Automatica 49 (2013) 829–833

Contents lists available at SciVerse ScienceDirect

Automatica

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

Brief paper

Accelerated gradient methods and dual decomposition in distributed model predictive control ☆

Pontus Giselsson [a,1], Minh Dang Doan [b], Tamás Keviczky [b], Bart De Schutter [b], Anders Rantzer [a]

[a] Department of Automatic Control, Lund University, Sweden

[b] Delft Center for Systems and Control, Delft University of Technology, The Netherlands

ARTICLE INFO

Article history:
Received 21 September 2011
Received in revised form
6 July 2012
Accepted 22 October 2012
Available online 11 February 2013

Keywords:
Distributed control
Gradient methods
Predictive control

ABSTRACT

We propose a distributed optimization algorithm for mixed $\mathcal{L}_1/\mathcal{L}_2$-norm optimization based on accelerated gradient methods using dual decomposition. The algorithm achieves convergence rate $O(\frac{1}{k^2})$, where k is the iteration number, which significantly improves the convergence rates of existing duality-based distributed optimization algorithms that achieve $O(\frac{1}{k})$. The performance of the developed algorithm is evaluated on randomly generated optimization problems arising in distributed model predictive control (DMPC). The evaluation shows that, when the problem data is sparse and large-scale, our algorithm can outperform current state-of-the-art optimization software CPLEX and MOSEK.

1. Introduction

Gradient-based optimization methods are known for their simplicity and low complexity within each iteration. A limitation of classical gradient-based methods is the slow rate of convergence. It can be shown (Bertsekas, 1999; Nesterov, 2004) that for functions with a Lipschitz-continuous gradient, i.e., smooth functions, classical gradient-based methods converge at a rate of $O(\frac{1}{k})$, where k is the iteration number. In Nemirovsky and Yudin (1983) it was shown that a lower bound on the convergence rate for gradient-based methods is $O(\frac{1}{k^2})$. Nesterov showed in his work (Nesterov, 1983) that an accelerated gradient algorithm can be constructed such that this lower bound on the convergence rate is achieved when minimizing unconstrained smooth functions. This result has been extended and generalized in several publications to handle constrained smooth problems and smooth problems with an additional non-smooth term (Beck & Teboulle, 2009; Nesterov, 1988, 2005; Tseng, submitted for publication). Gradient-based methods are suitable for distributed optimization when they are used in combination with dual decomposition techniques.

Dual decomposition has been a well-established concept since around 1960 when Uzawa's algorithm (Arrow, Hurwicz, & Uzawa, 1958) was presented. Similar ideas were exploited in large-scale optimization (Danzig & Wolfe, 1961). Over the next decades, methods for decomposition and coordination of dynamic systems were developed and refined (Findeisen, 1980; Mesarovic, Macko, & Takahara, 1970; Singh & Titli, 1978) and used in large-scale applications (Carpentier & Cohen, 1993). In Tsitsiklis, Bertsekas, and Athans (1986) a distributed asynchronous method was studied. More recently dual decomposition has been applied in the distributed model predictive control literature in Doan, Keviczky, and De Schutter (2011); Doan, Keviczky, Necoara, Diehl, and De Schutter (2009), Giselsson and Rantzer (2010) and Negenborn, De Schutter, and Hellendoorn (2008) for problems with a strongly convex quadratic cost and arbitrary linear constraints. The above mentioned methods rely on gradient-based optimization, which suffers from slow convergence properties $O(\frac{1}{k})$. Also the step size parameter in the gradient scheme must be chosen appropriately to get good performance. Such information has not been provided or has been chosen conservatively in these publications.

In this work, we improve on the previously presented distributed optimization methods by using an accelerated gradient method to solve the dual problem instead of a classical gradient

☆ The second, third and fourth authors were supported by the European Union Seventh Framework STREP project "Hierarchical and distributed model predictive control (HD-MPC)", contract number INFSO-ICT-223854, and the European Union Seventh Framework Program [FP7/2007-2013] under grant agreement no. 257462 HYCON2 Network of Excellence. The material in this paper was not presented at any conference. This paper was recommended for publication in revised form by Associate Editor Wei Ren under the direction of Editor Frank Allgöwer.

E-mail addresses: pontusg@control.lth.se (P. Giselsson), m.d.doan@tudelft.nl (M.D. Doan), t.keviczky@tudelft.nl (T. Keviczky), b.deschutter@tudelft.nl (B.D. Schutter), rantzer@control.lth.se (A. Rantzer).

[1] Tel.: +46 46 222 42 87; fax: +46 46 13 81 18.

0005-1098/$ – see front matter © 2013 Elsevier Ltd. All rights reserved.
doi:10.1016/j.automatica.2013.01.009

method. We also extend the class of problems considered by allowing an additional sparse but non-separable 1-norm penalty. Such 1-norm terms are used as regularization term or as penalty for soft constraints (Savorgnan, Romani, Kozma, & Diehl, 2011). Further, we provide the optimal step size parameter for the algorithm, which is crucial for performance. The convergence rate for the dual function value using the accelerated gradient method is implicitly known from Beck and Teboulle (2009) and Tseng (submitted for publication). However, the convergence rate in the dual function value does not indicate the rate at which the primal iterate approaches the primal optimal solution. In this paper we also provide convergence rate results for the primal variables.

Related to our work is the method presented in Necoara and Suykens (2008) for systems with a (non-strongly) convex cost. It is based on the smoothing technique presented by Nesterov in Nesterov (2005). Other relevant work is presented in Kögel and Findeisen (2011) and Richter, Jones, and Morari (2009) in which optimization problems arising in model predictive control (MPC) are solved in a centralized fashion using accelerated gradient methods. These methods are, however, restricted to handle only box-constraints on the control signals.

To evaluate the proposed distributed algorithm, we solve randomly generated large-scale and sparse optimization problems arising in distributed MPC and compare the execution times to state-of-the-art optimization software for large-scale optimization, in particular CPLEX and MOSEK. We also evaluate the performance loss obtained when suboptimal step lengths are used.

The paper is organized as follows. In Section 2, the problem setup is introduced. The dual problem to be solved is introduced in Section 3 and some properties of the dual function are presented. The distributed solution algorithm for the dual problem is presented in Section 4. In Section 5 a numerical example is provided, followed by conclusions drawn in Section 6.

2. Problem setup

In this paper we present a distributed algorithm for optimization problems with cost functions of the form

$$J(x) = \frac{1}{2}x^T Hx + g^T x + \gamma \|Px - p\|_1. \tag{1}$$

The full decision vector, $x \in \mathbb{R}^n$, is composed of local decision vectors, $x_i \in \mathbb{R}^{n_i}$, according to $x = [x_1^T, \dots, x_M^T]^T$. The quadratic cost matrix $H \in \mathbb{R}^{n\times n}$ is assumed separable, i.e., $H = \text{blkdiag}(H_1, \dots, H_M)$ where $H_i \in \mathbb{R}^{n_i\times n_i}$. Further, H is assumed positive definite with $\underline{\sigma}(H)I \preceq H \preceq \bar{\sigma}(H)I$, where $0 < \underline{\sigma}(H) \le \bar{\sigma}(H) < \infty$. The linear part $g \in \mathbb{R}^n$ consists of local parts, $g = [g_1^T, \dots, g_M^T]^T$ where $g_i \in \mathbb{R}^{n_i}$. Further, $P \in \mathbb{R}^{m\times n}$ is composed of $P = [P_1, \dots, P_m]^T$, where each $P_r = [P_{r1}^T, \dots, P_{rM}^T]^T \in \mathbb{R}^n$ and $P_{ri} \in \mathbb{R}^{n_i}$. We do not assume that the matrix P should be block-diagonal which means that the cost function J is not separable. However, we assume that the vectors P_r have sparse structure. Sparsity refers to the property that for each $r \in \{1, \dots, m\}$ there exist some $i \in \{1, \dots, M\}$ such that $P_{ri} = 0$. We also have $p = [p_1, \dots, p_m]^T$ and $\gamma > 0$. This gives the following equivalent formulation of (1)

$$J(x) = \sum_{i=1}^{M}\left[\frac{1}{2}x_i^T H_i x_i + g_i^T x_i\right] + \sum_{r=1}^{m}\left|\sum_{i=1}^{M} P_{ri}^T x_i - p_r\right|. \tag{2}$$

Minimization of (1) is subject to linear equality and inequality constraints

$$A_1 x = B_1, \qquad A_2 x \le B_2$$

where $A_1 \in \mathbb{R}^{q\times n}$ and $A_2 \in \mathbb{R}^{(s-q)\times n}$ contain $a_l \in \mathbb{R}^n$ as $A_1 = [a_1, \dots, a_q]^T$ and $A_2 = [a_{q+1}, \dots, a_s]^T$. Further, each $a_l = [a_{l1}^T, \dots, a_{lM}^T]^T$ where $a_{li} \in \mathbb{R}^{n_i}$. Further we have $B_1 \in \mathbb{R}^q$ and $B_2 \in \mathbb{R}^{s-q}$ where $B_1 = [b_1, \dots, b_q]^T$ and $B_2 = [b_{q+1}, \dots, b_s]^T$. We assume that the matrices A_1 and A_2 are sparse. By introducing the auxiliary variables y and the constraint $Px - p = y$ we get the following optimization problem

$$\begin{aligned} \min_{x,y} \quad & \frac{1}{2}x^T Hx + g^T x + \gamma \|y\|_1 \\ \text{s.t.} \quad & A_1 x = B_1 \\ & A_2 x \le B_2 \\ & Px - p = y. \end{aligned} \tag{3}$$

The objective of the optimization routine is to solve (3) in a distributed fashion using several computational units, where each computational unit computes the optimal local variables, denoted x_i^*, only. Each computational unit is assigned a number of constraints in (3) for which it is responsible. We denote the set of equality constraints that unit i is responsible for by $\mathcal{L}_i^1$, the set of inequality constraints by $\mathcal{L}_i^2$ and the set of constraints originating from the 1-norm by $\mathcal{R}_i$. This division is obviously not unique but all constraints should be assigned to one computational unit. Further for $l \in \mathcal{L}_i^1$ and $l \in \mathcal{L}_i^2$ we require that $a_{li} \neq 0$ and for $r \in \mathcal{R}_i$ that $P_{ri} \neq 0$. Now we are ready to define two sets of neighbors to computational unit i

$$\begin{aligned} \mathcal{N}_i &= \big\{j \in \{1, \dots, M\} \mid \exists l \in \mathcal{L}_i^1 \text{ s.t. } a_{lj} \neq 0 \\ &\quad \text{or } \exists l \in \mathcal{L}_i^2 \text{ s.t. } a_{lj} \neq 0 \text{ or } \exists r \in \mathcal{R}_i \text{ s.t. } P_{rj} \neq 0\big\}, \\ \mathcal{M}_i &= \big\{j \in \{1, \dots, M\} \mid \exists l \in \mathcal{L}_j^1 \text{ s.t. } a_{li} \neq 0 \\ &\quad \text{or } \exists l \in \mathcal{L}_j^2 \text{ s.t. } a_{li} \neq 0 \text{ or } \exists r \in \mathcal{R}_j \text{ s.t. } P_{ri} \neq 0\big\}. \end{aligned}$$

Through the introduction of these sets, the constraints that are assigned to unit i can equivalently be written as

$$a_l^T x = b_l \Leftrightarrow \sum_{j\in\mathcal{N}_i} a_{lj}^T x_j = b_l, \quad l \in \mathcal{L}_i^1 \tag{4}$$

$$a_l^T x \le b_l \Leftrightarrow \sum_{j\in\mathcal{N}_i} a_{lj}^T x_j \le b_l, \quad l \in \mathcal{L}_i^2 \tag{5}$$

and the 1-norm term can equivalently be written as

$$|P_r^T x - p_r| = \left|\sum_{j\in\mathcal{N}_i} P_{rj}^T x_j - p_r\right|, \quad r \in \mathcal{R}_i. \tag{6}$$

In the following section, the dual function to be maximized is introduced. First, we state an assumption that will be useful in the continuation of the paper.

Assumption 1. We assume that there exists a vector $\bar{x}$ such that $A_1\bar{x} = b_1$ and $A_2\bar{x} < b_2$. Further, we assume that a_l, $l = 1, \dots, q$ and P_r, $r = 1, \dots, m$ are linearly independent.

Remark 2. Assumption 1 is known as the Mangasarian–Fromovitz constraint qualification (MFCQ). In Gauvin (1977) it was shown that MFCQ is equivalent to the set of optimal dual variables being bounded. For convex problems, MFCQ is equivalent to Slater's constraint qualification with the additional requirement that the vectors defining the equality constraints should be linearly independent.

3. Dual problem

In this section we introduce a dual problem to (3) from which the primal solution can be obtained. We show that this dual problem has the properties required to apply accelerated gradient methods.