Automatica 55 (2015) 209–216

Contents lists available at ScienceDirect

Automatica

journal homepage: www.elsevier.com/locate/automatica

Brief paper

On linear convergence of a distributed dual gradient algorithm for linearly constrained separable convex problems[☆]

Ion Necoara [1], Valentin Nedelcu

Automatic Control and Systems Engineering Department, University Politehnica Bucharest, Spl. Independentei 313, 060042 Bucharest, Romania

ARTICLE INFO

Article history:
Received 10 February 2014
Received in revised form
30 September 2014
Accepted 29 January 2015
Available online 31 March 2015

Keywords:
Separable convex problems
Error bound
Dual decomposition
Distributed gradient algorithm
Linear convergence

ABSTRACT

In this paper we propose a fully distributed dual gradient algorithm for minimizing linearly constrained separable convex problems and analyze its rate of convergence. In particular, we prove that under the assumption of strong convexity and Lipschitz continuity of the gradient of the primal objective function we have a global error bound type property for the dual problem. Using this error bound property we devise a fully distributed dual gradient scheme, i.e. a gradient scheme based on a weighted step size, for which we derive global linear rate of convergence for both dual and primal suboptimality and for primal feasibility violation. Numerical simulations are also provided to confirm our theory.

1. Introduction

Nowadays, many engineering applications which appear in the context of communications networks or networked systems (e.g. distributed model predictive control (DMPC) Necoara, Nedelcu, & Dumitrache, 2011, network utility maximization (NUM) Beck, Nedic, Ozdaglar, & Teboulle, 2014, and direct current optimal power flow for a power system (DC-OPF) Bakirtzis & Biskas, 2003) can be posed as large scale linearly constrained separable convex problems. Due to the large dimension and the separable structure of these problems, distributed optimization methods have become an appropriate tool for solving them. Distributed methods are based on decomposition (Necoara et al., 2011), which consists in dividing the original large problem into smaller subproblems. Decomposition methods can be divided into two main classes: primal and dual decomposition. While in the primal decomposition methods the optimization problem is solved using the original formulation and variables, in dual decomposition the constraints are moved into the cost using the Lagrange multipliers and then the dual problem is solved (Necoara et al., 2011). In many applications, e.g. (DMPC), (NUM) and (DC-OPF), when the constraints set is complicated (i.e. the projection on this set is hard to compute), dual decomposition is more effective since a primal approach would require at each iteration a numerically expensive projection.

First order decomposition methods for solving dual problems have been extensively studied in the literature. Convergence rate analysis for the dual subgradient method is given e.g. in Nedic and Ozdaglar (2009), where estimates of order $\mathcal{O}(1/\sqrt{k})$ for suboptimality and feasibility violation of an averaged primal sequence are provided, with k denoting the iteration counter. In Necoara and Nedelcu (2014) (Giselsson, Doan, Keviczky, De Schutter, & Rantzer, 2013 and Patrinos & Bemporad, 2014) the authors propose inexact (exact) dual fast gradient algorithms for which estimates of order $\mathcal{O}\left(1/k^2\right)$ in an averaged primal sequence are provided for primal suboptimality and feasibility violation. To our knowledge, the first result on the linear convergence of dual gradient method was provided in Luo and Tseng (1993). However, the linear convergence was valid only *locally* using a local error bound condition that estimates the distance from the dual optimal solution set in terms of the norm of a proximal residual. Finally, very few results were known in the literature on distributed implementations of dual gradient type methods since most of the papers enumerated above require a centralized step size. Recently, in Beck et al. (2014) and Necoara and Clipici (2013), distributed (dual fast) gradient algorithms are given, where the step size is chosen in a distributed fashion.

Contributions: In this paper we propose a fully distributed dual gradient method generating approximate primal feasible and optimal solutions, but improving the convergence rate w.r.t. existing

[☆] The material in this paper was not presented at any conference. This paper was recommended for publication in revised form by Associate Editor Constantino M. Lagoa under the direction of Editor Richard Middleton.

E-mail addresses: ion.necoara@acse.pub.ro (I. Necoara), valentin.nedelcu@acse.pub.ro (V. Nedelcu).

[1] Tel.: +40 21 402 9195.

http://dx.doi.org/10.1016/j.automatica.2015.02.038

results. Under the assumptions of strong convexity and Lipschitz continuity of the gradient of the primal objective function, which are often satisfied in practical applications (e.g. (DMPC), (NUM) or (DC-OPF)), we prove that the corresponding dual problem satisfies a certain global error bound property that estimates the distance from the dual optimal solution set in terms of the norm of a proximal residual. In these settings we analyze the convergence behavior of a distributed dual gradient algorithm, for which we are able to provide *global linear* convergence rate on primal suboptimality and feasibility violation for the last primal iterate. Moreover, our algorithm is fully distributed since is based on a weighted step size, as opposed to typical dual distributed schemes existing in literature, where a centralized step size is used and sublinear convergence is proved (Giselsson et al., 2013; Meinel, Ulbrich, & Albrecht, 2014; Necoara & Nedelcu, 2014; Necoara et al., 2011; Nedic & Ozdaglar, 2009). Note that some proofs are left out due to space restrictions (see Necoara & Nedelcu, 2013 for a complete presentation of our results).

Notations. For $z, y \in \mathbb{R}^n$ we denote the Euclidean inner product (norm) $\langle z, y\rangle = z^T y$ ($\|z\| = \sqrt{\langle z, z\rangle}$) and the infinity norm $\|z\|_\infty = \max_i |z_i|$. For a matrix $G \in \mathbb{R}^{m\times n}$, $\|G\|$ denotes its spectral norm. Also, we denote the orthogonal projection onto the nonnegative orthant $\mathbb{R}^n_+$ by $[z]_+$ and the orthogonal projection onto the convex set D by $[z]_D$. For a positive definite matrix W we define norm $\|z\|_W = \sqrt{z^T W z}$ and the projection of vector z onto a convex set D w.r.t. norm $\|\cdot\|_W$ by $[z]^W_D$. For a matrix A, A_i is its ith (block) column.

2. Problem formulation

We consider the following large scale linearly constrained separable convex optimization problem:

$$f^* = \min_{z_i \in \mathbb{R}^{n_i}} f(z) \left(= \sum_{i=1}^{M} f_i(z_i) \right) \quad (1)$$
$$\text{s.t.:} \quad Az = b, \qquad Cz \le c,$$

where $f_i : \mathbb{R}^{n_i} \to \mathbb{R}$ are convex functions, $z = \left[z_1^T \cdots z_M^T\right]^T \in \mathbb{R}^n$, $A \in \mathbb{R}^{p\times n}$, $C \in \mathbb{R}^{q\times n}$, $b \in \mathbb{R}^p$ and $c \in \mathbb{R}^q$. To our problem (1) we associate a bipartite communication graph $\mathcal{G} = (V_1, V_2, E)$, where $V_1 = \{1, \dots, M\}$, $V_2 = \left\{1, \dots, \bar{M}\right\}$ and $E \in \{0, 1\}^{\bar{M}\times M}$ represents the incidence matrix. e.g., in the context of (NUM) and (DC-OPF), V_1 denotes the set of sources, V_2 the set of links between sources and the incidence matrix E models the way sources interact. In (DMPC), $V_1 = V_2$ represents the set of interacting subsystems, while the incidence matrix E indicates the dynamic couplings between these subsystems. We assume that A and C are block matrices with the blocks $A_{ji} \in \mathbb{R}^{p_j\times n_i}$ and $C_{ji} \in \mathbb{R}^{q_j\times n_i}$, where $\sum_{i=1}^M n_i = n$, $\sum_{j=1}^{\bar{M}} p_j = p$ and $\sum_{j=1}^{\bar{M}} q_j = q$. We also assume that if $E_{ji} = 0$, then both blocks A_{ji} and C_{ji} are zero. In these settings we allow a block A_{ji} or C_{ji} to be zero even if $E_{ji} = 1$. We also introduce the index sets: $\bar{\mathcal{N}}_i = \left\{j \in V_2 : E_{ji} \ne 0\right\}$ and $\mathcal{N}_j = \left\{i \in V_1 : E_{ji} \ne 0\right\}$ for all $i \in V_1, j \in V_2$, which describe the local information flow in the graph. Note that the cardinality of the sets $\bar{\mathcal{N}}_i$ and $\mathcal{N}_j$ can be viewed as a measure for the degree of separability of problem (1). Further, we make the following assumptions regarding optimization problem (1):

Assumption 2.1. (a) The functions f_i have Lipschitz continuous gradient with constants L_i and are σ_i-strongly convex w.r.t. the Euclidean norm $\|\cdot\|$ on $\mathbb{R}^{n_i}$ (Nesterov, 2004).
(b) Matrix A has full row rank and there exists a feasible point $\tilde{z}$ for problem (1) such that $A\tilde{z} = b$ and $C\tilde{z} < c$.

Assumption 2.1 implies that strong duality holds:

$$f^* = \max_{\nu\in\mathbb{R}^p, \mu\in\mathbb{R}^q_+} d(\nu, \mu), \quad (2)$$

where $d(\nu, \mu)$ denotes the dual function of (1):

$$d(\nu, \mu) = \min_{z\in\mathbb{R}^n} \mathcal{L}(z, \nu, \mu), \quad (3)$$

with the Lagrangian function $\mathcal{L}(z, \nu, \mu) = f(z) + \langle \nu, Az - b\rangle + \langle \mu, Cz - c\rangle$. For simplicity of the exposition we introduce the following notations: $G = [A^T\ C^T]^T$ and $g = [b^T\ c^T]^T$. Since f_i are strongly convex functions, then f is also strongly convex w.r.t. the Euclidean norm, with convexity parameter e.g. $\sigma_{\mathrm{f}} = \min_{i=1,\dots,M} \sigma_i$. Further, the dual function d is differentiable and its gradient w.r.t. (ν, μ) is given by the following expression (Necoara & Nedelcu, 2014): $\nabla d(\nu, \mu) = Gz(\nu, \mu) - g$, where $z(\nu, \mu)$ denotes the unique optimal solution of the inner problem (3), i.e.:

$$z(\nu, \mu) = \arg\min_{z\in\mathbb{R}^n} \mathcal{L}(z, \nu, \mu). \quad (4)$$

Moreover, the gradient ∇d of the dual function is Lipschitz continuous w.r.t. Euclidean norm $\|\cdot\|$, with constant (Necoara & Nedelcu, 2014): $L_{\mathrm{d}} = \|G\|^2/\sigma_{\mathrm{f}}$. If we denote by $\nu_{\bar{\mathcal{N}}_i} = \left[\nu_j\right]_{j\in\bar{\mathcal{N}}_i}$ and by $\mu_{\bar{\mathcal{N}}_i} = \left[\mu_j\right]_{j\in\bar{\mathcal{N}}_i}$ we can observe that the dual function can be written in the following separable form: $d(\nu, \mu) = \sum_{i=1}^M d_i(\nu_{\bar{\mathcal{N}}_i}, \mu_{\bar{\mathcal{N}}_i}) - \langle \nu, b\rangle - \langle \mu, c\rangle$, with

$$\begin{aligned} d_i(\nu_{\bar{\mathcal{N}}_i}, \mu_{\bar{\mathcal{N}}_i}) &= \min_{z_i\in\mathbb{R}^{n_i}} f_i(z_i) + \langle \nu, A_i z_i\rangle + \langle \mu, C_i z_i\rangle \\ &= \min_{z_i\in\mathbb{R}^{n_i}} f_i(z_i) + \sum_{j\in\bar{\mathcal{N}}_i} \left\langle A_{ji}^T \nu_j + C_{ji}^T \mu_j, z_i\right\rangle. \end{aligned} \quad (5)$$

In these settings, we have that the gradient ∇d_i is:

$$\nabla d_i(\nu_{\bar{\mathcal{N}}_i}, \mu_{\bar{\mathcal{N}}_i}) = G_{\bar{\mathcal{N}}_i} z_i(\nu_{\bar{\mathcal{N}}_i}, \mu_{\bar{\mathcal{N}}_i}),$$

where $G_{\bar{\mathcal{N}}_i} = \begin{bmatrix} [A_{ji}]_{j\in\bar{\mathcal{N}}_i} \\ [C_{ji}]_{j\in\bar{\mathcal{N}}_i} \end{bmatrix}$ and $z_i(\nu_{\bar{\mathcal{N}}_i}, \mu_{\bar{\mathcal{N}}_i})$ denotes the unique optimal solution in (5). Note that ∇d_i is Lipschitz continuous w.r.t. Euclidean norm $\|\cdot\|$, with constant (Necoara & Nedelcu, 2014): $L_{\mathrm{d}_i} = \left\|G_{\bar{\mathcal{N}}_i}\right\|^2/\sigma_i$. For simplicity, we also consider the notation $\lambda = \left[\nu^T \mu^T\right]^T$ and we denote the effective domain of the dual function by $\mathbb{D} = \mathbb{R}^p \times \mathbb{R}^q_+$. The following lemma, which is a distributed version of descent lemma (see e.g. Beck et al., 2014, Necoara & Clipici, 2013 and Necoara & Nedelcu, 2013 for a proof), is central in our derivations of a distributed dual algorithm.

Lemma 2.2. *Let Assumption 2.1(a) hold. Then, the following inequality is valid:*

$$d(\lambda) \ge d(\bar{\lambda}) + \left\langle \nabla d(\bar{\lambda}), \lambda - \bar{\lambda}\right\rangle - \frac{1}{2}\|\lambda - \bar{\lambda}\|^2_W \quad \forall \lambda, \bar{\lambda} \in \mathbb{D}, \quad (6)$$

where the matrix $W = diag(W_\nu, W_\mu)$ *with the matrices* $W_\nu = diag\left(\sum_{i\in\mathcal{N}_j} L_{\mathrm{d}_i} I_{p_j}; j \in V_2\right)$ *and* $W_\mu = diag\left(\sum_{i\in\mathcal{N}_j} L_{\mathrm{d}_i} I_{q_j}; j \in V_2\right)$.

We denote by Λ^* the set of optimal solutions of dual problem (2). According to Hiriart-Urruty and Lemarechal (1996, Theorem 2.3.2), if Assumption 2.1(b) holds for our original problem (1), then Λ^* is nonempty, convex and bounded. Then, for any $\lambda \in \mathbb{R}^{p+q}$, we can define the following finite quantity:

$$\mathcal{R}(\lambda) = \min_{\lambda^*\in\Lambda^*} \|\lambda^* - \lambda\|_W. \quad (7)$$

In the following sections, we analyze the structural properties of the dual problem (2) and then we propose a fully distributed dual gradient method for solving this problem which exploits the separability of the dual function and allow us to recover a suboptimal and nearly feasible solution for our original problem (1) in linear time.