Quantization design for distributed optimization with time-varying parameters

Ye Pu, Melanie N. Zeilinger and Colin N. Jones

***Abstract*— We consider the problem of solving a sequence of distributed optimization problems with time-varying parameters and communication constraints, i.e. only neighbour-to-neighbour communication and a limited amount of information exchanged. By extending previous results and employing a warm-starting strategy, we propose a on-line algorithm to solve the optimization problems under the given constraints and show that there exists a trade-off between the number of iterations for solving each problem in the sequence and the accuracy achieved by the algorithm. For a given accuracy ϵ, we can find a number of iterations K, which guarantees that for each step of the sequence the sub-optimal solution given by the algorithm satisfies the accuracy. We apply the method to solve a distributed model predictive control problem by considering the state measurement at each sampling time as the time-varying parameter and show that the simulation supports the theoretical results.**

I. Introduction

The problem of solving a sequence of distributed optimization problems with slowly varying parameters is central to many engineering problems, e.g. on-line resource allocation, distributed estimation and distributed optimal control problems. The main challenge for the distributed algorithms is how to solve a global optimization problem in a distributed fashion subject to communication constraints, i.e. only neighbour-to-neighbour communication and a limited amount of information exchanged, while providing high efficiency and using less computation and communication to achieve a given accuracy.

Inexact distributed optimization methods are attracting increasing attention, since these techniques have the potential to deal with errors, for instance the inexact solution of local problems, as well as noise caused by unreliable or limited communication such as transmission failures and quantization errors. Previous work has aimed at addressing the questions of how such errors affect the algorithm and under what conditions the convergence of the distributed algorithms can be guaranteed. In [5], an inexact decomposition algorithm for solving distributed optimization problems was proposed, which employs smoothing techniques and an excessive gap condition. In our previous work [10], we have proposed an inexact splitting method, named the inexact fast alternating minimization algorithm, and have applied it to distributed optimization problems, where local computation errors as well as errors resulting from limited communication were allowed, and convergence conditions on the errors were derived based on a complexity upper-bound. Some other related references for inexact optimization algorithms include [4], [7] and [12].

In this paper, we consider a sequence of distributed optimization problems parameterized by a slowly varying parameter. For each distributed problem, there are M subproblems with local cost functions that involves local and neighbouring variables, local varying parameters, and local constraints. We consider the following two challenges: 1. to solve each problem in a distributed manner with only local communication, i.e. between neighbouring sub-systems, and with limited communication bandwidth, where at each iteration only a limited number of bits can be transmitted; 2. to optimize the problems within some particular accuracy requirement sequentially and efficiently, i.e. to use less computation and communication to achieve the given accuracy. In order to meet the limited communication bandwidth, the information exchanged between the neighbouring subsystems is quantized. The quantization process results in inexact iterations throughout the distributed optimization algorithm, which effects its convergence. Related work includes [1], [6], [13] and [8], which study the effects of quantization on the performance of averaging or distributed optimization algorithms. Regarding the second challenge, previous work has shown that for on-line parameterized optimization there exists a trade-off between the solution accuracy and the complexity, i.e. the cost of computation and communication, represented by the the number of iterations K of the algorithms. The related work includes [14], where the authors considered a framework for sequentially solving stochastic optimization problems and presented the relationship between the accuracy and the complexity of the algorithm.

We propose an optimization method with a progressive quantization scheme to solve the distributed optimization problems sequentially. The idea is to extend the algorithm in [11] and [9] to a quantization design for parametric distributed optimization. By employing a warm-starting strategy, we improve the performance of the algorithm and show that there exists a trade-off between the accuracy and the number of iterations K. In particular, the paper makes the following main contributions:

Y. Pu and C.N. Jones are with the Automatic Control Lab, École Polytechnique Fédérale de Lausanne, EPFL-STI-IGM-LA Station 9 CH-1015 Lausanne, Switzerland, e-mail: {y.pu,colin.jones}@epfl.ch.

M.N. Zeilinger is with the Empirical Inference Department, Max Planck Institute for Intelligent Systems, 72076 Tübingen, Germany, e-mail: melanie.zeilinger@tuebingen.mpg.de.

This work has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP/2007-2013)/ ERC Grant Agreement n. 307608. The research of M. N. Zeilinger has received funding from the EU FP7 under grant agreement no. PIOF-GA-2011-301436-"COGENT".

- We extend the progressive quantization design for distributed optimization in [11] and [9] to the problem of optimizing a sequence of distributed problems with time-varying parameters and present the conditions on the quantizers, which guarantee that for all steps the values exchanged in the network always fall inside the quantization intervals and the quantization errors decrease linearly.
- By employing a warm-starting strategy, we improve the convergence speed of the algorithm and present a relationship between the solution accuracy and the cost of computation and communication represented by the number of iterations K. We show that for a given accuracy ϵ, there always exists a K guaranteeing that the sub-optimality of each solution in the sequence is upper-bounded by ϵ.
- We demonstrate the proposed method for solving a distributed model predictive control problem by considering the initial state measurement at each sampling time as the varying parameters and compare the simulation results with the theoretical bound.

II. Preliminaries

A. Notation

Let $v \in \mathbb{R}^{n_v}$ be a vector. $\|v\|$ denotes the l_2. Let $\mathbb{C}$ be a subset of $\mathbb{R}^{n_v}$. The projection of any point $v \in \mathbb{R}^{n_v}$ onto the set $\mathbb{C}$ is denoted by $\text{Proj}_{\mathbb{C}}(v) := \text{argmin}_{\mu \in \mathbb{C}} \|\mu - v\|$. Let $f : \Theta \to \Omega$ be a strongly convex function; σ_f denotes the convexity modulus $f(v) \geq f(\mu) + \langle \partial f(\mu), v - \mu \rangle + \frac{\sigma_f}{2}\|v - \mu\|^2$ for any $v, \mu \in \Theta$, where $\partial f(\cdot)$ denotes the set of sub-gradients of the function f at a given point. $L(f)$ denotes a Lipschitz constant of the function f, i.e. $\|f(v) - f(\mu)\| \leq L(f)\|v - \mu\|$, $\forall v, \mu \in \Theta$.

B. Uniform quantizer

Let x be a real number. A uniform quantizer with a quantization step-size Δ and the mid-value $\bar{x}$ can be expressed as

$$Q(x) = \bar{x} + \text{sgn}(x - \bar{x}) \cdot \Delta \cdot \left\lfloor \frac{\|x - \bar{x}\|}{\Delta} + \frac{1}{2} \right\rfloor \quad , \quad (1)$$

where $\text{sgn}(\cdot)$ is the sign function. The parameter Δ is equal to $\Delta = \frac{l}{2^n}$, where l represents the size of the quantization interval and n is the number of bits sent by the quantizer. In this paper, we assume that n is a fixed number, which means that the quantization interval is set to be $[\bar{x} - \frac{l}{2}, \bar{x} + \frac{l}{2}]$. The quantization error is upper-bounded by

$$x - Q(x) \leq \frac{\Delta}{2} = \frac{l}{2^{n+1}} \quad . \quad (2)$$

For the case that the input of the quantizer and the mid-value are not real numbers, but vectors of dimension n_x, the quantizer Q is composed of n_x independent scalar quantizers in (1) with the same quantization interval l and corresponding mid-value. In this paper, we design a uniform quantizer denoted as $Q^k(\cdot)$ with changing quantization interval l^k and mid-value $\bar{x}^k$ at every iteration k of the optimization algorithm.

C. Parametric distributed optimization problem

In this paper, we consider a parametric distributed optimization problem on a network of M sub-systems (nodes). The sub-systems communicate according to a fixed undirected graph $G = (\mathcal{V}, \mathcal{E})$. The vertex set $\mathcal{V} = \{1, 2, \cdots, M\}$ represents the sub-systems and the edge set $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ specifies pairs of sub-systems that can communicate. If $(i, j) \in \mathcal{E}$, we say that sub-systems i and j are neighbours, and we denote by $\mathcal{N}_i = \{j | (i, j) \in \mathcal{E}\}$ the set of the neighbours of sub-system i. Note that $\mathcal{N}_i$ includes i. We denote d as the degree of G. The optimization variable of sub-system i and the global variable are denoted by x_i and $x = [x_1^T, \cdots, x_M^T]^T$, respectively. For each sub-system i, the local variable has a local constraint $x_i \in \mathbb{C}_i \subseteq \mathbb{R}^{n_{m_i}}$. The constraint on the global variable x is denoted by $\mathbb{C} = \prod_{1 \leq i \leq M} \mathbb{C}_i$. The dimension of the local variable x_i is denoted by m_i and the maximum dimension of the local variables is denoted by $\bar{m}$, i.e. $\bar{m} := \max_{1 \leq i \leq M} m_i$. The concatenation of the variable of sub-system i and the variables of its neighbours is denoted by $x_{\mathcal{N}_i}$, and the corresponding constraint on $x_{\mathcal{N}_i}$ is denoted by $\mathbb{C}_{\mathcal{N}_i} = \prod_{j \in \mathcal{N}_i} \mathbb{C}_j$. With the selecting matrices E_i and F_{ji}, they can be represented as $x_{\mathcal{N}_i} = E_i x$ and $x_i = F_{ji} x_{\mathcal{N}_j}$, $j \in \mathcal{N}_i$, which implies the relation between the local variable x_i and the global variable x, i.e. $x_i = F_{ji} E_j x$, $j \in \mathcal{N}_i$. Note that E_i and F_{ji} are selecting matrices, and therefore $\|E_i\| = \|F_{ji}\| = 1$. The parametric distributed optimization problem is given in Problem II.1.

Problem II.1.

$$\begin{aligned} \min_{x,\, x_{\mathcal{N}_i}} \quad & f(x, \eta^t) = \sum_{i=1}^{M} f_i(x_{\mathcal{N}_i}, \eta_i^t) \\ s.t. \quad & x_i \in \mathbb{C}_i \ , \ x_i = F_{ji} x_{\mathcal{N}_j} \ , \ j \in \mathcal{N}_i \ , \\ & x_{\mathcal{N}_i} = E_i x \ , \ i = 1, 2, \cdots, M \ . \end{aligned}$$

The vector $\eta_i^t \in \Xi_i \subseteq \mathbb{R}^{n_{\eta_i}}$ is a time-dependent parameter, for $i = 1, 2, \cdots, M$. We denote $\eta^t = [\eta_1^{t^T}, \cdots, \eta_M^{t^T}]$.

Assumption II.2. *We assume that for all $\eta_i^t \in \Xi_i$ the global cost function $f(\cdot)$ is strongly convex with a convexity modulus σ_f and Lipschitz continuous gradient with a Lipschitz constant L, i.e. $\|\nabla f(x_1) - \nabla f(x_2)\| \leq L\|x_1 - x_2\|$ for any x_1 and x_2.*

Assumption II.3. *We assume that for all $t \leq 0$ every local cost function $f_i(\cdot)$ has Lipschitz continuous gradient with a Lipschitz constant L_i, and denote L_{max} as the maximum Lipschitz constant of the local functions, i.e. $L_{max} := \max_{1 \leq i \leq M} L_i$.*

Assumption II.4. *The local constraint $\mathbb{C}_i$ is a convex set, for $i = 1, \cdots, M$.*

Model predictive control is one application resulting in a parametric optimization problem, which generally satisfies Assumption II.2, Assumption II.3 and Assumption II.4. This will be discussed in more detail in the example in Section IV.

D. Distributed optimization with limited communication

In [11] and [9], a distributed optimization algorithms with progressive quantization design based on the inexact proximal gradient method in [12] has been proposed. The results address the challenge that the communication in the distributed optimization algorithms is limited and the information exchanged in the network needs to be quantized. The proposed progressive quantizer with changing parameters satisfies the communication limitations, while ensuring that the errors induced by quantization satisfy the conditions for convergence.

Algorithm 1 presents the distributed algorithm with the progressive quantization design for Problem II.1 with a fixed parameter η^t. For every sub-system i, there are two uniform quantizers $Q_{\alpha,i}^{t,k}$ and $Q_{\beta,i}^{t,k}$ using the formulation introduced in (1) with a fixed number of bits n, changing quantization intervals $l_{\alpha,i}^{t,k}$ and $l_{\beta,i}^{t,k}$ and changing mid-values $\bar{x}_{\alpha,i}^{t,k}$ and $\bar{\nabla} f_{\beta,i}^k$ for transmitting $x_i^{t,k}$, and ∇f_i^k at every iteration k. At iteration k, the quantization intervals are set to be $l_{\alpha,i}^{t,k} = C_\alpha^t \kappa^k$ and $l_{\beta,i}^{t,k} = C_\beta^t \kappa^k$, and the mid-values are set to be the previous quantized values $\bar{x}_{\alpha,i}^{t,k} = \hat{x}_i^{t,k-1}$ and $\bar{\nabla} f_{\beta,i}^k = \hat{\nabla} f_i^{k-1}$. The two parameters $C_\alpha^t = l_{\alpha,i}^{t,0}$ and $C_\beta^t = l_{\beta,i}^{t,0}$ denote the initial quantization intervals.

In this paper, $\hat{\cdot}$ is used to denote a quantized value, e.g. $\hat{x}_i^{t,k} = Q_{\alpha,i}^{t,k}(x_i^{t,k})$ and $\tilde{\cdot}$ is used to denote a re-projected value, e.g. $\tilde{x}_{\mathcal{N}_i}^{t,k} = \text{Proj}_{\mathbb{C}_{\mathcal{N}_i}}(\hat{x}_{\mathcal{N}_i}^{t,k})$. The quantization errors are denoted by $\alpha_i^{t,k} = \hat{x}_i^{t,k} - x_i^{t,k}$ and $\beta_i^{t,k} = \hat{\nabla} f_i^k - \nabla f_i^k$.

Algorithm 1 Distributed algorithm with quantization refinement

Require: Initialize $\hat{x}_i^{t,-1} = x_i^0(\eta^t)$, $\hat{\nabla} f_i^{-1} = \nabla f_i(\text{Proj}_{\mathbb{C}_{\mathcal{N}_i}}(x_{\mathcal{N}_i}^0(\eta^t)))$, $(1-\gamma) < \kappa < 1$, $\tau < \frac{1}{L}$, the initial quantization intervals C_α^t and C_β^t and the number of iterations K.

for $k = 0, 1, 2, \cdots, K$ **do**

For sub-system i, $i = 1, 2, \cdots, M$:

1: Update the parameters of quantizer $Q_{\alpha,i}^{t,k}$: $l_{\alpha,i}^{t,k} = C_\alpha^t \kappa^k$ and $\bar{x}_{\alpha,i}^{t,k} = \hat{x}_i^{t,k-1}$

2: Quantize the state: $\hat{x}_i^{t,k} = Q_{\alpha,i}^{t,k}(x_i^{t,k}) = x_i^{t,k} + \alpha_i^{t,k}$.

3: Send $\hat{x}_i^{t,k}$ to all the neighbours of sub-system i

4: Compute the projection of $\hat{x}_{\mathcal{N}_i}^{t,k}$: $\tilde{x}_{\mathcal{N}_i}^{t,k} = \text{Proj}_{\mathbb{C}_{\mathcal{N}_i}}(\hat{x}_{\mathcal{N}_i}^{t,k})$

5: Compute $\nabla f_i^k = \nabla f_i(\tilde{x}_{\mathcal{N}_i}^{t,k})$

6: Update the parameters of quantizer $Q_{\beta,i}^{t,k}$: $l_{\beta,i}^{t,k} = C_\beta^t \kappa^k$ and $\bar{\nabla} f_{\beta,i}^k = \hat{\nabla} f_i^{k-1}$

7: Quantize the gradient: $\hat{\nabla} f_i^k = Q_{\beta,i}^{t,k}(\nabla f_i^k) = \nabla f_i^k + \beta_i^{t,k}$.

8: Send $\hat{\nabla} f_i^k$ to all the neighbours of sub-system i

9: Update the state:$x_i^{t,k+1} = \text{Proj}_{\mathbb{C}_i}(x_i^{t,k} - \tau \sum_{j \in \mathcal{N}_i} F_{ji} \hat{\nabla} f_j^k)$

end for

In [11], we consider an unconstrained optimization problem, and Theorem 3.10 in [11] states the convergence results for the unconstrained case, whereas [9] provides an extension to the constrained case. The following result can be found in [9].

Assumption II.5. *Consider the quantizers $Q_{\alpha,i}^{t,k}$ and $Q_{\beta,i}^{t,k}$ in Algorithm 1. We assume that the parameters of the quantizers, i.e. the number of bits n and the initial quantization intervals C_α^t and C_β^t satisfy*

$$a_1 \cdot \|x^0(\eta^t) - x^\star(\eta^t)\| + a_2 \frac{C_\alpha^t}{2^{n+1}} + a_3 \frac{C_\beta^t}{2^{n+1}} \leq \frac{C_\alpha^t}{2} \quad (3)$$

$$b_1 \cdot \|x^0(\eta^t) - x^\star(\eta^t)\| + b_2 \frac{C_\alpha^t}{2^{n+1}} + b_3 \frac{C_\beta^t}{2^{n+1}} \leq \frac{C_\beta^t}{2} \ . \quad (4)$$

Remark II.6. *The parameters of the quantizers n, C_α^t and C_β^t are all positive constants. Assumption II.5 can always be satisfied by increasing n, C_α^t and C_β^t.*

Theorem II.7. *For any $t \geq 0$, if Assumptions II.2, II.3 and II.5 hold and $(1-\gamma) < \kappa < 1$, then for $0 \leq k \leq K$ the sequence $\{x^{t,k+1}\}$ generated by Algorithm 1 converges to the optimum linearly with the constant κ and satisfies*

$$\|x^{t,k+1} - x^\star(\eta^t)\| \leq \kappa^{k+1} \Big[\|x^0(\eta^t) - x^\star(\eta^t)\| + \frac{(C_1^t + \sqrt{2L} C_2^t)\kappa}{L(\kappa + \gamma - 1)(1-\gamma)} \Big] \ .$$

with $C_1^t = \frac{M\sqrt{m}(L_{max} d C_\alpha^t + \sqrt{d} C_\beta^t)}{2^{n+1}}$ and $C_2^t = \frac{\sqrt{2}}{2} \cdot \frac{M\sqrt{m} C_\alpha^t}{2^{n+1}}$.

Theorem II.7 states that with the proposed quantization design, the linear convergence of the algorithm is preserved, but the constant of the convergence rate has to be enlarged from $1-\gamma$ to κ in order to compensate for the deficiencies arising from limited communication.

III. Parametric distributed optimization with limited communication

We extend Algorithm 1 and the results in [11] to solve the parametric distributed optimization Problem II.1. By employing a warm-starting strategy to initialize the starting sequence for t, i.e. $x^0(\eta^t) = x^K(\eta^{t-1})$, we show that there exists a relationship between the number of iterations K and the accuracy ϵ. For a given ϵ, we can always find a K guaranteeing that for all $t \geq 0$ the sub-optimal solution $x^K(\eta^t)$ satisfies the accuracy ϵ, i.e. $\|x^K(\eta^t) - x^\star(\eta^t)\| \leq \epsilon$. The distributed optimization algorithm with quantization refinement for the parametric optimization problem is presented in Algorithm 2.

Assumption III.1. *We assume that the optimal solution satisfies*

$$\|x^\star(\eta^t) - x^\star(\eta^{t+1})\| \leq \rho \quad (5)$$

for all $t \geq 0$.

Assumption III.2. *We assume that at the first step $t = 0$ the initial solution of the algorithm is a sub-optimal solution satisfying*

$$\|x^0(\eta^0) - x^\star(\eta^0)\| \leq \epsilon \ . \quad (6)$$

Remark III.3. *A sub-optimal solution $x^0(\eta^0)$ satisfying Assumption III.2 can be computed off-line.*

$$a_1 = \frac{(\kappa+1)}{\kappa} \quad , \; a_2 = \frac{M\sqrt{\bar{m}}\kappa(\kappa+1)(dL_{\max}+\sqrt{L}) + M\sqrt{\bar{m}}L(\kappa+\gamma-1)(1-\gamma)}{L\kappa(\kappa+\gamma-1)(1-\gamma)} \;,$$
$$a_3 = \frac{M\sqrt{d\bar{m}}(\kappa+1)}{L(\kappa+\gamma-1)(1-\gamma)} \quad , \; b_3 = \frac{L_{\max}M\sqrt{d\bar{m}}\kappa(\kappa+1) + L\sqrt{d\bar{m}}(\kappa+\gamma-1)(1-\gamma)}{L\kappa(\kappa+\gamma-1)(1-\gamma)} \;,$$
$$b_1 = \frac{L_{\max}(\kappa+1)}{\kappa} \quad , \; b_2 = \frac{L_{\max}M\sqrt{\bar{m}}\kappa(\kappa+1)(dL_{max}+\sqrt{L}) + L_{max}d\sqrt{\bar{m}}L(\kappa+1)(\kappa+\gamma-1)(1-\gamma)}{L\kappa(\kappa+\gamma-1)(1-\gamma)} \;.$$

Assumption III.4. *We assume that the two parameters C_α and C_β in Algorithms 2 satisfy*

$$a_1\cdot(\epsilon+\rho) + a_2\frac{C_\alpha}{2^{n+1}} + a_3\frac{C_\beta}{2^{n+1}} \le \frac{C_\alpha}{2} \tag{7}$$

$$b_1\cdot(\epsilon+\rho) + b_2\frac{C_\alpha}{2^{n+1}} + b_3\frac{C_\beta}{2^{n+1}} \le \frac{C_\beta}{2} \;. \tag{8}$$

Remark III.5. *We want to point out that the two conditions on the initial quantization intervals C_α^t and C_β^t in Assumption III.4 do not vary with t, i.e. they are independent from the parameters η^t. Therefore, we can compute the initial intervals C_α and C_β satisfying (7) and (8) off-line and set C_α^t and C_β^t to the same values for all $t \ge 0$.*

Algorithm 2 Parametric distributed algorithm with quantization refinement

Require: An initial solution $x^0(\eta^0)$ and the number of iterations K

for $t = 0, 1, 2, \cdots$ **do**

1: Initialize quantization intervals $C_\alpha^t = C_\alpha$ and $C_\beta^t = C_\beta$.

2: Solve Problem II.1 with the parameter η^t by Algorithm 1 with the initial solution $x^0(\eta^t)$ and the number of iterations K.

3: $x^0(\eta^{t+1}) \leftarrow x^{K+1}(\eta^t)$

end for

Remark III.6. *In Algorithm 2, we use a warm-starting strategy to solve the optimization problem at each step $t \ge 1$, i.e. in Step 3 we initialize the solution with $x^0(\eta^{t+1}) \leftarrow x^{K+1}(\eta^t)$.*

Theorem III.7. *For a given $\epsilon > 0$, if Assumptions II.2, II.3, III.1, III.2 and III.4 hold, $(1-\gamma) < \kappa < 1$ and the number of iterations K satisfies*

$$K \ge \left\lceil \log_\kappa \frac{\epsilon(1-\kappa)}{\rho+\delta+(1-\kappa)(\epsilon+\delta)} \right\rceil - 1 \;, \tag{9}$$

with $\delta = \frac{\kappa(C_1+\sqrt{2L}C_2)}{L(\kappa+\gamma-1)(1-\gamma)}$, $C_1 = \frac{M\sqrt{\bar{m}}(L_{max}dC_\alpha+\sqrt{d}C_\beta)}{2^{n+1}}$ and $C_2 = \frac{\sqrt{2}}{2}\cdot\frac{M\sqrt{\bar{m}}C_\alpha}{2^{n+1}}$. Then the sub-optimality of the solution $x^{K+1}(\eta^t)$ satisfies:

$$\|x^{K+1}(\eta^t) - x^\star(\eta^t)\| \le \epsilon \;, \tag{10}$$

for all $t \ge 0$.

Proof: We will prove Theorem III.7 by induction.

- Base case: At $t = 0$, Assumption III.2 and Assumption III.4 imply that Assumption II.5 holds. Then all assumptions required by Theorem II.7 are satisfied and it follows that $\|x^{K+1}(\eta^0) - x^\star(\eta^0)\| \le \kappa^{K+1}(\|x^0(\eta^0) - x^\star(\eta^0)\| + \delta) \le \kappa^{K+1}(\epsilon+\delta)$. Using the condition in (9), we get $\kappa^{K+1}(\epsilon+\delta) \le \epsilon$. Hence, it holds that $\|x^{K+1}(\eta^0) - x^\star(\eta^0)\| \le \epsilon$.
- Induction step: Let $g \ge 0$ be given and suppose that $\|x^{K+1}(\eta^t) - x^\star(\eta^t)\| \le \epsilon$ for $t \le g$. We will prove that $\|x^{K+1}(\eta^{g+1}) - x^\star(\eta^{g+1})\| \le \epsilon$. By the warm-starting step in Step 3 in Algorithm 2, we know

$$\begin{aligned}\|x^0(\eta^{g+1}) - x^\star(\eta^{g+1})\| =& \|x^{K+1}(\eta^g) - x^\star(\eta^{g+1})\| \\ \le& \|x^{K+1}(\eta^g) - x^\star(\eta^g)\| \\ &+ \|x^\star(\eta^g) - x^\star(\eta^{g+1})\| \;.\end{aligned}$$

By the assumption of induction and Assumption III.1 , we obtain

$$\|x^0(\eta^{g+1}) - x^\star(\eta^{g+1})\| \le \epsilon + \rho \;.$$

Then Assumption III.4 implies that Assumption II.5 holds for $g+1$. It follows from Theorem II.7 that

$$\begin{aligned}&\|x^{K+1}(\eta^{g+1}) - x^\star(\eta^{g+1})\| \\ &\le \kappa^{K+1}(\|x^0(\eta^{g+1}) - x^\star(\eta^{g+1})\| + \delta) \;.\end{aligned}$$

By the warm-starting step in Step 3 in Algorithm 2, the above is upper-bounded by

$$\begin{aligned}\le& \kappa^{K+1}(\|x^{K+1}(\eta^g) - x^\star(\eta^{g+1})\| + \delta) \\ \le& \kappa^{K+1}(\|x^{K+1}(\eta^g) - x^\star(\eta^g)\| \\ &+ \|x^\star(\eta^g) - x^\star(\eta^{g+1})\| + \delta) \;,\end{aligned}$$

Assumption III.1 implies

$$\le \kappa^{K+1}(\|x^{K+1}(\eta^g) - x^\star(\eta^g)\| + \rho + \delta) \;.$$

Again by the warm-starting step, Assumption III.4 implies Assumption II.5. It follows from Theorem II.7 that the above is upper-bounded by

$$\begin{aligned}\le& \kappa^{K+1}(\rho+\delta) + (\kappa^{K+1})^2\cdot(\|x^0(\eta^g) - x^\star(\eta^g)\| + \delta) \\ \le& \kappa^{K+1}(\rho+\delta) \\ &+ (\kappa^{K+1})^2\cdot(\|x^{K+1}(\eta^{g-1}) - x^\star(\eta^{g-1})\| + \rho + \delta) \;.\end{aligned}$$

Sequentially, we get that the above is upper-bounded by

$$\le \sum_{p=1}^{g+1}(\kappa^{K+1})^p\cdot(\rho+\delta) + (\kappa^{K+1})^{g+2}\cdot(\epsilon+\delta) \;.$$

By Assumption III.2 and the property of geometric series, we have

$$\leq(\rho+\delta)\cdot\kappa^{K+1}\cdot\frac{1-(\kappa^{K+1})^{g+1}}{1-\kappa^{K+1}} + (\epsilon+\delta)\cdot(\kappa^{K+1})^{g+2} \ .$$

Using the fact that $0<\kappa<1$, we get

$$\|x^{K+1}(\eta^{g+1})-x^\star(\eta^{g+1})\| \leq \frac{\rho+\delta}{1-\kappa}\cdot\kappa^{K+1} + (\epsilon+\delta)\cdot\kappa^{K+1} \ .$$

Note that the inequality above holds for all $t\geq 0$. By the condition in (9), we conclude that

$$\|x^{K+1}(\eta^{g+1})-x^\star(\eta^{g+1})\|\leq\epsilon \ .$$

We conclude that by the principle of induction for all $t\geq 0$ the solution $x^{K+1}(\eta^t)$ satisfies $\|x^{K+1}(\eta^t)-x^\star(\eta^t)\|\leq\epsilon$. ■

IV. Numerical Example

This section illustrates the theoretical findings of the paper and demonstrates the performance of Algorithm 2. We consider a parametric distributed quadratic programming (QP) problem originating from the problem of regulating constrained distributed linear systems by model predictive control (MPC) in the form of Problem IV.1, where the initial state $\bar{z}_i^t$ is the time-varying parameter. For more information about distributed MPC, see e.g. [3], [2] and [10].

Problem IV.1.

$$\begin{aligned}\min_{z,u}\quad & \sum_{i=1}^{M}\sum_{g=0}^{N-1} l_i(z_i(g),u_i(g))+\sum_{i=1}^{M} l_i^f(z_i(N))\\ s.t.\quad & z_i(g+1)=A_{ii}z_i(g)+\sum_{j\in\mathcal{N}_i}B_{ij}u_j(g)\\ & u_i(g)\in\mathbb{U}_i \ , \ z_i(0)=\bar{z}_i^t \ , \ i=1,2,\cdots,M \ ,\end{aligned}$$

where M and N denote the number of subsystems and the horizon of the MPC problem, respectively. $\mathcal{N}_i$ denotes the set of the neighbours of subsystem i. The state and input sequences along the horizon of subsystem i are denoted by $z_i = [z_i^T(0), z_i^T(1), \cdots, z_i^T(N)]^T$ and $u_i = [u_i^T(0), u_i^T(1), \cdots, u_i^T(N-1)]^T$. The discrete-time linear dynamics of subsystem i are given by $z_i(g+1)=A_{ii}z_i(g)+\sum_{j\in\mathcal{N}_i}B_{ij}u_j(g)$, where A_{ii} and B_{ij} are the dynamic matrices. The initial state is a time-varying parameter denoted by $\bar{z}^t=[\bar{z}_1^{t^T},\bar{z}_2^{t^T},\cdots,\bar{z}_M^{t^T}]^t$. The control inputs of subsystem i are subject to local convex constraints $u_i(t)\in\mathbb{U}_i$. $l_i(\cdot,\cdot)$ and $l_i^f(\cdot)$ are strictly convex stage cost functions. From Problem IV.1, we can see that subsystem i is coupled with its neighbours in the linear dynamics $z_i(g+1)=A_{ii}z_i(g)+\sum_{j\in\mathcal{N}_i}B_{ij}u_j(g)$.

We randomly generate a distributed MPC problem in Problem IV.1. We first randomly generate a connected network with $M=40$ sub-systems. Each sub-system has three states and two inputs. The dynamical matrices A_{ii} and B_{ij} are randomly generated, i.e. generally dense, and the local systems are controllable and unstable. The input constraint $\mathbb{U}_i$ for sub-system i is set to $\mathbb{U}_i=\{u_i|-0.4\cdot\mathbf{1}\leq u_i(g)\leq 0.3\cdot\mathbf{1}\}$, where $\mathbf{1}$ denotes the all-ones vector with the same dimension as u_i. The horizon of the MPC problem is set to $N=11$. The local stage cost functions are chosen as quadratic functions $l_i(z_i(g),u_i(g))=z_i^T(g)Q_iz_i(g)+u_i^T(g)R_iu_i(g)$ and $l_i^f(z_i(N))=z_i^T(N)P_iz_i(N)$, where Q_i, R_i and P_i are identity matrices.

By eliminating all state variables, distributed MPC problems of this class can be reformulated as a parametric distributed QP of the form in Problem IV.2 $x_i=u_i$, $x_{\mathcal{N}_i}$, and the parameter $\bar{z}_i^t$. The matrix H_i is a dense and positive definite matrix and the constraint $\mathbb{C}_i=\mathbb{U}_i^N$ is a polytopic set.

Problem IV.2.

$$\begin{aligned}\min_{x\in\mathbb{R}^{n_x}}\ f(x,\bar{z}^t)&=\sum_{i=1}^{M}f_i(x_{\mathcal{N}_i})\\ &=\sum_{i=1}^{M}x_{\mathcal{N}_i}^TH_ix_{\mathcal{N}_i}+\bar{z}_{\mathcal{N}_i}^{t^T}h_ix_{\mathcal{N}_i}\\ s.t.\quad x_i&\in\mathbb{C}_i \quad .\end{aligned}$$

Note that the matrix H_i does not vary for different t, and the parameter appears in the linear term $\bar{z}_{\mathcal{N}_i}^{t^T}h_ix_{\mathcal{N}_i}$.

For Problem IV.2, the constants in Algorithm 2 are $\gamma=\frac{\sigma_f}{L}=0.1027$, the decrease rates of the quantization intervals $1-\gamma\leq\kappa=0.9692$, and the minimum number of bits required for convergence $n_{min}=13$.

In the simulation of Fig. 1, we set the number of steps to $t=50$ and the number of iterations K to 2, 10 and 30. The parameter $\bar{z}^t$ is randomly generated and satisfies $\|\bar{z}^t-\bar{z}^{t+1}\|\leq 3$. Fig. 1 shows the accuracy achieved by Algorithm 2 and Algorithm 2 without warm-starting strategy, i.e. setting $x^0(\eta^{t+1})=0$ (cold-starting) in Algorithm 2 for all $t\geq 0$. The curves show that warm-starting achieves significantly better accuracy for the same number of iterations.

In Fig. 2, we compute the average accuracy achieved by Algorithm 2 over all steps $t=50$ in Fig. 1 and calculate the corresponding number of iterations K satisfying the bound in Theorem III.7 for the average accuracy. Note that the parameter ρ in Assumption 5 is approximated by randomly sampling 500 initial states satisfying $\|\bar{z}^p-\bar{z}^{p+1}\|\leq 3$, for $1\leq p\leq 500$, and compute the largest $\rho=\max_{1\leq p\leq 500}\{\|x^\star(\bar{z}^p)-x^\star(\bar{z}^{p+1})\|\}$. We see that the bound gets tighter as the accuracy ϵ decreases.

References

[1] R. Carli, F. Fagnani, P. Frasca, T. Taylor, and R. Zampieri. Average consensus on networks with transmission noise or quantization. In *Proceedings of European Control Conference*, 2007.

[2] C. Conte, T. Summers, M.N. Zeilinger, M. Morari, and C.N. Jones. Computational aspects of distributed optimization in model predictive control. In *51th IEEE Conference on Decision and Control*, pages 6819–6824, 2012.

[3] C. Conte, N. R. Voellmy, M. N. Zeilinger, M. Morari, and C. N. Jones. Distributed synthesis and control of constrained linear systems. In *American Control Conference*, pages 6017–6022, 2012.

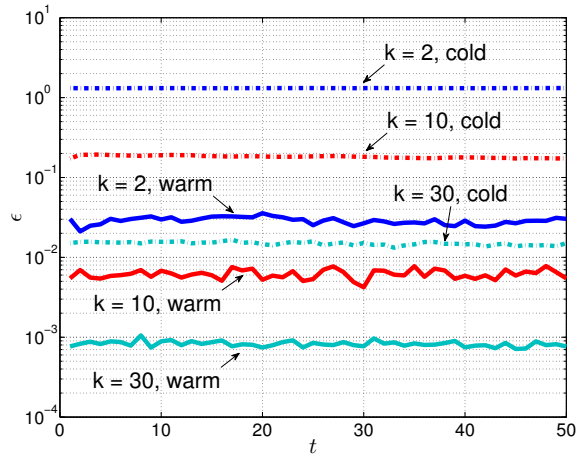

Fig. 1: Comparison of the accuracy achieved by Algorithm 2 to Algorithm 2 without warm starting for solving Problem IV.2 for different numbers of iterations K

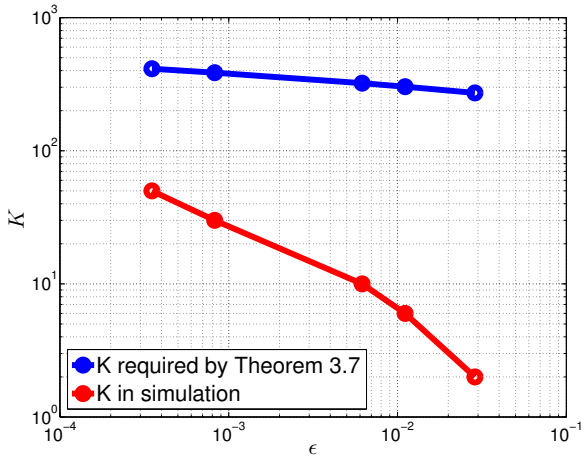

Fig. 2: Comparison of the number of iterations K required by Theorem III.7 and used in the simulation for solving Problem IV.2 for the same accuracy ϵ.

[4] O. Devolder, F. Glineur, and Y. Nesterov. First-order methods of smooth convex optimization with inexact oracle. *Mathematical Programming*, pages 1–39, 2013.

[5] Q. T. Dinh, I. Necoara, and M. Diehl. Fast inexact decomposition algorithms for large-scale separable convex optimization. *arXiv preprint arXiv:1212,4275*, 2012.

[6] A. Kashyap, T. Basar, and R. Srikant. Quantized consensus. *Automatica*, 43:1192–1203, 2007.

[7] V. Nedelcu, I. Necoara, and I. Dumitrache. Complexity of an inexact augmented lagrangian method: Application to constrained MPC. In *19th World Congress of the International Federation of Automatic Control*, 2014 to appear.

[8] A. Nedic, A. Olshevsky, A. Ozdaglar, and J.N. Tsitsiklis. Distributed subgradient methods and quantization effects. In *47th IEEE Conference on Decision and Control*, pages 4177–4184, 2008.

[9] Y. Pu, M.N. Zeilinger, and C. N. Jones. Quantization design for distributed optimization. *arXiv preprint*, 2015.

[10] Y. Pu, M.N. Zeilinger, and C. N. Jones. Inexact fast alternating minimization algorithm for distributed model predictive control. In *53th IEEE Conference on Decision and Control*, to appear 2014.

[11] Y. Pu, M.N. Zeilinger, and C. N. Jones. Quantization design for unconstrained distributed optimization. In *American Control Conference*, to appear, 2015.

[12] M. Schmidt, N. L. Roux, and F. Bach. Convergence rates of inexact proximal-gradient methods for convex optimization. In *25th Annual Conference on Neural Information Processing Systems*, pages 6819–6824, 2011.

[13] D. Thanou, E. Kokiopoulou, Y. Pu, and P. Frossard. Distributed average consensus with quantization refinement. *IEEE Transactions on Signal Processing*, 61:194–205, 2013.

[14] C. Wilson, V. Veeravalli, and A. Nedic. Dynamic stochastic optimization. In *IEEE 53rd Conference on Decision and Control*, pages 173–178, 2014.