

# On the Convergence of Overlapping Schwarz Decomposition for Nonlinear Optimal Control

Sen Na, Sungho Shin, Mihai Anitescu, and Victor M. Zavala

**Abstract**—We study the convergence properties of an overlapping Schwarz decomposition algorithm for solving nonlinear optimal control problems (OCPs). The algorithm decomposes the time domain into a set of overlapping subdomains, and solves all subproblems defined over subdomains in parallel. The convergence is attained by updating primal-dual information at the boundaries of overlapping subdomains. We show that the algorithm exhibits local linear convergence, and that the convergence rate improves exponentially with the overlap size. We also establish global convergence results for a general quadratic programming, which enables the application of the Schwarz scheme inside second-order optimization algorithms (e.g., sequential quadratic programming). The theoretical foundation of our convergence analysis is a sensitivity result of nonlinear OCPs, which we call “exponential decay of sensitivity” (EDS). Intuitively, EDS states that the impact of perturbations at domain boundaries (i.e. initial and terminal time) on the solution decays exponentially as one moves into the domain. Here, we expand a previous analysis available in the literature by showing that EDS holds for both primal and dual solutions of nonlinear OCPs, under uniform second-order sufficient condition, controllability condition, and boundedness condition. We conduct experiments with a quadrotor motion planning problem and a PDE control problem to validate our theory; and show that the approach is significantly more efficient than ADMM and as efficient as the centralized solver Ipopt.

**Index Terms**—Optimal Control; Nonlinear Programming; Decomposition Methods; Overlapping; Parallel algorithms

## I. INTRODUCTION

We study the nonlinear optimal control problem (OCP):

$$\min_{\{\mathbf{x}_k\}, \{\mathbf{u}_k\}} \sum_{k=0}^{N-1} g_k(\mathbf{x}_k, \mathbf{u}_k) + g_N(\mathbf{x}_N), \quad (1a)$$

$$\text{s.t. } \mathbf{x}_{k+1} = f_k(\mathbf{x}_k, \mathbf{u}_k) \quad (\boldsymbol{\lambda}_k), \quad (1b)$$

$$\mathbf{x}_0 = \bar{\mathbf{x}}_0 \quad (\boldsymbol{\lambda}_{-1}), \quad (1c)$$

where  $\mathbf{x}_k \in \mathbb{R}^{n_x}$  are the state variables;  $\mathbf{u}_k \in \mathbb{R}^{n_u}$  are the control variables;  $\boldsymbol{\lambda}_k \in \mathbb{R}^{n_x}$  are the dual variables associated with the dynamics (1b);  $\boldsymbol{\lambda}_{-1} \in \mathbb{R}^{n_x}$  are the dual variables associated with the initial conditions (1c);  $g_k : \mathbb{R}^{n_x} \times \mathbb{R}^{n_u} \rightarrow \mathbb{R}$

( $g_N : \mathbb{R}^{n_x} \rightarrow \mathbb{R}$ ) are the cost functions;  $f_k : \mathbb{R}^{n_x} \times \mathbb{R}^{n_u} \rightarrow \mathbb{R}^{n_x}$  are the dynamical constraint functions;  $N$  is the horizon length; and  $\bar{\mathbf{x}}_0 \in \mathbb{R}^{n_x}$  is the given initial state. We assume that  $f_k, g_k$  are twice continuously differentiable, nonlinear, and possibly nonconvex; as such, (1) is a nonconvex nonlinear program (NLP). The problem of interest has been studied extensively in the context of model predictive control [1] with applications in chemical process control [2], energy systems [3], production planning [4], autonomous vehicles [5], power systems [6], supply chains [7], and neural networks [8].

In this work, we are interested in solving OCPs with a large number of stages  $N$ . Such problems arise in the settings with long horizons, fine time discretization resolutions, and multiple timescales [9]. Temporal decomposition provides an approach to deal with such problems. In this approach, one partitions the time domain  $[0, N]$  into a set of subdomains  $\{[m_i, m_{i+1}]\}_{i=0}^{T-1}$ . One then solves more tractable subproblems over subdomains in parallel, and their solution trajectories are concatenated by using a coordination mechanism. Traditional coordination mechanisms include Lagrangian dual decomposition [10], alternating direction method of multipliers (ADMM) [11], dual dynamic programming [12], and Jacobi/Gauss-Seidel methods [13]. These decomposition approaches offer flexibility in that they can be implemented in different types of computing hardware that might have limitations on memory and processor speeds. This is critical because the performance of centralized nonlinear optimization solvers (e.g., Ipopt) degrades rapidly in resource-constrained computing environments [14]. Unfortunately, while Lagrangian dual decomposition, ADMM, and dual dynamic programming are guaranteed to converge under a variety of OCP settings, they often exhibit slow convergence [15]. This highlights the existence of a fundamental trade-off between the flexibility offered by distributed solvers and the efficiency offered by centralized solvers.

Direct decomposition approaches have also been studied for convex OCPs with long horizons. Specifically, such approaches have been used to decompose linear algebra systems inside interior-point solvers [16]–[23]. They also offer flexibility to enable limited-resource-hardware implementations and, since the methods are direct (as opposed to iterative), they do not suffer from convergence issues. However, direct approaches rely on reduction procedures (they are block elimination techniques), and such procedures suffer from scalability issues. For instance, parallel cyclic reduction, Schur, and Riccati decompositions do not scale well with the number of states and/or control variables. Moreover, we also highlight that iterative approaches such as ADMM and Lagrangian dual decomposition often offer more flexibility than direct decomposition methods in that the

S. Na and S. Shin equally contributed to this work.

S. Na is with the Department of Statistics, University of Chicago, 5747 South Ellis Avenue, Chicago, IL 60637, USA (e-mail: senna@uchicago.edu)

S. Shin is with the Department of Chemical and Biological Engineering, University of Wisconsin-Madison, Madison, WI 53706 USA (e-mail: sungho.shin@wisc.edu)

M. Anitescu is with the Mathematics and Computer Science Division, Argonne National Laboratory, Lemont, IL 60439, USA, and also with the Department of Statistics, University of Chicago, Chicago, IL 60637, USA (e-mail: anitescu@mcs.anl.gov)

V. M. Zavala is with the Department of Chemical and Biological Engineering, University of Wisconsin-Madison, Madison, WI 53706 USA and also with the Mathematics and Computer Science Division, Argonne National Laboratory, Lemont, IL 60439, USA (e-mail: victor.zavala@wisc.edu)

amount of communication needed is limited (thus preserving data privacy).

A recent study [24] has empirically tested the effectiveness of a different decomposition paradigm for OCPs. Specifically, the authors performed numerical tests with a *temporal decomposition scheme with overlaps* (see Fig. 1). Here, overlapping subdomains  $\{[n_i^1, n_i^2]\}_{i=0}^{T-1}$  are constructed by expanding the non-overlapping subdomains  $\{[m_i, m_{i+1}]\}_{i=0}^{T-1}$  by  $\tau$  stages on the left and right boundaries. Subproblems on the expanded subdomains are solved in parallel, and the resulting solution trajectories are concatenated by discarding the pieces of the trajectory in the overlapping regions. The authors observed that, as *the size of the overlap* increases, the approximation error of the concatenated solution trajectory drops rapidly. However, no quantitative analysis was provided. Subsequent work [25] provided the first rigorous error analysis of such overlapping decomposition scheme. The authors proved that, for strongly convex OCPs with linear dynamics and positive-definite quadratic stage costs that satisfy uniform controllability and boundedness conditions, the error of the concatenated trajectory decreases *exponentially* in  $\tau$ . This result requires a sensitivity property for convex OCPs that we call “exponential decay of sensitivity” (EDS). This property says that the impact of parametric perturbations on the primal solution trajectory  $\{(\mathbf{x}_k^*, \mathbf{u}_k^*)\}_k$  decays exponentially as one moves away from the perturbation stage. Unfortunately, the analysis in [25] does not apply for the general nonlinear OCP (1) and thus has limited applicability. Furthermore, we emphasize that the sensitivity on the dual solution (even for convex case) is not resolved in that work, and that the decomposition scheme analyzed there is only an approximation scheme (not a convergent algorithm).

Recent work [26] has applied the overlapping decomposition scheme for solving time-invariant nonlinear OCPs. This relies on the observation that such a decomposition scheme can be interpreted as a *single iteration* of an overlapping Schwarz decomposition scheme. In particular, for solving nonlinear OCPs, [26] partitions the time domain as in [24], [25], but utilizes both primal and dual information from adjacent subdomains to perform an iterative coordination to achieve the convergence. The authors of [26] conjectured that the effect of perturbations at two ends (that is initial and terminal stages) on the *primal and dual* trajectory  $\{(\mathbf{x}_k^*, \mathbf{u}_k^*, \boldsymbol{\lambda}_k^*)\}_k$  (not only for the primal trajectory as in [25]) decays *asymptotically*. Under this conjecture, they proved that the overlapping Schwarz scheme converges locally. The authors also provided empirical evidence with a nonlinear OCP that, the perturbation effect decays not only asymptotically, but indeed exponentially. That is, EDS empirically holds for nonlinear OCPs just like for convex quadratic OCPs as in [25], although a theoretical justification for such behavior was not provided.

The work in [27] investigated primal sensitivity for nonlinear OCPs. The authors showed that, under uniform second-order sufficient condition, controllability condition, and boundedness condition, EDS holds for primal solution of nonlinear OCPs. This result generalizes the convex setup in [25] to a general nonconvex nonlinear setup under the same conditions. The generalization relies on a convexification technique, which convexifies nonconvex problems to convex problems without

altering primal solutions (cf. Algorithm 1). However, the result in [27] is not sufficient for studying the convergence of Schwarz scheme in [26] because (i) a terminal perturbation is missing and (ii) the dual sensitivity is not formally analyzed.

This paper extends the related literature [24]–[27] in the following aspects. (i) We expand the results in [27] by enabling a terminal perturbation, and more importantly, complement [27] by showing that EDS also holds for the dual solution of nonlinear OCPs. We emphasize that obtaining dual sensitivity from primal sensitivity is not straightforward; and we emphasize that the former has not been studied even in the context of convex OCPs. To address this knowledge gap, we delve deeper into the convexification technique in [27]; and show that, although the dual solution is altered by convexification (not preserved like primal solution), it is shifted only by an affine transformation of the primal solution (cf. Theorem 3). With this relation, we further provide a *stagewise* closed form of the dual solution (cf. Theorem 4) and establish dual sensitivity (cf. Theorem 5). (ii) By sensitivity analysis, we enhance the existing overlapping decomposition and Schwarz schemes [24]–[26] by providing a convergence analysis for time-varying nonlinear OCPs, which cover a much wider range of applications than convex OCPs in [24], [25] and time-invariant OCPs in [26]. Furthermore, our primal-dual sensitivity analysis validates the conjecture in [26]. (iii) We prove that the overlapping Schwarz scheme enjoys *linear* convergence locally, provided the overlap size  $\tau$  is sufficiently large. We also show that the linear rate is given by  $C\rho^\tau$ , where  $C > 0$ ,  $\rho \in (0, 1)$  are constants independent of horizon length  $N$ . In other words, the linear rate improves *exponentially* with the overlap size. As a special case, we also show that the Schwarz scheme exhibits global convergence for a linear-quadratic OCP setting (but potentially with nonconvex objective). This result is of relevance, as it suggests that the Schwarz method can be used to solve quadratic programs and linear algebra systems inside second-order algorithms such as sequential quadratic programming and interior-point methods. Such a special case is still more general than [25] and requires a fundamentally different proof technique. Our theory explains favorable performance noticed in recent computational studies that use this approach [28].

It is worth mentioning that a recent work [29] made use of the established primal-dual sensitivity in this paper to study a real-time online model predictive control algorithm. Although this paper also solves nonlinear OCPs, there are significant differences in problem setup, techniques, and results with [29]. First, [29] solved (1) in an *online* fashion, where a single Newton step is performed to solve the subproblem *inexactly*, and then the system shifts to the next stage with a *new* subproblem to be targeted. Online algorithms are a special class of inexact methods for nonlinear predictive control problems, mostly used for systems that require a fast reaction to disturbances (for example, autonomous vehicles) [30]–[33]. In contrast, our approach solves a long-horizon problem (1) in an *offline* fashion with a parallel environment, where problems do not shift but are solved *to the optimality*. Second, [29] relied on the sensitivity (of linear-quadratic OCPs) to show a decay structure of KKT matrix inverse, based on which [29] explored Newton’s method and showed a linear-quadratic error recursion.

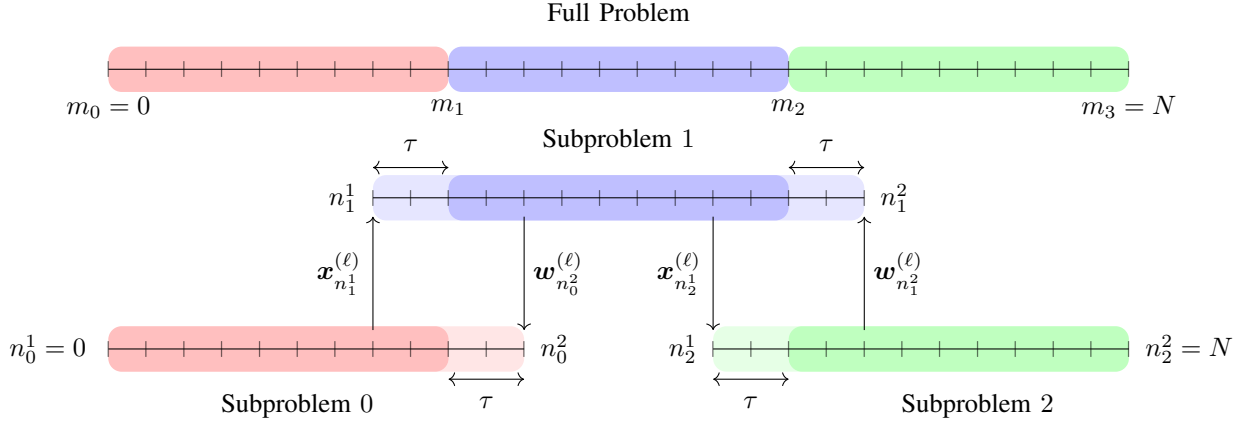


Fig. 1. Overlapping Schwarz decomposition scheme for OCPs. Here,  $m_i$  denotes non-overlapping subdomains boundaries;  $n_i^1, n_i^2$  denote left and right overlapping subdomains boundaries;  $\mathbf{x}_k^{(\ell)}$  is the state iterate at stage  $k$  in the  $\ell$ -th iteration;  $\mathbf{w}_k^{(\ell)} = (\mathbf{x}_k^{(\ell)}; \mathbf{u}_k^{(\ell)}; \boldsymbol{\lambda}_k^{(\ell)})$  is the primal-dual iterate at stage  $k$  in the  $\ell$ -th iteration. Each subproblem depends on the initial state iterate coming from the previous subproblem, and the terminal primal-dual iterate coming from the next subproblem.

In contrast, we rely on the sensitivity (of nonlinear OCPs) to have an increasingly more accurate boundary primal-dual iterates for subproblems and, hence, the subproblem solutions are increasingly closer to the *truncated full-horizon solution*. Third, [29] only showed the real-time iterates stably track the solution (i.e. stay in a neighborhood), while we show the offline iterates converge to the solution linearly with a quantitative relation between the convergence rate and the overlap size.

Our work focuses on the convergence properties of overlapping Schwarz scheme, which is a new and different paradigm for decomposing OCPs compared to traditional approaches [10]–[13]. This approach is interesting in that it spans a spectrum of algorithms that go from a fully centralized/sequential communication pattern (the overlap is the entire horizon) to a no-interaction communication pattern (no overlap). This iterative approach thus provides flexibility to enable different hardware implementations. The paper is motivated by the great success observed in practice [24]–[26], [28], [34]. Prior to this work, the convergence of overlapping decomposition schemes has only been explored for restrictive linear-quadratic convex cases. Here, we show that the Schwarz decomposition exhibits linear convergence for general nonconvex OCPs (cf. Theorem 8). This provides an advantage over the widely-used ADMM, whose convergence is established mostly for restrictive setups that do not apply for (1). For instance, the standard result only deals with convex problems [11], and the nonconvex results in [35, Equation (2.2)] and [36, Equation (1)] do not allow nonlinear dynamical constraints as in (1b). Furthermore, we numerically demonstrate that overlapping Schwarz has much faster convergence than ADMM and may be as efficient as Ipopt (a centralized solver). This observation is important because Ipopt is highly efficient but does not offer flexibility in hardware implementations. Establishing convergence theory for Schwarz scheme is also meaningful from a control practitioner’s stand-point, as it explains the performance observed in many recent computational studies. Moreover, the established theory provides insights on how the scheme will behave when tuning the overlap size  $\tau$ . Our primal-dual EDS result also provides a

foundation for analyzing the behavior of a variety of algorithms and approximations for predictive control [37]–[39].

The remainder of the paper is organized as follows. In Section II we establish primal-dual sensitivity results for (1). In Section III we describe the overlapping Schwarz scheme and its convergence analysis. Numerical results are shown in Section IV and conclusions are presented in Section V.

## II. PRIMAL-DUAL EXPONENTIAL DECAY OF SENSITIVITY

In this section, we enhance the analysis in [27] and establish a primal-dual sensitivity result for nonlinear OCPs that we call exponential decay of sensitivity (EDS). We use the following notation: for  $n, m \in \mathbb{Z}_{>0}$ , we let  $[n, m]$ ,  $[n, m)$ ,  $(n, m]$ ,  $(n, m)$  be the corresponding integer sets; also,  $[n] = [0, n]$ . Boldface symbols denote column vectors. For a set of vectors  $\{\mathbf{a}_i\}_{i=m}^n$ ,  $\mathbf{a}_{m:n} = (\mathbf{a}_m; \dots; \mathbf{a}_n)$  represents a long vector that stacks them together. For scalars  $a, b$ ,  $a \vee b = \max(a, b)$ ;  $a \wedge b = \min(a, b)$ . For a set of matrices  $\{A_i\}_{i=m}^n$ ,  $\prod_{i=m}^n A_i = A_n A_{n-1} \dots A_m$  if  $m \leq n$  and  $I$  otherwise. Without specification,  $\|\cdot\|$  denotes either  $\ell_2$  norm for a vector or spectral norm for a matrix. For a function  $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$ ,  $\nabla f \in \mathbb{R}^{n \times m}$  is its Jacobian.

### A. Sensitivity Analysis and Primal EDS Results

We begin by analyzing the sensitivity of the primal solution. Most of the results in this subsection are presented in [27], but we revisit them for completeness and to lay the groundwork for the new dual sensitivity in Section II-B. We rewrite (1) by explicitly expressing the dependence on external data (parameters) as

$$\min_{\{\mathbf{x}_k\}, \{\mathbf{u}_k\}} \sum_{k=0}^{N-1} g_k(\mathbf{x}_k, \mathbf{u}_k; \mathbf{d}_k) + g_N(\mathbf{x}_N; \mathbf{d}_N), \quad (2a)$$

$$\text{s.t. } \mathbf{x}_{k+1} = f_k(\mathbf{x}_k, \mathbf{u}_k; \mathbf{d}_k), \quad k \in [N-1], \quad (\boldsymbol{\lambda}_k) \quad (2b)$$

$$\mathbf{x}_0 = \bar{\mathbf{x}}_0, \quad (\boldsymbol{\lambda}_{-1}). \quad (2c)$$

Here  $\mathbf{d}_k \in \mathbb{R}^{n_d}$  and  $\mathbf{d}_{-1} = \bar{\mathbf{x}}_0$  are the external problem data. In what follows, we let  $\mathbf{z}_k = (\mathbf{x}_k; \mathbf{u}_k)$ ,  $\mathbf{w}_k = (\mathbf{z}_k; \boldsymbol{\lambda}_k)$  for



constraint qualification (LICQ) holds with any  $(z; d)$ . Thus, the results hold for any perturbation direction  $l$ .  $\square$

Let  $\xi_k = (\mathbf{p}_k; \mathbf{q}_k; \zeta_k)$  for  $k \in [N-1]$ , and  $\xi_{-1} = \zeta_{-1}$  and  $\xi_N = \mathbf{p}_N$ . Further,  $\xi = (\mathbf{p}, \mathbf{q}, \zeta) = \xi_{-1:N}$  (similar for  $\mathbf{p}, \mathbf{q}, \zeta$ ) is the full vector with variables being ordered by stages. From SOSC (cf. Assumption 1), LICQ, and [42, Lemma 16.1], we know that  $\xi^* = (\mathbf{p}^*, \mathbf{q}^*, \zeta^*)$  is unique global solution of (5). However, the indefiniteness of the Hessians  $H_k$  in Problem (5) brings difficulty in analyzing the solution of (5) obtained from the Riccati recursion. Thus, [27] relied on the convexification procedure proposed in [43], which transfers (5) into another linear-quadratic program whose new matrices  $\tilde{H}_k$  are positive definite. The procedure is displayed in Algorithm 1. One inputs quadratic matrices  $\{H_k, D_k, A_k, B_k, C_k\}$  of Problem (5), and then obtains new matrices  $\{\tilde{H}_k, \tilde{D}_k\}$ . The constraint matrices  $\{A_k, B_k, C_k\}$  need not be transformed. As shown in [27], with a proper set of  $\beta > 0$ , Algorithm 1 preserves the primal solution. We will show later that Algorithm 1 shifts the dual solution.

**Theorem 2** (Primal EDS). *Let Assumptions 1, 2, 3 hold at the solution  $\mathbf{w}^*$  of Problem (2). Then there exist constants  $\Upsilon > 0$ ,  $\rho \in (0, 1)$ , which only depend on constants in the assumptions and hence are independent of horizon length  $N$ , such that*

- (a) *if  $l = e_i, \forall i \in [N]$ , then  $\|\mathbf{p}_k^*\| \vee \|\mathbf{q}_k^*\| \leq \Upsilon \rho^{|k-i|}$  for  $k \in [N-1]$  and  $\|\mathbf{p}_N^*\| \leq \Upsilon \rho^{N-i}$ ;*
- (b) *if  $l = e_{-1}$ , then  $\|\mathbf{p}_k^*\| \vee \|\mathbf{q}_k^*\| \leq \Upsilon \rho^k$  for  $k \in [N-1]$  and  $\|\mathbf{p}_N^*\| \leq \Upsilon \rho^N$ .*

This is [27, Theorem 5.7]<sup>1</sup> and indicates that the impact of a perturbation on  $d_i$  on the primal solution  $\mathbf{z}_k^*$  at stage  $k$  decays exponentially fast as one moves away from stage  $i$ .

## B. Dual EDS Results

We now present dual sensitivity for (2) based on convexification procedure in Algorithm 1. As shown in [27], because of the positive definiteness of  $\tilde{H}_k$ , the convexified problem (obtained by replacing  $\{H_k, D_k\}$  in (5) with outputs  $\{\tilde{H}_k, \tilde{D}_k\}$ ) also has a unique global solution. Thus, we need to understand how the dual solutions are affected by convexification. We will show from the Karush-Kuhn-Tucker (KKT) conditions (i.e., the first-order necessary conditions) that Algorithm 1 shifts the dual solution by an affine transformation of the primal solution. In what follows, we use  $\mathcal{LQP}$  to denote Problem (5) defined with original matrices  $\{H_k, D_k\}$ , and  $\mathcal{CLQP}$  to denote Problem (5) defined with convexified matrices  $\{\tilde{H}_k, \tilde{D}_k\}$ . Furthermore,  $\xi^{c*} = (\mathbf{p}^{c*}, \mathbf{q}^{c*}, \zeta^{c*})$  denotes the (global) primal-dual solution of  $\mathcal{CLQP}$ . Recall that, by Theorem 1,  $(\mathbf{p}^*, \mathbf{q}^*, \zeta^*)$  is the global solution of  $\mathcal{LQP}$ .

The following result establishes a relationship between the solutions  $(\mathbf{p}^*, \mathbf{q}^*, \zeta^*)$  and  $(\mathbf{p}^{c*}, \mathbf{q}^{c*}, \zeta^{c*})$ .

**Theorem 3.** *Under Assumption 1, we execute Algorithm 1 with  $\beta \in (0, \gamma_H)$  for  $\mathcal{LQP}$ . We then have that*

$$\mathbf{p}^* = \mathbf{p}^{c*}, \quad \mathbf{q}^* = \mathbf{q}^{c*}, \quad \zeta^* = \zeta^{c*} - 2\bar{Q}\mathbf{p}^*, \quad (6)$$

<sup>1</sup>We note that Problem (2) is slightly different from the one in [27], in that [27] does not include the terminal data  $d_N$ . However, with fairly slight adjustment, their Theorem 5.7 can be extended to the case  $l = e_N$ .

## Algorithm 1 Convexification Procedure

---

```

1: Input:  $\{H_k, D_k\}_{k=0}^N, \{A_k, B_k, C_k\}_{k=0}^{N-1}, \beta > 0;$ 
2:  $\tilde{H}_N = \tilde{Q}_N = \beta I;$ 
3:  $\bar{Q}_N = Q_N - \tilde{Q}_N;$ 
4: for  $k = N-1, \dots, 0$  do
    $\begin{pmatrix} \tilde{Q}_k & \tilde{S}_k^T & \tilde{D}_{k1}^T \\ \tilde{S}_k & \tilde{R}_k & \tilde{D}_{k2}^T \\ \tilde{D}_{k1} & \tilde{D}_{k2} & * \end{pmatrix} = \begin{pmatrix} Q_k & S_k^T & D_{k1}^T \\ S_k & R_k & D_{k2}^T \\ D_{k1} & D_{k2} & \mathbf{0} \end{pmatrix}$ 
    $+ \begin{pmatrix} A_k^T \\ B_k^T \\ C_k^T \end{pmatrix} \bar{Q}_{k+1} \begin{pmatrix} A_k & B_k & C_k \end{pmatrix}$ 
6:  $\tilde{Q}_k = \tilde{S}_k^T \tilde{R}_k^{-1} \tilde{S}_k + \beta I$ 
7:  $\tilde{H}_k = \begin{pmatrix} \tilde{Q}_k & \tilde{S}_k^T \\ \tilde{S}_k & \tilde{R}_k \end{pmatrix}$ 
8:  $\bar{Q}_k = \tilde{Q}_k - \tilde{Q}_k;$ 
9: end for
10: Output:  $\{\tilde{H}_k\}_{k=0}^N, \{\tilde{D}_k\}_{k=0}^{N-1}, D_N (= \tilde{D}_N).$ 

```

---

where  $\bar{Q} = \text{diag}(\bar{Q}_0, \dots, \bar{Q}_N)$  with  $\{\bar{Q}_k\}_{k=0}^N$  is defined in Algorithm 1 recursively.

*Proof.* See Appendix V-A.  $\square$

Using (6), we first focus on  $\mathcal{CLQP}$  and establish the exponential decay result for  $\zeta^{c*}$ . Then we use relation (6) to bound  $\zeta^*$ . The motivation is that some nice properties of  $\{\tilde{H}_k, \tilde{D}_k\}$  does not hold for  $\{H_k, D_k\}$ , which brings difficulties to directly study  $\mathcal{LQP}$ .

The following theorem provides the *stagewise* closed form of the dual solution for linear-quadratic problems (either  $\mathcal{LQP}$  or  $\mathcal{CLQP}$ ). Our notation is the same as [27, Lemma 3.5], which provides the stagewise closed form of the primal solution.

**Theorem 4.** *Consider  $\mathcal{LQP}$  under Assumption 1. Suppose  $(\mathbf{p}^*, \mathbf{q}^*)$  is the primal solution. Then the dual solution  $\zeta^*$  at each stage is:*

$$\begin{aligned} \zeta_k^* = & -2K_{k+1}\mathbf{p}_{k+1}^* + 2 \sum_{i=k+1}^N (M_i^{k+1})^T \mathbf{l}_i \\ & + 2 \sum_{i=k+1}^{N-1} (V_i^{k+1})^T C_i \mathbf{l}_i, \quad \forall k \in [-1, N-1], \end{aligned} \quad (7)$$

with  $K_N = Q_N, D_{N1} = D_N, D_{N2} = \mathbf{0}$ , and  $\forall k \in [N-1]$ ,

$$\begin{aligned} W_k &= R_k + B_k^T K_{k+1} B_k, \\ K_k &= -(B_k^T K_{k+1} A_k + S_k)^T W_k^{-1} (B_k^T K_{k+1} A_k + S_k) \\ &\quad + Q_k + A_k^T K_{k+1} A_k, \\ P_k &= -W_k^{-1} (B_k^T K_{k+1} A_k + S_k), \\ E_k &= A_k + B_k P_k, \\ V_i^k &= -K_{i+1} \prod_{j=k}^i E_j, \quad \forall i \in [N-1], \end{aligned}$$

$$M_i^k = -(D_{i1} + D_{i2} P_i) \prod_{j=k}^{i-1} E_j, \quad \forall i \in [N].$$

We obtain a similar formula for  $\zeta^{c*}$  of  $\mathcal{CLQP}$ , where one replaces  $\{H_k, D_k\}$  in the above recursions by  $\{\tilde{H}_k, \tilde{D}_k\}$ .

*Proof.* See Appendix V-B.  $\square$

We now study the dual solution  $\zeta^{c*}$  of  $\mathcal{CLQP}$ . To enable concise notation, we abuse the notation  $K_k, M_i^k, V_i^k$ , and so on to denote the matrices computed by  $\{\tilde{H}_k, \tilde{D}_k\}$ . The following lemma establishes the exponential decay for  $\zeta^{c*}$ .

**Lemma 1.** *Let Assumptions 1, 2, 3 hold at the unperturbed solution  $w^*$  of Problem (2). We execute Algorithm 1 with  $\beta \in (0, \gamma_H)$ . Let  $\zeta^{c*}$  be the optimal dual solution of  $\mathcal{CLQP}$ . Then there exist constants  $\Upsilon' > 0$ ,  $\rho \in (0, 1)$ , independent of  $N$ , such that for any  $k \in [-1, N-1]$ ,*

- (a) if  $l = e_i$  for  $i \in [N]$ , then  $\|\zeta_k^{c*}\| \leq \Upsilon' \rho^{|k+1-i|}$ ;
- (b) if  $l = e_{-1}$ , then  $\|\zeta_k^{c*}\| \leq \Upsilon' \rho^{k+1}$ .

*Proof.* See Appendix V-C.  $\square$

We note that the constant  $\rho$  in this result is the same as the one used in Theorem 2. Combining Lemma 1 with Theorem 3, we can finally bound the dual solution for  $\mathcal{LQP}$ .

**Theorem 5 (Dual EDS).** *Let Assumptions 1, 2, 3 hold at the unperturbed solution  $w^*$  of Problem (2). Then for any  $k \in [-1, N-1]$ , Lemma 1 holds for  $\zeta_k^*$  with some constants  $\Upsilon'' > 0$ ,  $\rho \in (0, 1)$  that are determined by constants in the assumptions and hence are independent of  $N$ .*

*Proof.* See Appendix V-D.  $\square$

Combining Theorems 2 and 5, we get the desired primal-dual EDS result. The perturbation on the left and right boundaries  $\{-1, N\}$  are of particular interest in the following sections. Redefining  $\Upsilon \leftarrow \sqrt{3} \max(\Upsilon, \Upsilon'' \rho^{-1})$  yields the following<sup>2</sup>:

- (a) if  $l = e_N$ , then  $\|\xi_k^*\| \leq \Upsilon \rho^{N-k}$ ;
- (b) if  $l = e_{-1}$ , then  $\|\xi_k^*\| \leq \Upsilon \rho^k$ ,

where  $\xi_k^* = (p_k^*; q_k^*; \zeta_k^*)$  is defined in (4). The above exponential property plays a key role in the analysis of the algorithm.

### III. OVERLAPPING SCHWARZ DECOMPOSITION

In this section we introduce the overlapping Schwarz scheme and establish its convergence.

#### A. Setting

The full horizon of Problem (1) is  $[N]$ . Suppose  $T$  is the number of short horizons and  $\tau$  is the overlap size. Then we can decompose  $[N]$  into  $T$  consecutive intervals as

$$[N] = \bigcup_{i=0}^{T-1} [m_i, m_{i+1}],$$

where  $m_0 = 0 < m_1 < \dots < m_T = N$ . Moreover, we define the *expanded* (overlapping) boundaries:

$$n_i^1 = (m_i - \tau) \vee 0, \quad n_i^2 = (m_{i+1} + \tau) \wedge N. \quad (8)$$

Then we have  $[N] = \bigcup_{i=0}^{T-1} [n_i^1, n_i^2]$  and

$$[m_i, m_{i+1}] \subset [n_i^1, n_i^2], \quad \forall i \in [T-1].$$

<sup>2</sup>This is because  $\|\xi_k^*\| \leq \sqrt{3}(\|p_k^*\| \vee \|q_k^*\| \vee \|\zeta_k^*\|)$ . One can apply Theorem 2 for bounding  $\|p_k^*\| \vee \|q_k^*\|$  and Theorem 5 for bounding  $\|\zeta_k^*\|$ .

In the overlapping Schwarz scheme, the truncated approximation within the interval  $[m_i, m_{i+1}]$  is obtained by first solving a subproblem over an expanded interval  $[n_i^1, n_i^2]$ , then *discarding* the piece of the solution associated with the stages acquired from the expansion (8). We now introduce the subproblem for the expanded short horizon  $[n_i^1, n_i^2]$ . For any  $i \in [T-1]$ , the subproblem  $\mathcal{P}_i$  for the interval  $[n_i^1, n_i^2]$  is defined as

$$\min_{\{\mathbf{x}_k\}, \{\mathbf{u}_k\}} \sum_{k=n_i^1}^{n_i^2-1} g_k(\mathbf{x}_k, \mathbf{u}_k) + \tilde{g}_{n_i^2}(\mathbf{x}_{n_i^2}; \bar{\mathbf{w}}_{n_i^2}), \quad (9a)$$

$$\text{s.t. } \mathbf{x}_{k+1} = f_k(\mathbf{x}_k, \mathbf{u}_k), \quad k \in [n_i^1, n_i^2 - 1], \quad (\lambda_k) \quad (9b)$$

$$\mathbf{x}_{n_i^1} = \bar{\mathbf{x}}_{n_i^1}, \quad (\lambda_{n_i^1-1}). \quad (9c)$$

Here,  $\tilde{g}_{n_i^2}(\mathbf{x}_{n_i^2}; \bar{\mathbf{w}}_{n_i^2})$  is the terminal cost function adjusted from  $g_{n_i^2}(\mathbf{x}_{n_i^2}, \mathbf{u}_{n_i^2})$ . It is parameterized by  $\bar{\mathbf{w}}_{n_i^2} = (\bar{\mathbf{x}}_{n_i^2}; \bar{\mathbf{u}}_{n_i^2}; \bar{\lambda}_{n_i^2})$  and is formally defined as

$$\tilde{g}_{n_i^2}(\mathbf{x}_{n_i^2}; \bar{\mathbf{w}}_{n_i^2}) = \begin{cases} g_{n_i^2}(\mathbf{x}_{n_i^2}, \bar{\mathbf{u}}_{n_i^2}) - \bar{\lambda}_{n_i^2}^T f_{n_i^2}(\mathbf{x}_{n_i^2}, \bar{\mathbf{u}}_{n_i^2}) \\ \quad + \frac{\mu}{2} \|\mathbf{x}_{n_i^2} - \bar{\mathbf{x}}_{n_i^2}\|^2, & i \in [T-2], \\ g_N(\mathbf{x}_N), & i = T-1, \end{cases}$$

where  $\mu$  is a uniform penalty parameter that *does not depend* on  $i$ . In other words,  $\mu$  is set uniformly over all subproblems. When  $i \neq T-1$ , the terminal cost is adjusted by a dual penalty and a quadratic penalty on the state. Intuitively, the dual penalty reduces the KKT residuals, while the quadratic penalty ensures that SOSC holds for subproblems provided it holds for the full problem and  $\mu$  is set large enough (see Lemma 2). The formulation (9) is adopted from [29], which differs from the one in [26]. In particular, [26] imposed assumptions on subproblems, while we impose (standard) assumptions directly on the full problem, which is more reasonable. An alternative subproblem formulation can be found in [44, (5)], where the terminal adjustment on the cost is replaced by a terminal constraint.

We note that Problem (9) is a parametric subproblem with  $\mathcal{P}_i = \mathcal{P}_i(\bar{\mathbf{x}}_{n_i^1}, \bar{\mathbf{w}}_{n_i^2})$ . The parameter  $(\bar{\mathbf{x}}_{n_i^1}, \bar{\mathbf{w}}_{n_i^2})$  consists of the primal-dual data on both ends of the horizon (i.e. domain boundaries). Each time we have to first specify the parameter and then solve the subproblem. For  $i = T-1$ ,  $\bar{\mathbf{w}}_{n_i^2}$  is not necessary (see the definition of  $\tilde{g}_N(\cdot)$ ). The formal justification of the formulation in (9) will be given in Lemma 3.

**Definition 3.** *We define the following quantities for subproblem  $\mathcal{P}_i$  with  $i \in [T-1]$ :*

- (a) we let  $\mathbf{w}_{[i]}$  be the primal-dual variable of  $\mathcal{P}_i$ , i.e.  $\mathbf{w}_{[i]} = (\lambda_{n_i^1-1}; \mathbf{w}_{n_i^1:n_i^2-1}; \mathbf{x}_{n_i^2})$ .
- (b) we let  $\mathbf{w}_{(i)}$  be the primal-dual variable of  $\mathcal{P}_i$  on the **non-overlapping** subdomains, i.e.  $\mathbf{w}_{(i)} = \mathbf{w}_{m_i:m_{i+1}-1}$  (for the boundaries,  $\mathbf{w}_{(0)}$  and  $\mathbf{w}_{(T-1)}$ , are adjusted by letting  $\mathbf{w}_{(0)} = \mathbf{w}_{-1:m_1-1}$  and  $\mathbf{w}_{(T-1)} = \mathbf{w}_{m_{T-1}:m_T}$ ).
- (c) we let  $\mathbf{w}_{[-i]}$  be the parameter variable of  $\mathcal{P}_i$ , i.e.  $\mathbf{w}_{[-i]} = (\mathbf{x}_{n_i^1}; \mathbf{w}_{n_i^2})$  (the boundary  $\mathbf{w}_{[-(T-1)]}$  is adjusted by letting  $\mathbf{w}_{[-(T-1)]} = \mathbf{x}_{n_{T-1}^1}$ ).
- (d) we let  $n_{[i]}$ ,  $n_{(i)}$ ,  $n_{[-i]}$  be the corresponding dimensions of  $\mathbf{w}_{[i]}$ ,  $\mathbf{w}_{(i)}$ ,  $\mathbf{w}_{[-i]}$ .
- (e) we let  $\mathbf{w}_{[i]}^\dagger(\cdot) : \mathbb{R}^{n_{[-i]}} \rightarrow \mathbb{R}^{n_{[i]}}$  be the solution mapping of  $\mathcal{P}_i$ , i.e.  $\mathbf{w}_{[i]}^\dagger(\mathbf{w}_{[-i]})$  is a local solution of  $\mathcal{P}_i(\mathbf{w}_{[-i]})$ .

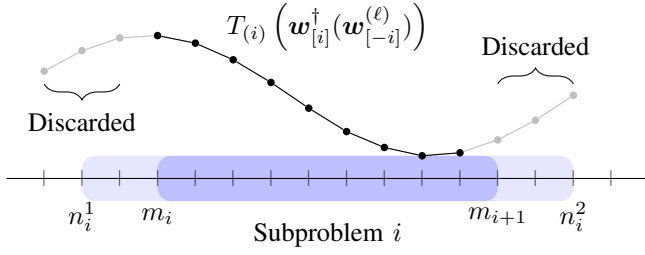


Fig. 2. Schematic of restriction operation.

(f) for  $k \in [n_i^1 - 1, n_i^2]$ , we use  $T_k(\mathbf{w}_{[i]})$  to extract the variable on stage  $k$  of  $\mathbf{w}_{[i]}$ ; we also use  $T_{(i)}(\mathbf{w}_{[i]})$  to extract variables of  $\mathbf{w}_{[i]}$  that are on non-overlapping subdomains.

The solution of  $\mathcal{P}_i(\cdot)$  may not be unique. The issues of the existence and uniqueness of the solution will be resolved in Theorem 6. For now, we assume that the solution  $\mathbf{w}_{[i]}^\dagger(\cdot)$  exists and consider this as one of the local solutions.

We now formally present the overlapping Schwarz scheme in Algorithm 2. Here, we use the superscript  $(\cdot)^{(\ell)}$  to denote its value at the  $\ell$ -th iteration. In addition, we suppose that the problem information (e.g.  $\{f_k\}_{k=0}^{N-1}$ ,  $\{g_k\}_{k=0}^N$ ) and the decomposition information (e.g.  $\{m_i\}_{i=0}^{T-1}$  and  $\{[n_i^1, n_i^2]\}_{i=0}^{T-1}$ ) are already given to the algorithm. Thus, the algorithm is well defined using only the initial guess  $\mathbf{w}^{(0)}$  of the full primal-dual solution as an input. Note that  $\mathbf{x}_0^{(0)}$  should match the initial state  $\bar{\mathbf{x}}_0$  given to the original problem (1).

Starting with  $\mathbf{w}^{(0)}$ , the procedure iteratively finds the primal-dual solution  $\mathbf{w}^{(\ell)}$  for (1). At each iteration  $\ell = 0, 1, \dots$ , the subproblems  $\{\mathcal{P}_i(\mathbf{w}_{[-i]}^{(\ell)})\}_{i=0}^{T-1}$  are solved to obtain the short-horizon solutions  $\mathbf{w}_{[i]}^\dagger(\mathbf{w}_{[-i]}^{(\ell)})$  over the expanded subdomains  $\{[n_i^1, n_i^2]\}_{i=0}^{T-1}$ . Here, we note that the previous primal-dual iterate enters into the subproblems as boundary conditions  $\mathbf{w}_{[-i]}^{(\ell)} = (\mathbf{x}_{n_i^1}^{(\ell)}; \mathbf{w}_{n_i^2}^{(\ell)})$ . This step is illustrated in Fig. 1. The solution is then *restricted* to the non-overlapping subdomains  $\{[m_i, m_{i+1}]\}_{i=0}^{T-1}$  by applying the operator  $T_{(i)}(\cdot)$  (cf. Definition 3(f)), which is illustrated in Fig. 2. After that, one concatenates the short-horizon solutions by  $\mathbf{w}^{(\ell+1)} = (\mathbf{w}_{(0)}^{(\ell+1)}; \dots; \mathbf{w}_{(T-1)}^{(\ell+1)})$ . We do not explicitly write this step in Algorithm 2 since updating the subvectors  $\mathbf{w}_{(i)}^{(\ell)}$  of  $\mathbf{w}^{(\ell)}$  over  $i \in [T-1]$  effectively concatenates the short-horizon solutions. In Proposition 1 we provide stopping criteria for the scheme.

Observe that, unless  $\tau = 0$ , the boundary conditions  $\mathbf{w}_{[-i]}^{(\ell)}$  of subproblem  $i$  are set by the output of other subproblems (in particular, adjacent ones if  $\tau$  does not exceed the horizon

---

### Algorithm 2 Overlapping Schwarz Decomposition

---

```

1: Input:  $\mathbf{w}^{(0)}$ 
2: for  $\ell = 0, 1, \dots$  do
3:   for (in parallel)  $i = 0, 1, \dots, T-1$  do
4:      $\mathbf{w}_{(i)}^{(\ell+1)} = T_{(i)}(\mathbf{w}_{[i]}^\dagger(\mathbf{w}_{[-i]}^{(\ell)}))$ ;
5:   end for
6: end for
7: Output:  $\mathbf{w}^{(\ell)}$ 

```

---

lengths of the adjacent problems). Thus, the procedure aims to achieve coordination across the subproblems by exchanging the primal-dual solution information. Furthermore, one can observe that  $n_i^1, n_i^2$  are at least  $\tau$  stages apart from  $[m_i, m_{i+1})$ . This guarantees all iterates improve (as shown in the next section) because the adverse effect of misspecification of boundary conditions has enough stages to be damped. We can hence anticipate that having larger  $\tau$  makes the convergence faster at the cost of having moderately larger subproblems.

We emphasize that Algorithm 2 can be implemented in parallel, although the subproblems are coupled. This is because subproblems are parameterized by boundary variables  $\{\mathbf{w}_{[-i]} = (\mathbf{x}_{n_i^1}; \mathbf{w}_{n_i^2})\}_i$  (as defined in Problem (9)) and, in each iteration, once  $\{\mathbf{w}_{[-i]}^{(\ell)}\}_i$  are specified all subproblems can be solved independently. To ensure convergence, we require information exchange between subproblems after each iteration. This is similar in spirit to traditional Jacobi and Gauss-Seidel schemes.

### B. Convergence Analysis

We now establish convergence for Algorithm 2. A sketch of convergence analysis is as follows. (i) We extend Assumptions 1, 2, 3 to a neighborhood of a local solution  $\mathbf{w}^*$  of Problem (1). (ii) We show that  $\mathbf{w}_{[i]}^\dagger(\mathbf{w}_{[-i]}^*) = \mathbf{w}_{[i]}^*$ . (iii) We bound the difference of solutions  $\mathbf{w}_{[i]}^\dagger(\bar{\mathbf{w}}_{[-i]})$  and  $\mathbf{w}_{[i]}^\dagger(\mathbf{w}_{[-i]}^*)$  by the difference in boundary conditions  $\|\bar{\mathbf{w}}_{[-i]} - \mathbf{w}_{[-i]}^*\|$  using primal-dual EDS results in Section II. (iv) We combine (i)-(iii) and derive an explicit estimate of the local convergence rate.

**Definition 4.** We define  $\|(\cdot)\|_{\mathbf{w}}$  as the stagewise max  $\ell_2$ -norm of the variable  $(\cdot)$ . For example,

$$\begin{aligned} \|\mathbf{z}_k\|_{\mathbf{w}} &= \|\mathbf{x}_k\| \vee \|\mathbf{u}_k\|, & \|\mathbf{w}_k\|_{\mathbf{w}} &= \|\mathbf{z}_k\|_{\mathbf{w}} \vee \|\boldsymbol{\lambda}_k\|, \\ \|\mathbf{w}\|_{\mathbf{w}} &= \max_{k \in [-1, N]} \|\mathbf{w}_k\|_{\mathbf{w}}, & \|\mathbf{w}_{[-i]}\|_{\mathbf{w}} &= \|\mathbf{x}_{n_i^1}\| \vee \|\mathbf{w}_{n_i^2}\|_{\mathbf{w}}, \\ \|\mathbf{w}_{[i]}\|_{\mathbf{w}} &= \|\boldsymbol{\lambda}_{n_i^1-1}\| \vee \max_{k \in [n_i^1, n_i^2]} \|\mathbf{w}_k\|_{\mathbf{w}} \vee \|\mathbf{x}_{n_i^2}\|. \end{aligned}$$

Further, we define  $\mathcal{N}_\varepsilon(\cdot)$  as the (closed)  $\varepsilon$ -neighborhood of the variable  $(\cdot)$  based on the norm  $\|(\cdot)\|_{\mathbf{w}}$ . For example,

$$\begin{aligned} \mathcal{N}_\varepsilon(\mathbf{z}_k^*) &= \{\mathbf{z}_k \in \mathbb{R}^{n_{z_k}} : \|\mathbf{z}_k - \mathbf{z}_k^*\|_{\mathbf{w}} \leq \varepsilon\}, \\ \mathcal{N}_\varepsilon(\mathbf{w}_k^*) &= \{\mathbf{w}_k \in \mathbb{R}^{n_{w_k}} : \|\mathbf{w}_k - \mathbf{w}_k^*\|_{\mathbf{w}} \leq \varepsilon\}. \end{aligned}$$

Similarly we have  $\mathcal{N}_\varepsilon(\mathbf{w}^*)$ ,  $\mathcal{N}_\varepsilon(\mathbf{w}_{[-i]}^*)$ ,  $\mathcal{N}_\varepsilon(\mathbf{w}_{[i]}^*)$ .

The norm  $\|(\cdot)\|_{\mathbf{w}}$  in Definition 4 takes the maximum of the  $\ell_2$ -norms of stagewise state, control, and dual variables over the corresponding horizon. One can verify that it is indeed a vector norm (triangle inequality, absolute homogeneity, and positive definiteness hold). With Definition 4, here we extend assumptions in Section II, which are stated for a local solution point, to the *neighborhood* of such a local solution. In what follows, we inherit the notation in Definition 1 but drop the reference variable  $\mathbf{d}$ . We denote  $A_k, B_k$  to be the Jacobian of  $f_k(\mathbf{x}_k, \mathbf{u}_k)$  with respect to  $\mathbf{x}_k$  and  $\mathbf{u}_k$ , respectively.  $H_k$  is the Hessian of the Lagrange function with respect to  $(\mathbf{x}_k, \mathbf{u}_k)$ .

**Assumption 4.** For a local primal-dual solution  $\mathbf{w}^*$  of Problem (1), we assume that there exists  $\varepsilon > 0$  such that:

(a) There exists a uniform constant  $\gamma_H > 0$  such that

$$\text{Re}H(\mathbf{w}) \succeq \gamma_H I,$$

for any  $\mathbf{w} = \mathbf{w}_{-1:N}$  with  $\mathbf{w}_k \in \mathcal{N}_\varepsilon(\mathbf{w}_k^*)$  for  $k \in [n_i^1, n_i^2]$  for some  $i \in [T-1]$ , and  $\mathbf{w}_k = \mathbf{w}_k^*$  otherwise.

(b) There exists a uniform constant  $\Upsilon_{\text{upper}}$  such that  $\forall k \in [N]$ :

$$\|H_k(\mathbf{w}_k)\| \vee \|A_k(\mathbf{z}_k)\| \vee \|B_k(\mathbf{z}_k)\| \leq \Upsilon_{\text{upper}},$$

for any  $\mathbf{w}_k \in \mathcal{N}_\varepsilon(\mathbf{w}_k^*)$ .

(c) There exist uniform constants  $\gamma_C$ ,  $t > 0$  such that  $\forall k \in [N-t]$  and some  $t_k \in [1, t]$ :

$$\Xi_{k,t_k}(\mathbf{z}_{k:k+t_k-1}) \Xi_{k,t_k}(\mathbf{z}_{k:k+t_k-1})^T \succeq \gamma_C I,$$

for any  $\mathbf{z}_{k:k+t_k-1} \in \mathcal{N}_\varepsilon(\mathbf{z}_{k:k+t_k-1}^*)$ .

We now show that the subproblem's solution recover the truncated full-horizon solution if perfect boundary conditions are given. We first state the following lemma, which adapts [29, Lemma 1 and Theorem 1] and suggests that the subproblem formulation (9) inherits uniform SOSC from the full problem, provided  $\mu$  is sufficiently large.

**Lemma 2.** *Let Assumption 4 hold for the local solution  $\mathbf{w}^*$  of Problem (1) and  $\varepsilon > 0$ . There exists  $\bar{\mu}$  depending on  $(\Upsilon_{\text{upper}}, \gamma_C, t)$  only such that if  $\mu \geq \bar{\mu}$ ,  $\mathbf{w}_{[i]} \in \mathcal{N}_\varepsilon(\mathbf{w}_{[i]}^*)$ , and  $\bar{\mathbf{w}}_{[-i]} \in \mathcal{N}_\varepsilon(\mathbf{w}_{[-i]}^*)$ , then the reduced Hessian of subproblem  $\mathcal{P}_i(\bar{\mathbf{w}}_{[-i]})$  at  $\mathbf{w}_{[i]}^*$  is lower bounded by  $\gamma_H$  as well. That is,*

$$\text{Re}H^i(\mathbf{w}_{[i]}; \bar{\mathbf{w}}_{[-i]}) \succeq \gamma_H I,$$

where  $\text{Re}H^i(\mathbf{w}_{[i]}; \bar{\mathbf{w}}_{[-i]})$  denotes the reduced Hessian of  $\mathcal{P}_i(\bar{\mathbf{w}}_{[-i]})$  evaluated at  $\mathbf{w}_{[i]}^*$ , defined similarly as in Definition 1.

More precisely, we see from [29, (15)] that

$$\bar{\mu} = \bar{\mu}(\Upsilon_{\text{upper}}, \gamma_C, t) := \frac{16\Upsilon_{\text{upper}}(\Upsilon_{\text{upper}}^{6t} - \Upsilon_{\text{upper}}^{4t})}{\gamma_C^2}. \quad (10)$$

However, the above expression is only a conservative bound for  $\mu$ . In our experiments, we will see that  $\mu = 1$  works well for different nonlinear OCPs.

**Lemma 3.** *Let Assumption 4 hold for the local solution  $\mathbf{w}^*$  of Problem (1) and  $\varepsilon > 0$ . We choose  $\mu \geq \bar{\mu}$  defined in Lemma 2. Then, for any  $i \in [T-1]$ ,  $\mathbf{w}_{[i]}^*$  is a local solution of  $\mathcal{P}_i(\mathbf{w}_{[-i]}^*)$ .*

*Proof.* By Lemma 2, we know that the reduced Hessian of  $\mathcal{P}_i(\mathbf{w}_{[-i]}^*)$  evaluated at  $\mathbf{w}_{[i]}^*$  is lower bounded by  $\gamma_H$ . Thus, it suffices to show that  $\mathbf{w}_{[i]}^*$  satisfies the KKT conditions for  $\mathcal{P}_i(\mathbf{w}_{[-i]}^*)$ . First, we write the KKT systems for Problem (1):

$$\mathbf{0} = \begin{cases} \nabla_{\mathbf{x}_k} g_k(\mathbf{z}_k) + \boldsymbol{\lambda}_{k-1} - A_k^T(\mathbf{z}_k)\boldsymbol{\lambda}_k, & \forall k \in [N-1], \\ \nabla_{\mathbf{u}_k} g_k(\mathbf{z}_k) - B_k^T(\mathbf{z}_k)\boldsymbol{\lambda}_k, & \forall k \in [N-1], \\ \nabla_{\mathbf{x}_N} g_N(\mathbf{x}_N) + \boldsymbol{\lambda}_{N-1}, & \\ \mathbf{x}_{k+1} - f_k(\mathbf{z}_k), & \forall k \in [N-1], \\ \mathbf{x}_0 - \bar{\mathbf{x}}_0. & \end{cases} \quad (11)$$

Analogously, the KKT system of  $\mathcal{P}_i(\bar{\mathbf{w}}_{[-i]})$  is

$$\mathbf{0} = \begin{cases} \nabla_{\mathbf{x}_k} g_k(\mathbf{z}_k) + \boldsymbol{\lambda}_{k-1} - A_k^T(\mathbf{z}_k)\boldsymbol{\lambda}_k, & \forall k \in [n_i^1, n_i^2], \\ \nabla_{\mathbf{u}_k} g_k(\mathbf{z}_k) - B_k^T(\mathbf{z}_k)\boldsymbol{\lambda}_k, & \forall k \in [n_i^1, n_i^2], \\ \nabla_{\mathbf{x}_{n_i^2}} \tilde{g}_{n_i^2}(\mathbf{x}_{n_i^2}; \bar{\mathbf{w}}_{n_i^2}) + \boldsymbol{\lambda}_{n_i^2-1}, & \\ \mathbf{x}_{k+1} - f_k(\mathbf{z}_k), & \forall k \in [n_i^1, n_i^2], \\ \mathbf{x}_{n_i^1} - \bar{\mathbf{x}}_{n_i^1}, & \end{cases} \quad (12)$$

where  $\nabla_{\mathbf{x}_{n_i^2}} \tilde{g}_{n_i^2}(\mathbf{x}_{n_i^2}; \bar{\mathbf{w}}_{n_i^2})$  is

$$= \begin{cases} \nabla_{\mathbf{x}_{n_i^2}} g_{n_i^2}(\mathbf{x}_{n_i^2}, \bar{\mathbf{u}}_{n_i^2}) - A_{n_i^2}^T(\mathbf{x}_{n_i^2}, \bar{\mathbf{u}}_{n_i^2}) \bar{\boldsymbol{\lambda}}_{n_i^2} \\ \quad + \mu(\mathbf{x}_{n_i^2} - \bar{\mathbf{x}}_{n_i^2}), & i \in [T-2], \\ \nabla_{\mathbf{x}_N} g_N(\mathbf{x}_N), & i = T-1. \end{cases}$$

One can see from the satisfaction of KKT system (11) for the full problem (1) with  $\mathbf{w}^*$  that (12) is satisfied with  $\mathbf{w}_{[i]}^*$  when  $\bar{\mathbf{w}}_{[-i]} = \mathbf{w}_{[-i]}^*$ . This completes the proof.  $\square$

We now estimate errors for the short-horizon solutions. The next theorem characterizes the existence and uniqueness of the local mapping  $\mathbf{w}_{[i]}^\dagger(\cdot)$  from the boundary variable  $\mathbf{w}_{[-i]}$  to the local solution of  $\mathcal{P}_i(\mathbf{w}_{[-i]})$ .

**Theorem 6.** *Let Assumption 4 hold for the local solution  $\mathbf{w}^*$  of Problem (1) and  $\varepsilon > 0$ . We choose  $\mu \geq \bar{\mu}$  defined in Lemma 2. Then, for any  $i \in [T-1]$ , there exist  $\delta > 0$ ,  $\varepsilon' \in (0, \varepsilon)$  and a continuously differentiable function  $\mathbf{w}_{[i]}^\dagger: \mathcal{N}_\delta(\mathbf{w}_{[-i]}^*) \rightarrow \mathcal{N}_{\varepsilon'}(\mathbf{w}_{[i]}^*)$  such that, if boundary variable  $\bar{\mathbf{w}}_{[-i]} \in \mathcal{N}_\delta(\mathbf{w}_{[-i]}^*)$ , then  $\mathbf{w}_{[i]}^\dagger(\bar{\mathbf{w}}_{[-i]})$  is a unique local solution of  $\mathcal{P}_i(\bar{\mathbf{w}}_{[-i]})$  in the neighborhood  $\mathcal{N}_{\varepsilon'}(\mathbf{w}_{[i]}^*)$ .*

Theorem 6 is a specialization of the classical result of [45, Theorem 2.1]. Since  $\mathbf{w}_{[i]}^\dagger$  is differentiable, we analogize the directional derivatives definitions in (4) and, for any point  $\bar{\mathbf{w}}_{[-i]}$  and perturbation direction  $\mathbf{l}$ , define

$$\boldsymbol{\xi}_{[i]}^\dagger(\bar{\mathbf{w}}_{[-i]}, \mathbf{l}) = \lim_{h \searrow 0} \frac{\mathbf{w}_{[i]}^\dagger(\bar{\mathbf{w}}_{[-i]} + h\mathbf{l} + o(h)) - \mathbf{w}_{[i]}^\dagger(\bar{\mathbf{w}}_{[-i]})}{h}$$

as the directional derivatives of  $\mathbf{w}_{[i]}^\dagger(\bar{\mathbf{w}}_{[-i]})$ . Here, we disregard the perturbation for stages  $[n_i^1, n_i^2]$  since in the formulation of subproblems (9), only stages  $n_i^1-1$  and  $n_i^2$  are perturbed. Implementing the exact computation of  $\mathbf{w}_{[i]}^\dagger(\bar{\mathbf{w}}_{[-i]})$  is challenging. In practice, one resorts to solving  $\mathcal{P}_i(\bar{\mathbf{w}}_{[-i]})$  using a generic NLP solver, and the solver may return a local solution outside of the neighborhood  $\mathcal{N}_{\varepsilon'}(\mathbf{w}_{[i]}^*)$ . Strictly preventing this is difficult in general. Fortunately, by warm-starting the solver with the previous iterate, one may reduce the chance that the solver returns a solution that is far from the previous iterate. Thus, in our numerical implementation, we implement Algorithm 2 by using the warm-start strategy.

The next result characterizes the stagewise difference between  $\mathbf{w}_{[i]}^\dagger(\bar{\mathbf{w}}_{[-i]})$  and  $\mathbf{w}_{[i]}^\dagger(\mathbf{w}_{[-i]}^*)$ .

**Theorem 7.** *Let Assumption 4 hold for the local solution  $\mathbf{w}^*$  of Problem (1) and  $\varepsilon > 0$ . We choose  $\mu \geq \bar{\mu}$  defined in Lemma 2, and  $\delta > 0$  defined in Theorem 6. For  $i \in [T-1]$ , if boundary*

variable  $\bar{\mathbf{w}}_{[-i]} \in \mathcal{N}_\delta(\mathbf{w}_{[-i]}^*)$ , then there exist constants  $\Upsilon > 0$  and  $\rho \in (0, 1)$  independent from  $N$  and  $i$ , such that

$$\|T_k\{\mathbf{w}_{[i]}^\dagger(\bar{\mathbf{w}}_{[-i]}) - \mathbf{w}_{[i]}^*\}\|_{\mathbf{w}} \leq \Upsilon\{\rho^{k-n_i^1}\|\bar{\mathbf{x}}_{n_i^1} - \mathbf{x}_{n_i^1}^*\| + \rho^{n_i^2-k}\|\bar{\mathbf{w}}_{n_i^2} - \mathbf{w}_{n_i^2}^*\|\|_{\mathbf{w}}\}, \quad \forall k \in [n_i^1 - 1, n_i^2]. \quad (13)$$

*Proof.* We define  $\Delta\mathbf{x}_{n_i^1} = \mathbf{x}_{n_i^1}^* - \bar{\mathbf{x}}_{n_i^1}$ ,  $\Delta\mathbf{w}_{n_i^2} = \mathbf{w}_{n_i^2}^* - \bar{\mathbf{w}}_{n_i^2}$ , and an intermediate boundary variable  $\tilde{\mathbf{w}}_{[-i]} = (\mathbf{x}_{n_i^1}^*, \bar{\mathbf{w}}_{n_i^2})$ .

Then  $\tilde{\mathbf{w}}_{[-i]} \in \mathcal{N}_\delta(\mathbf{w}_{[-i]}^*)$  and thus  $\mathbf{w}_{[i]}^\dagger(\tilde{\mathbf{w}}_{[-i]})$  exist. We have

$$\begin{aligned} & \mathbf{w}_{[i]}^\dagger(\bar{\mathbf{w}}_{[-i]}) - \mathbf{w}_{[i]}^\dagger(\mathbf{w}_{[-i]}^*) \\ &= \{\mathbf{w}_{[i]}^\dagger(\bar{\mathbf{w}}_{[-i]}) - \mathbf{w}_{[i]}^\dagger(\tilde{\mathbf{w}}_{[-i]})\} + \{\mathbf{w}_{[i]}^\dagger(\tilde{\mathbf{w}}_{[-i]}) - \mathbf{w}_{[i]}^\dagger(\mathbf{w}_{[-i]}^*)\}. \end{aligned} \quad (14)$$

The first term corresponds to the perturbation of the initial stage, while the second term corresponds to the perturbation of the terminal stage. Let us define two directions  $\mathbf{l}_1$  and  $\mathbf{l}_2$  as:

$$\mathbf{l}_1 = \begin{cases} \mathbf{0} & \text{if } \|\Delta\mathbf{x}_{n_i^1}\| = 0, \\ \frac{\Delta\mathbf{x}_{n_i^1}}{\|\Delta\mathbf{x}_{n_i^1}\|} & \text{otherwise,} \end{cases} \quad \mathbf{l}_2 = \begin{cases} \mathbf{0} & \text{if } \|\Delta\mathbf{w}_{n_i^2}\| = 0, \\ \frac{\Delta\mathbf{w}_{n_i^2}}{\|\Delta\mathbf{w}_{n_i^2}\|} & \text{otherwise,} \end{cases}$$

and accordingly the perturbation paths  $P_1 = \{\bar{\mathbf{w}}_{[-i]} + s\mathbf{l}_1 : s \in [0, \|\Delta\mathbf{x}_{n_i^1}\|]\}$  and  $P_2 = \{\tilde{\mathbf{w}}_{[-i]} + s\mathbf{l}_2 : s \in [0, \|\Delta\mathbf{w}_{n_i^2}\|]\}$ . Along  $P_1$  the boundary variable changes from  $\bar{\mathbf{w}}_{[-i]}$  to  $\tilde{\mathbf{w}}_{[-i]}$ ; and along  $P_2$  the boundary variable changes from  $\tilde{\mathbf{w}}_{[-i]}$  to  $\mathbf{w}_{[-i]}^*$ . We can easily verify from Definition 4 that any points on  $P_1 \cup P_2$  are in the neighborhood  $\mathcal{N}_\delta(\mathbf{w}^*)$ . Thus, by Theorem 6,  $\mathbf{w}_{[i]}^\dagger(P_1 \cup P_2)$  exists and lies in  $\mathcal{N}_{\varepsilon'}(\mathbf{w}_{[i]}^*)$ .

The perturbations along the path  $P_1 \cup P_2$  will be analyzed using Theorems 2 and 5. We first check that Assumptions 1, 2, 3 hold at  $\mathbf{w}_{[i]}^\dagger(\mathbf{w}_{[-i]})$  over  $\mathbf{w}_{[-i]} \in P_1 \cup P_2$ . Assumption 1 holds at each  $\mathbf{w}_{[i]}^\dagger(\mathbf{w}_{[-i]})$  by Assumption 4(a) and Lemma 2. Assumption 2 holds at each  $\mathbf{w}_{[i]}^\dagger(\mathbf{w}_{[-i]})$  by Assumption 4(c). For Assumption 3, we know  $H_k, A_k, B_k$  for  $k \in [n_i^1, n_i^2]$  are upper bounded by Assumption 4(b); further, one can verify that  $H_{n_i^2}, C_{n_i^1}$ , and  $D_{n_i^2}$  are also uniformly bounded by inspecting  $\mathcal{P}_i(\cdot)$  and noting that  $\mu$  is a parameter independent of  $N$  and  $i$ . Thus, Assumptions 1, 2, 3 hold at each  $\mathbf{w}_{[i]}^\dagger(\mathbf{w}_{[-i]})$  over  $\mathbf{w}_{[-i]} \in P_1 \cup P_2$ .

By Lemma 3, for any  $i \in [T-1]$  and  $k \in [n_i^1, n_i^2]$ , we have

$$\begin{aligned} & \|T_k\{\mathbf{w}_{[i]}^\dagger(\bar{\mathbf{w}}_{[-i]}) - \mathbf{w}_{[i]}^*\}\|_{\mathbf{w}} \\ &= \|T_k\{\mathbf{w}_{[i]}^\dagger(\bar{\mathbf{w}}_{[-i]}) - \mathbf{w}_{[i]}^\dagger(\mathbf{w}_{[-i]}^*)\}\|_{\mathbf{w}} \\ &\stackrel{(14)}{\leq} \|T_k\{\mathbf{w}_{[i]}^\dagger(\bar{\mathbf{w}}_{[-i]}) - \mathbf{w}_{[i]}^\dagger(\tilde{\mathbf{w}}_{[-i]})\}\|_{\mathbf{w}} \\ &\quad + \|T_k\{\mathbf{w}_{[i]}^\dagger(\tilde{\mathbf{w}}_{[-i]}) - \mathbf{w}_{[i]}^\dagger(\mathbf{w}_{[-i]}^*)\}\|_{\mathbf{w}}. \end{aligned} \quad (15)$$

Rewriting the first term of the right-hand side of the inequality in (15) and using the integral of line derivative yields

$$\begin{aligned} & \|T_k\{\mathbf{w}_{[i]}^\dagger(\bar{\mathbf{w}}_{[-i]}) - \mathbf{w}_{[i]}^\dagger(\tilde{\mathbf{w}}_{[-i]})\}\|_{\mathbf{w}} \\ &= \left\| \int_0^{\|\Delta\mathbf{x}_{n_i^1}\|} T_k\{\xi_{[i]}^\dagger(\bar{\mathbf{w}}_{[-i]} + s\mathbf{l}_1; \mathbf{l}_1)\} ds \right\|_{\mathbf{w}} \\ &\leq \int_0^{\|\Delta\mathbf{x}_{n_i^1}\|} \|T_k\{\xi_{[i]}^\dagger(\bar{\mathbf{w}}_{[-i]} + s\mathbf{l}_1; \mathbf{l}_1)\}\|_{\mathbf{w}} ds \end{aligned}$$

$$\leq \|\Delta\mathbf{x}_{n_i^1}\| \cdot \Upsilon\rho^{k-n_i^1},$$

where the second inequality follows from triangle inequality of integrals and the third inequality follows from Theorems 2 and 5. Similarly, the second term in (15) can be bounded by  $\Upsilon\rho^{n_i^2-k}\|\Delta\mathbf{w}_{n_i^2}\| \leq \sqrt{3}\Upsilon\rho^{n_i^2-k}\|\Delta\mathbf{w}_{n_i^2}\|_{\mathbf{w}}$  where the inequality is due to

$$\begin{aligned} \|\Delta\mathbf{w}_{n_i^2}\| &\leq \sqrt{3}(\|\Delta\mathbf{x}_{n_i^2}\| \vee \|\Delta\mathbf{u}_{n_i^2}\| \vee \|\Delta\lambda_{n_i^2}\|) \\ &= \sqrt{3}\|\Delta\mathbf{w}_{n_i^2}\|_{\mathbf{w}}. \end{aligned}$$

Thus, combining these results with (15) and letting  $\Upsilon \leftarrow \sqrt{3}\Upsilon$ , we obtain (13). This completes the proof.  $\square$

Theorem 7 provides a proof for the conjecture made in [26, Property 1]. It shows that the effect of the perturbation of the boundary variable  $\mathbf{w}_{[-i]}$  of  $\mathcal{P}_i$  on the solution  $\mathbf{w}_{[i]}^\dagger(\mathbf{w}_{[-i]})$  decays exponentially as moving away from two boundary ends. Here, the unperturbed data is  $\mathbf{w}_{[-i]}^*$ , and by Lemma 3, the truncated full-horizon solution is the unperturbed subproblem solution, i.e.  $\mathbf{w}_{[i]}^\dagger(\mathbf{w}_{[-i]}^*) = \mathbf{w}_{[i]}^*$ . However, since  $\mathbf{w}_{[-i]}^*$  is unknown, the algorithm uses the previous iterate to specify the boundary variable, which results in a perturbation of  $\mathbf{w}_{[-i]}^*$ . Suppose we perturb  $\mathbf{w}_{[-i]}^*$  to  $\bar{\mathbf{w}}_{[-i]}$ , which corresponds to the perturbation of both initial and terminal stages. Theorem 7 makes use of the EDS property in Theorems 2 and 5 and shows that the stagewise error brought by the perturbation decays exponentially. In particular, for  $k \in [m_{i-1}, m_i]$ , we note that  $k - n_i^1 \wedge n_i^2 - k \geq \tau$ , which implies the stagewise error within  $[m_{i-1}, m_i]$  has been improved by at least a factor  $2\Upsilon\rho^\tau$ . Thus, if  $\tau \geq \log(2\Upsilon)/\log(1/\rho)$ , we can observe a clear improvement for middle stages  $[m_i, m_{i+1}]$ . This justifies discarding iterates for the subdomain overlaps and concatenating iterates of the non-overlapping subdomains.

We are now in a position to establish our main convergence result. Based on Lemma 3 and Theorem 7, we establish the local convergence of Algorithm 2.

**Theorem 8.** *Let Assumption 4 hold for the local solution  $\mathbf{w}^*$  of Problem (1) and  $\varepsilon > 0$ . There exist parameters  $\bar{\mu}, \bar{\tau} > 0$ , and a constant  $\delta > 0$  such that if  $\mu \geq \bar{\mu}$ ,  $\tau \geq \bar{\tau}$ , and  $\mathbf{w}^{(0)} \in \mathcal{N}_\delta(\mathbf{w}^*)$ , the following holds for  $\ell = 0, 1, \dots$ :*

$$\|\mathbf{w}^{(\ell)} - \mathbf{w}^*\|_{\mathbf{w}} \leq \alpha^\ell \|\mathbf{w}^{(0)} - \mathbf{w}^*\|_{\mathbf{w}}, \quad (16)$$

where  $\alpha = 2\Upsilon\rho^\tau$  is independent of  $N$ .

*Proof.* We choose  $\bar{\mu}$  in Lemma 2,  $\bar{\tau} = \lceil \log(2\Upsilon)/\log(1/\rho) \rceil + 1$ , and  $\delta$  defined in Theorem 6. Then  $\alpha < 1$ . We first show  $\mathbf{w}^{(\ell)} \in \mathcal{N}_\delta(\mathbf{w}^*)$  for  $\ell = 0, 1, \dots$  using mathematical induction. The claim trivially holds for  $\ell = 0$  from the assumption. Assume that the claim holds for  $\ell$ ; thus  $\mathbf{w}_{[-i]}^{(\ell)} \in \mathcal{N}_\delta(\mathbf{w}_{[-i]}^*)$ . From Theorem 7 and noting that for any  $i \in [T-1]$  and any  $\forall k \in [m_i, m_{i+1}]$ ,  $k - n_i^1 \wedge n_i^2 - k \geq \tau$ , we obtain

$$\|\mathbf{w}_k^{(\ell+1)} - \mathbf{w}_k^*\|_{\mathbf{w}} \leq \alpha \|\mathbf{w}_{[-i]}^{(\ell)} - \mathbf{w}_{[-i]}^*\|_{\mathbf{w}} \leq \alpha \|\mathbf{w}^{(\ell)} - \mathbf{w}^*\|_{\mathbf{w}}.$$

Taking maximum over  $k$  on the left hand side,

$$\|\mathbf{w}^{(\ell+1)} - \mathbf{w}^*\|_{\mathbf{w}} \leq \alpha \|\mathbf{w}^{(\ell)} - \mathbf{w}^*\|_{\mathbf{w}}. \quad (17)$$

From  $\alpha < 1$ , we have  $\mathbf{w}^{(\ell+1)} \in \mathcal{N}_\delta(\mathbf{w}^*)$  and the induction is complete. Recursively using (17), we obtain (16).  $\square$

Theorem 8 establishes local linear convergence of Algorithm 2. In summary, if SOSC, controllability condition, and boundedness condition are satisfied around the neighborhood of the local primal-dual solution of interest, the algorithm converges to the solution at a linear rate (provided that the overlap size  $\tau$  is sufficiently large). Furthermore, the convergence rate  $\alpha$  decays exponentially in  $\tau$ . One may observe that the overlap size may reach the maximum (i.e.,  $[n_i^1, n_i^2] = [0, N]$  for  $i \in [T-1]$ ) before  $\alpha < 1$  is achieved. In that case, the algorithm converges in one iteration (it becomes a centralized algorithm). However, since  $\Upsilon$  and  $\rho$  are parameters independent of  $N$ , when a problem with a sufficiently long horizon is considered, one can always obtain the exponential improvement of the convergence rate before reaching the maximum overlap. The argument states the existence of  $\bar{\mu}$ ,  $\bar{\tau}$  and  $\delta$ . The expression of  $\bar{\mu}$  is provided in (10);  $\bar{\tau}$  is selected by enforcing  $\alpha = 2\Upsilon\rho^{\bar{\tau}} < 1$  so that  $\bar{\tau} = \lceil \log(2\Upsilon)/\log(1/\rho) \rceil + 1$ . Here,  $\Upsilon, \rho$  are constants from Theorem 7, originating from the sensitivity analysis presented in the end of Section II. Their expressions in terms of the constants in Assumption 4 can be found in [27], but we do not present them here observing that they are often conservative in practice. As typical for local convergence analysis for NLP, the local radius  $\delta$  is not accessible due to the intrinsic nonlinearity of the problem (e.g., see Theorems 11.2 and 18.4 in the textbook [42] for comparison), while it is also not required for performing our algorithm.

As stated in Theorem 8, Algorithm 2 has two tuning parameters  $\mu$  and  $\tau$ , where the former is the penalty parameter for the subproblem (cf. (9)) and the latter is the overlap size. We note that the quadratic penalty is only added to the terminal state variable. Thus a very large  $\mu$  is equivalent to fixing the terminal state  $\mathbf{x}_{n_i^2}$  at  $\bar{\mathbf{x}}_{n_i^2}$ . We can let  $\mu$  be large to ensure the condition  $\mu \geq \bar{\mu}$  to be satisfied, though our experiments show that a moderate  $\mu$  also works well in practice. On the other hand, a larger  $\tau$  implies faster convergence, but also results in longer subproblems. In practice, we may tune  $\tau$  to balance the fast convergence rate and the increased subproblem complexity. Moreover, one benefit from our analysis is that tuning  $\mu$  and  $\tau$  is independent from horizon length  $N$ . Thus, we only target the subproblems with fixed, short horizons to tune  $\mu$  and  $\tau$ , and the same parameters work even when  $N$  is extremely large.

The convergence of Algorithm 2 can be monitored by checking the KKT residuals of (1). However, a more convenient surrogate of the full KKT residuals can be derived as follows.

**Proposition 1.** *Let  $\{\mathbf{w}^{(\ell)}\}_{\ell=0}^{\infty}$  be the sequence generated by Algorithm 2 with  $\tau \geq 1$ . Any limit point of the sequence satisfies the KKT conditions (11) for the full problem (1) if the following is satisfied for  $i \in (0, T)$  as  $\ell \rightarrow \infty$ :*

$$\begin{cases} T_{m_i}(\mathbf{x}_{[i-1]}^{\dagger}(\mathbf{w}_{[-(i-1)]}^{(\ell)})) - \mathbf{x}_{m_i}^{(\ell+1)} \rightarrow 0, \\ T_{m_i-1}(\boldsymbol{\lambda}_{[i]}^{\dagger}(\mathbf{w}_{[-i]}^{(\ell)})) - \boldsymbol{\lambda}_{m_i-1}^{(\ell+1)} \rightarrow 0. \end{cases}$$

*Proof.* Recalling that each  $\mathbf{w}_{[i]}^{\dagger}(\mathbf{w}_{[-i]}^{(\ell)})$  satisfies the KKT conditions of  $\mathcal{P}(\mathbf{w}_{[-i]}^{(\ell)})$ , one can observe that the KKT conditions (11) for the full problem (1) are violated only in the first equation of (11) over  $k \in \{m_i\}_{i=1}^{T-1}$  and in the fourth equation of (11) over  $k \in \{m_i - 1\}_{i=1}^{T-1}$ ; and the

residuals at iteration  $\ell + 1$  are  $T_{m_i-1}(\boldsymbol{\lambda}_{[i]}^{\dagger}(\mathbf{w}_{[-i]}^{(\ell)})) - \boldsymbol{\lambda}_{m_i-1}^{(\ell+1)}$  and  $T_{m_i}(\mathbf{x}_{[i-1]}^{\dagger}(\mathbf{w}_{[-(i-1)]}^{(\ell)})) - \mathbf{x}_{m_i}^{(\ell+1)}$ , respectively. Therefore, by the given condition, we have that (11) is satisfied for any limit points of the sequence.  $\square$

Accordingly, we define the monitoring metrics by

$$\begin{aligned} \epsilon_{\text{pr}}^{(\ell)} &= \max_{i \in (0, T)} \|T_{m_i}(\mathbf{x}_{[i-1]}^{\dagger}(\mathbf{w}_{[-(i-1)]}^{(\ell)})) - \mathbf{x}_{m_i}^{(\ell+1)}\|, \\ \epsilon_{\text{du}}^{(\ell)} &= \max_{i \in (0, T)} \|T_{m_i-1}(\boldsymbol{\lambda}_{[i]}^{\dagger}(\mathbf{w}_{[-i]}^{(\ell)})) - \boldsymbol{\lambda}_{m_i-1}^{(\ell+1)}\|, \end{aligned}$$

and then set the convergence criteria by

$$\text{stop if: } \epsilon_{\text{pr}}^{(\ell)} \leq \epsilon_{\text{pr}}^{\text{tol}} \text{ and } \epsilon_{\text{du}}^{(\ell)} \leq \epsilon_{\text{du}}^{\text{tol}},$$

for the given tolerance values  $\epsilon_{\text{pr}}^{\text{tol}}, \epsilon_{\text{du}}^{\text{tol}}$ .

Before showing numerical results, we discuss the global behavior of Schwarz decomposition. In general, there is no guarantee for the scheme to converge globally for nonlinear OCPs. As shown in Theorem 6, the solution mapping  $\mathbf{w}_{[i]}^{\dagger}$  exists only in a neighborhood of  $\mathbf{w}_{[-i]}^*$  and our main rate of convergence result, Theorem 8, is strictly local in nature. Outside of the neighborhood,  $\mathcal{P}_i(\mathbf{w}_{[-i]})$  may have infinite solutions or may have no solution due to nonlinearity. In general, we would need a merit function and a line search to ensure local convergence to a stationary point globally [42]. However, when we have more structure in the problem, the scheme can converge globally. For example, we show in the next theorem that Algorithm 2 converges globally for linear-quadratic OCPs. More general results using merit functions will be investigated in future research.

**Theorem 9.** *Let us consider linear-quadratic problems with*

$$\begin{aligned} g_k(\mathbf{x}_k, \mathbf{u}_k) &= \begin{pmatrix} \mathbf{x}_k \\ \mathbf{u}_k \end{pmatrix}^T \underbrace{\begin{pmatrix} Q_k & S_k^T \\ S_k & R_k \end{pmatrix}}_{H_k} \begin{pmatrix} \mathbf{x}_k \\ \mathbf{u}_k \end{pmatrix} + \begin{pmatrix} \mathbf{r}_k \\ \mathbf{s}_k \end{pmatrix}^T \begin{pmatrix} \mathbf{x}_k \\ \mathbf{u}_k \end{pmatrix} \\ g_N(\mathbf{x}_N) &= \mathbf{x}_N^T Q_N \mathbf{x}_N + \mathbf{r}_N^T \mathbf{x}_N \end{aligned}$$

and  $f_k(\mathbf{x}_k, \mathbf{u}_k) = A_k \mathbf{x}_k + B_k \mathbf{u}_k + \mathbf{v}_k$ . Suppose Assumption 4 holds for  $\{H_k, A_k, B_k\}$ . Then there exist parameters  $\bar{\mu}, \bar{\tau} > 0$  such that if  $\mu \geq \bar{\mu}$ ,  $\tau \geq \bar{\tau}$ , the linear convergence to the unique global solution  $\mathbf{w}^*$  of Problem (1), shown in (16), holds for any initial point  $\mathbf{w}^{(0)}$ .

We note that  $\{H_k, A_k, B_k\}$  do not depend on  $\mathbf{w}_k$ , and linear terms  $\{\mathbf{r}_k, \mathbf{s}_k, \mathbf{v}_k\}$  do not affect the Lagrangian Hessian and constraint Jacobian. Thus, Assumption 4 holds in the whole space.

*Proof.* First, each subproblem (9) is still a linear quadratic problem with any boundary variable  $\mathbf{w}_{[-i]}$ . By Lemma 2, there exists  $\bar{\mu}$  such that  $\mu \geq \bar{\mu}$  implies  $\text{Re}H^i \succeq \gamma_H I$ . Moreover, by LICQ of (9) (i.e. the Jacobian  $G$  has full row rank, see Definition 1), we know from [42, Lemma 16.1] that each subproblem has a unique, global solution for any  $\mathbf{w}_{[-i]}$ , denoted by  $\mathbf{w}_{[i]}^{\dagger}(\mathbf{w}_{[-i]})$ . Thus, Theorem 6 is applicable with the stated neighborhoods being the entire space.

Second, from Lemma 3, we know  $\mathbf{w}_{[i]}^{\dagger}(\mathbf{w}_{[-i]}^*) = \mathbf{w}_{[i]}^*$ . Moreover, the one-step error recursion in (13) holds directly following the same proof as in Theorem 7, and finally (17) holds as well. This shows (16) holds.  $\square$

Although Theorem 9 is limited to the linear-quadratic settings, we allow for the possibility of negative curvature in the objective. Specifically, note that Assumption 4(a) only requires  $ReH \succeq \gamma_H I$ , while existing results [25] require the strong convexity  $H \succeq \gamma_H I$ . This benefit comes from the convexification procedure in Algorithm 1 that our EDS results are based on.

It is worthwhile to mention that computation of the Newton step (search direction) for nonlinear OCPs effectively reduces to solving a linear-quadratic OCP. Accordingly, the overlapping Schwarz scheme can be used as a method to compute the search directions within second-order methods (such as interior-point approaches). Theorem 9 provides the global convergence proof for overlapping Schwarz-based step computations. We acknowledge that there is a wide range of decomposition methods for linear-quadratic OCPs (that include both iterative and direct approaches). Studying and comparing the scalability of these methods with that of the Schwarz scheme is an interesting direction of future work.

#### IV. NUMERICAL EXPERIMENTS

We apply the overlapping Schwarz scheme to a quadrotor control problem (governed by equations of motions) and to a thin plate temperature control problem (governed by nonlinear PDEs). Here, we aim to illustrate the convergence behavior of the Schwarz scheme and also to compare performance against state-of-the-art approaches such as ADMM and a centralized interior-point solver (Ipopt). We will demonstrate that Schwarz provides flexibility (as ADMM) and efficiency (as Ipopt). Our results also illustrate how EDS property arises in applications.

##### A. Quadrotor Control

We consider a quadrotor model studied in [46], [47]:

$$\begin{aligned} \frac{d^2 X}{dt^2} &= a(\cos \gamma \sin \beta \cos \alpha + \sin \gamma \sin \alpha) \\ \frac{d^2 Y}{dt^2} &= a(\cos \gamma \sin \beta \sin \alpha - \sin \gamma \cos \alpha) \\ \frac{d^2 Z}{dt^2} &= a \cos \gamma \cos \beta - \bar{g} \\ \frac{d\gamma}{dt} &= (\omega_X \cos \gamma + \omega_Y \sin \gamma) / \cos \beta \\ \frac{d\beta}{dt} &= -\omega_X \sin \gamma + \omega_Y \cos \gamma \\ \frac{d\alpha}{dt} &= \omega_X \cos \gamma \tan \beta + \omega_Y \sin \gamma \tan \beta + \omega_Z. \end{aligned}$$

Here,  $(X, Y, Z)$  are the positions,  $(\gamma, \beta, \alpha)$  are the angles, and  $\bar{g} = 9.8$  is the gravitational acceleration. We regard  $\mathbf{x} = (X, \dot{X}, Y, \dot{Y}, Z, \dot{Z}, \gamma, \beta, \alpha)$  as the state variables and  $\mathbf{u} = (a, \omega_X, \omega_Y, \omega_Z)$  as the control variables. The dynamics are discretized with an explicit Euler scheme with time step  $\Delta t = 0.005$  to obtain an OCP of the form of interest. Furthermore, the stage cost function is  $g_k(\mathbf{x}_k, \mathbf{u}_k) = \frac{1}{2} \|\mathbf{x}_k - \mathbf{x}_k^{\text{ref}}\|_Q^2 + \frac{1}{2} \|\mathbf{u}_k\|_R^2$ ;  $g_N(\mathbf{x}_N) = \frac{1}{2\Delta t} \|\mathbf{x}_N - \mathbf{x}_N^{\text{ref}}\|_Q^2$ ;  $Q = \text{diag}(1, 0, 1, 0, 1, 0, 1, 1, 1)$ ;  $R = \text{diag}(1/10, 1/10, 1/10, 1/10)$ ;  $\{\mathbf{x}_k^{\text{ref}}\}_{k=1}^N$  is generated from a sinusoidal function;  $N = 24,000$  (full problem; corresponds to 60 secs horizon);  $\mu = 1$ ; and  $\bar{\mathbf{x}}_0 = (0; 0; 0; 0; 0; 0; 0; 0; 0)$ .

##### B. Thin Plate Temperature Control

We consider a thin plate temperature control problem [48]:

$$\min_{x,u} \int_0^T \int_{w \in \Omega} \frac{1}{2} (x(w,t) - d(w,t))^2 + \frac{1}{2} ru(w,t)^2 dw dt \quad (18a)$$

$$\begin{aligned} \text{s.t. } \frac{\partial x(w,t)}{\partial t} &= -\Delta x(w,t) + \frac{2h_c}{\kappa t_z} (x(w,t) - \bar{T}) \\ &+ \frac{2\epsilon\sigma}{\kappa t_z} (x(w,t)^4 - \bar{T}^4) - \frac{1}{\kappa t_z} u(w,t), \\ w &\in \Omega, t \in [0, T] \end{aligned} \quad (18b)$$

$$x(w,t) = \bar{T}, \quad w \in \partial\Omega, \quad (18c)$$

where  $\Omega = [0, 1] \times [0, 1] \subseteq \mathbb{R}^2$  is the 2-dimensional domain of interest;  $x : \Omega \times [0, T] \rightarrow \mathbb{R}$  is the temperature;  $u : \Omega \times [0, T] \rightarrow \mathbb{R}$  is the control;  $\Delta$  is the Laplacian operator;  $\partial\Omega$  is the boundary of  $\Omega$ ;  $d : \Omega \times [0, T] \rightarrow \mathbb{R}$  is the desired temperature;  $r = 0.1$ ,  $\kappa = 400$ ,  $t_z = 0.01$ ,  $h_c = 1$ ,  $\epsilon = 0.5$ ,  $\sigma = 5.67 \times 10^{-8}$ , and  $\bar{T} = 300$  are the problem parameters (see [48]). The desired temperature data are generated from a sinusoidal function. The PDE in (18b) is governed by the heat equation which consists of convection, radiation, and forcing terms, and the Dirichlet boundary condition is enforced. We consider a discretized version of the problem: we discretize  $\Omega$  by a  $10 \times 10$  mesh and  $T = 24$  hour prediction horizon with  $\Delta t = 10$  secs.

##### C. Methods and Results

We first present a numerical verification of primal-dual EDS (Theorem 7) using the quadrotor problem. We first obtain the reference primal-dual solution trajectory by solving the full problem. Then, the perturbed trajectories are obtained by solving the problem with the perturbation on the given initial state  $\bar{\mathbf{x}}_0$ , and on the terminal state  $\mathbf{x}_N^{\text{ref}}$ . In particular, we solved the full problem with random perturbations  $\Delta \bar{\mathbf{x}}_0$  and  $\Delta \mathbf{x}_N^{\text{ref}}$  drawn from a zero-mean normal distribution.

The reference trajectory and 30 samples of the perturbed trajectories are shown in Fig. 3. One can see that the solution trajectories coalesce in the middle of the time domain and increase the spread at two boundaries. This result indicates that the sensitivity is decreasing toward the middle of the interval and verifies our theoretical results.

TABLE I  
ITERATIONS AND SOLUTION TIME AS A FUNCTION OF OVERLAP SIZE.

	$\tilde{\tau} = 0.3$	$\tilde{\tau} = 0.5$	$\tilde{\tau} = 1.0$
Iterations	31	11	7
Solution Time (sec)	6.03	2.41	2.46

We now illustrate the convergence of the Schwarz scheme for the quadrotor problem with 3 subdomains. The evolution of KKT errors with different overlap sizes are plotted in Fig. 4. Here, we expand the domain until the size of the extended domain reaches  $\tilde{\tau}$  times the original non-overlapping domain. Such relative criteria are often more practical because the scaling of the problem changes with discretization mesh size. We observe that the convergence rate improves dramatically as  $\tilde{\tau}$  increases. This result verifies Theorem 8. Fig. 5 further

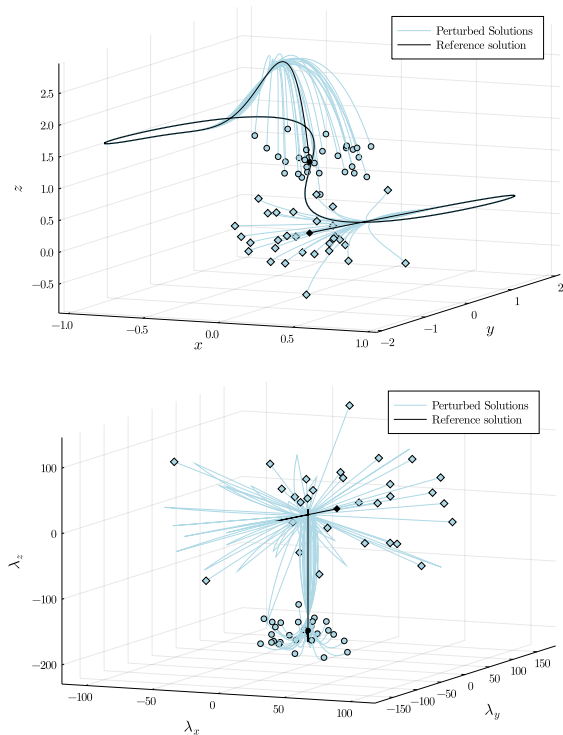


Fig. 3. Primal-dual exponential decay of sensitivity for quadrotor problem. The black line represents the reference trajectory; the light blue lines represent the perturbed trajectories; the diamonds represent the initial state; and the circles represent the terminal state. Observe the collapse of the perturbed trajectories to the reference one in the middle of the time intervals.

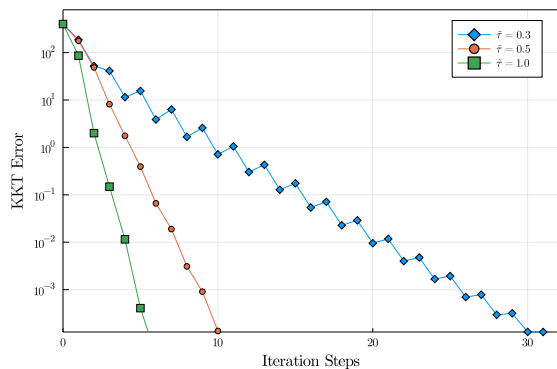


Fig. 4. Convergence of KKT residuals for overlapping Schwarz scheme.

illustrates convergence of the trajectories for  $\tilde{\tau} = 0.1$ . At the first iteration, the error is large at the boundaries and small in the middle of the domain. The error decays rapidly as the high-error components of the solution are discarded and the low-error components are kept. This behavior illustrates why EDS is central to achieve convergence. A computational trade-off exists for the Schwarz scheme when increasing  $\tilde{\tau}$  (since the subproblem complexity increases with  $\tilde{\tau}$ ). This trade-off is revealed from time per iteration (see Table I): we find the scheme takes 0.219 sec/iter when  $\tilde{\tau} = 0.5$  and 0.352 sec/iter when  $\tilde{\tau} = 1.0$ .

We also benchmark the Schwarz scheme against a centralized solver (Ipopt) and against a popular decomposition scheme

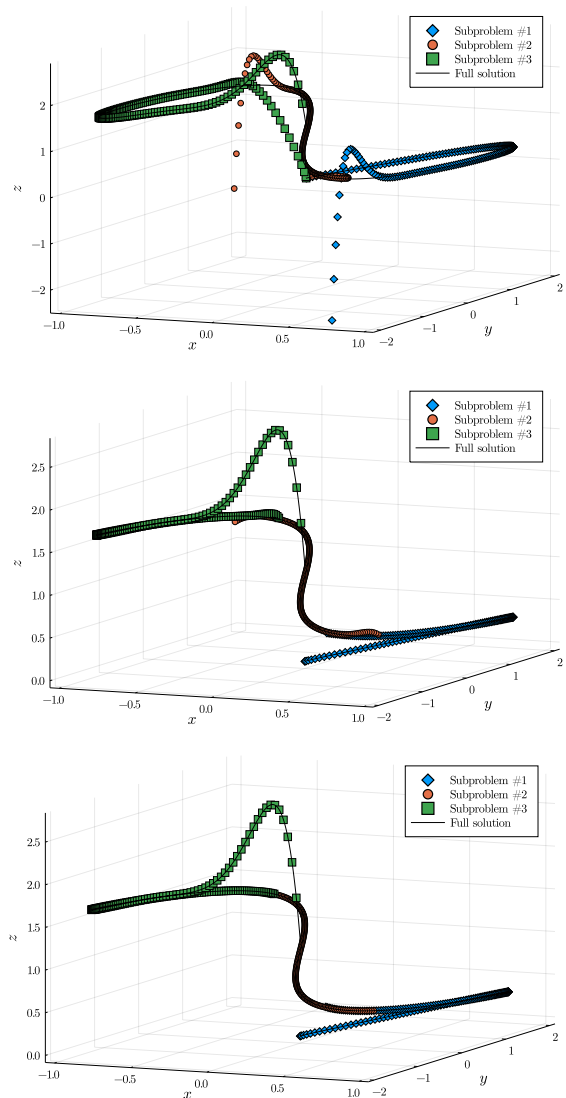


Fig. 5. Convergence of primal trajectory with  $\tilde{\tau} = 0.1$ . Top-to-bottom: iterations 1, 2, and 3; blue, red, green markers are solutions from subproblems 1, 2, and 3, respectively; black line is solution trajectory.

(ADMM) for solving the above two problems. For both the Schwarz and ADMM schemes, we partition the domain into 20 intervals with the same length. For the Schwarz scheme, we expand each interval by (8) with the relative size of overlap  $\tilde{\tau} = 1.0$ . For ADMM, subproblems are formulated by introducing duplicate variables and decomposing on the time domain [49]. To ensure consistency, subproblems in the Schwarz and ADMM schemes are all solved with Ipopt [50], configured with the sparse solver MA27 [51]. The study is run on a multicore parallel computing server (shared memory and 40 cores of Intel Xeon Gold 6140 CPU running at 2.30GHz) using the multi-thread parallelism in Julia. For both the Schwarz and ADMM schemes, we vary the penalty parameter as indicated in Figure 6. All results can be reproduced using the provided scripts in <https://github.com/zavalab/JuliaBox/tree/master/SchwarzOCP>.

One can see that, for both problems, the overlapping Schwarz scheme has much faster convergence than ADMM (Fig. 6)

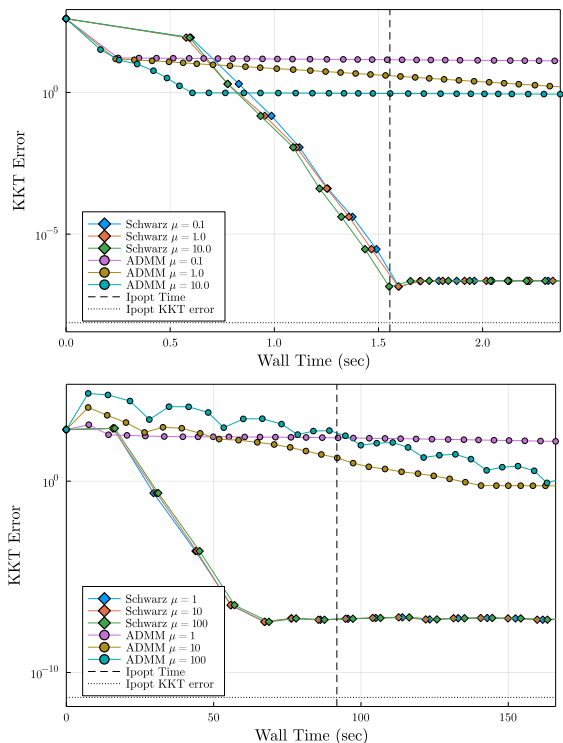


Fig. 6. Benchmark of overlapping Schwarz against Ipopt and ADMM. Top: Quadrotor control problem; bottom: Thin plate temperature control problem.

regardless of the choice of  $\mu$ . One can also observe that the performance of ADMM is sensitive to the choice of  $\mu$ , while the performance Schwarz scheme is insensitive to it. We note that ADMM tends to decrease the overall error but eventually the error settles to a rather high value. We also emphasize that ADMM does not have convergence guarantees for the general nonconvex OCPs considered here (as discussed in Section I). In contrast, the overlapping Schwarz scheme converges almost as fast as centralized solver Ipopt. The final accuracy of the Schwarz scheme ( $10^{-6}$ ) is much higher than that achieved by ADMM, but not as that achieved by Ipopt (less than  $10^{-8}$ ). The difference between the accuracy between Schwarz and Ipopt are due to the fact that Schwarz is an iterative scheme, while the linear algebra performed inside Ipopt uses a direct linear solver (MA27). Direct solvers are known for delivering high accuracy. We highlight, however, that in many control applications there is often flexibility to deliver moderate accuracy, so being able to deliver moderately accurate solution in a reasonably short time can be a favorable characteristic. For the thin plate temperature control problem, the accuracy of ADMM is notably worse than that achieved with the Schwarz scheme.

To sum up, the overlapping Schwarz scheme is an efficient method to solve OCPs and offers flexibility to be implemented in different computing hardware architectures.

## V. CONCLUSIONS

We established the convergence properties of an overlapping Schwarz decomposition scheme for general nonlinear optimal control problems. Under standard SOSC, controllability, and boundedness conditions for the full problem, we showed that

the scheme enjoys linear convergence locally, with a linear rate that improves exponentially with the size of the overlap. Central to our convergence proof is a primal-dual parametric sensitivity result that we call exponential decay of sensitivity. We also provided a global convergence proof for the Schwarz scheme for the linear-quadratic OCP case. This result is of relevance, as it suggests that the scheme could be used to solve linear systems inside NLP solvers. Computational results reveal that the Schwarz scheme is significantly more efficient than ADMM and as efficient as the centralized NLP solver Ipopt. In future work, we will seek to expand our results to alternative problem structures (e.g., networks and stochastic programs). Moreover, it will be interesting to compare performance in different hardware architectures (e.g., embedded systems) and against different decomposition schemes (e.g., Riccati and block cyclic reduction).

## ACKNOWLEDGMENT

This material is based upon work supported by the U.S. Department of Energy, Office of Science, Office of Advanced Scientific Computing Research (ASCR) under Contract DE-AC02-06CH11347 and by NSF through award CNS-1545046 and ECCS-1609183. We also acknowledge partial support from the National Science Foundation under award NSF-ECCS-1609183.

## APPENDIX

The missing proofs in Section II-B are presented here.

### A. Proof of Theorem 3

Under Assumption 1, we know that  $(\mathbf{p}^*, \mathbf{q}^*, \zeta^*)$  is a unique global solution of  $\mathcal{LQP}$ . When executing Algorithm 1 with  $\beta \in (0, \gamma_H)$ , we know from [27, Theorem 3.8] that  $\tilde{H}_k \succ 0$  (i.e.  $\tilde{H}_k(\beta)$  in their notation). Thus,  $(\mathbf{p}^{c*}, \mathbf{q}^{c*}, \zeta^{c*})$  is also a unique global solution of  $\mathcal{CLQP}$ . By [27, Lemma 3.4], we know that  $\mathcal{LQP}$  and  $\mathcal{CLQP}$  have the same objective. Since Algorithm 1 does not change constraint matrices  $A_k, B_k, C_k$  of (5), we have  $\mathbf{p}^* = \mathbf{p}^{c*}$  and  $\mathbf{q}^* = \mathbf{q}^{c*}$ . We now establish the relation of dual solutions  $\zeta^*$  and  $\zeta^{c*}$  by studying KKT conditions of Problem (5). To simplify notation, we denote the  $k$ -th component of the objective by

$$O_k(\mathbf{p}_k, \mathbf{q}_k) = \begin{pmatrix} \mathbf{p}_k \\ \mathbf{q}_k \\ \mathbf{l}_k \end{pmatrix}^T \begin{pmatrix} Q_k & S_k^T & D_{k1}^T \\ S_k & R_k & D_{k2}^T \\ D_{k1} & D_{k2} & \mathbf{0} \end{pmatrix} \begin{pmatrix} \mathbf{p}_k \\ \mathbf{q}_k \\ \mathbf{l}_k \end{pmatrix}, \forall k \in [N-1],$$

$$O_N(\mathbf{p}_N) = \begin{pmatrix} \mathbf{p}_N \\ \mathbf{l}_N \end{pmatrix}^T \begin{pmatrix} Q_N & D_N^T \\ D_N & \mathbf{0} \end{pmatrix} \begin{pmatrix} \mathbf{p}_N \\ \mathbf{l}_N \end{pmatrix}.$$

Similarly, we define  $\tilde{O}_k(\mathbf{p}_k, \mathbf{q}_k)$  and  $\tilde{O}_N(\mathbf{p}_N)$  by replacing  $H_k, D_k$  by  $\tilde{H}_k, \tilde{D}_k$ . The KKT system of  $\mathcal{LQP}$  is then given by

$$\mathbf{0} = \begin{cases} \nabla_{\mathbf{p}_k} O_k(\mathbf{p}_k, \mathbf{q}_k) + \zeta_{k-1} - A_k^T \zeta_k, & \forall k \in [N-1], \\ \nabla_{\mathbf{q}_k} O_k(\mathbf{p}_k, \mathbf{q}_k) - B_k^T \zeta_k, & \forall k \in [N-1], \\ \nabla_{\mathbf{p}_N} O_N(\mathbf{p}_N) + \zeta_{N-1}, \\ \mathbf{p}_{k+1} - (A_k \mathbf{p}_k + B_k \mathbf{q}_k + C_k \mathbf{l}_k), & \forall k \in [N-1], \\ \mathbf{p}_0 - \mathbf{l}_{-1}. \end{cases} \quad (19)$$

For the KKT system of  $\mathcal{CLQP}$ , we replace  $\nabla O_k$  by  $\nabla \tilde{O}_k$  and  $\zeta$  by  $\zeta^c$  in (19) since two problems have the same linear-quadratic form. By Algorithm 1, we know that  $\forall k \in [N-1]$ ,

$$\begin{aligned} \nabla_{\mathbf{p}_k} \tilde{O}_k(\mathbf{p}_k, \mathbf{q}_k) &= 2\tilde{Q}_k \mathbf{p}_k + 2\tilde{S}_k^T \mathbf{q}_k + 2\tilde{D}_{k1}^T \mathbf{l}_k \\ &= 2(\tilde{Q}_k - \bar{Q}_k) \mathbf{p}_k + 2\tilde{S}_k^T \mathbf{q}_k + 2\tilde{D}_{k1}^T \mathbf{l}_k \\ &= 2Q_k \mathbf{p}_k + 2S_k^T \mathbf{q}_k + 2D_{k1}^T \mathbf{l}_k - 2\bar{Q}_k \mathbf{p}_k \\ &\quad + 2A_k^T \bar{Q}_{k+1} (A_k \mathbf{p}_k + B_k \mathbf{q}_k + C_k \mathbf{l}_k) \\ &= \nabla_{\mathbf{p}_k} O_k(\mathbf{p}_k, \mathbf{q}_k) - 2\bar{Q}_k \mathbf{p}_k + 2A_k^T \bar{Q}_{k+1} \mathbf{p}_{k+1}, \end{aligned} \quad (20)$$

where the last equality results from definition of  $O_k$  and the  $k$ -th dynamic constraint. We can also show that

$$\begin{aligned} \nabla_{\mathbf{q}_k} \tilde{O}_k(\mathbf{p}_k, \mathbf{q}_k) &= \nabla_{\mathbf{q}_k} O_k(\mathbf{p}_k, \mathbf{q}_k) + 2B_k^T \bar{Q}_{k+1} \mathbf{p}_{k+1}, \\ \nabla_{\mathbf{p}_N} \tilde{O}_N(\mathbf{p}_N) &= \nabla_{\mathbf{p}_N} O_N(\mathbf{p}_N) - 2\bar{Q}_N \mathbf{p}_N. \end{aligned} \quad (21)$$

Plugging (20), (21) back into (19), we obtain that the KKT system of  $\mathcal{LQP}$  is equivalent to

$$\mathbf{0} = \begin{cases} \nabla_{\mathbf{p}_k} \tilde{O}_k + (\zeta_{k-1} + 2\bar{Q}_k \mathbf{p}_k) - A_k^T (\zeta_k + 2\bar{Q}_{k+1} \mathbf{p}_{k+1}), \\ \nabla_{\mathbf{q}_k} \tilde{O}_k - B_k^T (\zeta_k + 2\bar{Q}_{k+1} \mathbf{p}_{k+1}), & \forall k \in [N-1], \\ \nabla_{\mathbf{p}_N} \tilde{O}_N + (\zeta_{N-1} + 2\bar{Q}_N \mathbf{p}_N), \\ \mathbf{p}_{k+1} - (A_k \mathbf{p}_k + B_k \mathbf{q}_k + C_k \mathbf{l}_k), & \forall k \in [N-1], \\ \mathbf{p}_0 - \mathbf{l}_{-1}. \end{cases}$$

Comparing the above equation with the KKT system of  $\mathcal{CLQP}$ , and using the invariance of the primal solution, we see that  $(\mathbf{p}^{c*}, \mathbf{q}^{c*}, \zeta^* + 2\bar{Q}\mathbf{p}^*)$  satisfies the KKT system of  $\mathcal{CLQP}$ . Since LICQ holds for  $\mathcal{CLQP}$ , the dual solution is unique. This implies  $\zeta^{c*} = \zeta^* + 2\bar{Q}\mathbf{p}^*$  and we complete the proof.

### B. Proof of Theorem 4

First of all, the invertibility of  $W_k$  is guaranteed by Assumption 1, as directly shown in [27, Lemma 3.5(i)]. We use reverse induction to prove the formula of  $\zeta_k^*$ . According to (19), for  $k = N-1$  we have

$$\zeta_{N-1}^* = -\nabla_{\mathbf{p}_N} O_N(\mathbf{p}_N^*) = -2Q_N \mathbf{p}_N^* - 2D_N^T \mathbf{l}_N,$$

which satisfies (7) and proves the first induction step. Suppose  $\zeta_k^*$  satisfies (7). From (19), we have

$$\begin{aligned} \zeta_{k-1}^* &= A_k^T \zeta_k^* - \nabla_{\mathbf{p}_k} O_k(\mathbf{p}_k^*, \mathbf{q}_k^*) \\ &= A_k^T \zeta_k^* - 2Q_k \mathbf{p}_k^* - 2S_k^T \mathbf{q}_k^* - 2D_{k1}^T \mathbf{l}_k. \end{aligned}$$

Plugging the expression for  $\zeta_k^*$  from (7), we get

$$\begin{aligned} \zeta_{k-1}^* &= -2A_k^T K_{k+1} \mathbf{p}_{k+1}^* + 2A_k^T \left( \sum_{i=k+1}^N (M_i^{k+1})^T \mathbf{l}_i \right. \\ &\quad \left. + \sum_{i=k+1}^{N-1} (V_i^{k+1})^T C_i \mathbf{l}_i \right) - 2Q_k \mathbf{p}_k^* - 2S_k^T \mathbf{q}_k^* - 2D_{k1}^T \mathbf{l}_k \\ &= -2(A_k^T K_{k+1} A_k + Q_k) \mathbf{p}_k^* - 2(S_k + B_k^T K_{k+1} A_k)^T \mathbf{q}_k^* \\ &\quad + 2A_k^T \left\{ \sum_{i=k+1}^N (M_i^{k+1})^T \mathbf{l}_i + \sum_{i=k+1}^{N-1} (V_i^{k+1})^T C_i \mathbf{l}_i \right\} \\ &\quad - 2A_k^T K_{k+1} C_k \mathbf{l}_k - 2D_{k1}^T \mathbf{l}_k \\ &= -2(A_k^T K_{k+1} A_k + Q_k) \mathbf{p}_k^* + 2P_k^T W_k \mathbf{q}_k^* \end{aligned}$$

$$\begin{aligned} &+ 2A_k^T \left\{ \sum_{i=k+1}^N (M_i^{k+1})^T \mathbf{l}_i + \sum_{i=k+1}^{N-1} (V_i^{k+1})^T C_i \mathbf{l}_i \right\} \\ &- 2A_k^T K_{k+1} C_k \mathbf{l}_k - 2D_{k1}^T \mathbf{l}_k, \end{aligned}$$

where the second equality follows from  $\mathbf{p}_{k+1}^* - (A_k \mathbf{p}_k^* + B_k \mathbf{q}_k^* + C_k \mathbf{l}_k) = \mathbf{0}$ , and the third equality follows from the definition of  $P_k$ . By [27, Lemma 3.5(ii)], we have

$$\begin{aligned} \mathbf{q}_k^* &= P_k \mathbf{p}_k^* + W_k^{-1} B_k^T \sum_{i=k+1}^N (M_i^{k+1})^T \mathbf{l}_i - W_k^{-1} D_{k2}^T \mathbf{l}_k \\ &\quad + W_k^{-1} B_k^T \sum_{i=k+1}^{N-1} (V_i^{k+1})^T C_i \mathbf{l}_i - W_k^{-1} B_k^T K_{k+1} C_k \mathbf{l}_k. \end{aligned}$$

Combining the above two displays, we obtain

$$\begin{aligned} \zeta_{k-1}^* &= -2(A_k^T K_{k+1} A_k + Q_k) \mathbf{p}_k^* + 2P_k^T W_k \left( P_k \mathbf{p}_k^* \right. \\ &\quad \left. + W_k^{-1} B_k^T \left\{ \sum_{i=k+1}^N (M_i^{k+1})^T \mathbf{l}_i + \sum_{i=k+1}^N (V_i^{k+1})^T C_i \mathbf{l}_i \right\} \right. \\ &\quad \left. - W_k^{-1} B_k^T K_{k+1} C_k \mathbf{l}_k - W_k^{-1} D_{k2}^T \mathbf{l}_k \right) - 2D_{k1}^T \mathbf{l}_k \\ &\quad + 2A_k^T \left( \sum_{i=k+1}^N (M_i^{k+1})^T \mathbf{l}_i + \sum_{i=k+1}^{N-1} (V_i^{k+1})^T C_i \mathbf{l}_i \right) \\ &\quad - 2A_k^T K_{k+1} C_k \mathbf{l}_k \\ &= -2(A_k^T K_{k+1} A_k + Q_k - P_k^T W_k P_k) \mathbf{p}_k^* \\ &\quad + 2E_k^T \left( \sum_{i=k+1}^N (M_i^{k+1})^T \mathbf{l}_i + \sum_{i=k+1}^{N-1} (V_i^{k+1})^T C_i \mathbf{l}_i \right) \\ &\quad - 2E_k^T K_{k+1} C_k \mathbf{l}_k - 2(D_{k1} + D_{k2} P_k)^T \mathbf{l}_k \\ &= -2K_k \mathbf{p}_k^* + 2 \sum_{i=k}^N (M_i^k)^T \mathbf{l}_i + 2 \sum_{i=k}^N (V_i^k)^T C_i \mathbf{l}_i, \end{aligned}$$

where the second equality follows from the definition of  $E_k$  and the third equality follows from definitions of  $K_k$ ,  $M_i^k$ , and  $V_i^k$ . This verifies the induction step and finishes the proof.

### C. Proof of Lemma 1

We use the closed form of  $\zeta^{c*}$  established in Theorem 4. We mention that all matrices are calculated based on  $\{\tilde{H}_k, \tilde{D}_k\}$ . For any  $k \in [-1, N-1]$ , we consider two cases.

(a)  $\mathbf{l} = \mathbf{e}_i$  for  $i \in [N]$ . We then have three subcases.

(a1)  $i \leq k$ . In this case, we apply (7) and immediately have for some constant  $\Upsilon_1 > 0$  that

$$\|\zeta_k^{c*}\| = \|-2K_{k+1} \mathbf{p}_{k+1}^*\| \leq 2\|K_{k+1}\| \|\mathbf{p}_{k+1}^*\| \leq 2\Upsilon_1 \Upsilon \rho^{k+1-i}. \quad (22)$$

Here, the last inequality is due to Theorem 2 that  $\|\mathbf{p}_{k+1}^*\| \leq \Upsilon \rho^{k+1-i}$ , and the fact that  $\|K_{k+1}\| \leq \Upsilon_1$  for some constant  $\Upsilon_1 > 0$ , stated precisely in [27, (4.7)].

(a2)  $k+1 \leq i \leq N-1$ . We apply (7) and have for some constants  $\Upsilon_2, \Upsilon_3 > 0$  that

$$\begin{aligned} \|\zeta_k^{c*}\| &= \|-2K_{k+1} \mathbf{p}_{k+1}^* + 2(M_i^{k+1})^T \mathbf{e}_i + 2(V_i^{k+1})^T C_i \mathbf{e}_i\| \\ &\leq 2\Upsilon_1 \Upsilon \rho^{i-k-1} + 2\Upsilon_2 \rho^{i-k-1} + 2\Upsilon_3 \Upsilon_{\text{upper}} \rho^{i-k} \end{aligned}$$

$$\leq 2(\Upsilon_1\Upsilon + \Upsilon_2 + \Upsilon_3\Upsilon_{\text{upper}})\rho^{i-k-1}.$$

The second inequality is due to (22), the fact that  $\|C_i\| \leq \Upsilon_{\text{upper}}$  in Assumption 3, and the fact that

$$\|M_i^{k+1}\| \leq \Upsilon_2\rho^{i-k-1}, \quad \|V_i^{k+1}\| \leq \Upsilon_3\rho^{i-k}$$

for some constants  $\Upsilon_2, \Upsilon_3 > 0$ , stated precisely in [27, (5.11)].

(a3)  $i = N$ . Analogous to the derivations in (a2), we apply (7) and have

$$\begin{aligned} \|\zeta_k^{c*}\| &= \|-2K_{k+1}p_{k+1}^* + 2(M_N^{k+1})^T e_N\| \\ &\leq 2\Upsilon_1\Upsilon\rho^{N-k-1} + 2\Upsilon_2\rho^{N-k-1} = 2(\Upsilon_1\Upsilon + \Upsilon_2)\rho^{N-k-1}. \end{aligned}$$

(b)  $l = e_{-1}$ . Applying (7) and Theorem 2, we obtain that

$$\|\zeta_k^{c*}\| = \|-2K_{k+1}p_{k+1}^*\| \leq 2\Upsilon_1\Upsilon\rho^{k+1}.$$

Combining the above two cases, we let  $\Upsilon' = 2(\Upsilon_1\Upsilon + \Upsilon_2 + \Upsilon_3\Upsilon_{\text{upper}})$  and complete the proof.

#### D. Proof of Theorem 5

By Lemma 1 and Theorem 3 we have for all  $k \in [-1, N-1]$  that

$$\|\zeta_k^*\| = \|\zeta_k^* - 2\bar{Q}_{k+1}p_{k+1}^*\| \leq \|\zeta_k^*\| + 2\|\bar{Q}_{k+1}\| \|p_{k+1}^*\|.$$

By [27, Theorem 3.8, Claim 1] and [27, Lemma 4.3], we know  $\bar{Q}_{k+1}$  (which is  $\bar{Q}_{k+1}(\delta)$  in their context) satisfies  $\|\bar{Q}_{k+1}\| \leq \Upsilon_Q$  for some constant  $\Upsilon_Q$ . Thus, by Theorem 2, we can let  $\Upsilon'' = \Upsilon' + 2\Upsilon_Q\Upsilon$  to complete the proof.

#### REFERENCES

- [1] J. Rawlings, *Model predictive control: theory, computation, and design*. Madison, Wisconsin: Nob Hill Publishing, 2017, vol. 2. [Online]. Available: <http://www.nobhillpublishing.com/mpc/index-mpc.html>
- [2] S. Qin and T. A. Badgwell, "A survey of industrial model predictive control technology," *Control Engineering Practice*, vol. 11, no. 7, pp. 733–764, jul 2003. [Online]. Available: [https://doi.org/10.1016/S0967-0661\(02\)00186-7](https://doi.org/10.1016/S0967-0661(02)00186-7)
- [3] R. Kumar, M. J. Wenzel, M. N. ElBsat, M. J. Risbeck, K. H. Drees, and V. M. Zavala, "Stochastic model predictive control for central HVAC plants," *Journal of Process Control*, vol. 90, pp. 1–17, jun 2020. [Online]. Available: <https://doi.org/10.1016/j.jprocont.2020.03.015>
- [4] J. R. Jackson and I. E. Grossmann, "Temporal decomposition scheme for nonlinear multisite production planning and distribution models," *Industrial & Engineering Chemistry Research*, vol. 42, no. 13, pp. 3045–3055, jun 2003. [Online]. Available: <https://doi.org/10.1021/ie030070p>
- [5] P. Falcone, F. Borrelli, J. Asgari, H. E. Tseng, and D. Hrovat, "Predictive active steering control for autonomous vehicle systems," *IEEE Transactions on Control Systems Technology*, vol. 15, no. 3, pp. 566–580, may 2007. [Online]. Available: <https://doi.org/10.1109/TCST.2007.894653>
- [6] H. Shانهchi, N. Pariz, and E. Vaahedi, "General nonlinear modal representation of large scale power systems," *IEEE Transactions on Power Systems*, vol. 18, no. 3, pp. 1103–1109, aug 2003. [Online]. Available: <https://doi.org/10.1109/tpwrs.2003.814883>
- [7] W. B. Dunbar, "Distributed receding horizon control of dynamically coupled nonlinear systems," *IEEE Transactions on Automatic Control*, vol. 52, no. 7, pp. 1249–1263, jul 2007. [Online]. Available: <https://doi.org/10.1109/TAC.2007.900828>
- [8] J.-Q. Huang and F. Lewis, "Neural-network predictive control for nonlinear dynamic systems with time-delay," *IEEE Transactions on Neural Networks*, vol. 14, no. 2, pp. 377–389, mar 2003. [Online]. Available: <https://doi.org/10.1109/tmn.2003.809424>
- [9] R. Kumar, M. J. Wenzel, M. J. Ellis, M. N. ElBsat, K. H. Drees, and V. M. Zavala, "Handling long horizons in MPC: A stochastic programming approach," in *2018 Annual American Control Conference (ACC)*, IEEE. IEEE, jun 2018, pp. 715–720. [Online]. Available: <https://doi.org/10.23919/acc.2018.8430780>
- [10] A. Beccuti, T. Geyer, and M. Morari, "Temporal lagrangian decomposition of model predictive control for hybrid systems," in *2004 43rd IEEE Conference on Decision and Control (CDC) (IEEE Cat. No.04CH37601)*, vol. 3, IEEE. IEEE, 2004, pp. 2509–2514. [Online]. Available: <https://doi.org/10.1109/cdc.2004.1428793>
- [11] S. Boyd, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends® in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2010. [Online]. Available: <https://doi.org/10.1561/22000000016>
- [12] R. Kumar, M. J. Wenzel, M. J. Ellis, M. N. ElBsat, K. H. Drees, and V. M. Zavala, "A stochastic dual dynamic programming framework for multiscale MPC," *IFAC-PapersOnLine*, vol. 51, no. 20, pp. 493–498, 2018. [Online]. Available: <https://doi.org/10.1016/j.ifacol.2018.11.041>
- [13] V. M. Zavala, "New architectures for hierarchical predictive control," *IFAC-PapersOnLine*, vol. 49, no. 7, pp. 43–48, 2016. [Online]. Available: <https://doi.org/10.1016/j.ifacol.2016.07.214>
- [14] N.-Y. Chiang, R. Huang, and V. M. Zavala, "An augmented lagrangian filter method for real-time embedded optimization," *IEEE Transactions on Automatic Control*, vol. 62, no. 12, pp. 6110–6121, dec 2017. [Online]. Available: <https://doi.org/10.1109/tac.2017.2694806>
- [15] A. Kozma, C. Conte, and M. Diehl, "Benchmarking large-scale distributed convex quadratic programming algorithms," *Optimization Methods and Software*, vol. 30, no. 1, pp. 191–214, may 2014. [Online]. Available: <https://doi.org/10.1080/10556788.2014.911298>
- [16] I. Nielsen and D. Axehill, "An o (log n) parallel algorithm for newton step computation in model predictive control," *IFAC Proceedings Volumes*, vol. 47, no. 3, pp. 10505–10511, 2014. [Online]. Available: <https://doi.org/10.3182/20140824-6-ZA-1003.01577>
- [17] —, "A parallel structure exploiting factorization algorithm with applications to model predictive control," in *2015 54th IEEE Conference on Decision and Control (CDC)*, IEEE. IEEE, dec 2015, pp. 3932–3938. [Online]. Available: <https://doi.org/10.1109/CDC.2015.7402830>
- [18] F. Laine and C. Tomlin, "Parallelizing LQR computation through endpoint-explicit riccati recursion," in *2019 IEEE 58th Conference on Decision and Control (CDC)*, IEEE. IEEE, dec 2019, pp. 1395–1402. [Online]. Available: <https://doi.org/10.1109/CDC40024.2019.9029974>
- [19] S. J. Wright, "Solution of discrete-time optimal control problems on parallel computers," *Parallel Computing*, vol. 16, no. 2-3, pp. 221–237, dec 1990. [Online]. Available: [https://doi.org/10.1016/0167-8191\(90\)90060-M](https://doi.org/10.1016/0167-8191(90)90060-M)
- [20] C. V. Rao, S. J. Wright, and J. B. Rawlings, "Application of interior-point methods to model predictive control," *Journal of Optimization Theory and Applications*, vol. 99, no. 3, pp. 723–757, dec 1998. [Online]. Available: <https://doi.org/10.1023/A:1021711402723>
- [21] W. Wan, J. P. Eason, B. Nicholson, and L. T. Biegler, "Parallel cyclic reduction decomposition for dynamic optimization problems," *Computers & Chemical Engineering*, vol. 120, pp. 54–69, jan 2019. [Online]. Available: <https://doi.org/10.1016/j.compchemeng.2017.09.023>
- [22] J. Kang, N. Chiang, C. D. Laird, and V. M. Zavala, "Nonlinear programming strategies on high-performance computers," in *2015 54th IEEE Conference on Decision and Control (CDC)*, IEEE. IEEE, dec 2015, pp. 4612–4620. [Online]. Available: <https://doi.org/10.1109/CDC.2015.7402938>
- [23] G. Frison and J. B. Jorgensen, "Efficient implementation of the riccati recursion for solving linear-quadratic control problems," in *2013 IEEE International Conference on Control Applications (CCA)*, IEEE. IEEE, aug 2013, pp. 1117–1122. [Online]. Available: <https://doi.org/10.1109/CCA.2013.6662901>
- [24] C. Barrows, M. Hummon, W. Jones, and E. Hale, "Time domain partitioning of electricity production cost simulations," *National Renewable Energy Lab.(NREL), Golden, CO (United States)*, jan 2014. [Online]. Available: <https://doi.org/10.2172/1123223>
- [25] W. Xu and M. Anitescu, "Exponentially accurate temporal decomposition for long-horizon linear-quadratic dynamic optimization," *SIAM Journal on Optimization*, vol. 28, no. 3, pp. 2541–2573, jan 2018. [Online]. Available: <https://doi.org/10.1137/16M1081993>
- [26] S. Shin, T. Faulwasser, M. Zanon, and V. M. Zavala, "A parallel decomposition scheme for solving long-horizon optimal control problems," in *2019 IEEE 58th Conference on Decision and Control (CDC)*, IEEE, dec 2019, pp. 5264–5271. [Online]. Available: <https://doi.org/10.1109/cdc40024.2019.9030139>
- [27] S. Na and M. Anitescu, "Exponential decay in the sensitivity analysis of nonlinear dynamic programming," *SIAM Journal on Optimization*, vol. 30, no. 2, pp. 1527–1554, jan 2020. [Online]. Available: <https://doi.org/10.1137/19M1265065>
- [28] S. Shin, C. Coffrin, K. Sundar, and V. M. Zavala, "Graph-based modeling and decomposition of energy infrastructures," *arXiv preprint*

- arXiv:2010.02404*, 2020. [Online]. Available: <https://arxiv.org/abs/2010.02404>
- [29] S. Na and M. Anitescu, "Superconvergence of online optimization for model predictive control," *arXiv preprint arXiv:2001.03707*, 2020. [Online]. Available: <https://arxiv.org/abs/2001.03707>
- [30] T. Ohtsuka, "A continuation/GMRES method for fast computation of nonlinear receding horizon control," *Automatica*, vol. 40, no. 4, pp. 563–574, apr 2004. [Online]. Available: <https://doi.org/10.1016/j.automatica.2003.11.005>
- [31] M. Diehl, H. G. Bock, and J. P. Schlöder, "A real-time iteration scheme for nonlinear optimization in optimal feedback control," *SIAM Journal on Control and Optimization*, vol. 43, no. 5, pp. 1714–1736, jan 2005. [Online]. Available: <https://doi.org/10.1137/S0363012902400713>
- [32] V. M. Zavala and L. T. Biegler, "The advanced-step NMPC controller: Optimality, stability and robustness," *Automatica*, vol. 45, no. 1, pp. 86–93, jan 2009. [Online]. Available: <https://doi.org/10.1016/j.automatica.2008.06.011>
- [33] V. M. Zavala and M. Anitescu, "Real-time nonlinear optimization as a generalized equation," *SIAM Journal on Control and Optimization*, vol. 48, no. 8, pp. 5444–5467, jan 2010. [Online]. Available: <https://doi.org/10.1137/090762634>
- [34] D. Collet, M. Alamir, D. D. Domenico, and G. Sabiron, "Non quadratic smooth model of fatigue for optimal fatigue-oriented individual pitch control," in *Journal of Physics: Conference Series*, vol. 1618, no. 2, IOP Publishing. IOP Publishing, sep 2020, p. 022004. [Online]. Available: <https://doi.org/10.1088/1742-6596/1618/2/022004>
- [35] M. Hong, Z.-Q. Luo, and M. Razaviyayn, "Convergence analysis of alternating direction method of multipliers for a family of nonconvex problems," *SIAM Journal on Optimization*, vol. 26, no. 1, pp. 337–364, jan 2016. [Online]. Available: <https://doi.org/10.1137/140990309>
- [36] Y. Wang, W. Yin, and J. Zeng, "Global convergence of ADMM in nonconvex nonsmooth optimization," *Journal of Scientific Computing*, vol. 78, no. 1, pp. 29–63, jun 2018. [Online]. Available: <https://doi.org/10.1007/s10915-018-0757-z>
- [37] S. Shin and V. M. Zavala, "Diffusing-horizon model predictive control," *arXiv preprint arXiv:2002.08556*, 2020. [Online]. Available: <https://arxiv.org/abs/2002.08556>
- [38] L. Grüne, M. Schaller, and A. Schiela, "Efficient mpc for parabolic pdes with goal oriented error estimation," *arXiv preprint arXiv:2007.14446*, 2020. [Online]. Available: <https://arxiv.org/abs/2007.14446>
- [39] L. Grüne, M. Schaller, and A. Schiela, "Abstract nonlinear sensitivity and turnpike analysis and an application to semilinear parabolic PDEs," *ESAIM: Control, Optimisation and Calculus of Variations*, vol. 27, p. 56, 2021. [Online]. Available: <https://doi.org/10.1051/cocv/2021030>
- [40] S. S. Keerthi and E. G. Gilbert, "Optimal infinite-horizon feedback laws for a general class of constrained discrete-time systems: Stability and moving-horizon approximations," *Journal of Optimization Theory and Applications*, vol. 57, no. 2, pp. 265–293, may 1988. [Online]. Available: <https://doi.org/10.1007/bf00938540>
- [41] J. F. Bonnans and A. Shapiro, *Perturbation Analysis of Optimization Problems*. Springer New York, 2000. [Online]. Available: <https://doi.org/10.1007/978-1-4612-1394-9>
- [42] J. Nocedal and S. J. Wright, *Numerical Optimization*, 2nd ed., ser. Springer Series in Operations Research and Financial Engineering. Springer New York, 2006. [Online]. Available: <https://doi.org/10.1007/978-0-387-40065-5>
- [43] R. Verschueren, M. Zanon, R. Quirynen, and M. Diehl, "A sparsity preserving convexification procedure for indefinite quadratic programs arising in direct optimal control," *SIAM Journal on Optimization*, vol. 27, no. 3, pp. 2085–2109, jan 2017. [Online]. Available: <https://doi.org/10.1137/16m1081543>
- [44] M. Diehl, R. Findeisen, H. Bock, F. Allgöwer, and J. Schlöder, "Nominal stability of real-time iteration scheme for nonlinear model predictive control," *IEEE Proceedings - Control Theory and Applications*, vol. 152, no. 3, pp. 296–308, may 2005. [Online]. Available: <https://doi.org/10.1049/tp-cta:20040008>
- [45] S. M. Robinson, "Perturbed kuhn-tucker points and rates of convergence for a class of nonlinear-programming algorithms," *Mathematical Programming*, vol. 7, no. 1, pp. 1–16, dec 1974. [Online]. Available: <https://doi.org/10.1007/bf01585500>
- [46] M. Hehn and R. D'Andrea, "A flying inverted pendulum," in *2011 IEEE International Conference on Robotics and Automation*, IEEE. IEEE, may 2011, pp. 763–770. [Online]. Available: <https://doi.org/10.1109/icra.2011.5980244>
- [47] H. Deng and T. Ohtsuka, "A parallel newton-type method for nonlinear model predictive control," *Automatica*, vol. 109, p. 108560, nov 2019. [Online]. Available: <https://doi.org/10.1016/j.automatica.2019.108560>
- [48] "Nonlinear heat transfer in thin plate," <https://www.mathworks.com/help/pde/ug/nonlinear-heat-transfer-in-a-thin-plate.html>. [Online]. Available: <https://www.mathworks.com/help/pde/ug/nonlinear-heat-transfer-in-a-thin-plate.html>
- [49] J. S. Rodriguez, B. Nicholson, C. Laird, and V. M. Zavala, "Benchmarking ADMM in nonconvex NLPs," *Computers & Chemical Engineering*, vol. 119, pp. 315–325, nov 2018. [Online]. Available: <https://doi.org/10.1016/j.compchemeng.2018.08.036>
- [50] A. Wächter and L. T. Biegler, "On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming," *Mathematical Programming*, vol. 106, no. 1, pp. 25–57, apr 2005. [Online]. Available: <https://doi.org/10.1007/s10107-004-0559-y>
- [51] A. HSL, "collection of Fortran codes for large-scale scientific computation," See <http://www.hsl.rl.ac.uk>, 2007. [Online]. Available: <http://www.hsl.rl.ac.uk/>



**Sen Na** is a fifth-year Ph.D. student in the Department of Statistics at the University of Chicago under the supervision of Mihai Anitescu and Mladen Kolar. Before coming to UChicago, he received B.S. degree in mathematics from Nanjing University, China. His research interests lie in nonlinear dynamic programming, high-dimensional statistics, semiparametric modeling, and their interface. He is also serving as a reviewer of the SIAM Journal on Optimization, and Journal of Machine Learning Research.



**Sungho Shin** is a Ph.D. candidate in the Department of Chemical and Biological Engineering at the University of Wisconsin-Madison. He received his B.S. in chemical engineering and mathematics from Seoul National University, South Korea, in 2016. His research interests include control theory and optimization algorithms for complex networks.



**Mihai Anitescu** is a senior computational mathematician in the Mathematics and Computer Science Division at Argonne National Laboratory and a professor in the Department of Statistics at the University of Chicago. He obtained his engineer diploma (electrical engineering) from the Polytechnic University of Bucharest in 1992 and his Ph.D. in applied mathematical and computational sciences from the University of Iowa in 1997. He specializes in the areas of numerical optimization, computational science, numerical analysis, and uncertainty quantification. He is on the editorial board of the SIAM Journal on Optimization, and he is a senior editor for Optimization Methods and Software. He is a past member of the editorial boards of Mathematical Programming A and B, SIAM Journal on Scientific Computing, and SIAM/ASA Journal in Uncertainty Quantification.



**Victor M. Zavala** is the Baldovin-DaPra Associate Professor in the Department of Chemical and Biological Engineering at the University of Wisconsin-Madison. He holds a B.Sc. degree from Universidad Iberoamericana and a Ph.D. degree from Carnegie Mellon University, both in chemical engineering. He is an associate editor for the Journal of Process Control and for IEEE Transactions on Control and Systems Technology. He is also a technical editor of Mathematical Programming Computation. His research interests are in the areas of energy systems, high-performance computing, stochastic programming, and predictive control.