\, g(z) \\ \text{subject to} & x_i \,-\, x_j \,=\, 0, \quad (i,j) \in \mathcal{I} \\ & x_k \,-\, z \,=\, 0, \end{array} \quad (11)$$

where $\mathcal{I}$ is the set of indices between 1 and n such that any index appears in one pair of the set at least once. This set characterizes structure of the information exchange network between the agents. The interaction topology is given by a connected graph. Moreover, the index $k \in \{1, \ldots, n\}$ can be chosen arbitrarily. In what follows, we study one particular instance of problem (11).

Without loss of generality, we assume that $k = n$ and that the underlying communication network between different nodes in (11) is given by a path graph. By introducing the optimization variable X,

$$X \;:=\; \left[\begin{array}{ccc} x_1 & \cdots & x_n \end{array}\right] \,\in\, \mathbb{R}^{m \times n}$$

the column vector e_n,

$$\mathrm{e}_n \;=\; \left[\begin{array}{cccc} 0 & 0 & \cdots & 1 \end{array}\right]^T \,\in\, \mathbb{R}^n$$

and the matrix T,

$$T \;=\; \left[\begin{array}{cccc} 1 & 0 & \cdots & 0 \\ -1 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & -1 \end{array}\right] \,\in\, \mathbb{R}^{n \times (n-1)}$$

we can rewrite (11) as,

$$\begin{array}{ll} \underset{X,\,z}{\text{minimize}} & f(X) \,+\, g(z) \\ \text{subject to} & X\,T \,=\, 0, \\ & X\,\mathrm{e}_n \,-\, z \,=\, 0. \end{array} \tag{12}$$

In (12), the matrix T is the incidence matrix of an undirected path network that connects n nodes. We note that any connected network can be used to build the information exchange structure between the nodes. For an arbitrary connected graph with m edges, the matrix $T \in \mathbb{R}^{n\times m}$ has to satisfy the following properties,

$$T^T \mathbb{1} \,=\, 0, \quad T\,T^T \,=\, L,$$

where $\mathbb{1}$ is the vector of all ones and $L \in \mathbb{R}^{n\times n}$ is the Laplacian matrix of the underlying graph.

The augmented Lagrangian associated with (12) is given by

$$\mathcal{L}_{\mu_i}(X,z;\lambda,Y) \,=\, f(X) \,+\, g(z) \,+\, \langle \lambda, X\,\mathrm{e}_n \,-\, z\rangle \,+ \\ \langle Y, X\,T\rangle \,+\, \frac{1}{2\,\mu_1}\,\|X\,\mathrm{e}_n - z\|_2^2 \,+\, \frac{1}{2\,\mu_2}\,\|X\,T\|_F^2 \tag{13}$$

where $\lambda \in \mathbb{R}^m$ and $Y \in \mathbb{R}^{m\times(n-1)}$ are the Lagrange multipliers and μ_1 and μ_2 are positive parameters. The proximal augmented Lagrangian is obtained by evaluating the augmented Lagrangian on the manifold that results from the explicit minimization of $\mathcal{L}_{\mu_i}$ with respect to z [11]. This yields a function that is once but not twice continuously differentiable with respect to both the primal variable X and the dual variables λ and Y.

The proximal augmented Lagrangian associated with (12) is given by

$$\mathcal{L}_{\mu_i}(X;\lambda,Y) \,=\, \mathcal{L}_{\mu_i}(X, z^\star_{\mu_1}(X;\lambda);\lambda,Y) \,= \\ f(X) \,+\, M_{\mu_1 g}(X\,\mathrm{e}_n \,+\, \mu_1\,\lambda) \,-\, \frac{\mu_1}{2}\,\|\lambda\|_2^2 \,+\, \langle Y, X\,T\rangle \,+ \\ \frac{1}{2\,\mu_2}\,\|X\,T\|_F^2, \tag{14}$$

where $M_{\mu_1 g}$ is the Moreau envelope of the function $g(z)$ and $z^\star_{\mu_1}(X;\lambda)$ is given by

$$z^\star_{\mu_1}(X;\lambda) \,=\, \mathbf{prox}_{\mu_1 g}(X\,\mathrm{e}_n \,+\, \mu_1\,\lambda)$$

and by substituting $z^\star$ in the augmented Lagrangian (13), the proximal augmented Lagrangian can be written as (14).

IV. Arrow-Hurwicz-Uzawa gradient flow

The proximal augmented Lagrangian is a continuously differentiable function because the Moreau envelope is continuously differentiable. This facilitates the use of the Arrow-Hurwicz-Uzawa algorithm which is a primal-descent dual-ascent gradient flow method. In this algorithm, the primal variable X and the dual variables λ and Y are updated simultaneously. The gradient flow dynamics are given by

$$\begin{aligned} \dot{X} &= -\nabla_X \mathcal{L}_\mu(X;\lambda,Y) \\ \dot{\lambda} &= +\nabla_\lambda \mathcal{L}_\mu(X;\lambda,Y) \\ \dot{Y} &= +\nabla_Y \mathcal{L}_\mu(X;\lambda,Y). \end{aligned}$$

By taking the derivatives, the updates can be written as

$$\begin{aligned} \dot{X} &= -\,(\nabla f(X) \,+\, \nabla M_{\mu_1 g}(X\,\mathrm{e}_n \,+\, \mu_1\,\lambda)\,\mathrm{e}_n^T \,+ \\ &\quad \frac{1}{\mu_2}\,X\,T\,T^T \,+\, Y\,T^T) \\ \dot{\lambda} &= \mu_1\,\nabla M_{\mu_1 g}(X\,\mathrm{e}_n \,+\, \mu_1\,\lambda) \,-\, \mu_1\,\lambda \\ \dot{Y} &= X\,T. \end{aligned} \tag{15}$$

It is worth to note that if $f(x)$ is separable, i.e. $f(x) = \sum_i f_i(x_i)$, the gradient $\nabla f(X)$ is an $m\times n$ matrix and can be written as

$$\nabla f(X) \,=\, \begin{bmatrix} \nabla f_1(x_1) & \ldots & \nabla f_n(x_n) \end{bmatrix}.$$

Asymptotic convergence

Our subsequent developments are based on the following assumption

Assumption 1: The function f is continuously differentiable and convex, and the function g is proper, lower semicontinuous, and convex.

We show that under Assumption 1, dynamics (15) are globally asymptotically stable and converge to $(X^\star, \lambda^\star, Y^\star)$ where each of the columns of $X^\star$ is the optimal solution to (1). The optimal primal and dual points $(X^\star, \lambda^\star, Y^\star)$ satisfy the following first order optimality conditions

$$\nabla f(X^\star) \,+\, \lambda^\star\,\mathrm{e}_n^T \,+\, Y^\star\,T^T \,=\, 0 \tag{16a}$$
$$X^\star\,\mathrm{e}_n \,-\, z^\star \,=\, 0 \tag{16b}$$
$$\partial g(z^\star) \,-\, \lambda^\star \,\ni\, 0 \tag{16c}$$
$$X^\star\,T \,=\, 0 \tag{16d}$$

where ∂g is the subgradient of g.

Proposition 1: Let Assumption 1 hold. Then, gradient flow dynamics (15) are globally asymptotically stable, i.e., they converge globally to the optimal primal and dual points $(X^\star, \lambda^\star, Y^\star)$ of (12).

Proof: We introduce a change of variables

$$\tilde{X} \,:=\, X \,-\, X^\star, \quad \tilde{\lambda} \,:=\, \lambda \,-\, \lambda^\star, \quad \tilde{Y} \,:=\, Y \,-\, Y^\star$$

and a Lyapunov function

$$V(\tilde{X},\tilde{\lambda},\tilde{Y}) \,=\, \frac{1}{2}\left\langle \tilde{X},\tilde{X}\right\rangle \,+\, \frac{1}{2}\left\langle \tilde{\lambda},\tilde{\lambda}\right\rangle \,+\, \frac{1}{2}\left\langle \tilde{Y},\tilde{Y}\right\rangle$$

where $(\tilde{X},\tilde{\lambda},\tilde{Y})$ satisfy

$$\begin{aligned} \dot{\tilde{X}} &= \nabla f(X^\star) - \nabla f(X) \,-\, \frac{1}{\mu_1}\,\tilde{m}\,\mathrm{e}_n^T \,-\, \frac{1}{\mu_2}\,\tilde{X}\,T\,T^T \,- \\ &\quad \tilde{Y}\,T^T, \\ \dot{\tilde{\lambda}} &= -\mu_1\,\tilde{\lambda} \,+\, \tilde{m}, \quad \dot{\tilde{Y}} \,=\, \tilde{X}\,T, \end{aligned} \tag{17}$$

with

$$\tilde{m} := \mu_1 \left(\nabla M_{\mu_1 g}(X\, \mathrm{e}_n + \mu_1\, \lambda) - \nabla M_{\mu_1 g}(X^\star\, \mathrm{e}_n + \mu_1\, \lambda^\star) \right).$$

Based on [11, Lemma 2], we can write

$$P\,(\tilde{X}\mathrm{e}_n \,+\, \mu_1 \tilde{\lambda}) \;=\; \mathbf{prox}_{\mu_1 g}(X\mathrm{e}_n + \mu_1\,\lambda) \,-\, \mathbf{prox}_{\mu_1 g}(X^\star \mathrm{e}_n + \mu_1\,\lambda^\star), \tag{18}$$

where I is the identity matrix and P is a positive semidefinite matrix such that $P \preceq I$. Thus, from (9) we have,

$$\tilde{m} \;=\; (I \,-\, P)\,(\tilde{X}\mathrm{e}_n + \mu_1\,\tilde{\lambda}). \tag{19}$$

The derivative of the Lyapunov function candidate along the solutions of (17) is determined by

$$\begin{aligned}\dot{V} \;&=\; \left\langle \tilde{X}, \dot{\tilde{X}} \right\rangle + \left\langle \tilde{\lambda}, \dot{\tilde{\lambda}} \right\rangle + \left\langle \tilde{Y}, \dot{\tilde{Y}} \right\rangle \\ &=\; -\left\langle \tilde{X}, \nabla f(X) \,-\, \nabla f(X^\star) \right\rangle - \mu_1 \left\langle \tilde{\lambda}, P\,\tilde{\lambda} \right\rangle - \\ &\quad \frac{1}{\mu_1} \left\langle (I \,-\, P)\,\tilde{X}\mathrm{e}_n, \tilde{X}\,\mathrm{e}_n \right\rangle - \frac{1}{\mu_2} \left\langle \tilde{X}\,T\,T^T, \tilde{X} \right\rangle.\end{aligned}$$

Since f is convex, the first term in nonpositive. Moreover, by utilizing $0 \preceq P \preceq I$, it follows that $\dot{V} \le 0$. We next invoke LaSalle's invariance principle [19] to establish global asymptotic stability.

The points $(\tilde{X}, \tilde{\lambda}, \tilde{Y})$ in the set

$$S \;=\; \{(\tilde{X}, \tilde{\lambda}, \tilde{Y}) \mid \dot{V}(\tilde{X}, \tilde{\lambda}, \tilde{Y}) \;=\; 0\},$$

satisfy

$$\begin{aligned}\nabla f(X^\star + \tilde{X}) \;&=\; \nabla f(X^\star), && \text{(20a)}\\ \tilde{\lambda} \;&\in\; \ker(P), && \text{(20b)}\\ \tilde{X}\,\mathrm{e}_n \;&\in\; \ker(I \,-\, P), && \text{(20c)}\\ \tilde{X}^T \;&\in\; \mathrm{span}\{[\,\mathbb{1} \cdots \mathbb{1}\,]\}, && \text{(20d)}\end{aligned}$$

where $\ker(A)$ denotes the null space of the matrix A and $A \in \mathrm{span}\{[\,\mathbb{1} \cdots \mathbb{1}\,]\}$ signifies that each column of the matrix A is given by a scalar multiple of the vector of all ones, $\mathbb{1}$. From (19), for the points in this set, we have $\tilde{m} = \mu_1 \tilde{\lambda}$. In order to identify the largest invariant set in S, we evaluate dynamics (17) under constraints (20) to obtain

$$\dot{\tilde{X}} \;=\; -\,\tilde{\lambda}\,\mathrm{e}_n^T \,-\, \tilde{Y}\,T^T, \quad \dot{\tilde{\lambda}} \;=\; 0, \quad \dot{\tilde{Y}} \;=\; 0. \tag{21}$$

Thus, the invariant set is characterized by $\tilde{\lambda}\mathrm{e}_n^T + \tilde{Y}T^T = 0$ for constant $\tilde{\lambda}$ and $\tilde{Y}$. To complete the proof, we need to show that the largest invariant set in S yields

$$(X, \lambda, Y) \;=\; (X^\star, \lambda^\star, Y^\star) \;+\; (\tilde{X}, \tilde{\lambda}, \tilde{Y}),$$

that satisfy optimality conditions (16).

Points (X, λ, Y) satisfy optimality condition (16a) if

$$\nabla f(X) \,+\, (\tilde{\lambda} \,+\, \lambda^\star)\,\mathrm{e}_n^T \,+\, (\tilde{Y} \,+\, Y^\star)\,T^T \;=\; 0.$$

For any $(\tilde{X}, \tilde{\lambda}, \tilde{Y})$ in the invariant set, we can use (20a) to replace $\nabla f(X)$ with $\nabla f(X^\star)$. Furthermore, since $\tilde{\lambda}\mathrm{e}_n^T + \tilde{Y}T^T = 0$, the resulting (X, λ, Y) satisfy (16a). Moreover, the substitution of $P\tilde{\lambda} = 0$ and $(I - P)\tilde{X}\mathrm{e}_n = 0$ to (18) yields

$$\begin{aligned} X\mathrm{e}_n \,-\, X^\star \mathrm{e}_n \;&=\; \mathbf{prox}_{\mu_1 g}(X\mathrm{e}_n + \mu_1\,\lambda) - \\ &\quad\; \mathbf{prox}_{\mu_1 g}(X^\star \mathrm{e}_n + \mu_1\,\lambda^\star).\end{aligned}$$

The optimality condition (16b) leads to

$$X\mathrm{e}_n \;=\; \mathbf{prox}_{\mu_1 g}(X\mathrm{e}_n \,+\, \mu_1\,\lambda) \;=\; z$$

which implies that the pair (X, z) satisfies (16b). We next show that the optimality condition (16c) holds for any (z, λ) in this set. Taking sub-differential of the proximal operator of the function g in (8) yields

$$\partial g(z) \,+\, \tfrac{1}{\mu_1}(z \,-\, v) \;\ni\; 0$$

where v is an arbitrary vector. Choosing $v = X\mathrm{e}_n + \mu_1 \lambda$ and utilizing the fact that $X\mathrm{e}_n = z$ yields the third optimality condition (16c). Furthermore, $\tilde{X}T = 0$ yields $XT = 0$. Thus, X satisfies (16d) and the dynamics (15) converges asymptotically to the optimal points $(X^\star, \lambda^\star, Y^\star)$. ■

V. DISTRIBUTED IMPLEMENTATION

In this section, we exploit the structure of the problem and show that the gradient flow dynamics (15) is well-suited for distributed implementation. In this case, the underlying interconnection network is given by a path graph. We first discuss how the gradient flow of the primal variable X can be implemented in a distributed manner and then show that the dual variables can be also updated in a distributed fashion.

A. Primal update

The vector x_k denotes the kth column of the matrix X. Each of the columns from $k = 2, \ldots, (n-1)$ can be updated in a distributed manner as follows

$$\begin{aligned}\dot{x}_k \;&=\; -\,\nabla f_k(x_k) \,-\, \frac{1}{\mu_2}\,(2\,x_k \,-\, x_{k-1} \,-\, x_{k+1}) \,- \\ &\quad\; y_k \,+\, y_{k-1},\end{aligned} \tag{22a}$$

where the vector y_k is the kth columns of the matrix Y. Thus, each agent only uses its neighbors' states and the corresponding dual variables to update its own state. The updates for the first and last column of X which are x_1 and x_n are different than the other updates and can be written as follows

$$\dot{x}_1 \;=\; -\,\nabla f_1(x_1) \,-\, \frac{x_1 \,-\, x_2}{\mu_2} \,-\, y_1 \tag{22b}$$

$$\begin{aligned}\dot{x}_n \;&=\; -\,\nabla f_n(x_n) \,-\, \nabla M_{\mu_1 g}(x_n \,+\, \mu_1\,\lambda) \,- \\ &\quad\; \frac{1}{\mu_2}\,(x_n \,-\, x_{n-1}) \,+\, y_{n-1}.\end{aligned} \tag{22c}$$

Similarly, we can see only local information exchange and access to local dual variable is required for these two updates.

B. Dual updates

The dual variable λ is a column vector and its update can be done by using the following column update

$$\dot{\lambda} \;=\; \mu_1\,\nabla M_{\mu_1 g}(x_n \,+\, \mu_1\,\lambda) \,-\, \mu_1\,\lambda. \tag{22d}$$

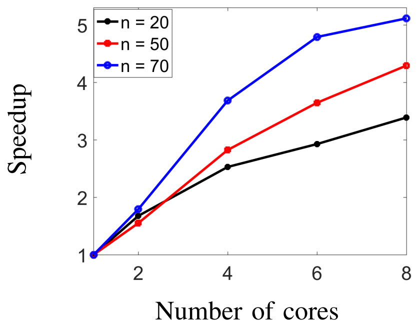

Fig. 1: Speedup ratio versus the number of the cores used for growing connected resistive networks with n nodes.

Thus, we only need the state of the nth agent to update its value. The second dual variable Y is an $m \times (n-1)$ matrix and the kth column of it, y_k, can be updated using only the states of x_k and x_{k+1} agents for $k = 1, \ldots, (n-1)$,

$$\dot{y}_k \;=\; x_k \,-\, x_{k+1}. \tag{22e}$$

VI. Computational experiments

In this section, we employ our algorithm for growing connected resistive Erdös-Rényi networks with edge probability $1.05\log(n)/n$ for different number of nodes using multiple cores. We choose the control weight matrix $R = I$ and a state weight matrix that penalizes the mean-square deviation from the network average, $Q = I - (1/n)\,\mathbb{1}\mathbb{1}^T$. Moreover, the incidence matrix of the controller is such that there are no joint edges between the plant and the controller graphs. This algorithm is implemented in a distributed fashion by splitting the problem into N separate subproblems over N different cores. We have provided a parallel implementation in MATLAB and have executed tests on a machine featuring an Intel Core i7-3770 with 16GB of RAM to measure the performance of the algorithm.

The solve times are averaged over 10 trials and the speedup relative to a single core is displayed in Fig 1. It demonstrates that the algorithm is scalable. In particular, multi-core execution outperforms running just on a single core. Moreover, the speed-up is even higher for larger networks since overheads of parallel execution are less and more time is spent on actual parallel computation.

VII. Concluding remarks

We have studied a class of convex nonsmooth composite optimization problems in which the objective function is a combination of differentiable and nondifferentiable functions. By exploiting the structure of the probelm, we have provided an equivalent consensus-based characterization and have developed an algorithm based on primal-descent dual-ascent gradient flow method. This algorithm exploits the separability of the objective function and is well-suited for distributed implementation. Convexity of the smooth part of the objective function is utilized to prove global asymptotic stability of our algorithm. Finally, by exploiting the structure of the $\mathcal{H}_2$ norm, we have employed this algorithm to design a sparse controller network that improves the performance of the closed-loop system in a large-scale undirected consensus network in a distributed manner. An example is provided to demonstrate the utility of the developed approach. We are currently working on implementing this algorithm in C++ and will use it to solve structured optimal control problems for large-scale systems in a distributed manner.

References

[1] M. Fardad, F. Lin, and M. R. Jovanović, "Sparsity-promoting optimal control for a class of distributed systems," in *Proceedings of the 2011 American Control Conference*, 2011, pp. 2050–2055.

[2] F. Lin, M. Fardad, and M. R. Jovanović, "Design of optimal sparse feedback gains via the alternating direction method of multipliers," *IEEE Trans. Automat. Control*, vol. 58, no. 9, pp. 2426–2431, 2013.

[3] S. Schuler, P. Li, J. Lam, and F. Allgöwer, "Design of structured dynamic output-feedback controllers for interconnected systems," *International Journal of Control*, vol. 84, no. 12, pp. 2081–2091, 2011.

[4] N. Matni and V. Chandrasekaran, "Regularization for design," *IEEE Trans. Automat. Control*, vol. 61, no. 12, pp. 3991–4006, 2016.

[5] M. R. Jovanović and N. K. Dhingra, "Controller architectures: tradeoffs between performance and structure," *Eur. J. Control*, vol. 30, pp. 76–91, July 2016.

[6] S. Hassan-Moghaddam and M. R. Jovanović, "Topology design for stochastically-forced consensus networks," *IEEE Trans. Control Netw. Syst.*, 2017, doi:10.1109/TCNS.2017.2674962; also arXiv:1506.03437v3.

[7] S. Hassan-Moghaddam, X. Wu, and M. R. Jovanović, "Edge addition in directed consensus networks," in *Proceedings of the 2017 American Control Conference*, Seattle, WA, 2017, pp. 5592–5597.

[8] S. Hassan-Moghaddam and M. R. Jovanović, "Distributed design of optimal structured feedback gains," in *Proceedings of the 56th IEEE Conference on Decision and Control*, 2017, pp. 6586–6591.

[9] M. Hong, Z. Q. Luo, and M. Razaviyayn, "Convergence analysis of alternating direction method of multipliers for a family of nonconvex problems," *SIAM J. Optimiz.*, vol. 26, no. 1, pp. 337–364, 2016.

[10] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundation and Trends in Machine Learning*, vol. 3, no. 1, pp. 1–124, 2011.

[11] N. K. Dhingra, S. Z. Khong, and M. R. Jovanović, "The proximal augmented Lagrangian method for nonsmooth composite optimization," *IEEE Trans. Automat. Control*, 2016, submitted; also arXiv:1610.04514.

[12] K. J. Arrow, L. Hurwicz, H. Uzawa, and H. B. Chenery, "Studies in linear and non-linear programming," 1958.

[13] J. Wang and N. Elia, "A control perspective for centralized and distributed convex optimization," in *Proceedings of the 50th IEEE Conference on Decision and Control and the 10th European Control Conference*, 2011, pp. 3800–3805.

[14] A. Cherukuri, E. Mallada, and J. Cortés, "Asymptotic convergence of constrained primal–dual dynamics," *Syst. Control Lett.*, vol. 87, pp. 10–15, 2016.

[15] A. Cherukuri, B. Gharesifard, J., and Cortés, "Saddle-point dynamics: conditions for asymptotic stability of saddle points," *SIAM J. Control Optim.*, vol. 55, no. 1, pp. 486–511, 2017.

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

[17] F. Lin, M. Fardad, and M. R. Jovanović, "Augmented Lagrangian approach to design of structured optimal state feedback gains," *IEEE Trans. Automat. Control*, vol. 56, no. 12, pp. 2923–2929, 2011.

[18] N. Parikh and S. Boyd, "Proximal algorithms," *Foundations and Trends in Optimization*, vol. 1, no. 3, pp. 123–231, 2013.

[19] H. K. Khalil, "Noninear systems," *Prentice-Hall, New Jersey*, vol. 2, no. 5, pp. 5–1, 1996.