Accelerated Distributed MPC of Linear Discrete-Time Systems With Coupled Constraints

Zheming Wang and Chong-Jin Ong, *Senior Member, IEEE*

***Abstract*—This paper proposes a distributed model predictive control (MPC) approach for a family of discrete-time linear systems with local (uncoupled) and global (coupled) constraints. The proposed approach is based on the dual problem of an overall MPC optimization problem involving all systems, which is then solved distributively using a modified distributed Nesterov-accelerated-gradient algorithm. To further reduce the computational requirement, this approach allows for early termination of the distributed gradient algorithm. This is made possible via a consensus algorithm that determines the satisfaction of the termination condition and by appropriate tightening of the coupled constraints. Under reasonable assumptions, the approach is able to produce a suboptimal solution as long as the network of the systems is connected while ensuring recursive feasibility and exponential stability of the closed-loop system. The performance of the proposed approach is demonstrated by a numerical example.**

***Index Terms*—Consensus algorithm, coupled constraints, distributed model predictive control, Nesterov's method.**

I. Introduction

THE problem of distributed model predictive control (DMPC) of M discrete-time linear dynamical systems is considered in this work. Each of the M system has the form

$$x^i(t+1) = A^i x^i(t) + B^i u^i(t) \tag{1}$$

$$x^i(t) \in X^i,\ u^i(t) \in U^i,\ i = 1, \ldots, M \tag{2}$$

and all of them have to satisfy a global constraint of the form

$$\sum_{i=1}^{M} \left(\Psi_x^i x^i(t) + \Psi_u^i u^i(t)\right) \leq \mathbf{1}_p, \quad \text{for all } t \tag{3}$$

where $X^i \subset \mathbb{R}^{n_i}, U^i \subset \mathbb{R}^{m_i}$ are constraint sets of states and controls of the ith system respectively, $\Psi_x^i \in \mathbb{R}^{p\times n_i}$ and $\Psi_u^i \in \mathbb{R}^{p\times m_i}$ are matrices that define the coupled constraints with $\mathbf{1}_p$ being the p-vector of all ones.

The study of DMPC has received considerable attention recently [1]–[5]. However, the literature on approaches for (1)–(3) is somewhat limited, possibly due to the presence of the global constraint. When (3) is present, a popular approach [6]–[9] is to optimize one system (or a group of systems) at a time while holding all others constant; this is followed sequentially by another system (another group) so that all M systems are optimized once at every M sampling periods. When one system (or a group of systems) is being optimized, the other systems (outside the group) will follow their predicted controls for the next time step. While these approaches are reasonable, the optimality of the overall system is unclear.

Approaches [10], [11] that address the optimality of the overall system rely on the dual problem of the Lagrangian function. In this setting, dual variables associated with (3) are needed and they are treated as consensus variables in a distributed consensus optimization problem. Typically, consensus of the dual variables is ensured (see [11, ch. 6]) using a central/master node and such an approach is adopted in many decomposition methods [12]–[15]. However, the need of a central node can be a strong requirement on the network. More recently, Wang and Ong [16], [17] solve the dual problem of DMPC with consensus variables in a distributed manner using the alternating direction multiplier method (ADMM). It avoids the use of a central node and allows for tolerable differences among the local copies of the consensus variables. However, the ADMM algorithm converges at a rate of $O(\frac{1}{k})$ where k is the number of iterations. Hence, the convergence can be slow, especially when high accuracy solutions are required.

This works proposes a faster convergence approach than that of [17]. It uses a modified distributed implementation of the standard stand-alone Nestrov gradient method [18], [19], which is known to have $O(\frac{1}{k^2})$ convergence. It requires more information exchanges than [17] and hence, is most useful when the systems are better connected. The use of Nestrov gradient method for MPC is not new. It has been used for the control of a single system [20], [21] as well as distributedly [22], [23] for standard one-off optimization problems. But it has never been used in a MPC setting where the optimization problem is solved repeatedly with considerations for both stability and performance.

Manuscript received March 6, 2017; revised March 6, 2017 and October 11, 2017; accepted January 15, 2018. Date of publication February 28, 2018; date of current version October 25, 2018. Recommended by Associate Editor L. Zhang. *(Corresponding author: Chong-Jin Ong.)*

The authors are with the Department of Mechanical Engineering, National University of Singapore, Singapore 117576 (e-mail: wang.zheming@u.nus.edu; mpeongcj@nus.edu.sg).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TAC.2018.2810511

There are also other nontrivial differences. The proposed update law for the consensus variable differs from those in [22], [23]. In [22], a constant step-size is used but such a choice may not converge to the optimal. The work of [23] solves an unconstrained optimization problem using an inner loop for better convergence of the consensus variables. However, the number of steps in the inner loop grows with the index of the outer loop, leading to a significant increase in the number of information exchanges. Moreover, its applicability in the presence of constraints is unclear. Similar to [23], this paper uses several steps in the inner-loop to achieve consensus of the local copies of the consensus variable. However, the number of consensus steps is fixed and does not grow with the outer-loop index. To further reduce the computational requirement, premature termination of the optimal solution at each iteration is allowed, a feature that is not found in [22] or [23]. Another important difference of this work to [17] is that the local copies of the dual variable reach exact consensus. This, together with premature termination allows the proposed approach to have a simplified stopping condition.

Convergence of the proposed approach under the premature termination condition, together with recursive feasibility and stability of the closed-loop system are proven. Consistent with the $O(\frac{1}{k^2})$ convergence rate of Nesterov method, the numerical examples show that the proposed approach converges faster than the approach of [17] for the same accuracy, with about 30% to 50% of the iterations needed by [17].

The rest of this paper is organized as follows. This section ends with a description of the notations used. Section II reviews some results of the standard stand-alone MPC for a single system and discusses the formulation of the overall MPC problem. This section is kept short as it is similar to the corresponding section in [17] as the underlying problem is the same. Section III presents the proposed approach, including the discussion of the coupled constraint, its dual and the convergence of the proposed distributed fast dual gradient algorithm. The recursive feasibility and stability results are given in Section V. The performance of the approach is illustrated by a numerical example in Section VI with the conclusions given in Section VII. For presentation of the main ideas, proofs of all Lemmas and Theorems are given in the appendix.

The notations used in this paper are standard. The symbols $\mathbb{R}^n, \mathbb{R}^{m\times n}, \mathbb{Z}_0^+$, and $\mathbb{Z}^+$ refer to the n-dimension Euclidean space, space of m by n matrices with real entries, nonnegative and positive integer sets, respectively. Let $\ell, h \in \mathbb{Z}_0^+$ with $h \geq \ell$, $\mathbb{Z}^\ell := \{1, 2, \ldots, \ell\}$ and $\mathbb{Z}_\ell^h := \{\ell, \ell+1, \ldots, h\}$. The identity matrix in n dimension is I_n, the n-column vector of all ones is $\mathbf{1}_n$ (subscript omitted when the dimension is clear) and the cardinality of an index set S is $|S|$. Given $\sigma > 0$, $X \subset \mathbb{R}^n$ containing 0 in its interior, $\sigma X = \{\sigma x : x \in X\}$. For a square matrix Q, $Q \succ (\succeq) 0$ means Q is positive definite (semi-definite). The ℓ_p-norm, $p = 1, 2, \infty$, of $x \in \mathbb{R}^n$ is $\|x\|_p$ while $\|x\|_Q^2 = x^T Q x$ for $Q \succ 0$. Given a matrix $H \in \mathbb{R}^{m\times n}$, its (i, j) element is denoted by $H_{i,j}$. Several representations of the states and controls are needed: $x^i(t)$, $u^i(t)$ refer to the state and control of the ith system at time t; x_k^i, u_k^i are the kth predicted state and control of the ith system; $x = (x^1, x^2, \ldots, x^M)$, $u = (u^1, u^2, \ldots, u^M)$ are the collections of x^i and u^i over the M systems; boldface $\boldsymbol{x}^i = (x_1^i, x_2^i, \ldots, x_N^i)$, $\boldsymbol{u}^i = (u_0^i, u_1^i, \ldots, u_{N-1}^i)$ are respectively the collections of the N predicted states and predicted controls over the horizon (of length N) for the ith system; in situation where the reference to time is needed, x_k^i, u_k^i can be written as $x_{k|t}^i$ and $u_{k|t}^i$. Hence, $x_{0|t}^i = x^i(t)$ and $u_{0|t}^i = u^i(t)$. Additional notations are introduced as required in the text.

II. Preliminaries and Problem Formulation

Consider a stand-alone single system represented by one choice of $i \in \mathbb{Z}^M$ in (1) with corresponding cost

$$\min_{\boldsymbol{u}^i} J^i(x^i, \boldsymbol{u}^i) := \sum_{\ell=0}^{N-1} (\|x_\ell^i\|_{Q^i}^2 + \|u_\ell^i\|_{R^i}^2) + \|x_N^i\|_{P^i}^2 \quad (4)$$

where N is the horizon length, $\boldsymbol{u}^i$ and $\boldsymbol{x}^i$ are the predicted controls and predicted states respectively satisfying $x_{\ell+1}^i = A^i x_\ell^i + B^i u_\ell^i$ with $x_0^i = x^i$, and $J^i(x^i, \boldsymbol{u}^i)$ is the standard quadratic costs parameterized by $(x^i, \boldsymbol{u}^i)$ defined by (4). Let P^i be the solution to the discrete-time algebraic Riccatti equation (DARE) with weights $Q^i \succ 0, R^i \succ 0$ and $K^i = -(R^i + (B^i)^T P^i B^i)^{-1}(B^i)^T P^i A^i$. Define

$$\mathcal{U}_T^i(x^i) := \{\boldsymbol{u}^i \in \mathbb{R}^{m_i N} : x_{\ell+1}^i = A^i x_\ell^i + B^i u_\ell^i, x_0^i = x^i \\ x_\ell^i \in X^i, u_\ell^i \in U^i, x_N^i \in T_f^i, \ell \in \mathbb{Z}_0^{N-1}\} \quad (5)$$

where T_f^i is some appropriate terminal set satisfying

$$A_K^i x^i \in T_f^i, \ K^i x^i \in U^i \text{ for all } x^i \in T_f^i \quad (6)$$

where $A_K^i := A^i + B^i K^i$. Using the above, the overall MPC optimization problem over the M systems incorporating (3) at state $x = \{x^1, \ldots, x^M\}$ is given by

$$\mathbb{P}(x): \quad V(x) := \min_{\{\boldsymbol{u}^i, i\in\mathbb{Z}^M\}} \sum_{i=1}^M J^i(x^i, \boldsymbol{u}^i) \quad (7a)$$

$$\text{s.t. } \boldsymbol{u}^i \in \mathcal{U}_T^i(x^i), \ \forall i \in \mathbb{Z}^M \quad (7b)$$

$$\sum_{i=1}^M \Psi_x^i x_\ell^i + \Psi_u^i u_\ell^i \leq \mathbf{1}_p, \forall \ell \in \mathbb{Z}_0^{N-1} \quad (7c)$$

where (7c) refers to the satisfaction of the coupled constraints at each predicted time step of the horizon. The conditions of (6) do not include the effect of the coupled constraint. As (3) has to be satisfied at all time, T_f^i has to satisfy additional constraints of

$$\sum_{i=1}^M \bar{\Psi}^i x^i := \sum_{i=1}^M (\Psi_x^i + \Psi_u^i K^i) x^i \leq \mathbf{1}_p \quad (8)$$

for all $x^i \in T_f^i$ and $i \in \mathbb{Z}^M$.

A. Tightening the Constraints

It is well known that little benefit is gained in obtaining accurate solution of the MPC optimization problem when the current iterate is far from the optimal. An approach to avoid these unnecessary computations is to allow early termination of the

optimization problem. However, early termination may lead to errors or infeasibility in the subsequent optimization problems. To avoid such issues, constraints (7) and (8) have to be tightened to account for errors arising from this premature termination, in the form of

$$\sum_{i=1}^{M} \Psi_x^i x_\ell^i + \Psi_u^i u_\ell^i \leq (1 - \epsilon M(\ell+1))\mathbf{1}_p \ \forall \ell \in \mathbb{Z}_0^{N-1} \quad (9)$$

$$\sum_{i=1}^{M} \bar{\Psi}^i x^i \leq (1 - MN\epsilon)\mathbf{1}_p, \forall x^i \in T_f^i \quad (10)$$

where ϵ is a user-defined acceptable tolerance to the violation of the constraint in question arising from the premature termination. This is similar to the tightening technique used in [21]. Correspondingly, the tightened DMPC formulation is

$$\min_{\{\boldsymbol{u}^i, i\in\mathbb{Z}^M\}} \left\{ \sum_{i=1}^{M} J^i(x^i, \boldsymbol{u}^i) : (7b) \text{ and } (9) \right\}. \quad (11a)$$

For some appropriate T_f^i that satisfy (6) and (10). To characterize T_f^i, suppose the following assumptions hold for all $i \in \mathbb{Z}^M$:

(**A1**): (A^i, B^i) is stabilizable and $x^i(t)$ is measurable, and

(**A2**): X^i, U^i are polytopes containing the origins in their respective interiors.

Then, $T_f^i = \sigma_\epsilon^i \mathcal{X}_f^i, \ \forall i \in \mathbb{Z}^M$ where $\mathcal{X}_f^i$ is the maximal constraint admissible invariant set [24] for the stand-alone ith system without the global constraints (3). The effect of (3) or (10) is then accounted for by the values of σ_ϵ^i. Let $\bar{\Psi}_r^i$ be the rth row of matrix $\bar{\Psi}^i$. Then, (10) holds if

$$\sum_{i=1}^{M} \max_{x^i \in \sigma^i \mathcal{X}_f^i} \bar{\Psi}_r^i x^i = \sum_{i=1}^{M} h_{\sigma^i \mathcal{X}_f^i}(\bar{\Psi}_r^i) \leq 1 - MN\epsilon, \forall r \in \mathbb{Z}^p \quad (12)$$

where $h_S(v) := \max\{v^T x : x \in S\}$ is the support function of set S along the direction of v. Since $h_{\sigma^i \mathcal{X}_f^i}(\bar{\Psi}_r^i) = \sigma^i h_{\mathcal{X}_f^i}(\bar{\Psi}_r^i)$, condition (12) is equivalent to

$$\sum_{i=1}^{M} \sigma^i h_{\mathcal{X}_f^i}(\bar{\Psi}_r^i) \leq (1 - MN\epsilon), \forall r \in \mathbb{Z}^p. \quad (13)$$

Using (13), an obvious choice of σ to enforce (12) is $(\sigma_\epsilon^1, \ldots, \sigma_\epsilon^M) = \arg\min_{0\leq\sigma^i\leq 1}\{\sum_{i=1}^{M}(1-\sigma^i)^2 : (13)\}$. This quadratic programming problem can be solved in a distributed manner using the proposed algorithm in this paper with σ^i as a variable of the ith system and (13) being the global constraint. If this is done at $t = 0$, the ensuing individual optimization problem requires only knowledge of $\mathcal{X}_f^i$ for system i. Further reduction in conservatism can also be obtained by making $(\sigma_\epsilon^1, \ldots, \sigma_\epsilon^M)$ as variables to be solved at each time step. However, doing so incurs additional communication overhead and is not discussed here as the focus is on issues associated with distributed implementation of the accelerated gradient approach. Let

$$\mathcal{U}^i(x^i) := \mathcal{U}_T^i(x^i) \text{ when } T_f^i = \sigma_\epsilon^i X_f^i \quad (14)$$

and express (9) in terms of $(x^i, \boldsymbol{u}^i)$ for each $i \in \mathbb{Z}^M$. The tightened DMPC formulation can be represented as

$$\mathbb{P}_\epsilon(x): \quad V_\epsilon(x) := \min_{\{\boldsymbol{u}^i, i\in\mathbb{Z}^M\}} \sum_{i=1}^{M} J^i(x^i, \boldsymbol{u}^i) \quad (15a)$$

$$\text{s.t.} \ \ \boldsymbol{u}^i \in \mathcal{U}^i(x^i) \quad i \in \mathbb{Z}^M \quad (15b)$$

$$\sum_{i=1}^{M} f^i(x^i, \boldsymbol{u}^i) \leq b(\epsilon) \quad (15c)$$

where $b^T(\epsilon) := [(1 - M\epsilon)\mathbf{1}_p^T, (1 - 2M\epsilon)\mathbf{1}_p^T, \ldots, (1 - NM\epsilon)\mathbf{1}_p^T]^T$ and

$$f^i(x^i, \boldsymbol{u}^i) := F^i \boldsymbol{u}^i + H^i x^i \quad (16)$$

with $F^i \in \mathbb{R}^{Np\times Nm_i}$ and $H^i \in \mathbb{R}^{Np\times n_i}$ being appropriate matrices from (9) by rewriting x_ℓ^i in terms of x^i and $\boldsymbol{u}^i$ using knowledge of Ψ_x^i and Ψ_u^i only.

B. Network Description

The connectivity of the M systems in this work is described by an undirected graph $G = (\mathcal{V}, \mathcal{E})$ with vertex set $\mathcal{V} = \{1, 2, \ldots, M\}$ and edge set $\mathcal{E} \subset \mathcal{V} \times \mathcal{V}$. The adjacency matrix $\mathcal{A}$ of G is the $M \times M$ matrix whose (i, j) and (j, i) entries are 1 if $(i, j) \in \mathcal{E}$ and 0 otherwise. The set of neighbors of the ith system is $N_i := \{j \in \mathcal{V} : (i, j) \in \mathcal{E}, i \neq j\}$ with $d_i = |N_i|$ and $D = diag\{d_1, d_2, \ldots, d_M\}$. It is assumed that (**A3**) G is connected.

III. The Main Results

A. The Dual Form

The Lagrangian of (15) is $\mathcal{L}(\{\boldsymbol{u}^i\}, \lambda) = \sum_{i=1}^{M} J^i(x^i, \boldsymbol{u}^i) + \lambda^T(\sum_{i=1}^{M} f^i(x^i, \boldsymbol{u}^i) - b(\epsilon))$ for all $\boldsymbol{u}^i \in \mathcal{U}^i(x^i), i \in \mathbb{Z}^M$ where $\lambda \in \mathbb{R}^{Np}$ is the dual variable of (15c). Its dual problem is

$$\max_{\lambda\geq 0} \Phi(x, \lambda) \quad (17)$$

where $\Phi(x, \lambda) = \min_{\{\boldsymbol{u}^i \in \mathcal{U}^i(x^i), i\in\mathbb{Z}^M\}} \mathcal{L}(\{\boldsymbol{u}^i\}, \lambda)$. This dual problem is also equivalent to

$$\min_{\lambda\geq 0} \max_{\{\boldsymbol{u}^i \in \mathcal{U}^i(x^i), i\in\mathbb{Z}^M\}} -\mathcal{L}(\{\boldsymbol{u}^i\}, \lambda) = \min_{\lambda\geq 0} \sum_{i=1}^{M} g^i(\lambda) \quad (18)$$

where

$$g^i(\lambda) := \max_{\boldsymbol{u}^i \in \mathcal{U}^i(x^i)} -J^i(x^i, \boldsymbol{u}^i) - \lambda^T \left(f^i(x^i, \boldsymbol{u}^i) - \frac{b(\epsilon)}{M} \right). \quad (19)$$

Let $\boldsymbol{u}^i(\lambda) = \arg\max_{\boldsymbol{u}^i \in \mathcal{U}^i(x^i)} g^i(\lambda)$. Then, it can be verified that $g^i(\lambda)$ is convex and the gradient of $g^i(\lambda)$ is $\nabla g^i(\lambda) = -(f^i(x^i, \boldsymbol{u}^i(\lambda)) - \frac{b(\epsilon)}{M})$ (see Danskin's Theorem of [11]). In addition, $\nabla g^i(\lambda)$ is Lipschitz continuous with Lipschitz constant $\frac{\|F^i\|^2}{\mu^i}$, where $\mu_i > 0$ is such that $\nabla^2_{\boldsymbol{u}^i} J^i(x^i, \boldsymbol{u}^i) \geq \mu_i I$ for all x^i and $\boldsymbol{u}^i$ (see [25]). Let $L_g = \max_{i\in\mathbb{Z}^M}\{\frac{\|F^i\|^2}{\mu^i}\}$, existence of which is guaranteed by $Q^i, R^i \succ 0$. Note that while the optimal solution of (15) is unique as $J^i(x^i, \boldsymbol{u}^i)$ is stictly convex in $\boldsymbol{u}^i$,

the optimal λ of (18) may not be [20], [26] as the constraints may be degenerate at the optimal $\boldsymbol{u}^i$. Let

$$\Lambda(x) = \{\lambda : \lambda \text{ is an optimal solution of } (17)\} \tag{20}$$

be the collection of all possible optimal λ. Despite the nonuniqueness of λ, the proposed algorithm (Algorithm 1) will converge to an unique solution of (15), see Theorem 1.

B. Finite-Time Average Consensus

The proposed approach requires a finite-time consensus algorithm as a substep. This algorithm is now discussed. Consider the case of a consensus variable $\boldsymbol{z} = (z^1, z^2, \ldots, z^M) \in \mathbb{R}^M$ over the network G satisfying (A3) with z^i being the scalar variable associated with the ith system. Let L be the Laplacian matrix of network G: $L_{i,i} = d_i$, the degree of node i; $L_{i,j} = L_{j,i} = -1$ if $(i,j) \in \mathcal{E}$ and 0 otherwise such that the row sum of L is zero. The discrete-time consensus algorithm can be obtained from L by letting $W = I - \gamma L$ for some $0 < \gamma \leq \frac{1}{\max_i\{d_i\}}$ in the form of

$$\boldsymbol{z}(\ell+1) = W\boldsymbol{z}(\ell) \tag{21}$$

This choice of W can be shown to be a doubly stochastic matrix with 1 being a simple eigenvalue and its spectral radius. The corresponding eigenvector (both left and right) for eigenvalue of 1 is $\frac{1}{\sqrt{M}}\mathbf{1}$. Then

$$\lim_{\ell\to\infty} \boldsymbol{z}(\ell) = \frac{1}{M}\mathbf{1}\mathbf{1}^T\boldsymbol{z}(0) = \frac{1}{M}\mathbf{1}\sum_{i=1}^{M} z^i(0) \tag{22}$$

The first equality of (22) holds because $\boldsymbol{z}(\ell) = \sum_{i=1}^{M} {v_i}^\ell \xi_i\zeta_i^T\boldsymbol{z}(0)$ [27] is the solution of (21) where v_i is the ith eigenvalue of W and $\xi_i(\zeta_i)$ is the corresponding right (left) eigenvector. Since $|v_1| = 1$, $\xi_1 = \zeta_1 = \frac{1}{\sqrt{M}}\mathbf{1}$, and $|v_i| < 1$ for all $i = 2, \ldots, M$, $\lim_{\ell\to\infty}\boldsymbol{z}(\ell) = \frac{1}{M}\mathbf{1}\mathbf{1}^T\boldsymbol{z}(0)$. The expression of (22) can be further simplified using (21) as

$$\lim_{\ell\to\infty}\boldsymbol{z}(\ell) = \lim_{\ell\to\infty} W^\ell\boldsymbol{z}(0) = \left(\lim_{\ell\to\infty} W^\ell\right)\boldsymbol{z}(0) \tag{23}$$

$$= \left(\sum_{\ell=0}^{T-1}\tau_\ell W^\ell\right)\boldsymbol{z}(0) = \sum_{\ell=0}^{T-1}\tau_\ell\boldsymbol{z}(\ell) \tag{24}$$

for some $T \leq M$. Here, the first equality of (24) follows from the closure property of W^∞ via the minimal polynomial of W ($t^T + \pi_0 t^{T-1} + \cdots + \pi_{T-1}t^0 = 0$) with the coefficients $\{\tau_0, \ldots, \tau_{T-1}\}$ obtained from $\{\pi_0, \ldots, \pi_{T-1}\}$ using standard results from functions of square matrices [27], or as a special case from the solution of a Vandermonde matrix [28]. Such a representation is guaranteed to exist since, in the worst case, the characteristic polynomial becomes the minimal polynomial with $T = M$ and the closure property follows from the well-known Caley-Hamilton principle.

Combining (22), (23), and (24) means that $\frac{1}{M}\mathbf{1}\sum_{i=1}^{M} z^i(0) = \sum_{\ell=0}^{T-1}\tau_\ell\boldsymbol{z}(\ell)$, or, considering each element of this vector equation,

$$\frac{1}{M}\sum_{i=1}^{M} z^i(0) = \sum_{\ell=0}^{T-1}\tau_\ell z^i(\ell). \tag{25}$$

This equation shows that the ith system can obtain the value of $\frac{1}{M}\sum_{i=1}^{M} z^i(0)$ by computing its consensus state $z^i(\ell)$ for $\ell = 0, \ldots, T-1$ and evaluating the right hand side of (25). Note that this T steps of z^i is obtained in a distributed manner using the ith row of (21), or

$$z^i(\ell+1) = W^{ii}z^i(\ell) + \sum_{j\in N_i} W^{ij}z^j(\ell),\ i \in \mathbb{Z}^M. \tag{26}$$

The above development is for the case where z^i is a scalar. In the case where $z^i \in \mathbb{R}^\eta$ is a vector with $\boldsymbol{z} \in \mathbb{R}^{M\eta}$, the development from (22) till (26) holds with W replaced by $W \otimes I_\eta$ where $\otimes$ refers the Kronecker product of two matrices. The above development is also for the case where L (hence W) is the standard Laplacian matrix. It is possible to obtain a lower order minimum polynomial for the same network by considering weighted Laplacian matrix ($L_{i,j}$ are not necessary -1 for $(i,j) \in \mathcal{E}$) [28].

C. Distributed Fast Dual Gradient Algorithm

The standard stand-alone Nesterov gradient algorithm [18]–[20] for $\min_{\lambda\geq 0}\sum_{i=1}^{M} g^i(\lambda)$ of (18) consists of the following iterates:

$$\tilde{\lambda}^k = \lambda^k + \theta^k((\theta^{k-1})^{-1} - 1)(\lambda^k - \lambda^{k-1}) \tag{27a}$$

$$\lambda^{k+1} = \left[\tilde{\lambda}^k - \frac{1}{L_g}\sum_{i=1}^{M}\nabla g^i(\tilde{\lambda}^k)\right]_+ \tag{27b}$$

$$\theta^{k+1} = \left(\sqrt{(\theta^k)^4 + 4(\theta^k)^2} - (\theta^k)^2\right)/2 \tag{27c}$$

where $[x]_+ = \max\{0, x\}$, $\lambda^{-1} = \lambda^0 = 0$, and $\theta^{-1} = \theta^0 = 1$ are the needed initializations. From (27c), the sequence $\{\theta^k\}$ has to satisfy [20]

$$\frac{1-\theta^{k+1}}{(\theta^{k+1})^2} = \frac{1}{(\theta^k)^2}, \frac{1}{(\theta^k)^2} = \sum_{\ell=0}^{k}(\theta^\ell)^{-1}, \theta^k \leq \frac{2}{k+2} \tag{28}$$

Note that (27b) requires the gradients from all M systems. In order to implement a fully distributed computation, each system i makes a local copy of λ, λ^i. Correspondingly, (27a) and (27b) are replaced by, $\forall i \in \mathbb{Z}^M$

$$\tilde{\lambda}^{i,k} = \lambda^{i,k} + \theta^k((\theta^{k-1})^{-1} - 1)(\lambda^{i,k} - \lambda^{i,k-1}) \tag{29a}$$

$$\lambda^{i,k+1} = \left[\frac{1}{M}\sum_{i=1}^{M}\left(\tilde{\lambda}^{i,k} - \frac{1}{L_g}\nabla g^i(\tilde{\lambda}^{i,k})\right)\right]_+ \tag{29b}$$

where $\nabla g^i(\tilde{\lambda}^{i,k}) = -(f^i(x^i, \tilde{\boldsymbol{u}}^{i,k}) - \frac{b}{M})$ with

$$\tilde{\boldsymbol{u}}^{i,k} = \arg\min_{\boldsymbol{u}^i\in\mathcal{U}^i(x^i)} J^i(x^i, \boldsymbol{u}^i) + (\tilde{\lambda}^{i,k})^T\left(f^i(x^i, \boldsymbol{u}^i) - \frac{b}{M}\right). \tag{30}$$

For all $i \in \mathbb{Z}^M$, let

$$\bar{\boldsymbol{u}}^{i,k} := (\theta^k)^2 \sum_{\ell=0}^{k} (\theta^\ell)^{-1} \tilde{\boldsymbol{u}}^{i,\ell} = (1-\theta^k)\bar{\boldsymbol{u}}^{i,k-1} + \theta^k \tilde{\boldsymbol{u}}^{i,k} \quad (31)$$

where $\bar{\boldsymbol{u}}^{i,-1} = 0$ and $\tilde{\boldsymbol{u}}^{i,k}$ is obtained from (30). Note that (29b) requires the quantity $(\tilde{\lambda}^{i,k} - \frac{1}{L_g}\nabla g^i(\tilde{\lambda}^{i,k}))$ from all $i \in \mathbb{Z}^M$ and, hence, is not fully distributed. To handle this problem, the finite-time consensus algorithm mentioned in Section III-B is used to compute the quantity $\frac{1}{M}\sum_{i=1}^{M}(\tilde{\lambda}^{i,k} - \frac{1}{L_g}\nabla g^i(\tilde{\lambda}^{i,k}))$ in (29b). Specifically, for each $i \in \mathbb{Z}^M$ and $\ell \in \mathbb{Z}_0^{T-2}$, introduce variable $y^i(\ell, k)$ with

$$y^i(0,k) := \tilde{\lambda}^{i,k} - \frac{1}{L_g}\nabla g^i(\tilde{\lambda}^{i,k}) \quad (32)$$

$$y^i(\ell+1,k) = W^{ii} y^i(\ell,k) + \sum_{j \in N_i} W^{ij} y^j(\ell,k) \quad (33)$$

where T is the order of the minimal polynomial of W and (33) is the ith component of the consensus dynamics given by (21). From (25), $\sum_{\ell=0}^{T-1} \tau_\ell y^i(\ell,k) = \frac{1}{M}\sum_{i=1}^{M}(\tilde{\lambda}^{i,k} - \frac{1}{L_g}\nabla g^i(\tilde{\lambda}^{i,k}))$ for all $i \in \mathbb{Z}^M$. With this property, (29b) is replaced by

$$\lambda^{i,k+1} = \left[\sum_{\ell=0}^{T-1} \tau_\ell y^i(\ell,k)\right]_+, \quad i \in \mathbb{Z}^M. \quad (34)$$

This process is then repeated by incrementing k. The stopping criterion of this distributed fast dual gradient algorithm is discussed in Section IV and suppose it terminates at iteration $\bar{k}$. Then, the solution from this algorithm is $\bar{\boldsymbol{u}}^{i,\bar{k}} := \{\bar{u}_0^{i,\bar{k}}, \bar{u}_1^{i,\bar{k}}, \ldots, \bar{u}_{N-1}^{i,\bar{k}}\}$, as defined by (31), $i \in \mathbb{Z}^M$. Correspondingly, the MPC control law applied on the ith system is

$$\kappa^i(x) = \bar{u}_0^{i,\bar{k}}, i \in \mathbb{Z}^M. \quad (35)$$

D. Convergence Analysis

The convergence results of the distributed fast dual gradient algorithm are discussed in this section. Let the feasible domain of $\mathbb{P}_\epsilon(x)$ be

$$\mathcal{D}_\epsilon := \{x \in \mathbb{R}^n : \mathbb{P}_\epsilon(x) \text{ is feasible}\}. \quad (36)$$

The convergence to an optimal dual solution is stated as follows.

Lemma 1: For any $x \in \mathcal{D}_\epsilon$, let $\{\lambda^{i,k}, \tilde{\lambda}^{i,k}\}_{i=1}^M$ be generated from (29a) and (29b) with $\lambda^{i,-1} = \lambda^{i,0} = 0$. Then, for any $\lambda^* \in \Lambda(x)$, the following results hold.

(i) For any $k \geq 0$, the objective $\sum_{i=1}^M g^i(\lambda^{i,k+1})$ is bounded by

$$0 \leq \sum_{i=1}^{M} g^i(\lambda^{i,k+1}) - \sum_{i=1}^{M} g^i(\lambda^*) \leq \frac{L_g M(\theta^k)^2}{2}\|\lambda^*\|^2 \leq \frac{2L_g M}{(k+2)^2}\|\lambda^*\|^2. \quad (37)$$

(ii) Let $\bar{\lambda}^{i,k} := \lambda^{i,k-1} + (\theta^{k-1})^{-1}(\lambda^{i,k} - \lambda^{i,k-1})$ for all $k \geq 0$ and $i \in \mathbb{Z}^M$. The sequence $\{\bar{\lambda}^{i,k}\}$ satisfies

$$\sum_{i=1}^{M} \|\bar{\lambda}^{i,k+1} - \lambda^*\|^2 \leq M\|\lambda^*\|^2. \quad (38)$$

(iii) Consider the sequences $\{\tilde{\boldsymbol{u}}^{i,k}\}_{i=1}^M$ and $\{\bar{\boldsymbol{u}}^{i,k}\}_{i=1}^M$ generated from (30) and (31), respectively. For any $k \geq 0$, the coupled constraint using $\{\bar{\boldsymbol{u}}^{i,k}\}_{i=1}^M$ satisfies the inequality

$$\sum_{i=1}^{M} f^i(x^i, \bar{\boldsymbol{u}}^{i,k}) - b \leq \frac{\bar{L}_g\|\lambda^*\|}{(k+2)^2}\mathbf{1}_{Np} \quad (39)$$

where $\bar{L}_g = 4M(\sqrt{M}+1)L_g$.

Property (iii) of Lemma 1 provides the decreasing upper bound on the violation of the coupled constraint. On the basis of the convergence of the dual variable, the primal fast convergence result is stated in the following theorem, which is a modification of [20, Th. 5].

Theorem 1: For any $x \in \mathcal{D}_\epsilon$, suppose $\{\boldsymbol{u}^{i*}\}_{i=1}^M$ is the optimal solution of $\mathbb{P}_\epsilon(x)$. Then, for any $k \geq 0$ and $\lambda^* \in \Lambda(x)$, it holds that

$$-\frac{\bar{L}_g\|\lambda^*\|^2}{(k+2)^2} \leq \sum_{i=1}^{M} J^i(x^i, \bar{\boldsymbol{u}}^{i,k}) - \sum_{i=1}^{M} J^i(x^i, \boldsymbol{u}^{i*}) \leq 0. \quad (40)$$

E. Primal Suboptimality and Feasibility

As mentioned before, a premature termination condition is used to reduce the computational load for the solution of (17). For this purpose, the relaxed solution of (15) is defined as follows.

Definition 1: Given any $\varepsilon > 0$, the set $\{x^i, \boldsymbol{u}^i\}_{i=1}^M$ is a ε-relaxed solution of (15) if, $\forall i \in \mathbb{Z}^M$,

$$\boldsymbol{u}^i \in \mathcal{U}^i(x^i), \sum_{i=1}^{M} f^i(x^i, \boldsymbol{u}^i) - b(\epsilon) \leq \varepsilon M \mathbf{1}_{pN} \quad (41)$$

where $\{\boldsymbol{u}^{i*}\}_{i=1}^M$ is the optimal solution of (15). In addition, for any $\varepsilon, \delta > 0$, the set $\{x^i, \boldsymbol{u}^i\}_{i=1}^M$ is a (ε, δ)-suboptimal solution of (15) if it is a ε-relaxed solution and

$$\sum_{i=1}^{M} \left((J^i(x^i, \boldsymbol{u}^i) - J^i(x^i, \boldsymbol{u}^{i*})\right) \leq \delta. \quad (42)$$

The following lemma discusses the existence of the suboptimal solution.

Lemma 2: For any $x \in \mathcal{D}_\epsilon$, let $\{\tilde{\boldsymbol{u}}^{i,k}\}_{i=1}^M$ and $\{\bar{\boldsymbol{u}}^{i,k}\}_{i=1}^M$ be generated from (30) and (31), respectively. Then, it holds that:

1) there exists a finite k such that $\{x^i, \bar{\boldsymbol{u}}^{i,k}\}_{i=1}^M$ is a $(\epsilon, 0)$-suboptimal solution of (15);
2) $\{x^i, \bar{\boldsymbol{u}}^{i,k}\}_{i=1}^M$ is a $(\epsilon, 0)$-suboptimal if and only if it is a ϵ-relaxed solution.

The next theorem shows the existence of a ϵ-relaxed solution ensures the recursive feasibility of (15).

Theorem 2: Suppose $\{x^i, \boldsymbol{u}^i\}_{i=1}^M$ is a ϵ-relaxed solution of (15) as defined by Definition 1 with $\boldsymbol{u}^i = \{u_0^i, u_1^i, \ldots, u_{N-1}^i\}$ for all $i \in \mathbb{Z}^M$. Let the state sequence associated with this solution be $\{x_0^i, x_1^i, \ldots, x_N^i\}$, $x^{i+} = A^i x^i + B^i u_0^i$, and $\boldsymbol{u}^{i+} =$

Algorithm 1: Distributed Fast Dual Gradient Algorithm.

Input: x^i, $i \in \mathbb{Z}^M$
Output: $\bar{\boldsymbol{u}}^{i,\bar{k}}$, $i \in \mathbb{Z}^M$
Initialization: set $k = 0$, $\bar{\boldsymbol{u}}^{i,-1} = 0, \lambda^{i,-1} = \lambda^{i,0} = 0$, and $\theta^{-1} = \theta^0 = 1$, for all $i \in \mathbb{Z}^M$;
repeat
  for all $i \in \mathbb{Z}^M$ (in parallel) **do**
    Obtain $\tilde{\lambda}^{i,k}$ and $\tilde{\boldsymbol{u}}^{i,k}$ from (29a) and (30), respectively;
    Perform the finite-consensus steps in (33) with $y^i(0,k)$ being given in (32);
    Obtain $\lambda^{i,k+1}$ from (34);
  end for
  Set up $z^i(0,k)$ using $\bar{\boldsymbol{u}}^{i,k}$ which is defined in (31), $i \in \mathbb{Z}^M$
  for all $i \in \mathbb{Z}^M$ (in parallel) **do**
    Obtain $z^i(1,k), \ldots, z^i(T-1,k)$ using (43) and(44);
  end for
  $\theta^{k+1} = (\sqrt{(\theta^k)^4 + 4(\theta^k)^2} - (\theta^k)^2)/2$;
  $k \leftarrow k+1$
until $\sum_{\ell=0}^{T-1} \tau_\ell z^i(\ell,k) \le \epsilon$

$\{u_1^i, \ldots, u_{N-1}^i, K^i x_N^i\}$ for all $i \in Z^M$. Then, the following results hold.

1) $\{\boldsymbol{u}^{i+}\}_{i=1}^M$ is a feasible solution to $\mathbb{P}_\epsilon(x^+)$.
2) Consider the solution of $\mathbb{P}_\epsilon(x^+)$ and let $\{\bar{\boldsymbol{u}}^{i,k}\}_{i=1}^M$ be generated from (30) and (31) with the states $\{x^{i+}\}_{i=1}^M$. Then, there exists a finite k such that $\{x^{i+}, \bar{\boldsymbol{u}}^{i,k}\}_{i=1}^M$ is a ϵ-relaxed solution of (15).

IV. The Overall DMPC Scheme

The overall DMPC scheme is now presented in this section. First, a proper stopping criterion for the distributed fast dual gradient algorithm is needed. This condition is based on the results of Lemma 2. Specifically, the algorithm terminates at the first k, denoted as $\bar{k}$, such that a ϵ-relaxed solution is achieved. Following Definition 1, the stopping criterion corresponds to $\sum_{i=1}^M f^i(x^i, \bar{\boldsymbol{u}}^{i,k}) - b(\epsilon) \le \epsilon M \mathbf{1}_{pN}$. This condition should be checked in a fully distributed manner. Again, the finite-time consensus algorithm in Section III-B is used. For each $i \in \mathbb{Z}^M$, introduce the variable $z^i(\ell,k)$ with, $\forall \ell \in \mathbb{Z}_0^{T-2}$,

$$z^i(0,k) := f^i(x^i, \bar{\boldsymbol{u}}^{i,k}) - \frac{b(\epsilon)}{M} \tag{43}$$

$$z^i(\ell+1,k) = W^{ii} z^i(\ell,k) + \sum_{j \in N_i} W^{ij} z^j(\ell,k). \tag{44}$$

The stopping criterion is satisfied if $\sum_{\ell=0}^{T-1} \tau_\ell z^i(\ell,k) \le \epsilon$ as $\sum_{\ell=0}^{T-1} \tau_\ell z^i(\ell,k) = \frac{1}{M}(\sum_{i=1}^M f^i(x^i, \bar{\boldsymbol{u}}^{i,k}) - b(\epsilon))$ for all $i \in \mathbb{Z}^M$. The distributed fast dual gradient algorithm with the finite-time consensus is summarized in Algorithm 1.

The overall procedure of the DMPC algorithm is summarized in Algorithm 2.

Algorithm 2: The Overall DMPC Algorithm.

1: At time t, every system i measures it own state $x^i(t)$;
2: Every system i calls Algorithm 1 with $x^i(t)$ and obtain $\bar{\boldsymbol{u}}^{i,\bar{k}(t)}$ as its output.
3: Every system obtains $\kappa^i(x(t))$ from $\bar{\boldsymbol{u}}^{i,\bar{k}(t)}$ via (35) and apply $\kappa^i(x(t))$ to the ith system.
4: Wait for next sampling time, let $t = t+1$ and go to step 1.

V. Recursive Feasibility and Stability

This section discusses the recursive feasibility and stability results of the proposed DMPC formulation. The next lemma pertains to a property of the terminal set of the overall system and is needed for stability of the closed-loop MPC system.

Lemma 3: Let

$$\bar{\sigma}^i := \min\left\{\sigma_\epsilon^i, \frac{1}{M} \min_{\ell \in \mathbb{Z}^{N_p}} \{b_\ell(\epsilon)/h_{X_f^i}(\bar{F}_\ell^i)\}\right\} \tag{45}$$

where $h_{X_f^i}(\cdot)$ is the support function of X_f^i, $b_\ell(\epsilon)$ denotes the ℓth element of $b(\epsilon)$, $\bar{F}^i := F^i K_A^i + H^i$ from (16) with K_A^i defined by (46) below, and $\bar{F}_\ell^i$ denotes the ℓth row of $\bar{F}^i$. For any $x^i \in \bar{\sigma}^i X_f^i$, the optimal solution to Algorithm 1 for the ith system is

$$\begin{aligned}\bar{\boldsymbol{u}}^{i,\bar{k}} &= \{\bar{u}_0^{i,\bar{k}}, \ldots, \bar{u}_{N-1}^{i,\bar{k}}\} \\ &= \{K^i x^i, K^i A_K^i x^i, \ldots, K^i (A_K^i)^{N-1} x^i\} := K_A^i x^i\end{aligned} \tag{46}$$

with $\bar{k} = 0$.

The recursive feasible and stability results of the proposed DMPC approach are stated in the following theorem.

Theorem 3: Suppose (A1)–(A3) hold and $\mathbb{P}_\epsilon(x(t))$ of (15) has a feasible solution at time t and that the MPC law of (35) is applied to the ith system of (2) for all $i \in \mathbb{Z}^M$. Then, the following results hold.

1) $\mathbb{P}_\epsilon(x(t+1))$ has a feasible solution at time $t+1$.
2) For all $t \ge 0$, $\sum_{i=1}^M (J^i(x^i(t), \bar{\boldsymbol{u}}^{i,\bar{k}(t)}) - J^i(x^i(t), \boldsymbol{u}_t^{i*})) \le 0$, where $\{\boldsymbol{u}_t^{i*}\}_{i=1}^M$ is the optimal solution of (15).
3) The closed-loop system (1) with the MPC law (35) is exponentially stable.

Remark 1: It can be shown that the true linear quadratic (LQ) cost is upper bounded by the predicted cost of the initial state. For any $x \in \mathcal{D}_\epsilon$, let the infinite true LQ cost associated with the control law (35) be

$$J_\epsilon^\infty(x) = \sum_{t=0}^{\infty} \sum_{i=1}^{M} (\|x^i(t)\|_{Q^i}^2 + \|\kappa^i(x(t))\|_{R^i}^2) \tag{47}$$

where $x(0) = x$ and $x^i(t+1) = A^i x^i(t) + B^i \kappa^i(x(t))$ for all $i \in \mathbb{Z}^M$. From (71),

$$V_\epsilon(x(t+1)) \le V_\epsilon(x(t)) - \sum_{i=1}^{M} \left(\|x^i(t)\|_{Q^i}^i + \|\kappa^i(x(t))\|_{R^i}^2\right) \tag{48}$$

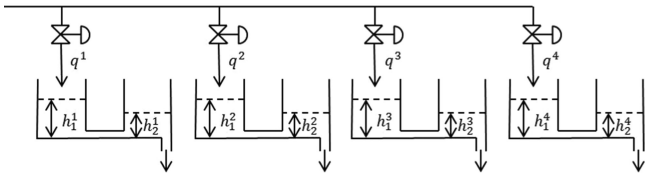

Fig. 1. Water tanks system.

which, when summed up from $t = 0$ to ∞, implies that $J_\epsilon^\infty(x) \leq V_\epsilon(x)$. This gives a performance bound for the infinite LQ cost of the closed-loop system.

VI. Numerical Results

The numerical example used to demonstrate the approach is a four-agent system where every agent is a coupled double-water tank system [29]. The objective is to regulate the water levels to some given references with a constraint on the total input flow rate. As shown in Fig. 1, q^i is the input flow and h_1^i and h_2^i are the water levels for system i. Suppose the desired water levels are $\tilde{h}_1^i$ and $\tilde{h}_2^i$ and the steady-state input flow rate is $\tilde{q}^i$ for system i. Let $x_1^i = h_1^i - \tilde{h}_1^i$, $x_2^i = h_2^i - \tilde{h}_2^i$ and $u^i = q^i - \tilde{q}^i$ for $i \in \mathbb{Z}^4$.

Given the parameters: $\tilde{h}_1^i = 1, \tilde{h}_2^i = 0.64, \tilde{q}^i = 0.3$, $i \in \mathbb{Z}^4$, the linearized and discretized model for each agent is

$$x^i(t+1) = \begin{pmatrix} 0.8750 & 0.1250 \\ 0.1250 & 0.8047 \end{pmatrix} x^i(t) + \begin{pmatrix} 0.3 \\ 0 \end{pmatrix} u^i(t).$$

All agents have the same local constraints of $X^i := \{x^i \in \mathbb{R}^2 : |x_1^i| \leq 1, |x_2^i| \leq 0.64\}$ and $U^i := \{u^i \in \mathbb{R} : |u^i| \leq 0.3\}$. Suppose the maximal total input flow rate is 2, the coupled constraint can be given by $\sum_{i=1}^4 q^i \leq 2$, which implies that $\sum_{i=1}^4 u^i \leq 2 - \sum_{i=1}^4 \tilde{q}^i = 0.8$. The values of K^i and P^i obtained from the discrete-time Algebraic Riccatti Equation (ARE), with $Q^i = 10I_2$ and $R^i = 1$, $i \in \mathbb{Z}^4$, are

$$K^i = \begin{pmatrix} -1.7916 & -0.7337 \end{pmatrix}, P^i = \begin{pmatrix} 31.7459 & 9.8300 \\ 9.8300 & 56.3415 \end{pmatrix}$$

for all $i \in \mathbb{Z}^4$. Consider the network connection of a ring and $W = I - 0.1L(G)$. The minimal polynomial of W is $t^3 - 2.4t^2 + 1.88t - 0.48 = 0$. Hence, value of T of (25) is 3 and the finite-time consensus in (25) can be obtained 2 steps. The initial conditions are: $x^1(0) = [-0.3241 \ -0.5977]^T$, $x^2(0) = [0.4390 \ -0.4667]^T$, $x^3(0) = [-0.4391 \ -0.5818]^T$, $x^4(0) = [-0.5337 \ -0.4347]^T$ and the horizon length $N = 8$. The performance of the proposed DMPC approach is presented for several choices of ϵ and comparisons are made between the results of the proposed approach and that obtained by solving (15) with $\epsilon = 0$ using a single centralized computer, known as the centralized MPC (CMPC) solution. The terminal sets $\{\sigma_s^i X_f^i\}_{i=1}^M$ in CMPC are obtained from $\min\{\sum_{i=1}^M (1-\sigma^i)^2 : (13)\}$ with $\epsilon = 0$: $\sigma_s^i = 0.6667, i \in \mathbb{Z}^4$. Consider the case of $\epsilon = 0.01$ in DMPC. The overall input is shown in Fig. 2. The MPC problems are solved by MOSEK [30] under the MATLAB interface. All numerical experiments are done on a Windows 7 PC with an Intel Xeon E5-1630 processor and 16 GB memory.

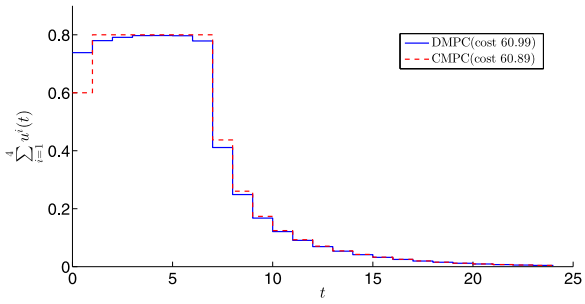

Fig. 2. Overall input trajectories: DMPC ($\epsilon = 0.01$) and CMPC.

TABLE I
The Values of $J_\epsilon^\infty(x(0))$ for Different Choices of ϵ

ϵ	0.01	0.005	0.001
$\sigma_\epsilon^i, i \in \mathbb{Z}^4$	0.4533	0.56	0.6453
$V_\epsilon(x(0))$	61.87	61.26	60.95
$J_\epsilon^\infty(x(0))$	60.99	60.91	60.89

TABLE II
The Number of Iterations Along the Trajectories Starting from $x(0)$ for Different Choices of ϵ

	ϵ	$t=0$	$t=1$	$t=2$	$t=3$	$t=4$	$t=5$	$t=6$	$t=7$
$\bar{k}_F(t)$	0.01	23	21	19	16	13	9	1	1
	0.005	24	25	23	20	17	12	1	1
	0.001	49	50	48	46	42	30	6	1
$\bar{k}_A(t)$	0.01	88	88	72	67	71	74	94	1
	0.005	96	96	80	79	79	83	105	1
	0.001	120	120	112	106	108	115	131	1

The following table gives the real LQ cost $J_\epsilon^\infty(x(0))$ (as defined in Remark 1) for different choices of ϵ. The values of $\{\sigma_\epsilon^i\}_{i\in\mathbb{Z}^4}$, obtained from $\min_{0\leq\sigma^i\leq 1}\{\sum_{i=1}^M (1-\sigma^i)^2 : (13)\}$, are also shown in Table I. It can be seen that the performance of the DMPC approach is close to that of CMPC because they have similar overall input trajectories and the degradation in cost is less than 0.2% even in the case of $\epsilon = 0.01$. The fact that $J_\epsilon^\infty(x(0))$ is upper bounded by $V_\epsilon(x(0))$ can also be verified in Table I.

A comparison between the proposed approach and the ADMM-based approach in [17] is presented next. The number of iterations of the proposed approach at each t along the trajectories is denoted by $\bar{k}_F(t) := \bar{k}(t) + 1$, whose values are shown in Table II together with $\bar{k}_A(t)$, the number of iterations of the ADMM-based approach. Notice that $\bar{k}(t) = 0$ for $t \geq 7$ because the global constraints are no longer active. It can be seen values of $\bar{k}_F(t)$ and $\bar{k}_A(t)$ of Table II that the number of iterations needed to reach the same accuracy by the proposed approach is about 30% to 50% of that needed by the ADMM-based approach with comparable numbers of communications: $2(T-1)\bar{k}_F(t)$ for this approach and $(T-1)\bar{k}_A(t)$ for the ADMM-based approach.

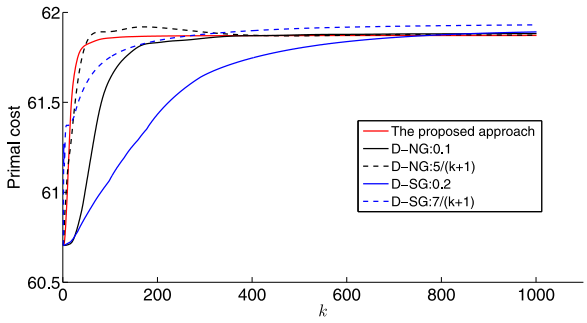

Fig. 3. Convergence curves of different distributed algorithms.

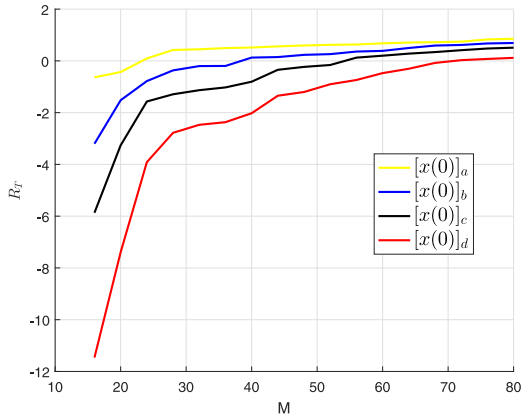

Fig. 4. Relative computational time difference between the local and centralized problems for different sizes of M and initial states with $\epsilon = 10^{-3}$: $[x(0)]_a = \mathbf{1}_M \otimes [-0.1178\, 0.005355]^T, [x(0)]_b = \mathbf{1}_M \otimes [-0.1863\, 0.1383]^T, [x(0)]_c = \mathbf{1}_M \otimes [-0.1220\ -0.07678]^T, [x(0)]_d = \mathbf{1}_M \otimes [-0.1296\ -0.2391]^T$.

This section also compares results of the proposed approach for problem (18) as a one-off (nonrecursive) optimization problem not in the context of MPC. The comparison is made with two other distributed algorithms in the literature: the distributed subgradient (D-SG) algorithm of [31] and the distributed Nesterov gradient (D-NG) algorithm of [22], [23]. As the step size affects the performance of D-SG and D-NG, reasonable step sizes are chosen after a few trials: 0.2 and $\frac{7}{k+1}$ for D-SG; 0.1 and $\frac{5}{k+1}$ for D-NG. The parameter $\epsilon = 0.01$ and x^i of (18) are those used in the earlier example. The plots of the convergence of these algorithms are shown in Fig. 3 and they show the faster convergence of the proposed approach, a result that is consistent with the theoretical convergence rate of D-SG and D-NG ($O(1/\sqrt{k})$ and $O(\log k/k)$, respectively, for the diminishing stepsize) and the proposed approach.

The rest of this section discusses the computational times of the proposed approach and compares them with those obtained from the CMPC approach of (15). Let T_o be the computational time of the CMPC problem. It is easy to see that the overall computational time of Algorithm 1 is $T_a + T_c$ where T_c is the total time for information exchange via communication and T_a is the total computational time for Steps 3 and 4 of Algorithm 1. In addition, $T_a = \bar{k}_F \max_i\{T_s^i : i \in Z^M\}$ where T_s^i is the average computational time for Steps 3 and 4 of Algorithm 1 since these steps are run in parallel in each system and $T_c = 2\bar{k}_F(T-1)\tau_c$ where τ_c is the communication time for each information exchange with the neighbors of each agent. Hence, as M increases, $T_o = T_a + T_c$ is the break even condition between CMPC and the proposed approach. Equivalently, the proposed approach is better than the CMPC when $\tau_c < (T_o - T_a)/(2\bar{k}_F(T-1))$. To appreciate this effect, the ratio of $(T_o - T_a)/T_o := R_T$ is plotted for various starting point and different values of M as given in Fig. 4. Hence, for small values of τ_c, the proposed distributed approach will eventually outperform the CMPC as M increases. This is not unexpected because the complexity of the centralized problem grows with M while the size of the local problems does not change if the communication time is reasonably small.

VII. Conclusion

A novel DMPC approach is proposed for a group of linear systems with local and global constraints. The proposed approach relies on the dual problem of the overall MPC problem and uses a distributed fast dual gradient algorithm for its solution. This is made possible by introducing local copies of the dual variables in individual system and enforcing all the local copies to achieve consensus at each iteration. Provision for computational expediency is made via early termination of the proposed algorithm where the inaccuracy depends on the user-defined allowable violation of the coupled constraint. Termination condition is checked using a finite-time consensus algorithm. Under mild assumptions, a suboptimal solution of the overall MPC problem can be obtained so long as the network of systems are connected. Recursive feasibility and exponential stability of the closed-loop system are ensured. The performance of the proposed approach is demonstrated by a four-tank networked system with a limited total input flow rate. Compared to the ADMM-based approach of the same problem, this approach achieves convergence of about 2 to 3 times faster and invokes fewer quadratic optimization solvers, but may require more communications among systems. This communication issue is minimized by the use of a finite-time consensus based on the minimal polynomial extracted from the network. Comparisons of convergence results are also made with the distributed subgradient algorithm and distributed Nesterov gradient algorithm. In both cases, the proposed method has faster convergence.

Appendix

A. Proof of Lemma 1

1) The problem (18) can be rewritten as

$$\min_{\lambda^i \geq 0, i \in \mathbb{Z}^M} \sum_{i=1}^{M} g^i(\lambda^i) \quad \text{s.t. } \lambda^1 = \lambda^2 = \cdots = \lambda^M \tag{49}$$

which is equivalent to $\min_{\boldsymbol{\lambda} \in \Omega} g(\boldsymbol{\lambda})$, where $g(\boldsymbol{\lambda}) := \sum_{i=1}^{M} g^i(\lambda^i)$, $\boldsymbol{\lambda} = (\lambda^1, \lambda^2, \ldots, \lambda^M)$ and $\Omega = \{\boldsymbol{\lambda} \geq 0 : \lambda^1 = \lambda^2 = \cdots = \lambda^M\}$. Then, (29a)–(29b) can be writ-

ten in a compact form

$$\tilde{\boldsymbol{\lambda}}^k = \boldsymbol{\lambda}^k + \theta^k((\theta^{k-1})^{-1} - 1)(\boldsymbol{\lambda}^k - \boldsymbol{\lambda}^{k-1}) \quad (50a)$$

$$\boldsymbol{\lambda}^{k+1} = P_\Omega\left[\tilde{\boldsymbol{\lambda}}^k - \frac{1}{L_g}\nabla g(\boldsymbol{\lambda}^k)\right]. \quad (50b)$$

It can be easily verified that $\nabla g(\boldsymbol{\lambda})$ is Lipschitz continuous with the constant L_g. For any $\lambda \geq 0$, from Proposition 6.9.2 in [19] and the fact that $\mathbf{1}_M \otimes \lambda \in \Omega$, it holds that

$$\sum_{i=1}^M g^i(\lambda^{i,k+1}) - \sum_{i=1}^M g^i(\lambda) + \frac{L_g(\theta^k)^2}{2} \times \sum_{i=1}^M \|\bar{\lambda}^{i,k+1} - \lambda\|^2 + (\theta^k)^2 \sum_{\ell=0}^k (\theta^\ell)^{-1} \times \sum_{i=1}^M (g^i(\lambda) - \Delta^i(\lambda, \tilde{\lambda}^{i,\ell})) \leq \frac{L_g M(\theta^k)^2}{2}\|\lambda\|^2 \quad (51)$$

where

$$\bar{\lambda}^{i,k} := \lambda^{i,k-1} + (\theta^{k-1})^{-1}(\lambda^{i,k} - \lambda^{i,k-1}) \quad (52)$$

$$\Delta^i(\lambda, \tilde{\lambda}) := g^i(\tilde{\lambda}) + (\nabla g^i(\tilde{\lambda}))^T(\lambda - \tilde{\lambda}). \quad (53)$$

Let λ in (51) be any $\lambda^* \in \Lambda(x)$. The first inequality of (37) holds because λ^* is a minimizer of (18) and the second inequality (37) holds due to the fact that $\Delta^i(\lambda^*, \tilde{\lambda}^{i,k}) \leq g^i(\lambda^*)$ from the convexity of $g^i(\cdot)$. The last inequality holds because $\theta^k \leq \frac{2}{k+2}$ from (28).

2) The inequality (38) follows from (51) because $\sum_{i=1}^M g^i(\lambda^{i,k+1}) - \sum_{i=1}^M g^i(\lambda^*) \geq 0$ and $\Delta^i(\lambda^*, \tilde{\lambda}^{i,k}) \leq g^i(\lambda^*)$.
3) Some intermediate results are needed to prove (39). Using the auxiliary variable $\bar{\lambda}^{i,k} = \lambda^{i,k-1} + (\theta^{k-1})^{-1}(\lambda^{i,k} - \lambda^{i,k-1})$, $\lambda^{i,k+1}$ can be written as $\lambda^{i,k+1} = \theta^k(\bar{\lambda}^{i,k+1} - \lambda^{i,k}) + \lambda^{i,k}$. This, together with $\tilde{\lambda}^{i,k} = \lambda^{i,k} + \theta^k((\theta^{k-1})^{-1} - 1)(\lambda^{i,k} - \lambda^{i,k-1})$, implies that

$$\begin{aligned}\lambda^{i,k+1} - \tilde{\lambda}^{i,k} &= \theta^k(\bar{\lambda}^{i,k+1} - \lambda^{i,k}) \\ &\quad - \theta^k((\theta^{k-1})^{-1} - 1)(\lambda^{i,k} - \lambda^{i,k-1}) \\ &= \theta^k(\bar{\lambda}^{i,k+1} - \bar{\lambda}^{i,k}). \end{aligned} \quad (54)$$

Now, we can prove (39). Consider that $\lambda^{1,k+1} = \lambda^{2,k+1} = \cdots = \lambda^{M,k+1}$ and $\lambda^{i,k+1} = [\sum_{\ell=0}^{T-1} \tau_\ell y^i(\ell, k)]_+ \geq \sum_{\ell=0}^{T-1} \tau_\ell y^i(\ell, k) = \frac{1}{M}\sum_{i=1}^M y^i(0, k)$, it implies that

$$\sum_{i=1}^M \lambda^{i,k+1} \geq \sum_{i=1}^M \tilde{\lambda}^{i,k} + \frac{1}{L_g}\left(\sum_{i=1}^M f^i(x^i, \tilde{\boldsymbol{u}}^{i,k}) - b(\epsilon)\right). \quad (55)$$

Using (54) and (55) can be rewritten as

$$\sum_{i=1}^M \bar{\lambda}^{i,k+1} \geq \sum_{i=1}^M \bar{\lambda}^{i,k} + (\theta^k)^{-1}\frac{1}{L_g}\left(\sum_{i=1}^M f^i(x^i, \tilde{\boldsymbol{u}}^{i,k}) - b(\epsilon)\right) \quad (56)$$

which implies that

$$\sum_{i=1}^M \bar{\lambda}^{i,k+1} \geq \sum_{\ell=1}^k (\theta^\ell)^{-1}\frac{1}{L_g}\left(\sum_{i=1}^M f^i(x^i, \tilde{\boldsymbol{u}}^{i,\ell}) - b(\epsilon)\right). \quad (57)$$

Using (28) and (31), the inequality above can be again rewritten

$$(\theta^k)^2 \sum_{i=1}^M \bar{\lambda}^{i,k+1} \geq \frac{1}{L_g}\left(\sum_{i=1}^M f^i(x^i, \bar{\boldsymbol{u}}^{i,k}) - b(\epsilon)\right). \quad (58)$$

From (38), we can know that $\|\bar{\lambda}^{i,k+1} - \lambda^*\| \leq \sqrt{M}\|\lambda^*\|$ for all $i \in \mathbb{Z}^M$, which from $\|\bar{\lambda}^{i,k+1} - \lambda^*\| \geq \|\bar{\lambda}^{i,k+1}\| - \|\lambda^*\|$ implies that $\|\bar{\lambda}^{i,k+1}\| \leq (\sqrt{M}+1)\|\lambda^*\|$. Consider that $\|\sum_{i=1}^M \bar{\lambda}^{i,k+1}\| \leq \sum_{i=1}^M \|\bar{\lambda}^{i,k+1}\|$. Therefore, from (58), we can get that

$$\begin{aligned}\frac{1}{L_g}\left\|\left[\sum_{i=1}^M f^i(x^i, \bar{\boldsymbol{u}}^{i,k}) - b(\epsilon)\right]_+\right\| &\leq (\theta^k)^2\left\|\sum_{i=1}^M \bar{\lambda}^{i,k+1}\right\| \\ &\leq (\theta^k)^2 M\left(\sqrt{M}+1\right)\|\lambda^*\|. \end{aligned} \quad (59)$$

This, together with $\theta^k \leq \frac{2}{k+2}$ and $\|[\sum_{i=1}^M f^i(x^i, \bar{\boldsymbol{u}}^{i,k}) - b(\epsilon)]_+\| \geq \|[\sum_{i=1}^M f^i(x^i, \bar{\boldsymbol{u}}^{i,k}) - b(\epsilon)]_+\|_\infty$, implies (39). ■

B. Proof of Theorem 1

Consider $g^i(\tilde{\lambda}^{i,k}) = -J^i(x^i, \tilde{\boldsymbol{u}}^{i,k}) - (\tilde{\lambda}^{i,k})^T(f^i(x^i, \tilde{\boldsymbol{u}}^{i,k}) - \frac{b(\epsilon)}{M})$ and $\nabla g^i(\tilde{\lambda}^{i,k}) = -(f^i(x^i, \tilde{\boldsymbol{u}}^{i,k}) - \frac{b(\epsilon)}{M})$ for all $i \in \mathbb{Z}^M$ and $k \geq 0$. For any $\lambda \geq 0$ in (51), it can be shown that $\Delta^i(\lambda, \tilde{\lambda}^{i,k}) = -J^i(x^i, \tilde{\boldsymbol{u}}^{i,k}) - \lambda^T(f^i(x^i, \tilde{\boldsymbol{u}}^{i,k}) - \frac{b(\epsilon)}{M})$ where the notation $\Delta^i(\cdot,\cdot)$ is given in (53). Substitute this into (51), it yields

$$\sum_{i=1}^M g^i(\lambda^{i,k+1}) + (\theta^k)^2\sum_{\ell=0}^k(\theta^\ell)^{-1}\sum_{i=1}^M\left(J^i(x^i, \tilde{\boldsymbol{u}}^{i,\ell}) + \lambda^T\left(f^i(x^i, \tilde{\boldsymbol{u}}^{i,\ell}) - \frac{b(\epsilon)}{M}\right)\right) \leq \frac{L_g M(\theta^k)^2}{2}\|\lambda\|^2 \quad (60)$$

by dropping the quadratic term on the right-hand side. Consider (31) and the fact that $(\theta^k)^2\sum_{\ell=0}^k(\theta^\ell)^{-1}\sum_{i=1}^M J^i(x^i, \tilde{\boldsymbol{u}}^{i,\ell}) \geq \sum_{i=1}^M J^i(x^i, \bar{\boldsymbol{u}}^{i,k})$, (60) implies

$$\sum_{i=1}^M g^i(\lambda^{i,k+1}) + \sum_{i=1}^M\left(J^i(x^i, \bar{\boldsymbol{u}}^{i,k}) + \lambda^T\left(f^i(x^i, \bar{\boldsymbol{u}}^{i,k}) - \frac{b(\epsilon)}{M}\right)\right) \leq \frac{L_g M(\theta^k)^2}{2}\|\lambda\|^2. \quad (61)$$

Let $\lambda = 0$ in (61). It holds that

$$\sum_{i=1}^M g^i(\lambda^{i,k+1}) + \sum_{i=1}^M J^i(x^i, \bar{\boldsymbol{u}}^{i,k}) \leq 0. \quad (62)$$

From (17) and (18), it can be easily verify from the dual problem (17) that $\sum_{i=1}^M g^i(\lambda^{i,k+1}) \geq -\Phi(x, \lambda^*) = -\sum_{i=1}^M J^i(x^i, \boldsymbol{u}^{i*})$ since $\lambda^{1,k+1} = \lambda^{2,k+1} = \cdots = \lambda^{M,k+1}$. Using $\theta^k \leq \frac{2}{k+2}$, the

second inequality of (40) holds. Now let us consider the proof of the first inequality of (40). Note that

$$\begin{aligned}\sum_{i=1}^{M} J^i(x^i,\bar{\boldsymbol{u}}^{i,k}) &= \mathcal{L}(\{\bar{\boldsymbol{u}}^{i,k}\},\lambda^*) - (\lambda^*)^T\left(\sum_{i=1}^{M} f^i(x^i,\bar{\boldsymbol{u}}^{i,k}) - b(\epsilon)\right)\\ &\geq \Phi(x,\lambda^*) - (\lambda^*)^T\left(\sum_{i=1}^{M} f^i(x^i,\bar{\boldsymbol{u}}^{i,k}) - b(\epsilon)\right)\\ &= \sum_{i=1}^{M} J^i(x^i,\boldsymbol{u}^{i*}) - (\lambda^*)^T\left(\sum_{i=1}^{M} f^i(x^i,\bar{\boldsymbol{u}}^{i,k}) - b(\epsilon)\right).\end{aligned} \tag{63}$$

Consider the following inequalities

$$\begin{aligned}(\lambda^*)^T\left(\sum_{i=1}^{M} f^i(x^i,\bar{\boldsymbol{u}}^{i,k}) - b\right) &\leq (\lambda^*)^T\left[\sum_{i=1}^{M} f^i(x^i,\bar{\boldsymbol{u}}^{i,k}) - b(\epsilon)\right]_+\\ &\leq \|\lambda^*\|\left\|\left[\sum_{i=1}^{M} f^i(x^i,\bar{\boldsymbol{u}}^{i,k}) - b(\epsilon)\right]_+\right\|\\ &\leq (\theta^k)^2 L_g M\left(\sqrt{M}+1\right)\|\lambda^*\|^2\end{aligned}$$

where the last inequality is from (59). This inequality, together with (63), implies the first inequality of (40). ■

C. Proof of Lemma 2

(i) Since $\tilde{\boldsymbol{u}}^{i,k} \in \mathcal{U}^i(x^i)$ for all $i \in \mathbb{Z}^M$ and k from (30), it holds that $\bar{\boldsymbol{u}}^{i,k} \in \mathcal{U}^i(x^i)$ for all $i \in \mathbb{Z}^M$ and k. From property (iii) of Lemma 1, there always exits a k such that (41) is satisfied.

(ii) The equivalence between the $(\epsilon,0)$-suboptimal solution and the ϵ-relaxed solution follows from Definition 1 and Theorem 1. ■

D. Proof of Theorem 2

(i) Since $\{x^i,\boldsymbol{u}^i\}_{i=1}^M$ is a ϵ-relaxed solution, it satisfies

$$\boldsymbol{u}^i \in \mathcal{U}^i(x^i), i \in \mathbb{Z}^M, \sum_{i=1}^{M} f^i(x^i,\boldsymbol{u}^i) - b(\epsilon) \leq \epsilon M \mathbf{1}_{pN}. \tag{64}$$

Rewriting $f^i(x^i,\boldsymbol{u}^i)$ and $b(\epsilon)$ back in terms of $\{x_0^i,x_1^i,\ldots,x_N^i\}$ and $\{u_0^i,u_1^i,\ldots,u_{N-1}^i\}$ (note that $f^i(x^i,\boldsymbol{u}^i) - b(\epsilon)$ are simplified expression of (9)), (64) is equivalent to

$$\begin{aligned}\sum_{i=1}^{M} \Psi_x^i x_\ell^i + \Psi_u^i u_\ell^i &\leq (1-\epsilon M(\ell+1))\mathbf{1}_p + \epsilon M \mathbf{1}_p\\ &= (1-\epsilon M(\ell+1))\mathbf{1}_p,\ \forall \ell \in \mathbb{Z}_0^{N-1}.\end{aligned} \tag{65}$$

For all $i \in \mathbb{Z}^M$, let a feasible control to the ith system at next time instant be chosen as

$$\begin{aligned}\boldsymbol{u}^{i+} &:= \{u_0^{i+},u_1^{i+},\ldots,u_{N-1}^{i+}\}\\ &:= \{u_1^i,u_2^i,\ldots,u_{N-1}^i,K^i x_N^i\}\end{aligned} \tag{66}$$

and the associated state sequence $\{x_0^{i+},x_1^{i+},\ldots,x_N^{i+}\} := \{x_1^i,x_2^i,\ldots,x_N^i,(A^i+B^iK^i)x_N^i\}$. It follows from this choice of $\boldsymbol{u}^{i+}$ and (65) that, $\forall \ell \in \mathbb{Z}_0^{N-2}$,

$$\begin{aligned}\sum_{i=1}^{M} \Psi_x^i x_\ell^{i+} + \Psi_u^i u_\ell^{i+} &= \sum_{i=1}^{M} \Psi_x^i x_{\ell+1}^i + \Psi_u^i u_{\ell+1}^i\\ &\leq (1-\epsilon M(1+\ell))\mathbf{1}_p\end{aligned} \tag{67}$$

and where $\ell = N-1$,

$$\sum_{i=1}^{M}(\Psi_x^i x_N^i + \Psi_u^i K^i x_N^i) = \sum_{i=1}^{M} \bar{\Psi} x_N^i \leq (1-\epsilon MN)\mathbf{1}_p \tag{68}$$

where the last inequality follows the fact that $x_N^i \in \sigma_\epsilon^i \mathcal{X}_f^i$ and (10) (with $T_f^i = \sigma_\epsilon^i \mathcal{X}_f^i$). In addition, $u_\ell^{i+} \in U^i$ for $\ell \in \mathbb{Z}_0^{N-2}$ since $u_{\ell+1}^i \in U^i$ because of (64). The last control, $u_{N-1}^{i+} = K^i x_N^i \in U^i$ because $x_N^i \in \sigma_\epsilon^i \mathcal{X}_f^i$ and $\sigma_\epsilon^i \mathcal{X}_f^i$ satisfies (6). The constraints of $x_\ell^{i+} \in X^i, \ell \in \mathbb{Z}_0^{N-1}$ and $x_N^{i+} \in \sigma_\epsilon^i \mathcal{X}_f^i$ follow similar argument. These properties implies $\boldsymbol{u}^{i+} \in \mathcal{U}(x^{i+})$, $i \in \mathbb{Z}^M$ and $\sum_{i=1}^M f^i(x^{i+},\boldsymbol{u}^{i+}) \leq b(\epsilon)$.

(ii) This result follows from Lemma 2 since $x^+ \in \mathcal{D}_\epsilon$. ■

E. Proof of Lemma 3

By definition of $h_{\bar{\sigma}^i X_f^i}$, $\bar{F}_\ell^i x^i \leq \max_{y \in \bar{\sigma}^i X_f^i} \bar{F}_\ell^i y = h_{\bar{\sigma}^i X_f^i}(\bar{F}_\ell^i) = \bar{\sigma}^i h_{X_f^i}(\bar{F}_\ell^i)$ for any $x^i \in \bar{\sigma}^i X_f^i$ and any $\ell \in \mathbb{Z}^{Np}$. This fact, together with $\bar{\sigma}^i \leq \frac{1}{M} b_\ell(\epsilon)/h_{X_f^i}(\bar{F}_\ell^i)$ from the definition of $\{\bar{\sigma}^i\}_{i\in\mathbb{Z}^M}$, implies $\frac{b(\epsilon)}{M} - \bar{F}^i x^i \geq 0$. Since $\bar{\sigma}^i X_f^i \subseteq \sigma_\epsilon^i X_f^i$, $K_A^i x^i \in \mathcal{U}^i(x^i)$ from (5), (6), and (14). Hence, when $k=0$, $K_A^i x^i$ is a feasible solution to $\min_{\boldsymbol{u}^i \in \mathcal{U}^i(x^i)} J^i(x^i,\boldsymbol{u}^i)$, $i \in \mathbb{Z}^M$. Since K^i and P^i are obtained from ARE, $K_A^i x^i$ is the optimal solution to $\min_{\boldsymbol{u}^i} J^i(x^i,\boldsymbol{u}^i)$. Therefore, the solution when $k=0$ is $\tilde{\boldsymbol{u}}^{i,0} = K_A^i x^i$ for all $i \in \mathbb{Z}^M$. This suggests that $\bar{\boldsymbol{u}}^{i,0} = K_A^i x^i$ and $f^i(x^i,K_A^i x^i) - \frac{b(\epsilon)}{M} \leq 0$ for all $i \in \mathbb{Z}^M$, which means that $\{x^i,\bar{\boldsymbol{u}}^{i,0}\}_{i=1}^M$ is a ϵ-relaxed solution and Algorithm 1 terminates at $k=0$. ■

F. Proof of Theorem 3

(i) Consider the ϵ-relaxed solution $\{x^i(t),\bar{\boldsymbol{u}}^{i,\bar{k}(t)}\}_{i=1}^M$ at time t. Let $\boldsymbol{u}_t^i := \{u_{0|t}^i,u_{1|t}^i,\ldots,u_{N-1|t}^i\} = \bar{\boldsymbol{u}}^{i,\bar{k}(t)}$ with the associated predictive state sequence $\boldsymbol{x}_t^i := \{x_{0|t}^i,x_{1|t}^i,\ldots,x_{N|t}^i\}$ for all $i \in \mathbb{Z}^M$. Define the shifted predicted sequence at next time instant $\hat{\boldsymbol{u}}_{t+1}^i = \{u_{1|t}^i,\ldots,u_{N-1|t}^i,K^i x_{N|t}^i\}$, $i \in \mathbb{Z}^M$. From property (i) of Theorem 2, it suggests that $\{\hat{\boldsymbol{u}}_{t+1}^i\}_{i=1}^M$ is a feasible solution to $\mathbb{P}_\epsilon(x(t+1))$ because $x^i(t+1) = A^i x^i(t) + B^i u_{0|t}^i$ using the control law (35).

(ii) This result holds since $\{x^i(t),\bar{\boldsymbol{u}}^{i,\bar{k}(t)}\}_{i=1}^M$ is a $(\epsilon,0)$-upper-relaxed solution of $\mathbb{P}_\epsilon(x(t))$ for all $t \geq 1$.

(iii) Let $V_\epsilon(x(t)) = \sum_{i=1}^M J^i(x^i(t),\boldsymbol{u}_t^{i*})$ (where $\{\boldsymbol{u}_t^{i*}\}_{i=1}^M$ is the optimal solution of (15)) be the Lyapunov function of the closed-loop system of (1) with input $u^i(t) = \bar{u}_0^{i,\bar{k}(t)}$ given by

(35). When Algorithm 1 terminates at time t, $\boldsymbol{u}^{i,\bar{k}(t)} := \boldsymbol{u}_t^i$ and it follows property (iii) of Theorem 2

$$\sum_{i=1}^{M} J^i(x^i(t), \boldsymbol{u}_t^i) - V_\epsilon(x(t)) \leq 0. \tag{69}$$

Using the same shifted control sequence $\hat{\boldsymbol{u}}_{t+1}^i$ at time $t+1$ as defined in (i), it follows from the standard argument in MPC,

$$\begin{aligned} &J^i(x^i(t+1), \hat{\boldsymbol{u}}_{t+1}^i) - J^i(x^i(t), \boldsymbol{u}_t^i) \\ &\quad = -\|x^i(t)\|_{Q^i}^i - \|u^i(t)\|_{R^i}^2 + \|x_{N|t}^i\|_{Q^i}^2 \\ &\qquad + \|K^i x_{N|t}^i\|_{R^i}^2 + \|A_K^i x_{N|t}^i\|_{P^i}^2 - \|x_{N|t}^i\|_{P^i}^2 \\ &\quad = -\|x^i(t)\|_{Q^i}^i - \|u^i(t)\|_{R^i}^2 \end{aligned} \tag{70}$$

where the last equality is from the fact K^i, P^i satisfy the algebraic Riccatii equation of $(A_K^i)^T P^i A_K^i - P^i = -(Q^i + K^i R^i K^i)$. Since $\hat{\boldsymbol{u}}_{t+1}^i$ may not be the optimal at $t+1$,

$$\begin{aligned} V_\epsilon(x(t+1)) &\leq \sum_{i=1}^{M} J^i(x^i(t+1), \hat{\boldsymbol{u}}_{t+1}^i) \\ &= \sum_{i=1}^{M} \left(J^i(x^i(t), \boldsymbol{u}_t^i) - \|x^i(t)\|_{Q^i}^i - \|u^i(t)\|_{R^i}^2 \right) \\ &= V_\epsilon(x(t)) - \sum_{i=1}^{M} \left(\|x^i(t)\|_{Q^i}^2 + \|u^i(t)\|_{R^i}^2 \right) \end{aligned} \tag{71}$$

$$\leq V_\epsilon(x(t)) - \sum_{i=1}^{M} \|x^i(t)\|_{Q^i}^2 \tag{72}$$

where the equality condition follows from (70) and the last inequality is due to (69). Therefore, $x^i(t)$ goes to 0 as $t \to \infty$ for all $i \in \mathbb{Z}^M$. This means that there exists a finite t_f such that $x^i(t_f) \in \bar{\sigma}^i X_f^i$ for all $i \in Z^M$. When this happens, it follows from Lemma 3 that $\kappa^i(x(t_f)) = K^i x^i(t_f)$ and the closed-loop system becomes $x^i(t_f+1) = A_K^i x^i(t_f)$. Since $x^i(t_f+1) \in \bar{\sigma}^i X_f^i$ for any $x^i(t_f) \in \bar{\sigma}^i X_f^i$ from (6). As a result, $x^i(t+1) = A_K^i x^i(t)$ for all $t \geq t_f$ and the closed-loop system is exponentially stable. ■

REFERENCES

[1] T. Keviczky, F. Borrelli, and G. J. Balas, "Decentralized receding horizon control for large scale dynamically decoupled systems," *Automatica*, vol. 42, no. 12, pp. 2105–2115, 2006.

[2] R. Scattolini, "Architectures for distributed and hierarchical model predictive control - A review," *J. Process Control*, vol. 19, no. 5, pp. 723–731, 2009.

[3] C. Wang and C. J. Ong, "Distributed model predictive control of dynamically decoupled systems with coupled cost," *Automatica*, vol. 46, no. 12, pp. 2053–2058, 2010.

[4] P. D. Christofides, R. Scattolini, D. M. Peña, and J. Liu, "Distributed model predictive control: A tutorial review and future research directions," *Comput. Chem. Eng.*, vol. 51, pp. 21–41, 2013.

[5] J. M. Maestre and R. R. Negenborn, *Distributed Model Predictive Control Made Easy*. Berlin, Germany: Springer, 2014.

[6] A. Richards and J. P. How, "Robust distributed model predictive control," *Int. J. Control*, vol. 80, no. 9, pp. 1517–1531, 2007.

[7] P. Trodden and A. Richards, "Distributed model predictive control of linear systems with persistent disturbances," *Int. J. Control*, vol. 83, no. 8, pp. 1653–1663, 2010.

[8] P. Trodden and A. Richards, "Cooperative distributed MPC of linear systems with coupled constraints," *Automatica*, vol. 49, no. 2, pp. 479–487, 2013.

[9] P. Trodden, "Feasible parallel-update distributed MPC for uncertain linear systems sharing convex constraints," *Syst. Control Lett.*, vol. 74, pp. 98–107, 2014.

[10] D. Bertsekas and J. Tsitsiklis, *Parallel and Distributed Computation: Numerical Methods*. Belmont, MA, USA: Athena Scientific, 1997.

[11] D. Bertsekas, *Nonlinear Programming*. Belmont, MA, USA: Athena Scientific, 1999.

[12] I. Necoara and J. A. Suykens, "Application of a smoothing technique to decomposition in convex optimization," *IEEE Trans. Autom. Control*, vol. 53, no. 11, pp. 2674–2679, Dec. 2008.

[13] Q. T. Dinh, C. Savorgnan, and M. Diehl, "Combining lagrangian decomposition and excessive gap smoothing technique for solving large-scale separable convex optimization problems," *Comput. Optim. Appl.*, vol. 55, no. 1, pp. 75–111, 2013.

[14] I. Necoara and V. Nedelcu, "Distributed dual gradient methods and error bound conditions," arXiv:1401.4398, 2014.

[15] J. Li, Z. Wu, C. Wu, Q. Long, and X. Wang, "An inexact dual fast gradient-projection method for separable convex optimization with linear coupled constraints," *J. Optim. Theory Appl.*, vol. 168, no. 1, pp. 153–171, 2016.

[16] Z. Wang and C. J. Ong, "Distributed model predictive control of linear discrete-time systems with coupled constraints," in *Proc. 55th IEEE Conf. Decis. Control*, 2016, pp. 5226–5231.

[17] Z. Wang and C. J. Ong, "Distributed model predictive control of linear discrete-time systems with local and global constraints," *Automatica*, vol. 81, pp. 184–195, 2017.

[18] Y. Nesterov, "A method of solving a convex programming problem with convergence rate O (1/k2)," *Sov. Math. Dokl.*, vol. 27, pp. 372–376, 1983.

[19] D. P. Bertsekas, *Convex Optimization Theory*. Belmont, MA, USA: Athena Scientific, 2009.

[20] P. Patrinos and A. Bemporad, "An accelerated dual gradient-projection algorithm for embedded linear model predictive control," *IEEE Trans. Autom. Control*, vol. 59, no. 1, pp. 18–33, Jan. 2014.

[21] M. Rubagotti, P. Patrinos, and A. Bemporad, "Stabilizing linear model predictive control under inexact numerical optimization," *IEEE Trans. Autom. Control*, vol. 59, no. 6, pp. 1660–1666, Jun. 2014.

[22] D. Jakovetic, J. M. Moura, and J. Xavier, "Distributed Nesterov-like gradient algorithms," in *Proc. IEEE Conf. Decis. Control*, 2012, pp. 5459–5464.

[23] D. Jakovetic, J. Xavier, and J. M. Moura, "Fast distributed gradient methods," *IEEE Trans. Autom. Control*, vol. 59, no. 5, pp. 1131–1146, May 2014.

[24] E. Gilbert and K. T. Tan, "Linear systems with state and control constraints: The theory and application of maximal output admissible sets," *IEEE Trans. Autom. Control*, vol. 36, pp. 1008–1020, Sep. 1991.

[25] Y. Nesterov, "Smooth minimization of non-smooth functions," *Math. Program.*, vol. 103, no. 1, pp. 127–152, 2005.

[26] A. Bemporad, M. Morari, V. Dua, and E. N. Pistikopoulos, "The explicit linear quadratic regulator for constrained systems," *Automatica*, vol. 38, no. 1, pp. 3–20, 2002.

[27] C. Chen, *Linear System Theory and Design*. London, U.K.: Oxford Univ. Press, 1995.

[28] Z. Wang and C. J. Ong, "Speeding up finite-time consensus via minimal polynomial of a weighted graph-a numerical approach," *Automatica*, to appear.

[29] P. Wellstead, *Introduction to Physical System Modelling*. Cambridge, MA, USA: Academic Press, 1979.

[30] E. D. Andersen, C. Roos, and T. Terlaky, "On implementing a primal-dual interior-point method for conic quadratic optimization," *Math. Program.*, vol. 95, no. 2, pp. 249–277, 2003.

[31] A. Nedić, A. Ozdaglar, and P. A. Parrilo, "Constrained consensus and optimization in multi-agent networks," *IEEE Trans. Autom. Control*, vol. 55, no. 4, pp. 922–938, Apr. 2010.

Zheming Wang received the B.S. degree from the Department of Mechanical Engineering and Automation of Shanghai Jiao Tong University, Shanghai, China, in 2012 and the Ph.D. degree from the Department of Mechanical Engineering of National University of Singapore, Singapore, in 2016.

His main research interests include model predictive control and distributed optimization algorithms.

Chong-Jin Ong (SM'12) received the B.Eng. (Hons.) and M.Eng. degrees in mechanical engineering from the National University of Singapore, Singapore, in 1986 and 1988, respectively, and the M.S.E. and Ph.D. degrees in mechanical and applied mechanics from the University of Michigan, Ann Arbor, MI, USA, in 1992 and 1993, respectively.

He joined the National University of Singapore, in 1993 and he is currently an Associate Professor with the Department of Mechanical Engineering. His research interests include model predictive control, robust control, and machine learning.