52nd IEEE Conference on Decision and Control
December 10-13, 2013. Florence, Italy

# On the Convergence Rate of a Jacobi Algorithm for Cooperative Distributed MPC

Dominic Groß and Olaf Stursberg

***Abstract*— This paper investigates the convergence of an iterative distributed model predictive control (DMPC) scheme for linear systems interconnected by dynamics and costs. The DMPC scheme is based on a Jacobi-type iteration and exchange of primal variables. Previous results show that, in the limit, the scheme converges to the Pareto optimal solution but no results on the convergence rate are given. We will first establish a bound on the convergence rate and show that weights used in the scheme and strength of coupling between subsystems have a strong influence on this bound. Subsequently, two approaches to determine the weights are compared. Random numerical examples are used to compare the theoretical bound on the convergence rate with the actual convergence of the scheme.**

## I. INTRODUCTION

Distributed model predictive control of interconnected systems is a challenging problem with many applications, for example in power grids and process control. We consider cooperative distributed MPC (DMPC) for a class of discrete-time linear systems which are either coupled only by costs and subject to local constraints or coupled by states, inputs and costs and subject to local input constraints. Centralized MPC of distributed and large scale systems is usually neither feasible nor desirable because it requires fast transmissions of large amounts of measured data to and from a centralized location, is more sensitive to faults, and leads to large and complex computational problems. On the other hand, fully decentralized MPC schemes may result in significantly degraded performance even if a decentralized control law is feasible.

Distributed model predictive control (DMPC) schemes can be broadly categorized into non-iterative and iterative schemes, in which local problems are solved iteratively and communication occurs in every iteration. In non-iterative schemes, primal variables are optimized and exchanged with interconnected subsystems once per time step, either in parallel (cf. [1], [2]) or in a sequence [3]. In sequential schemes, the overall computation time or the fact that only one subsystem optimizes at each time step may be problematic. To avoid this drawback, parallel non-iterative methods such as [2] are used. However, in both cases convergence to the optimal solution cannot be guaranteed.

Recently iterative schemes based on dual decomposition which guarantee convergence to a feasible solution in a finite number of iterations have been proposed [4]. In [5] accelerated gradient methods and dual decomposition are used to solve mixed $\mathscr{L}_1/\mathscr{L}_2$-problems which result from MPC problems with sparse coupling by costs and constraints and achieves a convergence rate of $\mathcal{O}\left(\frac{1}{p^2}\right)$, with iteration number $p$, for the dual cost and the distance between the current primal iterate and optimal primal solution. However, the implications for the primal cost before the termination of the scheme are not directly clear. In particular, the primal cost may increase from one iteration to the next, which may be problematic if the optimization has to be terminated before the optimum is reached (e.g. due to the time constraints, communication delays, etc.).

In [6] and [7], a distributed MPC scheme which directly optimizes the primal variables for subsystems interconnected by states, inputs, and possibly input constraints (cf. [7]) is investigated. This scheme can be terminated after any iteration, because the cost is guaranteed to be non-increasing in each iteration $p$. While this scheme is guaranteed to converge to the optimal centralized solution of the MPC problem for $p \to \infty$, no convergence rates have been obtained to the best of the authors knowledge. After specifying the DMPC problem, this paper shows that for the DMPC scheme from [6] the primal cost converges at least with $\mathcal{O}(c^p)$ and $0 < c < 1$. Furthermore, these results highlight that weights used in the DMPC scheme play a crucial role. Subsequently, we will derive time-invariant weights, which maximize the convergence speed and depend on the strength of the coupling between subsystems. These results are compared with convergence rates resulting from online optimization of the weights in every iteration of the scheme.

## II. PROBLEM FORMULATION

Consider a set $\mathcal{N} = \{1, \dots, N_s\}$ of interconnected linear subsystems modeled in linear time-invariant discrete-time form. The overall system is given by:

$$x_{k+1} = Ax_k + Bu_k, \tag{1}$$

with $A \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{n\times m}$, and $x_k \in \mathbb{X} \subseteq \mathbb{R}^n$, $u_k \in \mathbb{U} \subseteq \mathbb{R}^m$ are the state and input vector of the overall system. The state and input vector can be partitioned according to the subsystems, such that $x_k = [(x_k^1)^T, \dots, (x_k^{N_s})^T]^T$, $u_k = [(u_k^1)^T, \dots, (u_k^{N_s})^T]^T$, where $x_k^i \in \mathbb{R}^{n_i}$ and $u_k^i \in \mathbb{R}^{m_i}$ are the states and inputs of subsystem $i$. A cost function with finite horizon $N \geq 2$ is used to encode a common objective:

$$V(\mathbf{x}_{k|k}, \mathbf{u}_{k|k}) = \sum_{l=0}^{N-1} \|x_{k+l|k}\|_Q^2 + \|u_{k+l|k}\|_R^2 + \|x_{k+N|k}\|_P^2, \tag{2}$$

where $x_{k+l|k}$ denotes the state at time $k+l$ predicted at time $k$ and $\|x_{k+N|k}\|_P^2 = x_{k+N|k}^T P x_{k+N|k}$. The weighting

Institute of Control and System Theory, Department of Electrical Engineering and Computer Science, University of Kassel (Germany), Email: {dgross,stursberg}@uni-kassel.de

matrices satisfy $Q = Q^T \succeq 0$, $R = R^T \succ 0$, $P = P^T \succ 0$, and the curled inequality symbols $\succ$,$\succeq$ denote strict and non-strict matrix inequalities. The stacked input vector is denoted by $\mathbf{u}_{k|k} = [(\mathbf{u}^1_{k|k})^T, \ldots, (\mathbf{u}^{N_s}_{k|k})^T]^T \in \mathbb{R}^{Nm}$ with $\mathbf{u}^i_{k|k} := [(u^i_{k|k})^T, \ldots, (u^i_{k+N-1|k})^T]^T$ and $\mathbf{x}_{k|k} = [(x_{k|k})^T, \ldots, (x_{k+N|k})^T]^T$ is the vector of predicted states. Based on this model and cost function, the following centralized MPC problem may be formulated:

$$V^*(x_k) := \min_{u_{k+l|k}, x_{k+l|k}} V(\mathbf{x}_{k|k}, \mathbf{u}_{k|k}) \tag{3}$$
$$\text{s.t. } x_{k+l+1|k} = Ax_{k+l|k} + Bu_{k+l|k}, \quad \forall l \in \{0, \ldots, N-1\}$$
$$x_{k+l|k} \in \mathbb{X}, u_{k+l|k} \in \mathbb{U}, \quad \forall l \in \{0, \ldots, N-1\}$$
$$x_{k+N|k} \in \mathbb{T} \subseteq \mathbb{X}$$

where $\mathbb{T}$ is a terminal region, $V^*(x_k)$ denotes the optimal centralized cost of the MPC problem and the corresponding optimizer is given by $\mathbf{u}^*_{k|k}$. We will build on the Jacobi iteration based DMPC scheme first proposed in [6]. Even though extensions of this scheme to more general problems like coupled constraints have been proposed (cf. [7]), the results presented in the following section focus on decoupled constraints and will hold for the following two cases:

**Case 1** [7] The dynamics may be fully coupled, and the pair $(A, B)$ is controllable and is transformed into a distributed model based on its Kalman canonical form. The terminal constraint $\mathbb{T}$ is an equality constraint, and $N$ needs to be sufficiently large to zero the unstable modes of the distributed model. The decoupled input constraints $\mathbb{U} := \mathbb{U}^1(x_k) \times \ldots \times \mathbb{U}^{N_s}(x_k)$ include so called "stability" constraints and are given by polytopes which contain 0 in their interior. Finally, no state constraints are present, i.e. $\mathbb{X} := \mathbb{R}^n$. It is important to note, that the endpoint and "stability" constraints do not result in any coupled input constraints (for details see [7]).

**Case 2** The costs are coupled, the dynamics are decoupled, i.e. $A = \text{blkdiag}(A^1, \ldots, A^{N_s})$ and $B = \text{blkdiag}(B^1, \ldots, B^{N_s})$, and $(A^i, B^i)$ are controllable. The input, state, and terminal constraints are decoupled, i.e. $\mathbb{U} := \mathbb{U}^1 \times \ldots \times \mathbb{U}^{N_s}$, $\mathbb{X} := \mathbb{X}^1 \times \ldots \times \mathbb{X}^{N_s}$, $\mathbb{T} := \mathbb{T}^1 \times \ldots \times \mathbb{T}^{N_s}$ and are given by polytopes containing 0 in their interior.

### A. Distributed MPC Scheme

By substituting the dynamics (1) into the cost function (2), the cost can be rewritten into:

$$V(x_k, \mathbf{u}_{k|k}) = \mathbf{u}^T_{k|k} H \mathbf{u}_{k|k} + x_k^T F \mathbf{u}_{k|k} + x_k^T C x_k, \tag{4}$$

with $H = H^T \succ 0$, $H \in \mathbb{R}^{Nm \times Nm}$, $F \in \mathbb{R}^{n \times Nm}$ and $C \in \mathbb{R}^{n \times n}$ (see Appendix A). Substituting the dynamics into the state and terminal constraints yields local input constraints $\mathbf{u}^i_{k|k} \in \mathbf{U}^i(x_k)$ over the whole prediction horizon for both Case 1 and Case 2. Within the DMPC scheme, a local problem is solved for each subsystem, in each time-step $k$, and in each iteration $p$:

$$\rho^i_{k|k,p} := \arg\min_{\mathbf{u}^i_{k|k}} \quad V(x_k, \mathbf{u}_{k|k}) \tag{5}$$
$$\text{s.t.} \quad \mathbf{u}^i_{k|k} \in \mathbf{U}^i(x_k),\ \mathbf{u}^j_{k|k} := \mathbf{u}^j_{k|k,p}, \forall j \in \mathcal{N} \setminus i,$$

**Algorithm 1:** DMPC iterations in $k$

1: Given $k$, $x_k$, $\mathbf{u}_{k|k,0}$:
2: **for** $p \in [0, \ldots, \tilde{p}_k]$ **do**
3: Solve (5) and communicate $\rho^i_{k|k,p}$ to all subsystems
4: Compute $\mathbf{u}_{k|k,p+1}$ according to (5)
5: **end for**

i.e. each subsystem only optimizes its local input. Afterwards the optimized inputs are exchanged and a Jacobi-type iteration is performed to obtain new input variables for all subsystems:

$$\mathbf{u}_{k|k,p+1} := \sum\nolimits_{i \in \mathcal{N}} w^i_{k,p} \tilde{\mathbf{u}}^i_{k|k,p} \tag{6}$$

where $\tilde{\mathbf{u}}^i_{k|k,p} := [(\mathbf{u}^1_{k|k,p})^T, \ldots, (\rho^i_{k|k,p})^T, \ldots, (\mathbf{u}^{N_s}_{k|k,p})^T]^T$, and the weights $w^i_{k,p} > 0$ have to satisfy $\sum_{i \in \mathcal{N}} w^i_{k,p} = 1$. At each time step the iterative scheme given in Algorithm 1 is performed. The cost is non-increasing and, given a feasible initial guess, all iterates are feasible and stabilizing. Therefore the scheme can be stopped at any iterate $\tilde{p}_k$. The initialization $\mathbf{u}_{k+1|k+1,0}$ for the next time step $k+1$ is obtained based on $\mathbf{u}_{k|k,\tilde{p}_k}$. Furthermore as $p \to \infty$, the cost converges to the centralized optimum $V^*(x_k)$ (cf. [7]). However, no convergence rate is given in [7] and it is not clear, how to choose the weights $w^i_{k,p}$, or how the weights affect the convergence.

## III. Convergence Rate

Let $V_d(x_k, \mathbf{u}_{k|k,p})$ denote the difference between the cost at the current iterate and the optimum:

$$V_d(x_k, \mathbf{u}_{k|k,p}) := V(x_k, \mathbf{u}_{k|k,p}) - V(x_k, \mathbf{u}^*_{k|k}). \tag{7}$$

Convexity of $V(x_k, \mathbf{u}_{k|k,p})$ and $\mathbf{U}^i(x_k)$ implies that:

$$\nabla V(x_k, \mathbf{u}_{k|k,p})^T (\mathbf{u}^*_{k|k} - \mathbf{u}_{k|k,p}) \leq V(x_k, \mathbf{u}^*_{k|k}) - V(x_k, \mathbf{u}_{k|k,p}).$$

This gradient towards the centralized optimum can be decomposed into:

$$\sum_{i \in \mathcal{N}} \nabla V(x_k, \mathbf{u}_{k|k,p})^T \underbrace{\begin{bmatrix} 0 \\ \vdots \\ \Delta \mathbf{u}^{i,i}_{k|k,p} \\ \vdots \\ 0 \end{bmatrix}}_{:= \Delta \mathbf{u}^i_{k|k,p}} \leq -V_d(x_k, \mathbf{u}_{k|k,p}), \tag{8}$$

with $\Delta \mathbf{u}^{i,i}_{k|k,p} := \mathbf{u}^{*i}_{k|k} - \mathbf{u}^i_{k|k,p}$, and $\mathbf{u}^*_{k|k} = [(\mathbf{u}^{*1}_{k|k})^T, \ldots, (\mathbf{u}^{*i}_{k|k})^T, \ldots, (\mathbf{u}^{*N_s}_{k|k})^T]^T$. Because $V(x_k, \mathbf{u}_{k|k,p})$ is convex, it directly follows that:

$$\nabla V(x_k, \mathbf{u}_{k|k,p})^T \Delta \mathbf{u}^i_{k|k,p} \leq 0, \tag{9}$$

and there exist (possibly non-unique) multipliers $\mu^i_{k,p} \geq 0$, $\sum_i \mu^i_{k,p} = 1$ such that:

$$\nabla V(x_k, \mathbf{u}_{k|k,p})^T \Delta \mathbf{u}^i_{k|k,p} \leq -\mu^i_{k,p} V_d(x_k, \mathbf{u}_{k|k,p}). \tag{10}$$

Furthermore, the Hessian $H$ can be rearranged to obtain:

$$\mathbf{u}_{k|k,p}^T H \mathbf{u}_{k|k,p} = \begin{bmatrix} \mathbf{u}_{k|k,p}^i \\ \bar{\mathbf{u}}_{k|k,p}^i \end{bmatrix}^T \begin{bmatrix} H^i & \bar{H}_c^i \\ (\bar{H}_c^i)^T & \bar{H}^i \end{bmatrix} \begin{bmatrix} \mathbf{u}_{k|k,p}^i \\ \bar{\mathbf{u}}_{k|k,p}^i \end{bmatrix}, \quad (11)$$

where $\bar{\mathbf{u}}_{k|k,p}^i \in \mathbb{R}^{N(m-m^i)}$ contains the inputs of all subsystem $j \neq i$.

**Proposition 1** *There exists a constant $c_\lambda^i \geq 1$, such that:*

$$c_\lambda^i (H^i - \bar{H}_c^i (\bar{H}^i)^{-1} (\bar{H}_c^i)^T) \succeq H^i \quad (12)$$

*Proof:* The Schur complement [8] and $H \succ 0$ imply that $H^i \succ 0$, $\bar{H}^i \succ 0$ and $H^i - \bar{H}_c^i (\bar{H}^i)^{-1} (\bar{H}_c^i)^T \succ 0$. Because of $\bar{H}_c^i (\bar{H}^i)^{-1} (\bar{H}_c^i)^T \succeq 0$ it holds that $H^i - \bar{H}_c^i (\bar{H}^i)^{-1} (\bar{H}_c^i)^T \preceq H^i$ and it directly follows that there exists $c_\lambda^i \geq 1$ such that (12) holds. ∎

Note that $\bar{H}_c^i = 0$ (i.e. no coupling in the Hessian) implies $c_\lambda^i = 1$ and $c_\lambda^i \to \infty$ as the strength of coupling increases. The next results states that a bound on the convergence rate can be established which depends on the weights $w_{k,p}^i$, the coupling strength $c_\lambda^i$, and on the multipliers $\mu_{k,p}^i$.

**Theorem 1** *The difference to the optimal cost $V_d(x_k, \mathbf{u}_{k|k,p})$ converges*

$$V_d(x_k, \mathbf{u}_{k|k,p+1}) \leq \beta_{k,p} V_d(x_k, \mathbf{u}_{k|k,p}), \quad (13)$$

*with rate*

$$\beta_{k,p} = 1 - \frac{1}{4} \sum\nolimits_{i \in \mathcal{N}} w_{k,p}^i \frac{(\mu_{k,p}^i)^2}{c_\lambda^i} \quad (14)$$

*and $0 < \beta_{k,p} < 1$.*

*Proof:* Considering the update of the inputs (6), the cost function at every iterate $p+1$ is given by:

$$V(x_k, \mathbf{u}_{k|k,p+1}) = V\left(x_k, \sum\nolimits_{i \in \mathcal{N}} w_{k,p}^i \tilde{\mathbf{u}}_{k|k,p}^i\right). \quad (15)$$

It follows from convexity of $V(x_k, \mathbf{u}_{k|k,p+1})$ that:

$$V(x_k, \mathbf{u}_{k|k,p+1}) \leq \sum\nolimits_{i \in \mathcal{N}} w_{k,p}^i V(x_k, \tilde{\mathbf{u}}_{k|k,p}^i). \quad (16)$$

Applying Taylor's Theorem to $V_d(x_k, \mathbf{u}_{k|k,p})$ developed in $\mathbf{u}_{k|k,p}^*$ yields:

$$V_d(x_k, \mathbf{u}_{k|k,p}) = \nabla V(x_k, \mathbf{u}_{k|k}^*)^T (\mathbf{u}_{k|k,p} - \mathbf{u}_{k|k}^*) + \frac{1}{2} (\mathbf{u}_{k|k,p} - \mathbf{u}_{k|k}^*)^T \nabla^2 V(x_k, \mathbf{u}_{k|k}^*) (\mathbf{u}_{k|k,p} - \mathbf{u}_{k|k}^*) \quad (17)$$

Optimality of $\mathbf{u}_{k|k}^*$ implies that $\nabla V(x_k, \mathbf{u}_{k|k}^*)^T (\mathbf{u}_{k|k,p} - \mathbf{u}_{k|k}^*) \geq 0$ (c.f. [8]) . Furthermore, $\nabla^2 V(x_k, \mathbf{u}_{k|k}^*) = 2H$ is independent of $x_k$ and $\mathbf{u}_{k|k}$ and it directly follows that:

$$V_d(x_k, \mathbf{u}_{k|k,p}) \geq (\mathbf{u}_{k|k,p} - \mathbf{u}_{k|k}^*)^T H (\mathbf{u}_{k|k,p} - \mathbf{u}_{k|k}^*)$$
$$V_d(x_k, \mathbf{u}_{k|k,p}) \geq (\Delta \mathbf{u}_{k|k,p})^T H \Delta \mathbf{u}_{k|k,p} \quad (18)$$

Next, $H$ can be rearranged as in (11) to obtain:

$$V_d(x_k, \mathbf{u}_{k|k,p}) \geq \inf_{\Delta \bar{\mathbf{u}}_{k|k,p}^i} \begin{bmatrix} \Delta \mathbf{u}_{k|k,p}^{i,i} \\ \Delta \bar{\mathbf{u}}_{k|k,p}^i \end{bmatrix}^T \begin{bmatrix} H^i & \bar{H}_c^i \\ (\bar{H}_c^i)^T & \bar{H}^i \end{bmatrix} \begin{bmatrix} \Delta \mathbf{u}_{k|k,p}^{i,i} \\ \Delta \bar{\mathbf{u}}_{k|k,p}^i \end{bmatrix},$$

where $\Delta \bar{\mathbf{u}}_{k|k,p}^i \in \mathbb{R}^{N(m-m^i)}$, and with the minimizer $\Delta \bar{\mathbf{u}}_{k|k,p}^i = -(\bar{H}^i)^{-1} (\bar{H}_c^i) \Delta \mathbf{u}_{k|k,p}^{i,i}$ it holds that:

$$V_d(x_k, \mathbf{u}_{k|k,p}) \geq (\Delta \mathbf{u}_{k|k,p}^{i,i})^T \left(H^i - \bar{H}_c^i (\bar{H}^i)^{-1} (\bar{H}_c^i)^T\right) \Delta \mathbf{u}_{k|k,p}^{i,i}.$$

From Proposition 1 it follows that:

$$c_\lambda^i V_d(x_k, \mathbf{u}_{k|k,p}) \geq (\Delta \mathbf{u}_{k|k,p}^{i,i})^T H^i \Delta \mathbf{u}_{k|k,p}^{i,i}, \quad (19)$$

The inputs can be parametrized by $\mathbf{v}_{k|k,p}^{i,i} = \mathbf{u}_{k|k,p}^i + \alpha^i (\mathbf{u}_{k|k}^{*i} - \mathbf{u}_{k|k,p}^i)$. Because of $\mathbf{u}_{k|k,p}^i \in \mathbf{U}^i(x_k)$ and $\mathbf{u}_{k|k}^{*i} \in \mathbf{U}^i(x_k)$ it holds that $\mathbf{v}_{k|k,p}^{i,i} \in \mathbf{U}^i(x_k), \forall \alpha^i \in [0,1]$. Applying Taylor's Theorem to $V(x_k, \mathbf{u}_{k|k})$ yields:

$$V(x_k, \mathbf{v}_{k|k,p}^i) = V(x_k, \mathbf{u}_{k|k,p}) + \alpha^i \nabla V(x_k, \mathbf{u}_{k|k,p})^T \Delta \mathbf{u}_{k|k,p}^i + (\alpha^i)^2 (\Delta \mathbf{u}_{k|k,p}^i)^T H \Delta \mathbf{u}_{k|k,p}^i,$$

where $\mathbf{v}_{k|k,p}^i = [(\mathbf{u}_{k|k,p}^1)^T, \ldots, (\mathbf{v}_{k|k,p}^{i,i})^T, \ldots, (\mathbf{u}_{k|k,p}^{N_s})^T]^T$. Considering (19) and (10), it holds that:

$$V(x_k, \mathbf{v}_{k|k,p}^i) \leq V(x_k, \mathbf{u}_{k|k,p}) - \alpha^i \mu_{k,p}^i V_d(x_k, \mathbf{u}_{k|k,p}) + (\alpha^i)^2 c_\lambda^i V_d(x_k, \mathbf{u}_{k|k,p}) \quad (20)$$

and

$$V(x_k, \mathbf{v}_{k|k,p}^i) \leq V(x_k, \mathbf{u}_{k|k,p}) + \theta_{k|k,p}^i V_d(x_k, \mathbf{u}_{k|k,p}), \quad (21)$$

with $\theta_{k|k,p}^i = (\alpha^i)^2 c_\lambda^i - \alpha^i \mu_{k,p}^i$. The minimum of $\theta_{k|k,p}^i$ is attained for $\alpha^{i*} = \frac{\mu_{k,p}^i}{2c_\lambda^i} \leq 0.5$, where the last inequality follows from $\mu_i \leq 1$. Substituting $\alpha^{i*}$ into (21) yields:

$$V(x_k, \mathbf{v}_{k|k,p}^i) \leq V(x_k, \mathbf{u}_{k|k,p}) - \frac{(\mu_{k,p}^i)^2}{4c_\lambda^i} V_d(x_k, \mathbf{u}_{k|k,p}).$$

Because of $V(x_k, \tilde{\mathbf{u}}_{k|k,p}^i) \leq V(x_k, \mathbf{v}_{k|k,p}^i)$, the following upper bound on the optimal cost of the local minimization problem (5) is obtained:

$$V(x_k, \tilde{\mathbf{u}}_{k|k,p}^i) \leq V(x_k, \mathbf{u}_{k|k,p}) - \frac{(\mu_{k,p}^i)^2}{4c_\lambda^i} V_d(x_k, \mathbf{u}_{k|k,p}). \quad (22)$$

Finally, substituting (22) back into (16) yields:

$$V(x_k, \mathbf{u}_{k|k,p+1}) \leq \sum_{i \in \mathcal{N}} w_{k,p}^i V(x_k, \mathbf{u}_{k|k,p}) - \frac{(\mu_{k,p}^i)^2}{4c_\lambda^i} V_d(x_k, \mathbf{u}_{k|k,p})$$
$$\leq V(x_k, \mathbf{u}_{k|k,p}) - \sum_{i \in \mathcal{N}} w_{k,p}^i \frac{(\mu_{k,p}^i)^2}{4c_\lambda^i} V_d(x_k, \mathbf{u}_{k|k,p})$$

Subtracting $V(x_k, \mathbf{u}_{k|k}^*)$ on both sides results in (13) and $\beta_{k,p} = \left(1 - \sum_{i \in \mathcal{N}} w_{k,p}^i \frac{(\mu_{k,p}^i)^2}{4c_\lambda^i}\right)$. Furthermore for any $\mu_{k,p}^i \geq 0$, $\sum_{i \in \mathcal{N}} \mu_{k,p}^i = 1$ and any choice of $w_{k,p}^i > 0$, $\sum_{i \in \mathcal{N}} w_{k,p}^i = 1$, it directly follows that $\beta_{k,p} < 1$. ∎

It can be seen that a smaller $\beta_{k,p}$ implies faster convergence and that $\beta_{k,p}$ may strongly depend on the choice of weights $w_{k,p}^i$, which may be time-varying. In the subsequent sections, the condition $w_{k,p}^i > 0$ is relaxed to $w_{k,p}^i \geq 0$ and bounds on the convergence rate $\beta_{k,p}$ for different choices of $w_{k,p}^i$ are discussed. Specifically, time-invariant weights and time-variant weights, which are optimized online, are considered.

## IV. Optimized Weights

### A. Time-Invariant Weights

In the work on the Jacobi-type DMPC scheme from Algorithm 1 no results are given on how to choose the weights $w_{k,p}^i$ to achieve good convergence. Given the convergence rate derived in the previous section, we can now analyze the impact of the weights on the convergence of the overall scheme and derive optimized time-invariant weights.

The convergence rate $\beta_{k,p}$ can be bounded by $\beta_{k,p} \leq \bar{\beta}$:

$$\bar{\beta} = \min_{w_{k,p}^i} \max_{\mu_{k,p}^i} 1 - \frac{1}{4} \sum\nolimits_{i\in\mathcal{N}} w_{k,p}^i \frac{(\mu_{k,p}^i)^2}{c_\lambda^i}, \tag{23}$$

i.e. $\bar{\beta}$ is guaranteed a-priori for all $\mu_{k,p}^i$ (worst-case). We will first solve problem (23) for time-invariant weights $w^i$, i.e. compute time-invariant weights which minimize the bound $\beta_{k,p}$ given in Theorem 1. Subsequently, we investigate a schme for optimizing the weights online.

**Theorem 2** *The optimal solution of* (23) *is given by* $w_{k,p}^i = \sqrt{c_\lambda^i} r^{-1}$ *with* $r = \sum_{i\in\mathcal{N}} \sqrt{c_\lambda^i}$. *The resulting bound on the convergence rate is*

$$\bar{\beta} = 1 - \frac{1}{4} \left( \sum\nolimits_{i\in\mathcal{N}} \sqrt{c_\lambda^i} \right)^{-2} \tag{24}$$

*Proof:* Let $\mu_{k,p} := [\mu_{k,p}^1, \ldots, \mu_{k,p}^{N_s}]^T$ denote the vector of multipliers $\mu_{k,p}^i$ and $w_{k,p} := [w_{k,p}^1, \ldots, w_{k,p}^{N_s}]^T$ denote the vector of weights. Instead of solving the inner maximization problem in (23), the maximizer can be computed by solving the following problem:

$$\min_{\mu_{k,p}} \sum\nolimits_{i\in\mathcal{N}} w_{k,p}^i \frac{(\mu_{k,p}^i)^2}{c_\lambda^i}, \text{ s.t. } \mathbf{1}_{N_s}^T \mu_{k,p} = 1, \ \mu_{k,p} \geq 0. \tag{25}$$

where $\mathbf{1}_{N_s} \in \mathbb{R}^{N_s}$ is a vector of ones. The KKT Conditions [8] with multipliers $\xi_i$ and $\nu$ for problem (25) are given by:

$$\frac{2w_{k,p}^i}{c_\lambda^i} \mu_{k,p}^i - \xi_i + \nu = 0, \quad \forall i \in \{1, \ldots, N_s\} \tag{26}$$

$$\xi_i \geq 0, \quad \forall i \in \{1, \ldots, N_s\} \tag{27}$$

$$\xi_i(-\mu_{k,p}^i) = 0, \quad \forall i \in \{1, \ldots, N_s\} \tag{28}$$

$$-\mu_{k,p}^i < 0, \quad \forall i \in \{1, \ldots, N_s\} \tag{29}$$

$$\mathbf{1}^T \mu_{k,p} = 1. \tag{30}$$

It can be verified that the KKT conditions hold for $\mu_{k,p}^i = z\frac{c_\lambda^i}{w_{k,p}^i}$ and $z = \left(\sum_{i\in\mathcal{N}} \frac{c_\lambda^i}{w_{k,p}^i}\right)^{-1}$, $\nu = -2z$ and $\xi_i = 0$. Since the problem is convex, the KKT conditions are sufficient for optimality. Thus, the parametric minimizer is given by $\mu_{k,p}^i(w_{k,p}^i, c_\lambda^i) = z\frac{c_\lambda^i}{w_{k,p}^i}$ and $z = \left(\sum_{i\in\mathcal{N}} \frac{c_\lambda^i}{w_{k,p}^i}\right)^{-1}$. Substituting $\mu_{k,p}^i = z\frac{c_\lambda^i}{w_{k,p}^i}$ into (23) yields:

$$\bar{\beta} = \min_{w_{k,p}} 1 - \frac{1}{4} z^2 \sum\nolimits_{i\in\mathcal{N}} \frac{c_\lambda^i}{w_{k,p}^i} \tag{31}$$

$$\bar{\beta} = \min_{w_{k,p}} 1 - \frac{1}{4 \sum_{i\in\mathcal{N}} \frac{c_\lambda^i}{w_{k,p}^i}}. \tag{32}$$

It remains to find weights which solve the minimization problem. Note that the optimizer $w_{k,p}^*$ of (32) is identical to:

$$w_{k,p}^* = \arg\min_{w_{k,p}} \sum_{i\in\mathcal{N}} \frac{c_\lambda^i}{w_{k,p}^i}, \text{ s.t. } \mathbf{1}_{N_s}^T w_{k,p} = 1, \ w_{k,p}^i \geq 0. \tag{33}$$

The KKT conditions of this problem are given by:

$$-\frac{c_\lambda^i}{(w_{k,p}^i)^2} - \xi_i + \nu = 0, \quad \forall i \in \{1, \ldots, N_s\} \tag{34}$$

$$\xi_i \geq 0, \quad \forall i \in \{1, \ldots, N_s\} \tag{35}$$

$$\xi_i(-w_{k,p}^i) = 0, \quad \forall i \in \{1, \ldots, N_s\} \tag{36}$$

$$-w_{k,p}^i \leq 0, \quad \forall i \in \{1, \ldots, N_s\} \tag{37}$$

$$\mathbf{1}_{N_s}^T w_{k,p} = 1. \tag{38}$$

The KKT conditions are satisfied for $w_{k,p}^i = \sqrt{c_\lambda^i} r^{-1}$ with $r = \sum_{i\in\mathcal{N}} \sqrt{c_\lambda^i}$, $\xi_i = 0$ and $\nu = r^2$. Substituting into (32) yields (24). ∎

### B. Online Optimization of Weights

To improve performance, one may optimize over $w_{k,p}^i$ online after exchanging the planned input of each subsystem. However, optimizing the weights online requires full information exchange, even if the subsystems are decoupled. In such an implementation, one would optimize the weights $w_{k,p}^i$ for given $\tilde{\mathbf{u}}_{k|k,p}^i$, i.e. solve:

$$w_{k,p}^* = \arg\min_{w_{k,p}} J\left(x_k, \sum\nolimits_{i\in\mathcal{N}} w_{k,p}^i \tilde{\mathbf{u}}_{k|k,p}^i\right). \tag{39}$$

**Theorem 3** *When optimizing the weights* $w_{k,p}^i$ *in every iteration, the bound on the convergence rate* $\beta_{k,p}$ *in the worst-case is also given by* (24).

*Proof:* In terms of the convergence rate $\beta_{k,p}$, optimizing the weights online results in the following problem:

$$\min_{w_{k,p}} 1 - \frac{1}{4} \sum_{i\in\mathcal{N}} w_{k,p}^i \frac{(\mu_{k,p}^i)^2}{c_\lambda^i}, \text{ s.t. } \mathbf{1}_{N_s}^T w_{k,p} = 1, \ w_{k,p}^i \geq 0,$$

for the parametric minimizers $w_{k,p}^i(\mu_{k,p}^i, c_\lambda^i)$ for all $i$. The corresponding KKT conditions are given by:

$$-\frac{(\mu_{k,p}^i)^2}{c_\lambda^i} - \xi_i + \nu = 0, \quad \forall i \in \{1, \ldots, N_s\} \tag{40}$$

$$\xi_i(-w_{k,p}^i) = 0, \quad \forall i \in \{1, \ldots, N_s\} \tag{41}$$

$$-w_{k,p}^i \leq 0, \quad \forall i \in \{1, \ldots, N_s\} \tag{42}$$

$$\mathbf{1}_{N_s}^T w_{k,p} = 1. \tag{43}$$

They hold for $\nu = (\mu_{k,p}^j)^2 (c_\lambda^j)^{-1}$, $\xi_s = -(\mu_{k,p}^s)^2 (c_\lambda^s)^{-1} + (\mu_{k,p}^j)^2 (c_\lambda^j)^{-1}$ , $w_s = 0$, for all $s \in \mathcal{N} \setminus j^*$ and $\xi_{j^*} = 0$, $w_{j^*} = 1$ and $j^* = \arg\max_i (\mu_{k,p}^i)^2 (c_\lambda^i)^{-1}$. Since the problem is convex, the KKT conditions are sufficient. Substituting into (23) yields:

$$\bar{\beta} = 1 - \frac{1}{4} \min_{\mu_{k,p}} \max_i \frac{(\mu_{k,p}^i)^2}{c_\lambda^i}, \text{ s.t. } \mathbf{1}_{N_s}^T \mu_{k,p} = 1, \mu_{k,p}^i \geq 0. \tag{44}$$

A feasible solution is given by $\mu^i_{k,p} = \sqrt{c^i_\lambda} r^{-1}$ and it follows that:

$$(\mu^1_{k,p})^2 (c^1_\lambda)^{-1} = \ldots = (\mu^{N_s}_{k,p})^2 (c^{N_s}_\lambda)^{-1} = r^{-2} > 0. \quad (45)$$

Next, consider multipliers $\tilde{\mu}^i_{k,p} = \sqrt{c^i_\lambda} r^{-1} + \epsilon^i$, in order to satisfy the constraint $\sum_i \tilde{\mu}^i_{k,p} = 1$ it is necessary that $\sum_i \epsilon^i = 0$. For any perturbation $\epsilon^i$, $\sum_i \epsilon^i = 0$ this implies:

$$\max \left\{ \frac{(\mu^1_{k,p})^2}{c^1_\lambda}, \ldots, \frac{(\mu^{N_s}_{k,p})^2}{c^{N_s}_\lambda} \right\} \leq \max \left\{ \frac{(\tilde{\mu}^1_{k,p})^2}{c^1_\lambda}, \ldots, \frac{(\tilde{\mu}^{N_s}_{k,p})^2}{c^{N_s}_\lambda} \right\}.$$

In other words, $\mu^i_{k,p} = \sqrt{c^i_\lambda} r^{-1}$ is the optimal solution. Substituting into (44) results in (24). ∎

### C. Benefits of Online Optimization

The results in the previous section show that optimizing the weights $w^i_{k,p}$ online does not result in an improved a-priori bound $\bar{\beta}$. However, performance may be improved by online optimization of the weights if the current multipliers $\mu^i_{k,p}$, which depend on the current state and input, allow for fast convergence. This is made precise in the following two Theorems:

**Theorem 4** *Given time-invariant weights $w^i_{k,p} = \sqrt{(}c^i_\lambda) r^{-1}$ the multipliers $\mu^{j^*}_{k,p} = 1$ with $j^* = \arg\max_j \frac{1}{r\sqrt{c^j_\lambda}}$ and $\mu^j_{k,p} = 0$ for all $j \in \mathcal{N} \setminus j^*$ minimize $\bar{\beta}$ and the minimum is given by:*

$$\bar{\beta}_{\min} = 1 - \frac{1}{4} \left( r \min_i \sqrt{c^i_\lambda} \right)^{-1}. \quad (46)$$

*Proof:* Substituting $w^i_{k,p} = \sqrt{c^i_\lambda} r^{-1}$ into (14) and minimizing the bound $\bar{\beta}$ for the multipliers $\mu^i_{k,p}$ instead of maximizing results in:

$$\mu^*_{k,p} = \arg\max_{\mu^i_{k,p}} \sum_{i \in \mathcal{N}} \frac{(\mu^i_{k,p})^2}{r\sqrt{c^i_\lambda}}, \text{ s.t. } \mathbf{1}^T \mu_{k,p} = 1, \mu_{k,p} \geq 0, \quad (47)$$

i.e. the maximization of a convex function, which is a non-convex problem. If the feasible set is a bounded polytope the optimal value is attained at a vertex of the feasible set[1]. Only the vertices of the box $0 \leq \mu^i_{k,p} \leq 1$ that also lie in the subspace $\mathbf{1}^T \mu_{k,p} = 1$ are vertices of the feasible set of (47). It directly follows that the vertices of the feasible set are given by:

$$v_j = \begin{cases} v^i_j = 0 & \text{if } i \neq j \\ v^i_j = 1 & \text{if } i = j \end{cases}, \quad \forall j \in \{1, \ldots, N_s\}. \quad (48)$$

Thus the maximum is given by $\mu^{j_1}_{k,p} = 1$ with $j_1 = \arg\max_i \frac{1}{r\sqrt{c^i_\lambda}}$ and substituting into (47) results in (46). ∎
Similarly, one can minimize the bound on the convergence rate in the case of online optimization of the weights.

[1]Proof: consider any point $\bar{v}$ on a line segment between two vertices $v_i$ and $v_j$, assume that the cost at $\bar{v}$ is higher than the cost at $v_i$ and $v_j$. This contradicts the convexity of the cost.

**Theorem 5** *If the weights are optimized online the bound $\bar{\beta}$ is minimized by $\mu^{j^*}_{k,p} = 1$ with $j^* = \arg\min_i c^i_\lambda$ and $\mu^i_{k,p} = 0$ for all $i \in \mathcal{N} \setminus j^*$. The minimum is given by:*

$$\bar{\beta}_{\min} = 1 - \frac{1}{4} \left( \min_i c^i_\lambda \right)^{-1}. \quad (49)$$

*Proof:* In the proof of Theorem 3 it was shown that the optimal weights are $w_s = 0$, for all $s \in \mathcal{N} \setminus j^*$ and $\xi_{j^*} = 0$, $w_{j^*} = 1$ and $j^* = \arg\max_i (\mu^i_{k,p})^2 (c^i_\lambda)^{-1}$ (cf. (40)-(43)). Substituting into (14) and minimizing for the multipliers $\mu^i_{k,p}$ results in:

$$\bar{\beta}_{\min} = 1 - \frac{1}{4} \max_{\mu^i_{k,p}} \max_i \frac{(\mu^i_{k,p})^2}{c^i_\lambda} \text{ s.t. } \mathbf{1}^T_{N_s} \mu_{k,p} = 1, \mu^i_{k,p} \geq 0.$$

It can be seen that the optimal solution is given by $\mu^j_{k,p} = 1$ with $j = \arg\min_i c^i_\lambda$. Substituting into (14) completes the proof. ∎

### D. Discussion

The results in the previous sections reveal a strong connection between the weights $w^i_{k,p}$, the strength of the coupling between subsystems indicated by $c^i_\lambda$, and the convergence rate $\beta_{k,p}$. Futhermore, due to the summation over $c^i_\lambda > 1$ in (24) the bounds on the convergence rate $\bar{\beta}$ depends on the number of subsystems. This could be expected, because of the condition $\sum_{i \in \mathcal{N}} w^i_{k,p} = 1$, which limits the progress of each subproblem as the number of subproblems increases. Together with the impact of $c^i_\lambda$, this shows that the convergence of the scheme can be improved by solving larger subproblems, which also optimize inputs of some interconnected subsystems. To the best of the authors' knowledge, no results have been published on how to effectively re-partition the overall subsystem to achieve faster convergence if there exists some freedom in choosing the partition of the overall system. The results of Theorem 1, in particular (19) and the preceeding analysis, link the convergence of the scheme to the structure of the Hessian – this gives some insight into the problem of partitioning the overall system.

With respect to optimizing the weights online, it can be seen that this does not result in better a-priori guarantees for the convergence. However, faster convergence may be achieved by optimizing $w^i_{k,p}$ online. In contrast to the time-invariant weights discussed above, the convergence of the scheme may no longer depend on the number of subsystems when optimizing the weights online (cf. Theorem 5). For instance, if the inputs of only a few subsystems are far away from the optimum. It has to be noted, that the results are somewhat conservative, especially in the case of Theorem 5. If there exists a subsystem $j$ with $c^j_\lambda = 1$ (i.e. no coupling), then $\beta_{k,p} \leq 0.75$. At the same time, if only the input of subsystem $j$ is not optimal than the cost will actually converge after one step. The best weight for this case, $w^j_{k,p} = 1$ and $w^s_{k,p} = 0$ for all $s \neq j$, is correctly identified in the proofs of Theorem 3 and 5. However, the possible improvement of the cost by subsystem $j$ in this case is underestimated through (22) by a factor of 4. Thus

the reason for the discrepancy in the convergence rate in this case are the bounds established in the proof of Theorem 1.

## V. Numerical Example

To compare the convergence rates obtained in the previous sections to the actual convergence rate, the scheme given in Algorithm 1 is applied to DMPC problems with decoupled dynamics, local constraints, and potentially fully coupled costs. Specifically, the subsystems are modeled by double-integrator dynamics with local input and state constraints. To evaluate the influence of the number of subsystems Algorithm 1 was used to solve problems ranging from $N_s = 2$ to $N_s = 25$ subsystems. For each number of subsystems $40$ problems with random cost function and initial conditions were generated. The scheme was initialized with a decentralized solution, i.e. by ignoring all interconnections and 20 iterations were performed for each problem. The actual rate of convergence $\gamma_{k,p} = \frac{V_d(x_k,\mathbf{u}_{k|k,p+1})}{V_d(x_k,\mathbf{u}_{k|k,p})}$ was computed and compared to the bound $\bar{\beta}$ given in the previous sections. In Table I the mean, minimal and maximal ratio (taken over all random examples and iterations) between the actual rate of convergence $\gamma_{k,p}$ and bound $\bar{\beta}$ is given for the case of time-invariant weights according to Theorem 2, and online optimization of the weights. It can be seen that the bound is rather conservative, if the number of susbsystems is small. For instance, for $N_s = 2$ the scheme converges approximately twice as fast as indicated by Theorem 2. However, as the number of subsystems increases, the actual convergence approaches the bound $\bar{\beta}$. Futhermore, optimizing the weights online on average only results in slightly faster convergence for the random examples considered here.

## VI. Conclusions and Future Works

In this paper, the convergence properties for a DMPC scheme for interconnected linear systems based on a nonlinear Jacobi-type iteration where investigated. A bound on the convergence rate was derived, which depends on weighting factors used in the scheme and the strength of the coupling in the Hessian of the centralized problem. Based on the convergence rate, the problem of finding time-invariant weights which give the best a-priori convergence guarantee is solved and compared to online optimization of the weights. The results suggest, that the performance of the scheme strongly depends on the number of subsystems. In particular, the convergence is slow when the number of subsystems is large. This observation is confirmed by numerical experiments, in which DMPC problems with random coupling by costs were solved for different numbers of subsystems. Reducing the communication requirements of the DMPC scheme by event-based approaches is seen as interesting area for future research.

TABLE I: Ratio $\frac{\bar{\beta}}{\gamma_{k,p}}$ between the bound $\bar{\beta}$ and actual convergence rate $\gamma_{k,p}$.

| $N_s$ | Time-invariant weights | | | Optimized weights | | |
|---|---|---|---|---|---|---|
| | max | mean | min | max | mean | min |
| 2 | 2.335 | 2.072 | 1.944 | 20.98 | 2.519 | 1.944 |
| 3 | 1.533 | 1.418 | 1.294 | 1.665 | 1.446 | 1.322 |
| 4 | 1.354 | 1.322 | 1.298 | 1.446 | 1.362 | 1.317 |
| 5 | 1.263 | 1.225 | 1.174 | 1.299 | 1.242 | 1.177 |
| 7 | 1.176 | 1.135 | 1.099 | 1.191 | 1.142 | 1.101 |
| 10 | 1.121 | 1.101 | 1.071 | 1.128 | 1.105 | 1.073 |
| 15 | 1.066 | 1.059 | 1.052 | 1.073 | 1.06 | 1.053 |
| 20 | 1.051 | 1.043 | 1.034 | 1.058 | 1.044 | 1.035 |
| 25 | 1.044 | 1.039 | 1.03 | 1.049 | 1.04 | 1.03 |

## VII. Acknowledgments

Partial financial support by the German Science Foundation (DFG) within the research priority program SPP 1305 *Control Theory of Digitally Networked Dynamical Systems* and through the EU-funded NoE HYCON 2 is gratefully acknowledged.

## Appendix

### A. Problem Transformation

Note, that all future states can be expressed as follows:

$$\underbrace{\begin{bmatrix} x_{k|k} \\ x_{k+1|k} \\ \vdots \\ x_{k+N|k} \end{bmatrix}}_{\mathbf{x}_{k|k}} = \underbrace{\begin{bmatrix} I_n \\ A \\ \vdots \\ A^N \end{bmatrix}}_{=:\mathbf{A}} x_k + \underbrace{\begin{bmatrix} 0 & 0 & \cdots & 0 \\ B & 0 & \cdots & 0 \\ AB & B & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ A^{N-1}B & A^{N-2}B & \cdots & B \end{bmatrix}}_{=:\mathbf{B}} \mathbf{Z}\, \mathbf{u}_{k|k,p}, \quad (50)$$

where $\mathbf{Z} \in \mathbb{R}^{Nm \times Nm}$ is a permutation matrix, such that $\mathbf{Z}\mathbf{u}_{k|k,p} = [(u_{k|k})^T, \ldots, (u_{k+N-1|k})^T]^T$. With the Kronecker product $\otimes$, the cost over the prediction horizon $N$ is given by $V(x_k, \mathbf{u}_{k|k,p}) = \mathbf{x}_{k|k}^T \mathbf{Q} \mathbf{x}_{k|k} + \mathbf{u}_{k|k}^T \mathbf{R} \mathbf{u}_{k|k}$, with $\mathbf{Q} := \text{blkdiag}(I_N \otimes Q, P)$, $\mathbf{R} := \mathbf{Z}^T (I_N \otimes R) \mathbf{Z}$, where $I_N$ denotes the identity matrix of dimension $N$. Substituting (50) one obtains $H := \mathbf{B}^T \mathbf{Q} \mathbf{B} + \mathbf{R}$, $F := 2\mathbf{A}^T \mathbf{Q} \mathbf{B}$, $C := \mathbf{A}^T \mathbf{Q} \mathbf{A}$.

## References

[1] D. Jia and B. Krogh, "Min-max feedback model predictive control for distributed control with communication," in *Proc. Amer. Contr. Conf.*, 2002, pp. 4507–4512.

[2] D. Groß and O. Stursberg, "Robust distributed predictive control of communicating and constrained systems," in *Proc. 18th IFAC World Congress*, 2011, pp. 8926–8932.

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

[4] M. D. Doan, T. Keviczky, and B. De Schutter, "A distributed optimization-based approach for hierarchical mpc of large-scale systems with coupled dynamics and constraints," in *Decision and Control and European Control Conference (CDC-ECC), 2011 50th IEEE Conference on*, dec. 2011, pp. 5236 –5241.

[5] P. Giselsson, M. D. Doan, T. Keviczky, B. D. Schutter, and A. Rantzer, "Accelerated gradient methods and dual decomposition in distributed model predictive control," *Automatica*, vol. 49, no. 3, pp. 829 – 833, 2013.

[6] A. Venkat, I. Hiskens, J. Rawlings, and S. Wright, "Distributed MPC strategies with application to power system automatic generation control," *IEEE Trans. Contr. Sys. Tech.*, vol. 16, no. 6, pp. 1192–1206, 2008.

[7] B. T. Stewart, A. N. Venkat, J. B. Rawlings, S. J. Wright, and G. Pannocchia, "Cooperative distributed model predictive control," *Systems & Control Letters*, vol. 59, no. 8, pp. 460 – 469, 2010.

[8] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.