SIAM J. OPTIM.
Vol. 28, No. 3, pp. 2541–2573

EXPONENTIALLY ACCURATE TEMPORAL DECOMPOSITION FOR LONG-HORIZON LINEAR-QUADRATIC DYNAMIC OPTIMIZATION*

WANTING XU† AND MIHAI ANITESCU‡

Abstract. In this work, we investigate a temporal decomposition approach to long-horizon dynamic optimization problems. The problems are discrete-time, linear, time dependent, and with box constraints on the control variables. We prove that an overlapping domains temporal decomposition, while inexact, approaches the solution of the long-horizon dynamic optimization problem exponentially fast in the size of the overlap. The resulting subproblems share no solution information and thus can be computed independently in parallel. Our findings are demonstrated with a small, synthetic production cost model with real demand data.

Key words. optimal control, temporal decomposition, controllability, sensitivity analysis

AMS subject classifications. 49N05, 49M27, 93B05, 49Q12

DOI. 10.1137/16M1081993

1. Introduction.

Long-horizon dynamic optimization problems appear in several application areas [1, 2, 7, 9, 10, 13, 16] and pose significant computational challenges because of the increase in the number of variables in proportion to the number of time periods considered. One very long horizon instance is optimal planning in the electrical power industry for transmission or generation expansion [1], which we now describe in some detail.

Such a planning analysis involves a production cost model (PCM). A PCM simulates the operation of generation and transmission systems by finding, during each time interval, the least-cost solution to generating sufficient energy to meet demand. As an abstraction, it is an optimal control problem, which can have nonlinear dynamics, control, and state constraints. Most studies require running a PCM on an hourly scale for 1–20 years under different scenarios in order to address the operation and reliability aspects of the proposed transmission or expansion plan [21]. Doing so can result in a very large number of periods. For example, if a PCM is run for 12 years with an hourly scale, the number of time periods would exceed 100,000. Added to this are the tens of thousands of degrees of freedom at one time point, which are characteristic for planning at the interconnect level, making the problem a daunting one to solve. As a result, many planning studies, which involve investments of billions of dollars, are done with multiple approximations to make them fit the computing resources [20].

Researchers have therefore sought to identify approaches for long-horizon dynamic optimization that result in efficient temporal parallelism to address this complexity by bringing to bear more computing power. Approaches have included temporal decomposition strategies using Lagrangian decomposition [2, 13, 16] and a two-level opti-

*Received by the editors June 28, 2016; accepted for publication (in revised form) March 21, 2018; published electronically September 20, 2018. Preprint ANL/MCS-P6022-0616.
http://www.siam.org/journals/siopt/28-3/M108199.html
Funding: This material is based upon work supported by the U.S. Department of Energy, Office of Science, under Contract No. DE-AC02-06CH11357.

†Department of Statistics, University of Chicago, Chicago, IL 60637 (wantingxu@uchicago.edu).
‡Mathematics and Computer Science Division, Argonne National Laboratory, Lemont, IL 60439 (anitescu@mcs.anl.gov).

mization formulation with the lower level derived from a decomposition approach [9]. These ideas create the opportunity for faster computation using parallelism. For instance, a heuristic decomposition algorithm is presented in [10] for scheduling a batch chemical plant. The problem is decomposed into more tractable subproblems that are solved to optimality. Empirical evidence suggests largely reduced computational efforts and reasonable accuracy. Strengths and weaknesses of a number of temporal decomposition methods are investigated in [2]. A multiperiod nonlinear programming model is developed in [16] for production planning and distribution. Temporal decomposition is used for the solution and is shown to generate faster computation and good accuracy of the optimal solutions.

A recent approach for PCMs is to partition the simulation horizon and turn the annual problem into multiple overlapping weekly/monthly problems [1] that compute the contribution of an *inner time interval only* to the overall objective and then add up all these contributions. While such an approach cannot be an exact decomposition, it can be computed in parallel without information exchange between the problems on each decomposition interval, and therefore the computation can be sped up. Moreover, researchers showed empirically in [1] that the error in the approach drops rapidly with the increase of the buffer region (the overlapping area) surrounding the inner time interval.

Our aim here is to provide theoretical support for approximate temporal decomposition of dynamic optimization problems with long horizons using overlapping intervals such as the work in [1]. A particular focus is on characterizing the error made by using such approximations.

For this initial foray, we will use a considerably simpler model than the PCMs in [1] or other planning models [16]. That is, our formulation is the following optimization problem:

$$\min_{u_{n_1:n_2-1},x_{n_1:n_2}} \quad \sum_{k=n_1}^{n_2-1} u_k^T R_k u_k + (x_k - d_k)^T Q_k (x_k - d_k) \tag{1.1a}$$

$$+(x_{n_2} - d_{n_2})^T Q_{n_2} (x_{n_2} - d_{n_2}) \tag{1.1b}$$

$$\text{subject to (s.t.)} \quad x_{k+1} = A_k x_k + B_k u_k, \quad x_{n_1} = x_{n_1}^0, \tag{1.1c}$$

$$l_k \le u_k \le b_k, \qquad n_1 \le k \le n_2 - 1, \tag{1.1d}$$

for some given initial value $x_{n_1}^0$. We call such a problem a linear-quadratic dynamic optimization problem. Such problems are known under various other names such as linear-quadratic (model predictive) control [12] or dynamic programming [4]. We choose the name *dynamic optimization* for problem (1.1) [8, 11] as we are interested in finding the solution of the optimization problem rather than computing the control rule or policy functions themselves. We will, however, use the terms control and dynamic programming as well when referring to the existing results and their interpretations. In (1.1) $[n_1, n_2]$ is the entire time horizon under consideration, and x_{n_1} is known. We respectively refer to x_k, u_k, and d_k as the supply or generation, control, and reference trajectory (also known as demand in PCM contexts). Problem (1.1) has a few simplifications and changes [1, 16]: our objective is quadratic and not linear, and we do not allow for integer variables. We note that these approximations are used in the target areas. Quadratic objectives are sometimes used instead of linear for the one-period cost function [18]. Economic dispatch, that is, a version of PCM where the scheduling decisions are all known in advance and thus no integer variables are present, is used in planning studies [3]. A more important approximation is that we

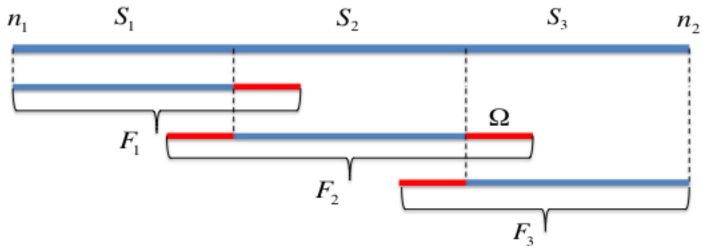

FIG. 1.1. *Illustration of the temporal decomposition scheme with three decomposition intervals. The entire horizon $[n_1, n_2]$ is decomposed into subintervals $S_{1:3}$, which are embedded in regions $F_{1:3}$ correspondingly. The red areas are buffer regions, each of length Ω.*

do not allow for hard path constraints. For example, supply and demand mismatch is penalized in the objective but not enforced to be zero. Approaches exist to accommodate supply equaling demand, at least in some circumstances, as will be done in our numerical example in section 4; we do not claim, however, that this can be done in general. Given the complexity of the analysis with even this simplified formulation, extensions that obtain results like ours under circumstances closer to [1, 16] will be investigated in future research.

Our approach, however, retains two important features from planning models that allow us to investigate approximate temporal decomposition: intertemporal constraints (1.1c) and box constraints on the control (1.1d). In particular, it allows us to substantiate a key insight that makes the temporal decomposition approach work efficiently. That is, when the system (1.1c) is controllable, the closed loop control law attached to the optimal active set results in an asymptotically stable policy [4]. In turn, the effect of perturbations of the parameters d_k and initial state x_{n_1} on the solution decreases exponentially with the distance in time between the perturbation moment and the index of the state. Hence, the system can forget its past and ignore its future exponentially fast with the distance from both.

This observation suggests the following temporal decomposition approach. Given a fixed time period $S_i \subset [n_1, n_2]$, we are interested in finding a shorter embedding interval F_i with $S_i \subset F_i \subset [n_1, n_2]$, so that the solution on S_i obtained by solving problem (1.1) on F_i is close to the one obtained by solving problem (1.1) on $[n_1, n_2]$. As a result, the entire horizon can be decomposed approximately, but with little error, into pieces like S_i, and the optimal solutions on each piece can be computed in parallel by solving problem (1.1) on F_i. Figure 1.1 illustrates this decomposition scheme (see color figure in online version). The temporal decomposition approach then consists of approximating the optimal value of problem (1.1) on $[n_1, n_2]$ by the sum of the optimal values on S_i obtained from solving (1.1) on F_i, over all i. A formal definition of this decomposition approach is presented in section 3.

Our goal is to estimate the error of this decomposition approach. In our proofs we will use several results from optimal control theory, which were done for the case in which $d_k = \mathbf{0}$ and in the absence of the bound constraints (1.1d), with respect to the notation in (1.1). In that case, the solution of (1.1) is provided by the linear-quadratic regulator (LQR), a feedback control law to achieve minimal cost. A derivation of the finite-horizon, discrete-time LQR based on the dynamic programming principle can be found in [4], which also shows that the resulting optimal trajectory tracks zero exponentially fast for time-independent linear systems. In this work, one particular control feature we will characterize and use is the rate of stabilization of the optimal trajectory for discrete-time, time-varying linear-quadratic dynamic optimization problems. To this end, Zhang, Hu, and Abate [23] derive some important properties for the finite-horizon and infinite-horizon value functions of the switched system discrete-time, linear-quadratic dynamic optimization. The authors show that under some mild assumptions the optimal trajectory stabilizes exponentially, and they give a workable estimate of that rate that we will use here. Some algorithms based on those theoretical results are also shown in [22] and [24]. A similar result to our Theorem 3.11, which upper bounds the approximation error of the optimal cost for the temporal decomposition approach, is given in [17]. The authors show that, for a class of constrained discrete-time systems, the infinite-horizon cost associated with the moving-horizon feedback law converges to the optimal infinite-horizon cost as the moving horizon is extended. Our work inherits similar temporal decomposition features as those in [17]. However, in this paper we additionally prove that, with a long but finite horizon, the *solutions* on the decomposition intervals converge as the embedding regions increase. Moreover, we characterize the convergence rates for the solutions and optimal cost as exponentially fast, which is crucial for the approach to be practical.

The rest of the article is organized as follows. Section 2 proves results about the box constrained control linear-quadratic problem. Section 3 describes the temporal decomposition approach and proves, based on the results derived in section 2, that the error of the temporal decomposition method decays exponentially in the size of the embedding interval. In section 4, we illustrate the theoretical findings by applying the temporal decomposition approach to a production cost model using real demand data. Proofs of results that are not central to the development of the main ideas are presented in Appendix A.

2. Box constrained control linear-quadratic problem. In this section, we derive results for a subproblem of the following box constrained control linear-quadratic problem:

$$\min_{u_{n_1:n_2-1}, x_{n_1:n_2}} \quad \Gamma_{n_1:n_2}(u_{n_1:n_2-1}, x_{n_1:n_2}) \triangleq \sum_{k=n_1}^{n_2-1} u_k^T R_k u_k + (x_k - d_k)^T Q_k (x_k - d_k) \tag{2.1a}$$

$$+ (x_{n_2} - d_{n_2})^T Q_{n_2} (x_{n_2} - d_{n_2}) \tag{2.1b}$$

$$\text{s.t.} \quad x_{k+1} = A_k x_k + B_k u_k, \qquad x_{n_1} = x_{n_1}^0, \tag{2.1c}$$

$$l_k \le u_k \le b_k, \qquad n_1 \le k \le n_2 - 1, \tag{2.1d}$$

where the initial state $x_{n_1}^0$ is given. Throughout the article, we have that $A_k \in \mathbb{R}^{n\times n}$, $B_k \in \mathbb{R}^{n\times m}$, and R_k, Q_k are positive definite matrices. We make the following uniform boundedness assumption about the system.

Assumption 2.1. For any n_1, n_2, $n_1 \le q \le n_2$, we have the following:

(a) $\|A_q\|_2 \le C_A$, $\|B_q\|_2 \le C_B$, $\|Q_q\|_2 \le C_Q$, $\|R_q\|_2 \le C_R$ for some C_A, C_B, C_Q, $C_R > 0$, and $C_A \ne 1$.

(b) $\lambda_{min}(Q_q) \ge \lambda_Q > 0$, $\lambda_{min}(R_q) \ge \lambda_R > 0$.

(c) $\|b_q\|$, $\|l_q\| \le U$ for some $U > 0$.

Note that in Assumption 2.1(a), since C_A is the uniform *upper* bound of $\|A_q\|_2$, we can assume $C_A \ne 1$ in general. The subproblem of (2.1) we consider is an equality constrained problem obtained by considering some active subset of the box control constraints (2.1d).

2.1. An equality constrained subproblem. To define the equality constrained subproblem, we let $I_k \subset \{1, \ldots, m\}$ be some index set for the elements of u_k attaining either the upper or lower bound of (2.1d), and define $N_k = I_k^{\mathsf{c}}$. Let e_i be the ith standard basis vector. We associate I_k with a selection matrix C_k and a vector $\bar{b}_k$ defined as follows:

(2.2)
$$C_k(i,:) = \begin{cases} e_{j_i}^T, & u_k(j_i) = l_k(j_i), \\ -e_{j_i}^T, & u_k(j_i) = b_k(j_i), \end{cases} \quad \bar{b}_k(i) = \begin{cases} l_k(j_i), & u_k(j_i) = l_k(j_i), \\ -b_k(j_i), & u_k(j_i) = b_k(j_i), \end{cases}$$

where $i = 1, \ldots, |I_k|$ and j_i is the ith element in I_k. With these definitions, the equality constraints corresponding to I_k can be expressed as $C_k u_k = \bar{b}_k$ and the equality constrained subproblem is defined as

$$\begin{aligned} \min_{u_{n_1:n_2-1}, x_{n_1:n_2}} \quad & \sum_{k=n_1}^{n_2-1} u_k^T R_k u_k + (x_k - d_k)^T Q_k (x_k - d_k) \\ & \quad + (x_{n_2} - d_{n_2})^T Q_{n_2} (x_{n_2} - d_{n_2}) \\ \text{s.t.} \quad & x_{k+1} = A_k x_k + B_k u_k, \qquad x_{n_1} = x_{n_1}^0, \\ & C_k u_k = \bar{b}_k, \qquad n_1 \le k \le n_2 - 1. \end{aligned} \tag{2.3}$$

Note that when I_k is the active set of problem (2.1) at optimality, problems (2.3) and (2.1) have the same solutions.

Problem (2.3) is the primary topic we consider in this section and will appear later in section 3 in a sensitivity analysis needed to prove temporal decomposition. In the rest of this subsection, we focus on deriving properties for the solution of problem (2.3) for some index set I_k. In particular, we will show the exponential decay property of the dependence of the solutions of problem (2.3) on the initial state and terminal reference under certain conditions. This is crucial in establishing the main temporal decomposition results in section 3. To start with, we note that a reduced problem can be obtained by eliminating the equality constraints of (2.3). We partition u_k, B_k, and R_k into blocks corresponding to I_k and N_k. Denote

$$\widetilde{u}_k = [u_k(i)]_{i \in I_k}, \quad \widetilde{B}_k = [B_k(:, i)]_{i \in I_k}$$

as the elements (or columns) of u_k (or B_k) corresponding to the equality index set I_k. Similarly, for N_k, correspondingly write

$$\hat{u}_k = [u_k(i)]_{i \in N_k}, \quad \hat{B}_k = [B_k(:, i)]_{i \in N_k}.$$

Also, R_k can be partitioned into blocks corresponding to the index sets as follows:

$$\hat{R}_k = [R_k(i,j)]_{i \in N_k, j \in N_k}, \quad \widetilde{R}_k = [R_k(i,j)]_{i \in I_k, j \in I_k}, \quad \bar{R}_k = [R_k(i,j)]_{i \in I_k, j \in N_k}.$$

Then we have that

$$\begin{aligned} B_k u_k &= \hat{B}_k \hat{u}_k + \widetilde{B}_k \widetilde{u}_k, \\ u_k^T R_k u_k &= \hat{u}_k^T \hat{R}_k \hat{u}_k + 2 \widetilde{u}_k^T \bar{R}_k \hat{u}_k + \widetilde{u}_k^T \widetilde{R}_k \widetilde{u}_k, \end{aligned}$$

and that the equality constraint $C_k u_k = \bar{b}_k$ is equivalent to $\widetilde{u}_k = \widetilde{b}_k$, where the ith element of $\widetilde{b}_k$ is $l_k(j_i)$ (or $b_k(j_i)$) if $u_k(j_i)$ attains the lower bound $l_k(j_i)$ (or the upper bound $b_k(j_i)$) for $i \in \{1, \ldots, |I_k|\}$, $j_i \in I_k$. Define a change of variable as follows:

$$\begin{aligned} v_k &= \hat{u}_k + \hat{R}_k^{-1} \bar{R}_k \widetilde{b}_k, \\ f_k &= \widetilde{B}_k \widetilde{b}_k - \hat{B}_k \hat{R}_k^{-1} \bar{R}_k \widetilde{b}_k. \end{aligned} \tag{2.4}$$

Note that $\hat{R}_k$ is invertible since R_k is positive definite. Then (u_k^*, x_k^*) is the solution of problem (2.3) if and only if (v_k^*, x_k^*), defined by (2.4), is the solution of the following problem:

$$\min_{v_{n_1:n_2-1}, x_{n_1:n_2}} \sum_{k=n_1}^{n_2-1} v_k^T \hat{R}_k v_k + (x_k - d_k)^T Q_k (x_k - d_k) \tag{2.5a}$$

$$+ (x_{n_2} - d_{n_2})^T Q_{n_2} (x_{n_2} - d_{n_2}) \tag{2.5b}$$

$$\text{s.t.} \quad x_{k+1} = A_k x_k + \hat{B}_k v_k + f_k, \quad n_1 \leq k \leq n_2 - 1, \quad x_{n_1} = x_{n_1}^0. \tag{2.5c}$$

One can easily verify that (2.4) defines a one-to-one correspondence between the feasible sets of problems (2.3) and (2.5) and that the objective functions differ by a constant for the corresponding elements in the feasible sets. Note that the optimal values of problems (2.3) and (2.5) differ by a constant. However, since we are only interested in the *solutions* of problem (2.3) with which the solutions of problem (2.5) have a one-to-one relationship (2.4), we thus solve problem (2.5) in order to investigate properties for the solutions of (2.3).

Problem (2.5) is a linear-quadratic optimal control problem for which we need a notion of controllability for the sequence pair $\{A_k, \hat{B}_k\}_{k=n_1:n_2}$. Note that $\hat{B}_k$ is uniquely determined by the index set I_k under consideration, and hence the choice of the index sets $\mathcal{I} \triangleq \{I_k\}$ will affect the controllability of the resulting $\{A_k, \hat{B}_k\}$. We make the following definition of controllability.

DEFINITION 2.2. *For some index sets* $\mathcal{I} = \{I_k\}_{k=n_1:n_2}$, *let* $\hat{B}_k = [B_k(:,i)]_{i \in I_k^c}$. *With some* $\lambda_C > 0$, $t > 0$ *both independent of* n_1 *and* n_2,

(a) *define the controllability matrix associated with time steps* $[q, q+t-1]$ *as*

$$C_{q,t}(\mathcal{I}) = \left[\hat{B}_{q+t-1} \quad A_{q+t-1}\hat{B}_{q+t-2} \quad \cdots \quad \left(\prod_{l=1}^{t-1} A_{q+l}\right)\hat{B}_q\right];$$

(b) *the index set* $\mathcal{I}$ *is uniformly completely controllable with parameters* λ_C *and* t, *denoted by* UCC(λ_C, t), *if the sequence pair* $\{A_k, \hat{B}_k\}$ *is uniformly completely controllable with parameters* λ_C *and* t [17, Definition 3.1], *i.e., for any* $n_1 \leq q \leq n_2$,

$$\lambda_{min}\left(C_{q,t}(\mathcal{I}) C_{q,t}^T(\mathcal{I})\right) \geq \lambda_C > 0.$$

Now we derive the optimal control law and optimal states for problem (2.5) using a dynamic programming approach. When $d_k \equiv 0$, $\forall k \in n_1 : n_2$, and $f_k \equiv 0$, $\forall k \in n_1 : (n_2 - 1)$, the solution to problem (2.5) is well known from classical dynamic programming references. For our temporal decomposition, however, the dependence on d_k is crucial, whereas $f_k \neq 0$ is needed as an artifact of the box constraints. To simplify our notation, we use a *reverse product* notation as follows.

DEFINITION 2.3. *We define*

$$\prod_{i=m}^{n} A_i = \begin{cases} A_n \cdots A_m, & n \geq m, \\ I, & n < m. \end{cases}$$

PROPOSITION 2.4. *For $n_1 \leq k \leq n_2 - 1$, the optimal control laws for problem* (2.5) *are*

$$\begin{aligned} v_k^*(x_k) = L_k x_k + W_k^{-1} \sum_{i=k+1}^{n_2} \hat{B}_k^T \left(M_i^{k+1}\right)^T d_i \\ + W_k^{-1} \sum_{i=k+1}^{n_2-1} \hat{B}_k^T \left(S_i^{k+1}\right)^T f_i - W_k^{-1} \hat{B}_k K_{k+1} f_k, \end{aligned} \tag{2.6}$$

where

$$K_{n_2} = Q_{n_2}, \tag{2.7a}$$

$$K_k = A_k^T (K_{k+1} - K_{k+1} \hat{B}_k W_k^{-1} \hat{B}_k^T K_{k+1}) A_k + Q_k, \qquad n_1 \leq k \leq n_2 - 1, \tag{2.7b}$$

$$W_k = \hat{R}_k + \hat{B}_k^T K_{k+1} \hat{B}_k, \qquad n_1 \leq k \leq n_2 - 1, \tag{2.7c}$$

$$L_k = -W_k^{-1} \hat{B}_k^T K_{k+1} A_k, \qquad n_1 \leq k \leq n_2 - 1, \tag{2.7d}$$

$$D_k = A_k + \hat{B}_k L_k, \qquad n_1 \leq k \leq n_2 - 1, \tag{2.7e}$$

$$M_i^k = Q_i \prod_{l=k}^{i-1} D_l, \qquad i \geq k, \quad n_1 \leq k \leq n_2, \tag{2.7f}$$

$$S_i^k = -K_{i+1} \prod_{l=k}^{i} D_l, \qquad i \geq k, \quad n_1 \leq k \leq n_2 - 1. \tag{2.7g}$$

Proof. See Appendix A.1. ◻

We note that (2.7b)–(2.7e) and the expression of v_k^* when $d_k \equiv 0$, $f_k \equiv 0$ are the results of classical LQ control.

DEFINITION 2.5. *For $n_1 \leq k \leq n_2 - 1$, define*

$$E_k = \hat{B}_k^T W_k^{-1} \hat{B}_k,$$

where W_k is defined in (2.7c).

PROPOSITION 2.6. *Let $x^*_{n_1+1:n_2}$ be the optimal states of* (2.5)*. Then we have that*

$$x_k^* = \left(\prod_{i=n_1}^{k-1} D_i \right) x_{n_1} + \sum_{i=n_1+1}^{n_2} C_i^k d_i + \sum_{i=n_1}^{n_2-1} F_i^k f_i, \tag{2.8}$$

where

$$\begin{aligned} C_i^k &= \sum_{s=n_1}^{\min(i,k)-1} \left(\prod_{l=s+1}^{k-1} D_l \right) E_s \left(M_i^{s+1}\right)^T, \\ F_i^k &= \sum_{s=n_1}^{\min(i,k)-1} \left(\prod_{l=s+1}^{k-1} D_l \right) E_s \left(S_i^{s+1}\right)^T + \left(\prod_{l=i+1}^{k-1} D_l \right) (I - E_i K_{i+1}) \mathbf{1}_{(k \geq i+1)}. \end{aligned} \tag{2.9}$$

Proof. See Appendix A.2. ◻

Next we investigate the properties of K_k defined by the Riccati recursion (2.7b) and the closed-loop matrices D_k defined in (2.7e). In the following, we only consider the index sets that are UCC(λ_C, t) according to Definition 2.2.

PROPOSITION 2.7. *Under Assumption* 2.1*, if the index set $\mathcal{I}$ is* UCC(λ_C, t)*, then for any $n_1 \le q \le n_2$ we have that $\|K_q\|_2 \le \beta$ for some $\beta > 0$ independent of n_1, n_2, and the particular choice of $\mathcal{I}$.*

Proof. We note from the definition (2.7b) of matrix K_q that, while it is a function of the quantities in Definition 2.2, it does not depend on the reference d_k or the shift f_k. We will thus consider it on the system for which d_k and f_k are 0. That is, for any $x_q \in \mathbb{R}^n$, consider the problem

$$\min_{u_{q:n_2-1}, x_{q:n_2}} \quad \sum_{k=q}^{n_2-1} u_k^T \hat{R}_k u_k + x_k^T Q_k x_k + x_{n_2}^T Q_{n_2} x_{n_2} \tag{2.10a}$$

$$\text{s.t.} \quad x_{k+1} = A_k x_k + \hat{B}_k u_k, \qquad q \le k \le n_2 - 1. \tag{2.10b}$$

For $k \ge q$, successively applying $x_{k+1} = A_k x_k + \hat{B}_k u_k$ gives that, for $j \ge 0$,

(2.11)

$$x_{q+j} - \left(\prod_{l=0}^{j-1} A_{q+l}\right) x_q = \left[\hat{B}_{q+j-1} \quad A_{q+j-1}\hat{B}_{q+j-2} \quad \cdots \quad \left(\prod_{l=1}^{j-1} A_{q+l}\right)\hat{B}_q\right] \begin{bmatrix} u_{q+j-1} \\ \vdots \\ u_q \end{bmatrix},$$

and for $j = t$ (2.11) reduces to

$$x_{q+t} - \left(\prod_{l=0}^{t-1} A_{q+l}\right) x_q = C_{q,t} \begin{bmatrix} u_{q+t-1} \\ \vdots \\ u_q \end{bmatrix}.$$

$\mathcal{I}$ being UCC(λ_C, t) implies that $C_{q,t}$ is uniformly completely controllable, and in particular that $C_{q,t}$ has full row rank. Then there exists $\hat{u} = (\hat{u}_q^T, \ldots, \hat{u}_{q+t-1}^T)^T$ so that

$$-\left(\prod_{l=0}^{t-1} A_{q+l}\right) x_q = C_{q,t} \begin{bmatrix} \hat{u}_{q+t-1} \\ \vdots \\ \hat{u}_q \end{bmatrix}. \tag{2.12}$$

Several $\hat{u}$ satisfy this relationship; we consider the one defined by

$$\hat{u} = -C_{q,t}^T (C_{q,t} C_{q,t}^T)^{-1} \left(\prod_{l=0}^{t-1} A_{q+l}\right) x_q. \tag{2.13}$$

Denote the corresponding states generated with $\hat{u}_{q:q+t-1}$ by $\hat{x}_{q:q+t}$; then $\hat{x}_{q+t} = \mathbf{0}$ by (2.12).

Assumption 2.1(a) implies that

$$\begin{aligned} &\max_{1\le j\le t} \left\| \left[\hat{B}_{q+j-1} \quad A_{q+j-1}\hat{B}_{q+j-2} \quad \ldots \left(\prod_{l=1}^{j-1} A_{q+l}\right)\hat{B}_q\right] \right\|_2 \\ &\le \max_{1\le j\le t} \left(C_B + C_A C_B + \ldots + C_A^{j-1} C_B\right) \\ &\le \frac{C_B\,(1 - C_A^t)}{1 - C_A} \triangleq M. \end{aligned}$$

Moreover, since $C_{q,t}^T(C_{q,t}C_{q,t}^T)^{-1}$ in (2.13) is the Moore–Penrose pseudoinverse $C_{q,t}^\dagger$ of $C_{q,t}$, we have that

$$\|\hat{u}\| \overset{\text{Ass. 2.1(a)}}{\leq} \|C_{q,t}^\dagger\|_2 C_A^t \|x_q\| = \frac{1}{\sigma_{min}(C_{q,t})} C_A^t \|x_q\| \overset{\text{Def. 2.2(b)}}{\leq} \frac{C_A^t\|x_q\|}{\sqrt{\lambda_C}}. \tag{2.14}$$

As a result, from (2.11), we have that, for $1 \leq j \leq t-1$,

$$\|\hat{x}_{q+j}\| \leq C_A^j\|x_q\| + M\|\hat{u}\| \leq \left(C_A^j + \frac{M}{\sqrt{\lambda_C}}C_A^t\right)\|x_q\|. \tag{2.15}$$

Now we let $\hat{u}_k = \mathbf{0}$ for $k \geq q+t$. Then it follows that $\hat{x}_k = \mathbf{0}$ for $k \geq q+t$. Also note that (2.10) is a standard linear-quadratic regulator problem, and the optimal value is given by $x_q^T K_q x_q$ [4]. As a result, we have that

$$\begin{aligned}
x_q^T K_q x_q &= \min_{u_k, x_k} \sum_{k=q}^{n_2-1} x_k^T Q_k x_k + u_k^T \hat{R}_k u_k + x_{n_2}^T Q_{n_2} x_{n_2} \\
&\leq \sum_{k=q}^{n_2-1} \hat{x}_k^T Q_k \hat{x}_k + \hat{u}_k^T \hat{R}_k \hat{u}_k + \hat{x}_{n_2}^T Q_{n_2} \hat{x}_{n_2} \\
&\leq \sum_{k=q}^{q+t-1} \hat{x}_k^T Q_k \hat{x}_k + \hat{u}_k^T \hat{R}_k \hat{u}_k \\
&\leq C_Q \sum_{k=q}^{q+t-1} \|\hat{x}_k\|^2 + C_R \sum_{k=q}^{q+t-1} \|\hat{u}_k\|^2 \\
&\overset{(2.14),(2.15)}{\leq} C_Q\left(1 + \sum_{i=1}^{t-1}\left(C_A^i + \frac{M}{\sqrt{\lambda_C}}C_A^t\right)^2\right)\|x_q\|^2 + C_R\frac{C_A^{2t}}{\lambda_C}\|x_q\|^2.
\end{aligned}$$

Note that the minimum above is also subject to the dynamics constraints (2.10b). Letting

$$\beta = C_Q\left(1 + \sum_{i=1}^{t-1}\left(C_A^i + \frac{M}{\sqrt{\lambda_C}}C_A^t\right)^2\right) + C_R\frac{C_A^{2t}}{\lambda_C}$$

completes the proof. Note that β only depends on the quantities in Definition 2.2 and Assumption 2.1, which are independent of n_1, n_2, and the particular choice of $\mathcal{I}$ given it is UCC(λ_C, t). ∎

In the following, we prove that the closed-loop system is asymptotically stable with an exponential decay rate. While the asymptotic result is well known, we need bounds on the decay rate at any time index; this is what we prove below. The proof is motivated by [23].

PROPOSITION 2.8. *Under Assumption* 2.1, *if the index set* $\mathcal{I}$ *is* UCC(λ_C, t), *then for any* $q \leq j \leq n_2 - 1$ *we have that*

$$\left\|\prod_{l=q}^{j} D_l\right\|_2 \leq C_1\rho^{j-q+1},$$

where $C_1 = \sqrt{\beta/\lambda_Q}$, $\rho = 1/\sqrt{1 + (\lambda_Q/\beta)}$, *and* C_1, ρ *are independent of* n_1, n_2, *and the particular choice of* $\mathcal{I}$.

Proof. It is shown in [4] that the recursion (2.7b) is equivalent to

$$K_k = {D_k}^T K_{k+1} D_k + Q_k + {L_k}^T \hat{R}_k L_k. \tag{2.16}$$

For $q \leq j \leq n_2 - 1$, define $x_{j+1} = D_j x_j$. Note that in this proof, x_j is a synthetic sequence, and not the solution of the problems (2.1) or (2.3). Therefore the properties of x_j defined here do not necessarily reflect those of the solution sequence. Then (2.16) and Proposition 2.7 imply that

$$\begin{aligned} x_j^T K_j x_j &\geq x_{j+1}^T K_{j+1} x_{j+1} + x_j^T Q_j x_j \\ &\geq x_{j+1}^T K_{j+1} x_{j+1} + \frac{\lambda_Q}{\beta} x_j^T K_j x_j \\ &\geq \left(1 + \frac{\lambda_Q}{\beta}\right) x_{j+1}^T K_{j+1} x_{j+1}. \end{aligned} \tag{2.17}$$

Here we used the bounds from Assumption 2.1 and the fact that

$$x_j^T K_j x_j \geq x_{j+1}^T K_{j+1} x_{j+1},$$

as implied by (2.16) and the positive definiteness of Q_k, $\hat{R}_k$. Also we have that

$$x_j^T K_j x_j \geq x_j^T Q_j x_j \geq \lambda_Q \|x_j\|^2. \tag{2.18}$$

As a result, for $n_2 - 1 \geq j \geq q$, we have the following:

$$\begin{aligned} \left\| \prod_{l=q}^{j} D_l x_q \right\|^2 = \|x_{j+1}\|^2 &\overset{(2.18)}{\leq} \frac{1}{\lambda_Q} x_{j+1}^T K_{j+1} x_{j+1} \\ &\overset{(2.17)}{\leq} \frac{1}{\lambda_Q (1 + \lambda_Q/\beta)} x_j^T K_j x_j \\ &\overset{(2.17)}{\leq} \frac{1}{\lambda_Q} \left(\frac{1}{1 + \lambda_Q/\beta} \right)^{j-q+1} x_q^T K_q x_q \\ &\overset{\text{Prop. 2.7}}{\leq} \frac{\beta}{\lambda_Q} \left(\frac{1}{1 + \lambda_Q/\beta} \right)^{j-q+1} \|x_q\|^2, \end{aligned}$$

where the third inequality is obtained by repeatedly applying (2.17). ∎

We have the following uniform boundedness result of matrices frequently used in the rest of this section.

LEMMA 2.9. *Under Assumption* 2.1, *if the index set* $\mathcal{I}$ *is* UCC(λ_C, t), *then for any* $n_1 \leq k \leq n_2 - 1$ *we have that*

$$\|E_k\|_2 \leq C_E, \qquad \|L_k\|_2 \leq C_L, \qquad \|f_k\|_2 \leq l_0$$

for some C_E, C_L, *and* l_0 *independent of* n_1, n_2, *and the particular choice of* $\mathcal{I}$. *Here* E_k *is defined in Definition* 2.5, L_k *in* (2.7d), *and* f_k *in* (2.4).

Proof. See Appendix A.3. ∎

Next, we investigate properties of the optimal states x_k^* and controls u_k^* for problem (2.3). Due to the one-to-one correspondence between solutions of problems (2.3) and (2.5), we first consider the optimal states of (2.5) obtained in Proposition 2.6. We have the following lemma characterizing the dependence of x_k^* on d_i and f_i.

LEMMA 2.10. *Let C_i^k and F_i^k be defined as in Proposition* 2.6. *Under Assumption* 2.1, *if the index set $\mathcal{I}$ is* UCC(λ_C, t), *we have that*

$$\|C_i^k\|_2 \leq C_2 \rho^{|i-k|}, \quad \|F_i^k\|_2 \leq C_F \rho^{|i-k|}$$

for some $C_2, C_F > 0$ independent of n_1, n_2, and the particular choice of $\mathcal{I}$. Here $\rho = 1/\sqrt{1+(\lambda_Q/\beta)}$ as in Proposition 2.8.

Proof. See Appendix A.4. ∎

Proposition 2.8 and Lemma 2.10 establish the exponential decay properties with respect to $|k-n_1|$ and $|i-k|$ for matrices $\prod_{i=n_1}^{k-1} D_i$ and C_i^k, which encode the dependencies of the optimal states x_k^* of problem (2.5) on the initial value and the reference d_i, respectively, by Proposition 2.6. This property is the key to proving the following main result of this section. Proposition 2.11 bounds the dependence of solutions x_k^*, u_k^* on the initial value x_{n_1} and terminal reference d_{n_2} with an exponential term. The importance of this result is shown in section 3 when we investigate the sensitivity of problem (2.1) to the initial value and terminal reference.

PROPOSITION 2.11. *Let x_k^* and u_k^* be the optimal states and controls of problem* (2.3). *Under Assumption* 2.1, *if the index set $\mathcal{I}$ is* UCC(λ_C, t), *then we have that*

$$\begin{aligned}
&\|\nabla_{x_{n_1}} x_k^*\|_2 \leq Z_1 \rho^{k-n_1}, \quad \|\nabla_{d_{n_2}} x_k^*\|_2 \leq Z_2 \rho^{n_2-k}, \quad n_1+1 \leq k \leq n_2,\\
&\|\nabla_{x_{n_1}} u_k^*\|_2 \leq Z_1 \rho^{k-n_1}, \quad \|\nabla_{d_{n_2}} u_k^*\|_2 \leq Z_2 \rho^{n_2-k}, \quad n_1 \leq k \leq n_2-1,
\end{aligned}$$

for some $Z_1, Z_2 > 0$ independent of n_1, n_2, and the particular choice of $\mathcal{I}$.

Proof. Due to the change of variable (2.4), the optimal states of problems (2.3) and (2.5) are the same, and the unconstrained parts of the optimal controls differ by a constant sequence. As a result, Proposition 2.6 and the optimal control law (2.6) give the following:

$$\begin{aligned}
\left\|\nabla_{x_{n_1}} x_k^*\right\|_2 &= \left\|\prod_{i=n_1}^{k-1} D_i\right\|_2 \overset{\text{Prop. 2.8}}{\leq} C_1 \rho^{k-n_1},\\
\left\|\nabla_{d_{n_2}} x_k^*\right\|_2 &= \left\|C_{n_2}^k\right\|_2 \overset{\text{Lem. 2.10}}{\leq} C_2 \rho^{n_2-k},\\
\left\|\nabla_{x_{n_1}} u_k^*\right\|_2 &= \left\|\nabla_{x_{n_1}} v_k^*\right\|_2 \overset{(2.6)}{=} \|L_k \nabla_{x_{n_1}} x_k^*\|_2 \overset{\text{Lem. 2.9}}{\leq} C_L \left\|\nabla_{x_{n_1}} x_k^*\right\|_2 \leq C_L C_1 \rho^{k-n_1},\\
\left\|\nabla_{d_{n_2}} u_k^*\right\|_2 &= \left\|\nabla_{d_{n_2}} v_k^*\right\|_2 \overset{(2.6)}{\leq} C_L \left\|\nabla_{d_{n_2}} x_k^*\right\|_2 + \left\|W_k^{-1} \hat{B}_k^T \left(\prod_{l=k+1}^{n_2-1} D_l\right)^T Q_{n_2}\right\|_2\\
&\overset{\text{Prop. 2.8}}{\leq} C_L \left\|\nabla_{d_{n_2}} x_k^*\right\|_2 + \frac{C_Q C_B}{\lambda_R} C_1 \rho^{n_2-k-1}\\
&\leq \quad C_L C_2 \rho^{n_2-k} + \frac{C_Q C_B}{\lambda_R} C_1 \rho^{n_2-k-1}.
\end{aligned}$$

Writing $Z_1 = \max(C_1, C_1 C_L)$ and $Z_2 = \max(C_2, C_L C_2 + C_Q C_B C_1/\lambda_R \rho)$ completes the proof. ∎

The next result gives an uniform upper bound for the solutions of problem (2.3) whose index set is UCC(λ_C, t). First, we make the following assumptions about the size of the initial value $x_{n_1}^0$ and the reference trajectory.

Assumption 2.12. For any n_1, n_2, and $n_1 \le q \le n_2$, we have that
(a) $\|x^0_{n_1}\|_2 \le u_0$ for some $u_0 > 0$,
(b) $\|d_q\|_2 \le m_0$ for some $m_0 > 0$.

Since the initial state is part of input to the system, we can reasonably assume that the values are taken in some compact set. Note that the reference trajectory models the demand in a PCM, which is our target application area. If we were to analyze asymptotics of our problem as $n_2 \to \infty$, uniformly bounded demand would be a tenuous assumption (though, with peak population scenarios currently considered, not impossible). The results here can be extended to polynomial increase of demand (as it will be compensated by exponential decays with rate ρ). To simplify the algebra, at this time we use Assumption 2.12(b), where the demand/reference trajectory is uniformly bounded over time.

LEMMA 2.13. *Let x^*_k and u^*_k be the optimal states and controls of problem* (2.3). *Under Assumptions* 2.1 *and* 2.12*, if the index set $\mathcal{I}$ is* UCC(λ_C, t)*, we have that*

$$\|x^*_k\|_2 \le C_g,\ n_1 + 1 \le k \le n_2; \quad \|u^*_k\|_2 \le C_u,\ n_1 \le k \le n_2 - 1$$

for some C_g, $C_u > 0$ independent of n_1, n_2, and the particular choice of $\mathcal{I}$.

Proof. Note again that, for problems (2.3) and (2.5), the optimal states are identical, and the unconstrained parts of the optimal controls satisfy the relation $v^*_k = \hat{u}^*_k + \hat{R}^{-1}_k \bar{R}_k \widetilde{b}_k$ by (2.4). Consequently, Proposition 2.6 and Lemmas 2.9 and 2.10 give the following:

$$\begin{aligned}\|x^*_k\|_2 &\le C_1 \rho^{k-n_1} u_0 + \sum_{i=n_1+1}^{n_2} C_2 \rho^{|k-i|} m_0 + \sum_{i=n_1}^{n_2-1} C_F \rho^{|k-i|} l_0 \\ &\le C_1 u_0 + 2 m_0 C_2 \sum_{s=0}^{\infty} \rho^s + 2 l_0 C_F \sum_{s=0}^{\infty} \rho^s \\ &= C_1 u_0 + \frac{2(m_0 C_2 + l_0 C_F)}{1-\rho} \triangleq C_g,\end{aligned}$$

where m_0 and u_0 are the bounds on the reference trajectory and initial state defined in Assumption 2.12. Note that (2.7c) gives that $\|W^{-1}_k\|_2 \le 1/\lambda_R$. The optimal control law (2.6) and Proposition 2.8 give the following:

$$\begin{aligned}\|u^*_k\|_2 &\le \|\widetilde{u}_k\|_2 + \|\hat{u}^*_k\|_2 = \|\widetilde{b}_k\|_2 + \|v^*_k - \hat{R}^{-1}_k \bar{R}_k \widetilde{b}_k\|_2 \le \|v^*_k\|_2 + \|\widetilde{b}_k\|_2 + \frac{C_R}{\lambda_R}\|b_k\|_2 \\ &\le C_L C_g + \frac{C_Q C_B}{\lambda_R} \sum_{i=k+1}^{n_2} \rho^{i-k-1} m_0 + \frac{\beta C_B}{\lambda_R} \sum_{i=k+1}^{n_2-1} \rho^{i-k} l_0 + \frac{\beta C_B}{\lambda_R} l_0 \\ &\quad + \left(2 + \frac{C_R}{\lambda_R}\right) U \\ &\le C_L C_g + \frac{m_0 C_Q C_B}{\lambda_R(1-\rho)} + \frac{l_0 \beta C_B}{\lambda_R(1-\rho)} + \left(2 + \frac{C_R}{\lambda_R}\right) U \triangleq C_u.\end{aligned}$$

This completes the proof. ∎

2.2. Box constrained control inequality problem. For the rest of this section, we return to the inequality constrained problem (2.1) and investigate properties of its solutions and Lagrange multipliers using the results derived for problem (2.3). We make the following assumption about the active set $\mathcal{A}$ of problem (2.1).

Assumption 2.14. The active set $\mathcal{A}$ of problem (2.1) is $\mathrm{UCC}(\lambda_C, t)$ as defined in Definition 2.2(b).

COROLLARY 2.15. *Let x_k^* and u_k^* be the optimal states and controls of problem* (2.1)*. Under Assumptions* 2.1*,* 2.12*, and* 2.14*, we have that*

$$\|x_k^*\|_2 \le C_g,\ n_1+1 \le k \le n_2; \quad \|u_k^*\|_2 \le C_u,\ n_1 \le k \le n_2 - 1$$

for C_g, $C_u > 0$ as in Lemma 2.13*.*

Proof. Note that when the index set defining the equality constrained problem (2.3) is the active set $\mathcal{A}$ of problem (2.1), problems (2.1) and (2.3) have the same solution. Since $\mathcal{A}$ is $\mathrm{UCC}(\lambda_C, t)$, Lemma 2.13 gives the conclusion. ◻

In the following, for problem (2.1), we investigate the adjoint variables which are the Lagrange multipliers associated with the constraints $x_{k+1} = A_k x_k + B_k u_k$.

PROPOSITION 2.16. *Let x_k^*, u_k^* be the solutions and ϕ_k^* be the optimal adjoint variables for problem* (2.1)*. For $n_1 \le k \le n_2 - 1$, we have that*

$$\phi_k^* = 2K_{k+1}x_{k+1}^* - 2\sum_{i=k+1}^{n_2} \left(M_i^{k+1}\right)^T d_i - 2\sum_{i=k+1}^{n_2-1} \left(S_i^{k+1}\right)^T f_i, \tag{2.19}$$

where K_k, M_i^k, S_i^k, and f_i are defined with respect to the active constraints $C_k u_k^* = \bar{b}_k$ of problem (2.1) *at optimality.*

Proof. See Appendix A.5. ◻

LEMMA 2.17. *Let ϕ_k^* be the optimal adjoint variables for problem* (2.1)*. Then under Assumptions* 2.1*,* 2.12*, and* 2.14*, for $n_1 \le k \le n_2 - 1$, we have that*

$$\|\phi_k^*\| \le C_\phi$$

for some $C_\phi > 0$ independent of n_1 and n_2.

Proof. See Appendix A.6. ◻

3. A temporal decomposition approach. In this section, we define a temporal decomposition approach to approximate the solutions and optimal values of problem (2.1). To partition the entire horizon, we decompose $[n_1, n_2]$ into n_0 subintervals of the same length $p = (n_2 - n_1)/n_0$. Denote the subintervals as

$$S_i = [n_1 + (i-1)p, n_1 + ip], \qquad i = 1, \ldots, n_0. \tag{3.1}$$

For some buffer size $0 < \Omega < p$, define an embedding region F_i for each S_i as $F_i = [n_1'(i), n_2'(i)]$, where

$$\begin{aligned} n_1'(i) &= \begin{cases} n_1, & i = 1, \\ n_1 + (i-1)p - \Omega, & i = 2, \ldots, n_0, \end{cases} \\ n_2'(i) &= \begin{cases} n_1 + ip + \Omega, & i = 1, \ldots, n_0 - 1, \\ n_2, & i = n_0. \end{cases} \end{aligned} \tag{3.2}$$

Note that $S_i \subset F_i$ for $i = 1, \ldots, n_0$. Figure 1.1 shows an illustration of such a decomposition scheme when $n_0 = 3$. We define the following parametrized problem.

Definition 3.1. *For $i = 1, \ldots, n_0$, let $\theta = (\theta^{(h)}, \theta^{(d)})$. We define the parametrized problem P_θ^i as follows:*

$$\min_{w_{n_1'(i):n_2'(i)-1}, h_{n_1'(i):n_2'(i)}} \sum_{k=n_1'(i)}^{n_2'(i)-1} w_k^T R_k w_k + (h_k - d_k)^T Q_k (h_k - d_k) \tag{3.3a}$$

$$+(h_{n_2'(i)} - d_{n_2'(i)})^T Q_{n_2'(i)} (h_{n_2'(i)} - d_{n_2'(i)}) \tag{3.3b}$$

$$\text{s.t.} \quad h_{k+1} = A_k h_k + B_k w_k, \qquad n_1'(i) \le k \le n_2'(i) - 1, \tag{3.3c}$$

$$l_k \le w_k \le b_k, \qquad n_1'(i) \le k \le n_2'(i) - 1, \tag{3.3d}$$

$$h_{n_1'(i)} = \theta^{(h)}, \quad d_{n_2'(i)} = \theta^{(d)}, \tag{3.3e}$$

where $d_{n_1'(i):n_2'(i)-1}$ are the same as those in problem (2.1).

Note that problem P_θ^i is problem (2.1) defined on a shorter interval F_i, but with a possibly different terminal reference vector $\theta^{(d)}$ and initial state $\theta^{(h)} = h^0_{n_1'(i)}$. For the latter, we invoke an assumption similar to Assumption 2.12.

Assumption 3.2. For $i = 1, \ldots, n_0$, let $h^0_{n_1'(i)}$ be the initial value of problem P_θ^i, then $\|h^0_{n_1'(i)}\|_2 \le u_0$, where u_0 is the same as that in Assumption 2.12.

Let $\theta_0(i) = (h^0_{n_1'(i)}, d_{n_2'(i)})$, where $h^0_{n_1'(i)}$ is any initial value satisfying Assumption 3.2 and $h^0_{n_1'(1)} = x^0_{n_1}$, and where $d_{n_2'(i)}$ is the reference in problem (2.1). Let $x^*_{n_1+1:n_2}$, $u^*_{n_1:n_2-1}$ be the optimal states and controls of problem (2.1), respectively, and let $h^*_{F_i}$ and $w^*_{F_i}$ be the optimal states and controls of problem $P^i_{\theta_0(i)}$. Define

$$J\left([m_1, m_2], [n_1, n_2], x^0_{n_1}\right) \triangleq \sum_{k=m_1}^{m_2-1} {u_k^*}^T R_k u_k^* + (x_k^* - d_k)^T Q_k (x_k^* - d_k) \\ + (x^*_{n_2} - d_{n_2})^T Q_{n_2} (x^*_{n_2} - d_{n_2}) \mathbf{1}_{(m_2 = n_2)}, \tag{3.4}$$

where $x^*_{m_1:m_2-1}$ and $u^*_{m_1:m_2-1}$ are respectively the optimal states and controls of problem (2.1) on $[n_1, n_2]$ with initial value $x^0_{n_1}$ restricted to $[m_1, m_2] \subset [n_1, n_2]$. Then the temporal decomposition approach consists of the approximation

$$J\left([n_1, n_2], [n_1, n_2], x^0_{n_1}\right) = \Gamma_{n_1:n_2}\left(u^*_{n_1:n_2-1}, x^*_{n_1:n_2}\right)$$

by $\sum_{i=1}^{n_0} J(S_i, F_i, h^0_{n_1'(i)})$.

In other words, the optimal value of problem (2.1) is approximated by solving problem $P^i_{\theta_0(i)}$ on each embedding region F_i and summing over the solutions restricted on the subintervals $S_i \subset F_i$. On the target intervals S_i, solving $P^i_{\theta_0(i)}$ results in the states $h^*_{S_i}$ and controls $w^*_{S_i}$. To bound the error of this approximation, we need to relate the solution of $P^i_{\theta_0(i)}$ to the solution of (2.1) when solved on the full horizon.

To this end we define a *modified problem* on the embedding horizon F_i, whose solution vector is the same as the restriction to F_i of the solution of (2.1) for the full horizon $[n_1, n_2]$. The modified problem is also an instance of (3.3), but its solution vector will be precisely the solution of (2.1) restricted to F_i. We have the following result.

PROPOSITION 3.3. *Let* $(u^*_{n_1:n_2-1}, x^*_{n_1+1:n_2})$ *be the solutions and* ϕ^*_k *be the adjoint variables of problem* (2.1). *For* $i = 1, \ldots, n_0$, *define*

$$\hat{h}_{n'_1(i)} = \begin{cases} x^0_{n_1}, & i = 1, \\ x^*_{n'_1(i)}, & i = 2, \ldots, n_0, \end{cases}$$
$$\hat{d}_{n'_2(i)} = \begin{cases} -Q^{-1}_{n'_2(i)} \phi^*_{n'_2(i)-1}/2 + x^*_{n'_2(i)}, & i = 1, \ldots, n_0 - 1, \\ d_{n_2}, & i = n_0. \end{cases} \tag{3.5}$$

Then $(u^*_{n'_1(i):n'_2(i)-1}, x^*_{n'_1(i)+1:n'_2(i)})$ *satisfies the KKT conditions and the second-order sufficient conditions of problem* $P^i_{\theta_1(i)}$ *with* $\theta_1(i) = (\hat{h}_{n'_1(i)}, \hat{d}_{n'_2(i)})$.

Proof. For $k = n_1, \ldots, n_2 - 1$, let $C_k u^*_k = \bar{b}_k$ be the active box constraints for problem (2.1) at optimality and let λ^*_k be the associated optimal Lagrange multipliers. The KKT conditions for problem (2.1) are

$$2R_k u^*_k - C^T_k \lambda^*_k + B^T_k \phi^*_k = 0, \qquad n_1 \le k \le n_2 - 1, \tag{3.6a}$$
$$2Q_k(x^*_k - d_k) + A^T_k \phi^*_k - \phi^*_{k-1} = 0, \qquad n_1 + 1 \le k \le n_2 - 1, \tag{3.6b}$$
$$2Q_{n_2}(x^*_{n_2} - d_{n_2}) - \phi^*_{n_2-1} = 0, \tag{3.6c}$$
$$x^*_{k+1} = A_k x^*_k + B_k u^*_k, \qquad n_1 \le k \le n_2 - 1, \qquad x_{n_1} = x^0_{n_1}, \tag{3.6d}$$
$$l_k \le u^*_k \le b_k, \qquad n_1 \le k \le n_2 - 1, \tag{3.6e}$$
$$\lambda^*_k \ge 0, \qquad n_1 \le k \le n_2 - 1. \tag{3.6f}$$

Then for problem $P^i_{\theta_1(i)}$ with parameters $\hat{h}_{n'_1(i)}$ and $\hat{d}_{n'_2(i)}$ defined in (3.5), the KKT conditions are satisfied by the same solutions $(u^*_{n'_1(i):n'_2(i)-1}, x^*_{n'_1(i)+1:n'_2(i)})$ with the same Lagrange multipliers λ^*_k, ϕ^*_k as follows:

$$2R_k u^*_k - C^T_k \lambda^*_k + B^T_k \phi^*_k = 0, \qquad n'_1(i) \le k \le n'_2(i) - 1, \tag{3.7a}$$
$$2Q_k(x^*_k - d_k) + A^T_k \phi^*_k - \phi^*_{k-1} = 0, \qquad n'_1(i) + 1 \le k \le n'_2(i) - 1, \tag{3.7b}$$
$$2Q_{n'_2(i)}(x^*_{n'_2(i)} - \hat{d}_{n'_2(i)}) - \phi^*_{n'_2(i)-1} = 0, \tag{3.7c}$$
$$x^*_{k+1} = A_k x^*_k + B_k u^*_k, \qquad n'_1(i) \le k \le n'_2(i) - 1, \qquad x_{n'_1(i)} = \hat{h}_{n'_1(i)}, \tag{3.7d}$$
$$l_k \le u^*_k \le b_k, \qquad n'_1(i) \le k \le n'_2(i) - 1, \tag{3.7e}$$
$$\lambda^*_k \ge 0, \qquad n'_1(i) \le k \le n'_2(i) - 1, \tag{3.7f}$$

where (3.7a)–(3.7b) and (3.7e)–(3.7f) directly follow from (3.6a)–(3.6b) and (3.6e)–(3.6f), respectively. Equations (3.7c) and (3.7d) follow from the definitions of $\hat{h}_{n'_1(i)}$ and $\hat{d}_{n'_2(i)}$, respectively. The second-order condition is satisfied by virtue of the strong convexity of the problem. ∎

Proposition 3.3 indicates that, in order for problem P^i_θ to have the same solutions as problem (2.1) on F_i, the modified parameters (3.5) need to incorporate information from (2.1) about the adjoint variables $\phi^*_{n'_2(i)-1}$, and about the states $x^*_{n'_1(i)}$, $x^*_{n'_2(i)}$. We note that problem $P^i_{\theta_1(i)}$ defined in Proposition 3.3 is only theoretically meaningful, since it cannot be set up without having solved the full horizon $[n_1, n_2]$ problem. However, the solution vector of $P^i_{\theta_1(i)}$ is *identical* to that of the full horizon problem restricted to F_i. In the following, we will prove that the solution of problem $P^i_{\theta_1(i)}$, on the subinterval S_i, is sufficiently close to that of $P^i_{\theta_0(i)}$. The latter problem is computable by using the reference trajectory corresponding only to the short interval F_i. Note that problem $P^i_{\theta_1(i)}$ can be viewed as the result of perturbing the parameter of problem $P^i_{\theta_0(i)}$. Therefore, to prove the relationship between the solutions of $P^i_{\theta_0(i)}$ and $P^i_{\theta_{1(i)}}$, we use the parametric sensitivity results derived from [6]. We note that our base problem (2.1) is a quadratic program, for which several results concerning the Lipschitz continuity with respect to parameters exist [6, 14]. Our aim, however, concerns more specific elements of the solution and seeks stronger results than directly using [6, 14] would allow. We aim to show that the entries corresponding to *a subset* of the solution vector components (the ones corresponding to the subintervals S_i in Figure 1.1) are Lipschitz continuous with respect to the initial state and terminal reference on the embedding regions F_i, *but with a Lipschitz constant L that decays exponentially in the buffer size Ω*. To achieve such an objective, we compute the directional derivative of the target components with respect to the perturbations, using results from [6], and then show that its value can be upper bounded using results such as Proposition 2.11. In turn, this gives the sought-after exponential decay result.

DEFINITION 3.4. *For $\theta \in \mathbb{R}^q$, define the one-sided directional derivative of $y(\theta)$ along a direction $p \in \mathbb{R}^q$ at θ_0 as*

$$D_p y(\theta_0) = \lim_{t \downarrow 0} \frac{y(\theta_0 + tp) - y(\theta_0)}{t},$$

given that the limit exists.

LEMMA 3.5. *Consider the parametrized quadratic programming problem*

$$\begin{aligned} &\min && f(y,\theta) \triangleq y^T G y/2 + y^T c(\theta) + \theta^T F \theta + y^T c_1 + \theta^T c_2 + C \\ &s.t. && Ay - r \le 0, \\ & && By - d(\theta) = 0, \end{aligned} \tag{3.8}$$

where G, F are positive definite, $\theta \in \mathbb{R}^q$ and $A^T = [a_1, \ldots, a_m] \in \mathbb{R}^{n\times m}$. Denote the solution of problem (3.8) by $y(\theta)$. When $\theta = \theta_0$, let $y_0 = y(\theta_0)$ and let the Lagrange multiplier corresponding to y_0 be $\bar{\lambda}$. Denote $I(y_0, \theta_0) = \{i : a_i^T y_0 = r_i, i = 1, \ldots, m\}$ be the set of active inequality constraints, $I_+(y_0, \theta_0, \bar{\lambda}) = \{i \in I(y_0, \theta_0) : \bar{\lambda}_i > 0\}$ and $I_0(y_0, \theta_0, \bar{\lambda}) = \{i \in I(y_0, \theta_0) : \bar{\lambda}_i = 0\}$. If the linear independence constraint qualification (LICQ) holds at $y(\theta_0)$, then for any $p \in \mathbb{R}^q$ we have that

$$D_p y(\theta_0) = \left(\left. \frac{dy^*_{I'(\theta_0)}(\theta)}{d\theta} \right|_{\theta=\theta_0} \right) p,$$

where $y^*_{I'(\theta_0)}(\theta)$ is the solution of the problem

$$\begin{aligned} &\min && f(y,\theta) = y^T G y/2 + y^T c(\theta) + \theta^T F \theta + y^T c_1 + \theta^T c_2 + C \\ &s.t. && A_{I'(\theta_0)} y - r' = 0, \\ & && By - d(\theta) = 0, \end{aligned} \tag{3.9}$$

and where $I'(\theta_0) = I_+(y_0, \theta_0, \bar{\lambda}) \cup I_1$ for some $I_1 \subset I_0(y_0, \theta_0, \bar{\lambda})$, and $A_{I'(\theta_0)} = [a_i^T]_{i\in I'(\theta_0)}$, $r' = [r_i]_{i \in I'(\theta_0)}$.

Proof. See Appendix A.7. ∎

With Lemma 3.5, we are now ready to investigate the effect on solutions of perturbing the parameters of problem P_θ^i. Since the proof for each subinterval is the same, for notational simplicity we suppress the dependence of $n'_1(i)$, $n'_2(i)$, and P_θ^i on i whenever the index of the subinterval under consideration is clear.

PROPOSITION 3.6. *Write $\theta_0 = (h^0_{n'_1}, d_{n'_2})$ and $\theta_1 = (\hat{h}_{n'_1}, \hat{d}_{n'_2})$, as defined in (3.5). For $\theta = (\theta^{(h)}, \theta^{(d)})$, let $y(\theta)$ be the solution of problem P_θ. We then have, for $s \in [0,1]$,*

$$D_{\theta_1-\theta_0} y\left(\theta_0 + s(\theta_1 - \theta_0)\right) = \left(\frac{dy_s^*(\theta)}{d\theta}\Big|_{\theta=\theta_0+s(\theta_1-\theta_0)}\right)(\theta_1 - \theta_0),$$

and $y_s^*(\theta)$ is the solution of the following equality constrained problem:

$$\begin{aligned}
&\min_{w_{n'_1:n'_2-1}, h_{n'_1:n'_2}} && \sum_{k=n'_1}^{n'_2-1} w_k^T R_k w_k + (h_k - d_k)^T Q_k (h_k - d_k) \\
&&& \quad + (h_{n'_2} - \theta_s^{(d)})^T Q_{n'_2} (h_{n'_2} - \theta_s^{(d)}) \\
&\text{s.t.} && h_{k+1} = A_k h_k + B_k w_k, \qquad n'_1 \le k \le n'_2 - 1, \qquad h_{n'_1} = \theta_s^{(h)}, \\
&&& C_k(s) w_k = \bar{b}_k(s), \qquad n'_1 \le k \le n'_2 - 1,
\end{aligned} \tag{3.10}$$

where rows of $C_k(s)$ and $\bar{b}_k(s)$ are, respectively, subsets of rows of $C'_k(s)$ and $\bar{b}'_k(s)$, which are the selection matrix and bound vector defined by (2.2) corresponding to the active set of P_{θ_s} at optimality, and $\theta_s = \theta_0 + s(\theta_1 - \theta_0)$. In other words, the equations $C_k(s) w_k = \bar{b}_k(s)$ represent a subset of the active constraints of P_{θ_s} at optimality.

Proof. For any $\theta \in \mathbb{R}^{2n}$, problem P_θ is an instance of problem (3.8) with the following parameters:

$$G = \begin{bmatrix} 2R_{n'_1} &&&&& \\ & \ddots &&&& \\ && 2R_{n'_2-1} &&& \\ &&& 2Q_{n'_1+1} && \\ &&&& \ddots & \\ &&&&& 2Q_{n'_2} \end{bmatrix}, \quad F = \begin{bmatrix} Q_{n'_1} & \\ & Q_{n'_2} \end{bmatrix},$$

$$A = \left[\begin{array}{c|c} I_{(n'_2-n'_1)m} & \\ & \mathbf{0}_{2(n'_2-n'_1)m\times(n'_2-n'_1)n} \\ -I_{(n'_2-n'_1)m} & \end{array}\right], \quad r = \begin{bmatrix} b_{n'_1} \\ \vdots \\ b_{n'_2-1} \\ -l_{n'_1} \\ \vdots \\ -l_{n'_2-1} \end{bmatrix},$$

$$c(\theta) = \begin{bmatrix} \mathbf{0}_{(n'_2-n'_1)m+(n'_2-n'_1-1)n} \\ -2Q_{n'_2}\theta^{(d)} \end{bmatrix},$$

$$B = \left[\begin{array}{ccc|ccccc} -B_{n'_1} &&& I &&&& \\ & \ddots && -A_{n'_1+1} & I &&& \\ &&&&& \ddots && \\ && -B_{n'_2-1} &&&& -A_{n'_2-1} & I \end{array}\right], \quad d(\theta) = \begin{bmatrix} A_{n'_1}\theta^{(h)} \\ \mathbf{0}_{(n'_2-n'_1-1)n} \end{bmatrix}.$$

Here $Ax \leq r$ and $Bx = d(\theta)$ correspond respectively to the box constraints (3.3d) and the system dynamics (3.3c). Note that A and B have the same number of columns. G and F are positive definite from Assumption 2.1. Let $\bar{A}$ be the matrix whose rows are subsets of rows of A corresponding to the active constraints of problem P_θ. Since an active constraint can achieve either lower or upper bound, but not both, the rows of $\bar{A}$ are linearly independent. Also, B has full row rank, and the rows of $\bar{A}$ are linearly independent of rows of B. As a result, LICQ holds for any $\theta \in \mathbb{R}^{2n}$ at optimality. For $s \in [0,1]$, directly applying Lemma 3.5 to problem P_{θ_s} gives the conclusion. □

Proposition 3.6 relates the directional derivative of the solution of problem P_{θ_s} with respect to the parameters θ to the solution of an equality constrained problem (3.10). Note that problem (3.10) has the same form as problem (2.3) for which we derive the exponential decay result Proposition 2.11 under some controllability conditions. Now we make similar controllability assumptions for the problems P_{θ_s}.

Assumption 3.7. For $i = 1, \ldots, n_0$ and any $s \in [0,1]$, let $\theta_0(i) = (h^0_{n_1'(i)}, d_{n_2'(i)})$, $\theta_1(i) = (\hat{h}_{n_1'(i)}, \hat{d}_{n_2'(i)})$ as defined in (3.5), and $\theta_s(i) = \theta_0(i) + s(\theta_1(i) - \theta_0(i))$; then the active sets of problems $P^i_{\theta_s(i)}$ at optimality are $\mathrm{UCC}(\lambda_C, t)$.

Note that Assumption 3.7 assumes $\mathrm{UCC}(\lambda_C, t)$ for the active sets of the continuously indexed family of problems $P^i_{\theta_s(i)}$ on each embedding region F_i, which is stronger than Assumption 2.14 for problem (2.1). We note, however, that Assumption 3.7 is only made for the active sets at optimality.

LEMMA 3.8. *Under Assumption* 3.7, *the index set for problem* (3.10) *is* $\mathrm{UCC}(\lambda_C, t)$ *for any* $s \in [0,1]$ *and* $i = 1, \ldots, n_0$.

Proof. Since the proof for each $i = 1, \ldots, n_0$ is the same, we suppress the dependence on i in the proof. The index set $\mathcal{I}_s$, which we use in the definition of problem (3.10) that we use to compute the directional derivative of the solution with respect to the parameter θ, is a subset of the active set $\mathcal{A}_s$ for problem P_{θ_s}. As a result, the columns of the controllability matrix $C_{q,t}(\mathcal{A}_s)$ are contained in those of $C_{q,t}(\mathcal{I}_s)$. Since $\lambda_{min}(C_{q,t}(\mathcal{A}_s)C^T_{q,t}(\mathcal{A}_s)) \geq \lambda_C$ by Assumption 3.7, we have that

$$\lambda_{min}(C_{q,t}(\mathcal{I}_s)C^T_{q,t}(\mathcal{I}_s)) \geq \lambda_{min}(C_{q,t}(\mathcal{A}_s)C^T_{q,t}(\mathcal{A}_s)) \geq \lambda_C,$$

and hence $\mathcal{I}_s$ is also $\mathrm{UCC}(\lambda_C, t)$. □

Together with Assumption 2.1, Lemma 3.8 justifies the application of the exponential decay result Proposition 2.11 to problem (3.10), and hence, combined with Proposition 3.6, it bounds the distance between solutions of P_{θ_0} and P_{θ_1} as follows.

PROPOSITION 3.9. *Let* $y(\theta_0) = (w^*_{n_1':n_2'-1}, h^*_{n_1'+1:n_2'})$ *be the solution of problem* P_{θ_0} *and let* $y(\theta_1) = (u^*_{n_1':n_2'-1}, x^*_{n_1'+1:n_2'})$ *be the solution of problem* P_{θ_1}. *From Proposition* 3.3, $y(\theta_1)$ *is also the solution of problem* (2.1) *restricted to the embedding region* F_i. *Under Assumptions* 2.1, 2.12, 2.14, 3.2 *and* 3.7, *for* $i = 1, \ldots, n_0$ *and* $k \in S_i$, *we have that*

$$\|x^*_k - h^*_k\|_2, \quad \|u^*_k - w^*_k\|_2 \leq Y\rho^\Omega$$

for some $Y > 0$ *independent of* n_1 *and* n_2, *where* ρ *is defined in Proposition* 2.8 *and* Ω *is the buffer size as in* (3.2).

Proof. From Leibniz–Newton, we have that

$$y(\theta_1) - y(\theta_0) = \int_0^1 D_{\theta_1 - \theta_0} y\left(\theta_0 + s(\theta_1 - \theta_0)\right) ds,$$

which gives that

$$x_k^* - h_k^* = \int_0^1 D_{\theta_1-\theta_0}\widetilde{p}_k^*(\theta_s)\,ds, \quad u_k^* - w_k^* = \int_0^1 D_{\theta_1-\theta_0}\widetilde{s}_k^*(\theta_s)\,ds, \tag{3.11}$$

where $(\widetilde{s}^*_{n'_1:n'_2-1}(\theta_s), \widetilde{p}^*_{n'_1+1:n'_2}(\theta_s))$ is the solution of problem P_{θ_s}. Proposition 3.6 implies that

$$D_{\theta_1-\theta_0}\widetilde{p}_k^*(\theta_s) = \begin{bmatrix}\nabla_{h_{n'_1}} p_k^*(\theta_s) & \nabla_{d_{n'_2}} p_k^*(\theta_s)\end{bmatrix}\begin{bmatrix}\hat{h}_{n'_1} - h^0_{n'_1} \\ \hat{d}_{n'_2} - d_{n'_2}\end{bmatrix},$$

$$D_{\theta_1-\theta_0}\widetilde{s}_k^*(\theta_s) = \begin{bmatrix}\nabla_{h_{n'_1}} s_k^*(\theta_s) & \nabla_{d_{n'_2}} s_k^*(\theta_s)\end{bmatrix}\begin{bmatrix}\hat{h}_{n'_1} - h^0_{n'_1} \\ \hat{d}_{n'_2} - d_{n'_2}\end{bmatrix},$$

where $(s^*_{n'_1:n'_2-1}(\theta_s), p^*_{n'_1+1:n'_2}(\theta_s))$ is the solution of the equality constrained problem (3.10). Note that each P_{θ_s} may have a different active set, which may also be different from that of problem (2.1). However, under Assumption 3.7, the active set of every P_{θ_s} is UCC(λ_C, t), and Lemma 3.8 implies that the index set for the corresponding problem (3.10) is UCC(λ_C, t) as well. In addition, the system parameters (e.g., R_k, Q_k, A_k, B_k) of problem (3.10) are bounded above by the corresponding quantities under Assumption 2.1. As a result, problem (3.10) satisfies all the conditions of Proposition 2.11, which can be applied to give that

$$\begin{aligned} &\|\nabla_{h_{n'_1}} p_k^*(\theta_s)\|_2, \quad \|\nabla_{h_{n'_1}} s_k^*(\theta_s)\|_2 \le Z_1\rho^{k-n'_1}, \\ &\|\nabla_{d_{n'_2}} p_k^*(\theta_s)\|_2, \quad \|\nabla_{d_{n'_2}} s_k^*(\theta_s)\|_2 \le Z_2\rho^{n'_2-k}. \end{aligned} \tag{3.12}$$

Note that as given in Proposition 2.11, Z_1, Z_2, and ρ are independent of the problem interval and the particular choice of the index set.

Assumptions 2.12 and 3.2 and Propositions 2.13 and 2.17 give that

$$\|\hat{h}_{n'_1} - h^0_{n'_1}\|_2 \le C_g + u_0, \quad \|\hat{d}_{n'_2} - d_{n'_2}\|_2 \le \frac{C_\phi}{2\lambda_Q} + C_g + m_0, \tag{3.13}$$

where u_0 and m_0 are defined in Assumptions 2.12 and 3.2. Combining (3.12) and (3.13), we have, for $k \in S_i$,

$$\begin{aligned} &\|D_{\theta_1-\theta_0}\widetilde{p}_k^*(\theta_s)\|_2\,,\ \|D_{\theta_1-\theta_0}\widetilde{s}_k^*(\theta_s)\|_2 \\ &\quad\le Z_1\left(C_g + u_0\right)\rho^{k-n'_1} + Z_2\left(\frac{C_\phi}{2\lambda_Q} + C_g + m_0\right)\rho^{n'_2-k} \\ &\quad\le \left(Z_1\left(C_g + u_0\right) + Z_2\left(\frac{C_\phi}{2\lambda_Q} + C_g + m_0\right)\right)\rho^{\Omega}. \end{aligned}$$

Letting $Y = Z_1(C_g + u_0) + Z_2(\frac{C_\phi}{2\lambda_Q} + C_g + m_0)$ and combining with (3.11) completes the proof. ∎

Proposition 3.9 is our key result. It proves the main hypothesis of this paper that solutions restricted to the subinterval S_i of (2.1) formulated over the long horizon $[n_1, n_2]$ are exponentially close to the solutions restricted to the interval S_i of the problem $P^i_{\theta_0(i)}$, which is set up and solved only on the embedding region F_i. The exponent is proportional to Ω, the buffer size. Now we derive the following error bound of the optimal values on each decomposition subinterval.

Proposition 3.10. *Under Assumptions* 2.1*,* 2.12*,* 2.14*,* 3.2*, and* 3.7*, we have*

$$\left|J(S_i,[n_1,n_2],x^0_{n_1}) - J(S_i,F_i,h^0_{n_1'(i)})\right| \le \frac{n_2-n_1}{n_0} X\rho^{\Omega}$$

for some $X>0$ independent of n_1 and n_2.

Proof. We let $(w^*_{n_1':n_2'-1}, h^*_{n_1'+1:n_2'})$ be the solution of problem $P^i_{\theta_0(i)}$ and we let $(u^*_{n_1':n_2'-1}, x^*_{n_1'+1:n_2'})$ be the solution of $P^i_{\theta_1(i)}$, which, by Proposition 3.3, is also the solution of problem (2.1) on F_i. Since the active set of problem $P^i_{\theta_0(i)}$ at optimality is UCC(λ_C,t) by Assumption 3.7, and the initial state is bounded by u_0 from Assumption 3.2, Lemma 2.13 gives that, for $j\in S_i$, $\|h^*_j\|_2 \le C_g$, $\|w^*_j\|_2 \le C_u$. Then we have

$$\begin{aligned}
&\left|(x^*_j-d_j)^TQ_j(x^*_j-d_j)-(h^*_j-d_j)^TQ_j(h^*_j-d_j)\right|\\
&\le \left|(x^*_j-d_j)^TQ_j(x^*_j-h^*_j)\right| + \left|(x^*_j-h^*_j)^TQ_j(h^*_j-d_j)\right|\\
&\le 2C_Q(C_g+m_0)\|x^*_j-h^*_j\|_2
\end{aligned} \tag{3.14}$$

and

$$\begin{aligned}
&\left|u^{*T}_jR_ju^*_j - w^{*T}_jR_jw^*_j\right|\\
&\le \left|(u^*_j-w^*_j)^TR_ju^*_j\right| + \left|w^{*T}_jR_j(u^*_j-w^*_j)\right|\\
&\le 2C_RC_u\|u^*_j-w^*_j\|_2.
\end{aligned} \tag{3.15}$$

Combining with Proposition 3.9, we have that

$$\begin{aligned}
&\left|J(S_i,[n_1,n_2],x^0_{n_1}) - J(S_i,F_i,h^0_{n_1'(i)})\right|\\
&\le \sum_{j\in S_i}\Bigg(\left|(x^*_j-d_j)^TQ_j(x^*_j-d_j)-(h^*_j-d_j)^TQ_j(h^*_j-d_j)\right|\\
&\qquad + \left|u^{*T}_jR_ju^*_j - w^{*T}_jR_jw^*_j\right|\Bigg)\\
&\le 2\frac{n_2-n_1}{n_0}\left(C_Q(C_g+m_0)+C_RC_u\right)Y\rho^{\Omega}.
\end{aligned}$$

Defining $X = 2(C_Q(C_g+m_0)+C_RC_u)Y$ completes the proof. ∎

Now we bound the total error of optimal values generated by the decomposition approach.

Theorem 3.11. *Under Assumptions* 2.1*,* 2.12*,* 2.14*,* 3.2*, and* 3.7*, we have that*

$$\left|J([n_1,n_2],[n_1,n_2],x^0_{n_1}) - \sum_{i=1}^{n_0} J(S_i,F_i,h^0_{n_1'(i)})\right| \le (n_2-n_1)X\rho^{\Omega},$$

where $X>0$ is the same as in Proposition 3.10.

Proof. Using Proposition 3.10, we have that

$$\begin{aligned}
&\left| J([n_1, n_2], [n_1, n_2], x^0_{n_1}) - \sum_{i=1}^{n_0} J(S_i, F_i, h^0_{n'_1(i)}) \right| \\
&\le \sum_{i=1}^{n_0} \left| J(S_i, [n_1, n_2], x^0_{n_1}) - J(S_i, F_i, h^0_{n'_1(i)}) \right| \\
&\le n_0 \frac{n_2 - n_1}{n_0} X \rho^{\Omega} \\
&= (n_2 - n_1) X \rho^{\Omega}.
\end{aligned}$$

∎

Theorem 3.11 upper bounds the total error induced by the decomposition approach by the product of an exponential term ρ^{Ω} and the length of the horizon $n_2 - n_1$. The rate of decay is eventually dominated by the exponential term. The exponential decay rate in the buffer size Ω enables the buffer regions to be chosen significantly shorter than the entire horizon while producing reasonable approximations under increasing horizon. Hence, when this approach is implemented in parallel, the computation time can be significantly reduced with little compromise on accuracy.

We also note that the techniques developed in section 2, particularly Proposition 2.11, and in the first part of section 3 can be used beyond proving our main result, Theorem 3.11. We believe they can be useful in other contexts, and in particular, in model predictive control. For example, it appears that one can show with similar techniques that the trajectory obtained from a receding horizon control approach converges exponentially to the solution of the full horizon problem (1.1). For instance, Proposition 3.9 can be applied to show that, if the short horizon problem has length Ω, then the first optimal control vector $u_{n'_1}$ and the second state vector $x_{n'_1+1}$ are exponentially close in Ω to the corresponding elements of the solution of the full horizon problem (1.1). Due to the space limit, we aim to develop this observation in future research.

4. Numerical results. In this section, we apply the temporal decomposition approach to a simplified production cost model in order to verify some of our theoretical findings. We employ the estimated hourly demand data in the northern Illinois region from year 2011 to 2015 provided by PJM Interconnection [15]. The model we are considering is the following:

$$\min_{u_{1:N}, x_{1:N+1}} \quad \sum_{k=1}^{N} c_1(x_k - d_k)^2 + c_2 x_k^2 + u_k^2 \tag{4.1a}$$

$$\text{s.t.} \quad x_{k+1} = x_k + u_k, \tag{4.1b}$$

$$-U \le u_k \le U. \tag{4.1c}$$

Here d_k is the electricity demand to be satisfied on hour k, described by the data from [15]. We assume this can be done by two fictitious generators: one with high quadratic cost, with parameter $c_1 = 10$, and one with low quadratic cost $c_2 = 5$. The cheaper generator has limited ability to change its output x_k, which is modeled by the box constraints (4.1c) (also called the ramp rate constraints [1]) combined with the dynamics (4.1b). The more expensive generator is fast and can thus serve all remaining load $d_k - x_k$. This situation models, for example, the situation in which one has a cheap but slow coal plant and a fast but expensive gas plant. Here the control is u_k, the amount of change at hour k of the generation level of the cheaper generator. We note that the formulation has the form from (1.1).

To define the temporal decomposition approach described in section 3, we partition the hourly scale five-year horizon into weeks, resulting in $n_0 = 261$ subintervals $S_1, \dots, S_{n_0}$. With buffer size Ω, we define the embedding regions $F_1, \dots, F_{n_0}$ as in (3.2). In order to apply our decomposition approach, we need to specify the initial state for the short horizon problems other than the earliest one, which uses the initial value of the full horizon problem. In general, finding a good initial state is difficult. In the case of the production cost models that motivated this research, however, the demand, while random, is fairly stable [1] for the same time of the day in the week. Moreover, optimal generation levels tend to be stable too, with a similar pattern [1]. Therefore, for production cost models, a good initial guess is readily available. For the general dynamic optimization problem, such a good guess may not be available. On the other hand, for a given policy of choosing it—for example, choosing the analytical center of the feasible set—Proposition 3.9 can be used on test problems to determine a good choice of the buffer size Ω that allows for the effect of the initial condition policy to be small enough for the tolerance sought. The fact that Proposition 3.9 establishes exponential decay of the error with respect to the buffer size Ω allows such trade-offs to be carried out. For our production cost model example, as the demand pattern is relatively predictable, as also indicated in [1], a good guess does exist at a given time of the day and week. As a consequence, we use as an initial state the average demand at each hour of a day for all of 2011. At the initial time point $n_1'(i)$ of F_i, we thus set the initial value $x^0_{n_1'(i)}$ to the average demand for that hour. Denote the optimal objective function value of problem (4.1) by J^*, namely,

$$J^* \triangleq J([1,N],[1,N],x_1^0),$$

as defined in (3.4), and define

$$J_i^* \triangleq J(S_i, F_i, x^0_{n_1'(i)})$$

for $i = 1, \dots, n_0$. We solve both long and short horizon versions of (4.1) using the Ipopt software [5]. The model was defined by using the Julia/JuMP interface [19].

We now analyze how well the sum of the computation on the short intervals, J_i^*, approximates the long horizon problem, J^*. Figure 4.1 shows the relative approximation error $|J^* - \sum_{i=1}^{n_0} J_i^*|/J^*$ as a function of the buffer size (measured in hours) for an increasing value of the ramping constraint bound U in (4.1c). For each U, the largest buffer size we experiment with is the smallest value that results in a relative approximation error less than 10^{-5}. The cost of such large-scale planning projects is usually on the order of billions of dollars. A relative error on the order of 10^{-5} corresponds to a discrepancy of less than a hundred thousand dollars, which is already well within the tolerance level of planning. We observe from Figure 4.1 that for all values of U the relative error decreases exponentially with Ω, which is the conclusion of our main result, Theorem 3.11. We conclude that Figure 4.1 validates the exponential decay of the approximation error of temporal decomposition with respect to the buffer size as proved in Theorem 3.11. We note that the target accuracy is achieved by buffer regions of less than 24 hours for all bounds U, although for different and larger PCM the results could be different. The order of magnitude of the buffer for which such accuracy is achieved is, however, of the same order—days—as in [1].

Figure 4.1 also shows that the decay rate of the approximation error increases with increasing bound U of controls. Note that the error bound in Theorem 3.11

TABLE 4.1
Longest period t (hour) for which the optimal controls of problem (4.1) are on the bound.

U	100	200	300	400	500	600	700	800	900	1000
t	91	62	48	34	14	12	10	8	7	5

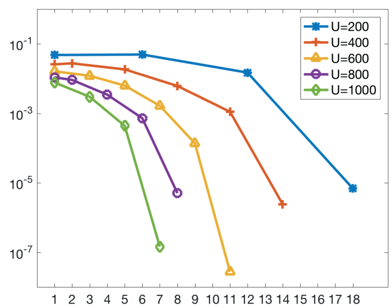

FIG. 4.1. *Relative error in approximation $|J^* - \sum_{i=1}^{n_0} J_i^*|/J^*$ at each buffer size (hour) for $U = 200, 400, 600, 800, 1000$.*

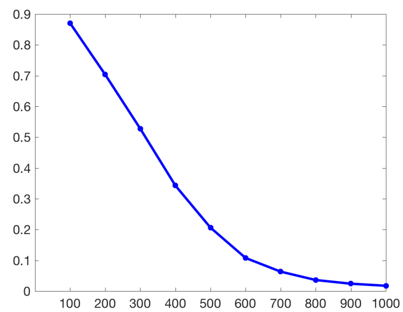

FIG. 4.2. *Proportions of optimal controls of problem (4.1) that are on the bound for $U = 100, \ldots, 1000$.*

depends on the controllability of problem (4.1) at optimality. We thus investigate numerically the longest period t for which the problem (4.1) is not controllable; t as used here carries the same meaning as in Definition 2.2. Since problem (4.1) is one dimensional, from Definition 2.2, it follows that t is simply the longest contiguous period for which the optimal controls are on the bound. Table 4.1 shows t in hours for different choices of the bound U. The longest period of uncontrollability decreases as the bound becomes larger. Figure 4.2 shows the proportions of optimal controls of problem (4.1) that are on the bound for increasing U. When $U = 100$, more than 85% of the optimal controls attain the bound, which approaches the controllability limit of the problem. Even for this tightest bound $U = 100$, however, the value of t is 91, which is about three to four days. This bound is certainly covered by our weekly partitioned subintervals and thus ensures that the controllability Assumption 2.14 holds on each embedding region F_i. Therefore the conditions of our main result Theorem 3.11 are satisfied.

5. Conclusions. Temporal decompositions are useful techniques for exposing parallelism in dynamic optimization problems. Such approaches are particularly useful for production cost simulations in electricity planning problems, where the calculations can have hundreds of thousands of time periods. The version of temporal decomposition discussed in this work approximates the solution over the entire horizon by the one obtained by patching the solution from multiple dynamic optimization problems with much shorter, overlapping, horizons initialized at some guess of the state. In turn, this transforms a sequential problem into one that is immediately amenable to parallel computing, thus massively reducing the time to solution. While used to great effect in [1], such temporal decomposition approaches were, up to our work, heuristic with no theoretical basis for their good approximating behavior.

In this work we prove that for the class of linear-quadratic dynamic optimization problems the temporal decomposition with overlap approaches the solution of the original problem exponentially fast in the size of the overlap. This approach partitions the entire horizon into subintervals and embeds each subinterval into the interval of interest plus a buffer region. The objective cost and, respectively, the solution on the entire horizon are then approximated by the the sum of the costs on each subinterval and, respectively, the solutions obtained from solving the corresponding problem on its embedding region. We prove that under some boundedness and controllability assumptions the approximation errors in both the solution and objective function decrease exponentially as a function of the buffer size. The exponential decay rate enables one to choose embedding regions much shorter than the length of the horizon, and since problems on each buffer region can be solved independently the time to solution is significantly reduced when the approach is implemented in parallel.

We validate our theoretical findings by using a numerical experiment that mimics a production cost evaluation over a five-year interval with hourly time periods and real data but with a simple, two-generator model. For all the cases, the relative error in the approximation of the objective function decreases exponentially with the buffer size. The decay rate decreases as more optimal controls attain the bound. In other words, the decay is slower when the system stays uncontrollable for longer periods. For this small experiment, even with the tightest bound on the controls and more than 85% of the optimal controls attaining the bounds, the buffer size needed for the relative error to be less than 10^{-5} is less than 24 periods—one day. Since the decomposed horizons have length only slightly more than one week, little extra effort has been added to solving problems when compared with the problem for the useful interval only, the one-week inner temporal region.

The class of dynamic optimization problems considered here is simplified when compared with [1] in that it is a linear-quadratic dynamic optimization problem with box control constraints. While our problem class does not include the complicating features of linear objective, integer variables, and path constraints, it includes the intertemporal constraints that make the analysis of error difficult. Consequently, our approach gives analytical support for the rapid convergence of the temporal decomposition with overlapping intervals. Future work will address extending the results for these complicating features as well as applying the techniques of this paper to model predictive control.

Appendix A. Proofs of results in sections 2 and 3.

A.1. Proof of Proposition 2.4. With dynamic programming, the "cost-to-go" value functions for a problem started at k with state x_k, $J_k(x_k)$, satisfy

$$\begin{aligned}J_{n_2}(x_{n_2}) &= (x_{n_2} - d_{n_2})^T Q_{n_2}(x_{n_2} - d_{n_2}),\\ J_k(x_k) &= \min_{v_k}\,(x_k - d_k)^T Q_k(x_k - d_k) + v_k^T \hat{R}_k v_k + J_{k+1}(A_k x_k + \hat{B}_k v_k + f_k)\end{aligned}$$

for $n_1 \le k \le n_2 - 1$. We claim that

$$J_k(x_k) = x_k^T K_k x_k - 2\sum_{i=k}^{n_2} d_i^T M_i^k x_k - 2\sum_{i=k}^{n_2-1} f_i^T S_i^k x_k + T_k, \qquad n_1 \le k \le n_2, \tag{A.1}$$

where T_k is some constant matrix.

Now we prove (A.1) by reverse induction and show that (2.6) holds whenever (A.1) holds at $k+1$. When $k = n_2$, (A.1) holds with $K_{n_2} = Q_{n_2}$ by definition. Suppose (A.1) holds for $J_{k+1}(x_{k+1})$ for some $n_1 \le k \le n_2 - 1$. Then, replacing the induction hypothesis formula in the cost-to-go function, we obtain

$$\begin{aligned} J_k(x_k) = \min_{v_k} \left\{ v_k^T W_k v_k + 2 r_k^T v_k \right\} + x_k^T \left(A_k^T K_{k+1} A_k + Q_k \right) x_k \\ - 2 \sum_{i=k+1}^{n_2} d_i^T M_i^{k+1} A_k x_k - 2 d_k^T Q_k x_k \\ - 2 \sum_{i=k+1}^{n_2-1} f_i^T S_i^{k+1} A_k x_k + 2 f_k^T K_{k+1} A_k x_k + T_k, \end{aligned}$$

where

$$\begin{aligned} W_k &= \hat{R}_k + \hat{B}_k^T K_{k+1} \hat{B}_k, \\ r_k^T &= x_k^T A_k^T K_{k+1} \hat{B}_k - \sum_{i=k+1}^{n_2} d_i^T M_i^{k+1} \hat{B}_k - \sum_{i=k+1}^{n_2-1} f_i^T S_i^{k+1} \hat{B}_k + f_k^T K_{k+1} \hat{B}_k. \end{aligned}$$

The optimal control law, which is the solution of the preceding optimization problem, is hence

$$\begin{aligned} v_k^*(x_k) &= -W_k^{-1} r_k \\ &= L_k x_k + W_k^{-1} \sum_{i=k+1}^{n_2} \hat{B}_k^T \left(M_i^{k+1} \right)^T d_i + W_k^{-1} \sum_{i=k+1}^{n_2-1} \hat{B}_k^T \left(S_i^{k+1} \right)^T f_i \\ &\quad - W_k^{-1} \hat{B}_k K_{k+1} f_k. \end{aligned}$$

Therefore (2.6) holds true at k.

Substituting $v_k^* = -W_k^{-1} r_k$, we obtain that $v_k^{*T} W_k v_k^* + 2 r_k^T v_k^* = -r_k^T W_k^{-1} r_k$. As a result,

$$\begin{aligned} J_k(x_k) = -r_k^T W_k^{-1} r_k + x_k^T \left(A_k^T K_{k+1} A_k + Q_k \right) x_k \\ - 2 \sum_{i=k+1}^{n_2} d_i^T M_i^{k+1} A_k x_k - 2 d_k^T Q_k x_k \\ - 2 \sum_{i=k+1}^{n_2-1} f_i^T S_i^{k+1} A_k x_k + 2 f_k^T K_{k+1} A_k x_k + T_k. \end{aligned} \tag{A.2}$$

Substituting the expression of r_k defined in (A.1), we have that, up to a constant term,

(A.3)

$$\begin{aligned} &- r_k^T W_k^{-1} r_k + x_k^T \left(A_k^T K_{k+1} A_k + Q_k \right) x_k \\ &= x_k^T \left(A_k^T (K_{k+1} - K_{k+1} \hat{B}_k W_k^{-1} \hat{B}_k^T K_{k+1}) A_k + Q_k \right) x_k \\ &\quad + 2 \left(\sum_{i=k+1}^{n_2} d_i^T M_i^{k+1} \hat{B}_k + \sum_{i=k+1}^{n_2-1} f_i^T S_i^{k+1} \hat{B}_k - f_k^T K_{k+1} \hat{B}_k \right) W_k^{-1} \hat{B}_k K_{k+1} A_k x_k \\ &= x_k^T K_k x_k - 2 \sum_{i=k+1}^{n_2} d_i^T M_i^{k+1} \hat{B}_k L_k x_k - 2 \sum_{i=k+1}^{n_2-1} f_i^T S_i^{k+1} \hat{B}_k L_k x_k + 2 f_k^T K_{k+1} \hat{B}_k L_k x_k, \end{aligned}$$

by (2.7b) and (2.7d). Now we substitute (A.3) back into (A.2) and have that

$$\begin{aligned}
J_k(x_k) &= x_k^T K_k x_k - 2\sum_{i=k+1}^{n_2} d_i^T M_i^{k+1}\hat{B}_k L_k x_k \\
&\quad - 2\sum_{i=k+1}^{n_2-1} f_i^T S_i^{k+1}\hat{B}_k L_k x_k + 2 f_k^T K_{k+1}\hat{B}_k L_k x_k \\
&\quad - 2\sum_{i=k+1}^{n_2} d_i^T M_i^{k+1} A_k x_k - 2 d_k^T Q_k x_k \\
&\quad - 2\sum_{i=k+1}^{n_2-1} f_i^T S_i^{k+1} A_k x_k + 2 f_k^T K_{k+1} A_k x_k + T_k \\
&\overset{(2.7e)}{=} x_k^T K_k x_k - 2\sum_{i=k+1}^{n_2} d_i^T M_i^{k+1} D_k x_k - 2 d_k^T Q_k x_k \\
&\quad - 2\sum_{i=k+1}^{n_2-1} f_i^T S_i^{k+1} D_k x_k + 2 f_k^T K_{k+1} D_k x_k + T_k \\
&= x_k^T K_k x_k - 2\sum_{i=k}^{n_2} d_i^T M_i^k x_k - 2\sum_{i=k}^{n_2-1} f_i^T S_i^k x_k + T_k,
\end{aligned}$$

by (2.7f) and (2.7g). This proves the induction hypothesis at (A.1) at step k. Since we proved above that (2.6) holds true at k, this completes the induction step and proves that both (A.1) and (2.6) hold for all k.

A.2. Proof of Proposition 2.6. The recursion (2.5c) and optimal control law (2.6) imply that x_k^* has the form of (2.8) for some C_i^k and F_i^k. When $k = n_1 + 1$, we have that

$$\begin{aligned}
x_{n_1+1}^* &= A_{n_1} x_{n_1} + \hat{B}_{n_1} v_{n_1}^* + f_{n_1} \\
&= D_{n_1} x_{n_1} + E_{n_1} \sum_{i=n_1+1}^{n_2} \left(M_i^{n_1+1}\right)^T d_i \\
&\quad + E_{n_1} \sum_{i=n_1}^{n_2-1} \left(S_i^{n_1+1}\right)^T f_i - E_{n_1} K_{n_1+1} f_{n_1} + f_{n_1}.
\end{aligned} \tag{A.4}$$

Applying recursion (2.5c) and the optimal control law (2.6) gives

$$\begin{aligned}
x_{k+1}^* = &\left(\prod_{i=n_1}^{k} D_i\right) x_{n_1} + \sum_{i=n_1+1}^{n_2} D_k C_i^k d_i + \sum_{i=n_1}^{n_2-1} D_k F_i^k f_i \\
&+ E_k \sum_{i=k+1}^{n_2} \left(M_i^{k+1}\right)^T d_i + E_k \sum_{i=k+1}^{n_2-1} \left(S_i^{k+1}\right)^T f_i - E_k K_{k+1} f_k + f_k.
\end{aligned}$$

Combining with (A.4), we obtain the recursions

$$\begin{aligned}
C_i^{k+1} &= \begin{cases} D_k C_i^k, & n_1 + 1 \le i \le k, \\ D_k C_i^k + E_k \left(M_i^{k+1}\right)^T, & k+1 \le i \le n_2, \end{cases} \\
C_i^{n_1+1} &= E_{n_1}\left(M_i^{n_1+1}\right)^T, \qquad n_1 + 1 \le i \le n_2,
\end{aligned} \tag{A.5}$$

and

$$F_i^{k+1} = \begin{cases} D_k F_i^k, & n_1 \le i \le k-1, \\ D_k F_i^k - E_k K_{k+1} + I, & i = k, \\ D_k F_i^k + E_k \left(S_i^{k+1}\right)^T, & k+1 \le i \le n_2 - 1, \end{cases} \tag{A.6}$$

$$F_i^{n_1+1} = \begin{cases} E_{n_1} \left(S_i^{n_1+1}\right)^T, & n_1 + 1 \le i \le n_2 - 1, \\ -E_{n_1} K_{n_1+1} + I, & i = n_1. \end{cases}$$

Now we prove that C_i^k and F_i^k defined in (2.9) satisfy the recursions (A.5) and (A.6), respectively.

(a) *Proof that* C_i^k *defined in* (2.9) *satisfies* (A.5).
When $k = n_1 + 1$, since $i \ge n_1 + 1$, we have that

$$C_i^{n_1+1} = \sum_{s=n_1}^{n_1} \left(\prod_{l=s+1}^{n_1} D_l \right) E_s \left(M_i^{s+1}\right)^T = E_{n_1} \left(M_i^{n_1+1}\right)^T,$$

which satisfies (A.5).

When $k > n_1 + 1$, if $i \le k$, then $i \le k+1$, and then we have that

$$\begin{aligned} C_i^{k+1} &= \sum_{s=n_1}^{i-1} \left(\prod_{l=s+1}^{k} D_l \right) E_s \left(M_i^{s+1}\right)^T \\ &= D_k \sum_{s=n_1}^{i-1} \left(\prod_{l=s+1}^{k-1} D_l \right) E_s \left(M_i^{s+1}\right)^T \\ &= D_k C_i^k, \end{aligned}$$

and if $i \ge k+1$, then $i \ge k$, so it follows that

$$\begin{aligned} C_i^{k+1} &= \sum_{s=n_1}^{k} \left(\prod_{l=s+1}^{k} D_l \right) E_s \left(M_i^{s+1}\right)^T \\ &= D_k \sum_{s=n_1}^{k-1} \left(\prod_{l=s+1}^{k-1} D_l \right) E_s \left(M_i^{s+1}\right)^T + E_k \left(M_i^{k+1}\right)^T \\ &= D_k C_i^k + E_k \left(M_i^{k+1}\right)^T, \end{aligned}$$

which both satisfy (A.5).

(b) *Proof that* F_i^k *defined in* (2.9) *satisfies* (A.6).
When $k = n_1 + 1$, if $i = n_1$, we have from (2.9) that

$$F_{n_1}^{n_1+1} = I - E_{n_1} K_{n_1+1},$$

and if $i \ge n_1 + 1 = k$, it follows that

$$F_i^{n_1+1} = \sum_{s=n_1}^{n_1} \left(\prod_{l=s+1}^{n_1} D_l \right) E_s \left(S_i^{s+1}\right)^T = E_{n_1} \left(S_i^{n_1+1}\right)^T,$$

which satisfies (A.6).

When $k > n_1 + 1$, if $i \le k-1$, then $i \le k \le k+1$, and then we have that

$$\begin{aligned}
F_i^{k+1} &= \sum_{s=n_1}^{i-1} \left(\prod_{l=s+1}^{k} D_l \right) E_s \left(S_i^{s+1}\right)^T + \left(\prod_{l=i+1}^{k} D_l \right) (I - E_i K_{i+1}) \\
&= D_k \left(\sum_{s=n_1}^{i-1} \left(\prod_{l=s+1}^{k-1} D_l \right) E_s \left(S_i^{s+1}\right)^T + \left(\prod_{l=i+1}^{k-1} D_l \right) (I - E_i K_{i+1}) \right) \\
&= D_k F_i^k.
\end{aligned}$$

If $i = k$, then $k+1 \ge i+1$, and hence

$$\begin{aligned}
F_i^{k+1} &= \sum_{s=n_1}^{i-1} \left(\prod_{l=s+1}^{k} D_l \right) E_s \left(S_i^{s+1}\right)^T + (I - E_i K_{i+1}) \\
&= D_k \sum_{s=n_1}^{i-1} \left(\prod_{l=s+1}^{k-1} D_l \right) E_s \left(S_i^{s+1}\right)^T + (I - E_k K_{k+1}) \\
&= D_k F_i^k + (I - E_k K_{k+1}),
\end{aligned}$$

and if $i \ge k+1$, then $i \ge k$, and we have that

$$\begin{aligned}
F_i^{k+1} &= \sum_{s=n_1}^{k} \left(\prod_{l=s+1}^{k} D_l \right) E_s \left(S_i^{s+1}\right)^T \\
&= D_k \sum_{s=n_1}^{k-1} \left(\prod_{l=s+1}^{k-1} D_l \right) E_s \left(S_i^{s+1}\right)^T + E_k \left(S_i^{k+1}\right)^T \\
&= D_k F_i^k + E_k \left(S_i^{k+1}\right)^T,
\end{aligned}$$

which all satisfy (A.6).

A.3. Proof of Lemma 2.9. For $n_1 \le k \le n_2 - 1$, the definition of E_k (2.5), Assumption 2.1, and the definition of W_k (2.7c) imply that

$$\|W_k^{-1}\| \le \frac{1}{\lambda_{min}(R_k)} \le \frac{1}{\lambda_R}$$

and thus

$$\|E_k\|_2 \le \frac{C_B^2}{\lambda_R} \triangleq C_E.$$

Proposition 2.7 and (2.7d) imply that $\|L_k\|_2 \le \beta C_A C_B / \lambda_R \triangleq C_L$. Lastly, Assumption 2.1 and (2.4) give that

$$\|f_k\|_2 \le \left(C_B + \frac{C_B C_R}{\lambda_R} \right) \|\widetilde{b}_i\|_2 \le 2 \left(C_B + \frac{C_B C_R}{\lambda_R} \right) U \triangleq l_0.$$

A.4. Proof of Lemma 2.10. Proposition 2.6 gives

$$
\begin{aligned}
C_i^k &= \sum_{s=n_1}^{\min(i,k)-1} \left(\prod_{l=s+1}^{k-1} D_l\right) E_s \left(\prod_{l=s+1}^{i-1} D_l\right)^T Q_i, \\
F_i^k &= -\sum_{s=n_1}^{\min(i,k)-1} \left(\prod_{l=s+1}^{k-1} D_l\right) E_s \left(\prod_{l=s+1}^{i} D_l\right)^T K_{i+1} \\
&\quad + \left(\prod_{l=i+1}^{k-1} D_l\right) (I - E_i K_{i+1})\, \mathbf{1}_{(k\geq i+1)},
\end{aligned}
$$

where $E_s = \hat{B}_s^T W_s^{-1} \hat{B}_s$ (Definition 2.5). Lemma 2.9 gives that $\|E_s\|_2 \leq C_E$. Using Proposition 2.8, the triangle inequality and properties of norms, we have that

$$
\begin{aligned}
\|C_i^k\|_2 &\leq \sum_{s=n_1}^{\min(i,k)-1} C_E C_Q C_1^2 \rho^{k-s-1} \rho^{i-s-1} \\
&\leq C_E C_Q C_1^2 \begin{cases} \rho^{k-i} \sum_{s=n_1}^{i-1} \rho^{2i-2s-2}, & i \leq k, \\ \rho^{i-k} \sum_{s=n_1}^{k-1} \rho^{2k-2s-2}, & k < i, \end{cases} \\
&= C_E C_Q C_1^2 \begin{cases} \rho^{k-i} \sum_{t=0}^{i-n_1-1} \rho^{2t}, & i \leq k, \\ \rho^{i-k} \sum_{t=0}^{k-n_1-1} \rho^{2t}, & k < i, \end{cases} \\
&\leq \frac{C_E C_Q C_1^2}{1-\rho^2} \rho^{|k-i|}.
\end{aligned}
$$

Similarly for F_i^k, we have that

$$
\begin{aligned}
\|F_i^k\|_2 &\leq \sum_{s=n_1}^{\min(i,k)-1} C_E C_1 \beta C_1^2 \rho^{k-s-1} \rho^{i-s-1} + (1 + C_E \beta) C_1 \rho^{k-i-1} \mathbf{1}_{(k\geq i+1)} \\
&\leq \frac{C_E C_1 \beta C_1^2}{1-\rho^2} \rho^{|k-i|} + \frac{C_1(1 + C_E\beta)}{\rho} \rho^{|k-i|}.
\end{aligned}
$$

Letting

$$
C_2 = \frac{C_E C_Q C_1^2}{1-\rho^2}, \quad C_F = \frac{C_E C_1 \beta C_1^2}{1-\rho^2} + \frac{C_1(1 + C_E\beta)}{\rho}
$$

completes the proof.

A.5. Proof of Proposition 2.16. The Karush–Kuhn–Tucker (KKT) conditions for problem (2.1) are

(A.7a) $\quad 2R_k u_k^* - C_k^T \lambda_k^* + B_k^T \phi_k^* = 0, \qquad n_1 \leq k \leq n_2 - 1,$

(A.7b) $\quad 2Q_k(x_k^* - d_k) + A_k^T \phi_k^* - \phi_{k-1}^* = 0, \qquad n_1 + 1 \leq k \leq n_2 - 1,$

(A.7c) $\quad 2Q_{n_2}(x_{n_2}^* - d_{n_2}) - \phi_{n_2-1}^* = 0,$

(A.7d) $\quad x_{k+1}^* = A_k x_k^* + B_k u_k^*, \qquad n_1 \leq k \leq n_2 - 1,$

(A.7e) $\quad l_k \leq u_k^* \leq b_k, \qquad n_1 \leq k \leq n_2 - 1,$

(A.7f) $\quad \lambda_k^* \geq 0, \qquad n_1 \leq k \leq n_2 - 1,$

where λ_k^* are the optimal Lagrange multipliers associated with the active constraints $C_k u_k^* = \underline{b}_k$.

We prove the result by induction starting from the rightmost endpoint. When $k = n_2 - 1$, KKT condition (A.7c) gives

$$\phi_{n_2-1}^* = 2Q_{n_2} x_{n_2}^* - 2Q_{n_2} d_{n_2},$$

which satisfies (2.19) because $M_{n_2}^{n_2} = Q_{n_2}$, as defined in (2.7f). Suppose (2.19) is true for k. Then, for $k-1$, (A.7b) gives

$$\phi_{k-1}^* = A_k^T \phi_k^* + 2Q_k(x_k^* - d_k).$$

Then by substituting the induction hypothesis and (A.7d), we have that

$$\begin{aligned}
\phi_{k-1}^* &= 2A_k^T K_{k+1}(A_k x_k^* + B_k u_k^*) + 2Q_k(x_k^* - d_k) \\
&\quad - 2A_k^T \left(\sum_{i=k+1}^{n_2} \left(M_i^{k+1}\right)^T d_i + \sum_{i=k+1}^{n_2-1} \left(S_i^{k+1}\right)^T f_i \right) \\
&= 2A_k^T K_{k+1}(A_k x_k^* + \hat{B}_k v_k^* + f_k) + 2Q_k(x_k^* - d_k) \\
&\quad - 2A_k^T \left(\sum_{i=k+1}^{n_2} \left(M_i^{k+1}\right)^T d_i + \sum_{i=k+1}^{n_2-1} \left(S_i^{k+1}\right)^T f_i \right),
\end{aligned}$$

since (2.4) implies that $\hat{B}_k v_k^* + f_k = \hat{B}_k \hat{u}_k^* + \widetilde{B}_k \widetilde{b}_k = B_k u_k^*$. Substituting v_k^* from the optimal control law (2.6) then gives the following:

$$\begin{aligned}
\phi_{k-1}^* &= 2\left(A_k^T K_{k+1} A_k + Q_k + A_k^T K_{k+1} \hat{B}_k L_k\right) x_k^* \\
&\quad - 2A_k^T \left(\sum_{i=k+1}^{n_2} \left(M_i^{k+1}\right)^T d_i + \sum_{i=k+1}^{n_2-1} \left(S_i^{k+1}\right)^T f_i \right) - 2Q_k d_k + 2A_k^T K_{k+1} f_k \\
&\quad + 2A_k^T K_{k+1} \hat{B}_k W_k^{-1} \hat{B}_k^T \left(\sum_{i=k+1}^{n_2} \left(M_i^{k+1}\right)^T d_i + \sum_{i=k+1}^{n_2-1} \left(S_i^{k+1}\right)^T f_i - K_{k+1} f_k \right) \\
&\overset{(2.7b),(2.7d)}{=} 2K_k x_k^* - 2A_k^T \left(\sum_{i=k+1}^{n_2} \left(M_i^{k+1}\right)^T d_i + \sum_{i=k+1}^{n_2-1} \left(S_i^{k+1}\right)^T f_i \right) \\
&\quad - 2Q_k d_k + 2A_k^T K_{k+1} f_k \\
&\quad - 2\left(\hat{B}_k L_k\right)^T \left(\sum_{i=k+1}^{n_2} \left(M_i^{k+1}\right)^T d_i + \sum_{i=k+1}^{n_2-1} \left(S_i^{k+1}\right)^T f_i - K_{k+1} f_k \right) \\
&\overset{(2.7e)}{=} 2K_k x_k^* - 2Q_k d_k - 2D_k^T \left(\sum_{i=k+1}^{n_2} \left(M_i^{k+1}\right)^T d_i + \sum_{i=k+1}^{n_2-1} \left(S_i^{k+1}\right)^T f_i - K_{k+1} f_k \right) \\
&= 2K_k x_k^* - 2\sum_{i=k}^{n_2} \left(M_i^k\right)^T d_i - 2\sum_{i=k}^{n_2-1} \left(S_i^k\right)^T f_i,
\end{aligned}$$

where the last equality follows from (2.7f) and (2.7g). This completes the proof.

A.6. Proof of Lemma 2.17. Propositions 2.8, 2.16, and Corollary 2.15 give

$$\begin{aligned}\|\phi_k^*\|_2 &\le 2\beta C_g + 2m_0 \sum_{i=k+1}^{n_2} \|M_i^{k+1}\|_2 + 2l_0 \sum_{i=k+1}^{n_2-1} \|S_i^{k+1}\|_2 \\ &\le 2\beta C_g + 2m_0 C_Q \sum_{i=k+1}^{n_2} C_1 \rho^{i-k-1} + 2l_0\beta \sum_{i=k+1}^{n_2-1} C_1\rho^{i-k} \\ &\le 2\beta C_g + \frac{2C_1(m_0 C_Q + \beta l_0)}{1-\rho} \triangleq C_\phi,\end{aligned}$$

where m_0 is the bound on the reference trajectory in Assumption 2.12 and l_0 is the bound on $\|f_i\|$ derived in Lemma 2.9.

A.7. Proof of Lemma 3.5. Let

$$\begin{aligned}L(y,\theta) = y^T G y/2 + y^T c(\theta) + \lambda^T(Ay - r) + \phi^T(By - d(\theta)) \\ + \theta^T F\theta + y^T c_1 + \theta^T c_2 + C\end{aligned} \tag{A.8}$$

be the Lagrangian of problem (3.8). Then we have that

$$\nabla^2_{(y,\theta)} L = \begin{bmatrix} G & \nabla_\theta c \\ \nabla_\theta^T c & * \end{bmatrix}.$$

Since G and F are positive definite and LICQ holds at y_0, from [6, Theorem 5.53] and [6, Remark 5.55] we have that

$$\begin{aligned}D_p y(\theta_0) &= \operatorname{argmin}_{h\in S} \begin{bmatrix} h^T & p^T\end{bmatrix} \left(\nabla^2_{(y,\theta)} L(y_0,\theta_0)\right) \begin{bmatrix} h \\ p\end{bmatrix} \\ &= \operatorname{argmin}_{h\in S}\, h^T G h/2 + p^T\left(\nabla_\theta^T c(\theta_0)\right) h,\end{aligned} \tag{A.9}$$

where S is the solution of linearized problem

$$\begin{aligned}\min_h \quad & (Gy_0 + c(\theta_0) + c_1)^T h + \left(\nabla_\theta^T c(\theta_0) y_0 + 2F\theta_0 + c_2\right)^T p \\ \text{s.t.} \quad & Bh - (\nabla_\theta d(\theta_0))\, p = 0, \\ & A_{I(y_0,\theta_0)} h \le 0,\end{aligned} \tag{A.10}$$

and S is given by

$$S = \left\{ h : \begin{bmatrix} B & -\nabla_\theta d(\theta_0)\end{bmatrix} \begin{bmatrix} h \\ p\end{bmatrix} = 0,\ \begin{bmatrix} A_{I_+(y_0,\theta_0,\bar\lambda)} & 0\end{bmatrix} \begin{bmatrix} h \\ p\end{bmatrix} = 0,\ \begin{bmatrix} A_{I_0(y_0,\theta_0,\bar\lambda)} & 0\end{bmatrix} \begin{bmatrix} h \\ p\end{bmatrix} \le 0 \right\}.$$

Thus the directional derivative $D_p y(\theta_0)$ of $y(\theta)$ along direction p at θ_0 is the solution of the problem

$$\begin{aligned}\min_h \quad & h^T G h/2 + p^T\left(\nabla_\theta^T c(\theta_0)\right) h \\ \text{s.t.} \quad & Bh - (\nabla_\theta d(\theta_0))\, p = 0, \\ & A_{I_+(y_0,\theta_0,\bar\lambda)} h = 0, \\ & A_{I_0(y_0,\theta_0,\bar\lambda)} h \le 0.\end{aligned} \tag{A.11}$$

Let I_1 be the set of active inequality constraints of problem (A.11). Then $I_1 \subset I_0(y_0,\theta_0,\bar\lambda)$ and let $I'(\theta_0) = I_1 \cup I_+(y_0,\theta_0,\bar\lambda)$. The KKT condition of problem (A.11) is hence

$$\widetilde{G} \triangleq \begin{bmatrix} G & A_{I'(\theta_0)}^T & B^T \\ A_{I'(\theta_0)} & 0 & 0 \\ B & 0 & 0 \end{bmatrix}, \qquad \widetilde{G} \begin{bmatrix} h^* \\ \phi_1^* \\ \phi_2^* \end{bmatrix} = \begin{bmatrix} -\nabla_\theta c(\theta_0)p \\ 0 \\ \nabla_\theta d(\theta_0)p \end{bmatrix}$$

for some Lagrange multipliers ϕ_1^* and ϕ_2^*. Since LICQ holds at y_0, rows of $A_{I'(\theta_0)}$ and B are linearly independent. Together with the fact that G is positive definite, we have that $\widetilde{G}$ is invertible. Denote the first row of $\widetilde{G}^{-1}$ to be $\begin{bmatrix} p_{11} & p_{12} & p_{13} \end{bmatrix}$. Then we have that

$$D_p y(\theta_0) = h^* = \left(-p_{11}\nabla_\theta c(\theta_0) + p_{13}\nabla_\theta d(\theta_0)\right) p.$$

On the other hand, for problem (3.9) with $I'(\theta_0)$ constructed above, the KKT condition is

$$\widetilde{G} \begin{bmatrix} y_{I'(\theta_0)}^*(\theta) \\ \psi_1^* \\ \psi_2^* \end{bmatrix} = \begin{bmatrix} -c(\theta) \\ r' \\ d(\theta) \end{bmatrix}$$

for some Lagrange multipliers ψ_1^* and ψ_2^*. Since $\widetilde{G}$ is invertible, we have that

$$y_{I'(\theta_0)}^*(\theta) = -p_{11}c(\theta) + p_{12}r' + p_{13}d(\theta).$$

It follows that

$$\left.\frac{dy_{I'(\theta_0)}^*(\theta)}{d\theta}\right|_{\theta=\theta_0} = -p_{11}\nabla_\theta c(\theta_0) + p_{13}\nabla_\theta d(\theta_0).$$

As a result, we have that

$$D_p y(\theta_0) = \left(\left.\frac{dy_{I'(\theta_0)}^*(\theta)}{d\theta}\right|_{\theta=\theta_0} \right) p,$$

which proves the claim.

Acknowledgments. We thank the anonymous referees for providing valuable comments and suggestions that greatly enhanced the content and presentation of the paper, and for pointing us to [17]. We thank the authors of [1] for clarifications regarding their work. Mihai Anitescu thanks his M^{2}ACS colleagues for countless discussions on such topics. We also thank Francois Gilbert for discussions and Julia software concerning temporal decomposition.

REFERENCES

[1] C. Barrows, M. Hummon, W. B. Jones, and E. T. Hale, *Time Domain Partitioning of Electricity Production Cost Simulations*, Technical report, National Renewable Energy Laboratory, Golden, CO, 2014.

[2] M. H. Bassett, J. F. Pekny, and G. V. Reklaitis, *Decomposition techniques for the solution of large-scale scheduling problems*, AIChE J., 42 (1996), pp. 3373–3387.

[3] P. E. Berry and R. M. Dunnett, *Contingency constrained economic dispatch algorithm for transmission planning*, IEE Generation Transm. Distrib., 136 (1989), pp. 238–244.

[4] D. P. Bertsekas, *Dynamic Programming and Optimal Control*, Approx. Dyn. Program. 1, Athena Scientific, Belmont, MA, 1995.

[5] L. T. Biegler and V. M. Zavala, *Large-scale nonlinear programming using Ipopt: An integrating framework for enterprise-wide dynamic optimization*, Comput. Chem. Eng., 33 (2009), pp. 575–582.

[6] J. F. Bonnans and A. Shapiro, *Perturbation Analysis of Optimization Problems*, Springer Ser. Oper. Res. Financ. En., Springer, New York, 2013.

[7] M. Bromberg, T.-S. Chang, and P. B. Luh, *Decomposition and coordination for non-convex optimal control problems with parallel algorithm*, in Procceedings of the 26th IEEE Conference on Decision and Control, IEEE Press, Piscataway, NJ, 1987, pp. 1468–1475.

[8] A. E. Bryson, *Dynamic Optimization*, Vol. 1, Prentice-Hall, Upper Saddle River, NJ, 1999.

[9] T.-S. Chang, S.-C. Chang, and P. B. Luh, *A hierarchical decomposition for large scale optimal control problems with parallel processing capability*, in Proceedings of the 1986 American Control Conference, IEEE Press, Piscataway, NJ, 1986, pp. 1995–2000; available at https://ieeexplore.ieee.org/xpl/mostRecentIssue.jsp?punumber=4788893.

[10] A. Elkamel, M. Zentner, F. Pekny, and G. V. Reklaitis, *A decomposition heuristic for scheduling the general batch chemical plant*, Eng. Optim., 28 (1997), pp. 299–330.

[11] J. Engwerda, *LQ Dynamic Optimization and Differential Games*, John Wiley, New York, 2005.

[12] C. E. Garcia, D. M. Prett, and M. Morari, *Model predictive control: Theory and practice—a survey*, Automatica J. IFAC, 25 (1989), pp. 335–348.

[13] I. Grossmann, *Enterprise-wide optimization: A new frontier in process systems engineering*, AIChE J., 51 (2005), pp. 1846–1857.

[14] W. W. Hager, *Lipschitz continuity for constrained processes*, SIAM J. Control Optim., 17 (1979), pp. 321–338.

[15] *Estimated Hourly Load Data*, PJM Interconnection, http://www.pjm.com/markets-and-operations/energy/real-time/loadhryr.aspx (accessed 28 May 2016).

[16] J. R. Jackson and I. E. Grossmann, *Temporal decomposition scheme for nonlinear multisite production planning and distribution models*, Ind. Eng. Chem. Res., 42 (2003), pp. 3045–3055.

[17] 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*, J. Optim. Theory Appl., 57 (1988), pp. 265–293.

[18] J. Khazaei, G. Zakeri, and S. S. Oren, *Market Clearing Mechanisms Under Uncertainty*, Technical report, University of California at Berkeley, 2014.

[19] M. Lubin and I. Dunning, *Computing in operations research using Julia*, INFORMS J. Comput., 27 (2015), pp. 238–248.

[20] F. D. Munoz, E. E. Sauma, and B. F. Hobbs, *Approximations in power transmission planning: Implications for the cost and performance of renewable portfolio standards*, J. Regulatory Econ., 43 (2013), pp. 305–338.

[21] Planning Division Transmission Planning Department, *PJM Region Transmission Planning Process: PJM Manual 14B*, PJM, Audubon, PA, 2017.

[22] W. Zhang, A. Abate, and J. Hu, *Efficient suboptimal solutions of switched LQR problems*, in Proceedings of the 2009 American Control Conference, IEEE Press, Piscataway, NJ, 2009, pp. 1084–1091; available at https://ieeexplore.ieee.org/xpl/mostRecentIssue.jsp?punumber=5089257.

[23] W. Zhang, J. Hu, and A. Abate, *On the value functions of the discrete-time switched LQR problem*, IEEE Trans. Automat. Control, 54 (2009), pp. 2669–2674.

[24] W. Zhang, J. Hu, and A. Abate, *A Study of the Discrete-Time Switched LQR Problem*, Technical report 384, Purdue University, West Lafayette, IN, 2009.