Preprint ANL/MCS-P1641-0609.

ON-LINE NONLINEAR PROGRAMMING AS A GENERALIZED EQUATION

VICTOR M. ZAVALA AND MIHAI ANITESCU*

Abstract. We establish results for the problem of tracking a time-dependent manifold arising in on-line nonlinear programming by casting this as a generalized equation. We demonstrate that if points along a solution manifold are consistently strongly regular, it is possible to track the manifold approximately by solving a single linear complementarity problem (LCP) at each time step. We derive sufficient conditions guaranteeing that the tracking error remains bounded to second order with the size of the time step, even if the LCP is solved only to first-order accuracy. We use these results to derive a fast, augmented Lagrangean tracking algorithm and demonstrate the developments through a numerical case study.

Key words. generalized equations, stability, nonlinear programming, on-line, complementarity

AMS subject classifications. 34B15, 34H05, 49N35, 49N90, 90C06, 90C30, 90C55, 90C59

1. Introduction. Advanced on-line optimization, control, and estimation strategies rely on repetitive solutions of nonlinear programming (NLP) problems. The structure of the NLP is normally fixed, but it depends on *time-dependent* data obtained at predefined sampling times (e.g. sensor measurements and model states).

Traditional on-line NLP strategies try to extend the sampling time (time step) as much as possible in order to accommodate the solution of the NLP to a fixed degree of accuracy. A problem with this approach is that it neglects the fact that the NLP solver is implicitly tracking a time-dependent solution manifold. For instance, insisting on obtaining a high degree of accuracy can translate into long sampling times and increasing distances between subsequent problems. In turn, the number of iterations required by the NLP solver increases. This inconsistency limits the application scope of on-line NLP to systems with slow dynamics.

Approximate on-line NLP strategies, on the other hand, try to minimize the time step by computing a cheap approximate solution within a fixed computational time. Since shortening the time step reduces the distance between neighboring problems, this approach also tends to reduce the tracking error. These strategies are particularly attractive for systems with fast dynamics. However, an important issue is to ensure that the tracking error will remain stable.

Approximate strategies such as real-time iterations and continuation schemes have been studied previously in the context of receding-horizon control and estimation. These strategies solve a single Newton-type step at each sampling time. In the real-time iteration strategy reported in [7], the model is used to predict the data (e.g., states) at the next step, and a perturbed quadratic programming (QP) problem is solved once the true data becomes available. In the absence of active-set changes, the perturbed QP reduces to a perturbed Newton step obtained from the solution of a linear system. It has been demonstrated that, by computing a single Newton step per time step, the tracking error remains bounded to second order with respect to the error between the predicted and the actual data. In order to prove this result, a specialized discrete-time, shrinking-horizon control setting was used. A limitation of this analysis is that the impact of the size of the time step gets lost, and the results cannot be applied directly in a more general setting. Furthermore, no error bounds have been provided for the case in which non smoothness effects are present along the manifold (e.g., active-set changes).

* Mathematics and Computer Science Division, Argonne National Laboratory, 9700 South Cass Ave, Argonne IL, 60439, USA (vzavala@mcs.anl.gov, anitescu@mcs.anl.gov)

The continuation scheme reported in [14] is a manifold tracking strategy in which the optimality conditions of the NLP are formulated as a differential equation. This permits a detailed numerical analysis of the tracking error as a function of the size of the time step. Sufficient conditions for the stability of the tracking error are derived. However, no order results are established. For implementation, the differential equation is linearized and discretized to derive the Newton step. The resulting linear system is solved approximately by using an iterative scheme such as generalized minimum-residual (GMRES). The use of an iterative linear solver is particularly attractive because it can be terminated early, as opposed to direct solvers. This is important in an on-line environment since it can significantly reduce the size of the time step. However, a limitation of continuation schemes is that active-set changes need to be handled indirectly using smoothing techniques (e.g., barrier functions [14, 17, 6]) which can introduce numerical instability.

In this work, we present a framework for the analysis of on-line NLP strategies based on generalized equation (GE) concepts. Our results are divided in two parts. First, we demonstrate that if points along a solution manifold are consistently strongly regular, it is possible to track the manifold approximately by solving a single linear complementarity problem (LCP) per time step. We derive sufficient conditions that guarantee that the tracking error remains bounded to second order with the size of the time step, even if the LCP is solved only approximately. These results generalize the approximation results in [7, 14] in the sense that we consider both equality and inequality constraints, with the possibility of changing the active-set along the manifold. In particular, the proposed approach does not require any smoothing, which makes it numerically more robust. Second, we derive an approximation approach where the NLP is reformulated using an augmented Lagrangean function. This permits the use of a matrix-free, projected successive over-relaxation (PSOR) algorithm to solve the LCP at each sampling time. We demonstrate that PSOR is particularly efficient because it can perform linear algebra and active-set identification tasks efficiently.

The paper is structured as follows. In the next section, we review basic concepts of parametric generalized equations. In Section 3 we will establish general approximation results and derive stability conditions for the tracking error. In Section 4 we will specialize these to the context of nonlinear programming. The augmented Lagrangean tracking algorithm and associated stability properties are presented in Section 5. A numerical case study is provided in Section 6. The paper closes with conclusions and directions of future work.

2. Generalized Equations.

In this section, we use the notation from [15, 4]. Consider the following parametric GE problem: For a given $t \in T \subseteq \Re$, find $w \in W \subseteq \Re^n$ such that

$$0 \in F(w,t) + \mathcal{N}_K(w). \tag{2.1}$$

Here, $F : W \times T \to Z$ is a continuously differentiable mapping in both arguments. The multifunction $\mathcal{N}_K : W \to 2^Z$ is the normal cone operator,

$$\mathcal{N}_K(w) = \begin{cases} \{\nu \in W' \mid (w-\alpha)^T \nu \geq 0,\, \forall\, \alpha \in K\} & \text{if } w \in K \\ \emptyset & \text{if } w \notin K \end{cases} \tag{2.2}$$

where $K \subseteq W$ is a polyhedral convex set and W' is the dual space of W. We denote the solution of (2.1) as w_t^*.

Our final goal is to create a discrete-time scheme $\bar{w}_{t_k}$ providing a cheap but stable approximation of the solution of (2.1), $w_{t_k}^*$. To achieve this, we will perform a single *truncated* Newton iteration for the *generalized equation* per time step.

2.1. The Nonlinear Equation Case.

While our work is concerned primarily with the situation when the generalized equation contains inequalities, a good intuition as to why a truncated approximation scheme works can be obtained by considering the case without inequality constraints $F(w,t) = 0$ or, equivalently, of $K = \Re^n$. In this particular case, the discrete-time scheme is obtained from

$$r_{t_k} + \nabla_w F(\bar{w}_{t_k}, t_k)(w - \bar{w}_{t_k}) + \nabla_t F(\bar{w}_{t_k}, t_k)\Delta t + F(\bar{w}_{t_k}, t_k) = 0. \tag{2.3}$$

Here, $\Delta t = t_{k+1} - t_k$, and r_{t_k} models the fact that the Newton iteration is solved inexactly. The solution of this linear system is $w = \bar{w}_{t_{k+1}}$. Assume that $||\nabla_w F||$, $\left|\left|\nabla_w F^{-1}\right|\right|$, $\nabla_t F$ and all the second derivatives of F are uniformly bounded. After inverting $\nabla_w F$ in (2.3) squaring the terms, and using the triangle inequality, we obtain that there exists $\gamma_0 = 3 \cdot \max\{\left|\left|\nabla_w F^{-1}\right|\right|^2, \left|\left|\nabla_w F^{-1}\right|\right|^2 ||\nabla_t F||^2\}$, independent of t and w, such that

$$\left|\left|\bar{w}_{t_{k+1}} - \bar{w}_{t_k}\right|\right|^2 \leq \gamma_0 \left(\Delta t^2 + ||r_{t_k}||^2 + ||F(\bar{w}_{t_k}, t_k)||^2\right). \tag{2.4}$$

Using Taylor's theorem, (2.3), and the triangle inequality, we obtain that there exists $\gamma_1 > 0$, dependent on the second derivatives of F, independent of t and w, such that

$$||F(\bar{w}_{t_k}, t_k)|| \leq ||r_{t_k}|| + \gamma_1 \left(\Delta t^2 + \left|\left|\bar{w}_{t_{k+1}} - \bar{w}_{t_k}\right|\right|^2\right). \tag{2.5}$$

Then, if we choose $\Delta t \leq \frac{1}{2}$ and solve the linear system (2.3) such that $||r_{t_k}|| \leq \gamma_2 ||\Delta t||^2$, it follows from (2.5), after replacing $\left|\left|\bar{w}_{t_{k+1}} - \bar{w}_{t_k}\right|\right|^2$ from (2.4), that

$$||F(\bar{w}_{t_k}, t_k)|| \leq \gamma_3 \Delta t^2 + \gamma_4 ||F(\bar{w}_{t_k}, t_k)||^2, \tag{2.6}$$

where $\gamma_3 = \gamma_2 + \gamma_1(1+\gamma_0) + \frac{1}{2}\gamma_0\gamma_1\gamma_2^2$, and $\gamma_4 = \gamma_0\gamma_1$. It follows that, as soon as $\Delta t \leq 1/\sqrt{4\gamma_4\gamma_3}$, we have that

$$||F(\bar{w}_{t_k}, t_k)|| \leq 2\gamma_3 \Delta t^2 \Rightarrow \left|\left|F(\bar{w}_{t_{k+1}}, t_{k+1})\right|\right| \leq 2\gamma_3 \Delta t^2.$$

By induction, we have that the preceding holds for all k. Therefore, if we solve the linearized problem sufficiently accurately, then for a sufficiently small time step we have that the residuals at all time steps will not exceed $O(\Delta t^2)$. Therefore, the solution manifold of $F(w,t) = 0$ can be tracked by the approximation scheme within $O(\Delta t^2)$. In particular, the discrete scheme remains stable as $\Delta t \to 0$.

Approximation results can also be established for the more general case including inequality constraints (i.e., the cone K is not trivial), but this task is not straightforward. The difficulties, as pointed out in [7], are technical and include the fact that, in the presence of inequality constraints, we cannot algebraically invert the solution mapping. In addition, nonsmoothness effects prevent the application of Taylor-like results. These are the difficulties we resolve in the following sections.

2.2. Linearized Generalized Equations. An important consequence of the structure of (2.1) is that it allows us to analyze the smooth and nonsmooth components independently. This greatly simplifies the task of establishing theoretical properties. As before, we are interested in constructing approximation schemes to track w_t^*. We start by defining the linearized generalized equation (LGE) at a given solution $w_{t_0}^*$,

$$r \in F(w_{t_0}^*, t_0) + F_w(w_{t_0}^*, t_0)(w - w_{t_0}^*) + \mathcal{N}_K(w) \tag{2.7}$$

where $F_w(w,t) := \nabla_w F(w,t) \in \Re^{n \times n}$. If $K = \Re^n_+$ (the nonnegativity orthant), solving the above LGE is equivalent to solving the perturbed linear complementarity problem,

$$w \geq 0, \quad \nu = F(w_{t_0}^*, t_0) + F_w(w_{t_0}^*, t_0)\Delta w - r \geq 0, \quad w^T \nu = 0. \tag{2.8}$$

If F_w is a symmetric matrix then (2.8) are, in turn, the optimality conditions of the quadratic programming (QP) problem,

$$\min_{\Delta w \geq -w^*_{t_0}} \quad \frac{1}{2}\Delta w^T F_w(w^*_{t_0}, t_0)\Delta w + F(w^*_{t_0}, t_0)^T \Delta w - r^T \Delta w. \tag{2.9}$$

We can rewrite (2.1) at any point w, t in the neighborhood of $w^*_{t_0}$ in terms of (2.7) by defining the *residual*,

$$r(w, t) = F(w^*_{t_0}, t_0) + F_w(w^*_{t_0}, t_0)(w - w^*_{t_0}) - F(w, t). \tag{2.10}$$

This gives, for any point satisfying (2.1),

$$r(w, t) \in F(w^*_{t_0}, t_0) + F_w(w^*_{t_0}, t_0)(w - w^*_{t_0}) + \mathcal{N}_K(w). \tag{2.11}$$

The above formulation will allow us to bound the distance between $w^*_{t_0}$ and neighboring points in terms of $r(w, t)$.

Central to this study is the inverse operator $\psi^{-1} : Z \to W$ of the perturbed LGE (2.11) which we define as

$$w \in \psi^{-1}[r] \quad \Leftrightarrow \quad r \in F(w^*_{t_0}, t_0) + F_w(w^*_{t_0}, t_0)(w - w^*_{t_0}) + \mathcal{N}_K(w). \tag{2.12}$$

In other words, the operator is a multifunction from the space of the residual (perturbation) of the LGE to the space of the solution. Note that the operator and the residual $r(w, t)$ depend implicitly on the linearization point $w^*_{t_0}$. This dependence will be made clear from the context, so we will not carry it in the notation. Some basic properties arising from the definition of the inverse operator are as follows:

$$w^*_{t_0} \in \psi^{-1}[r(w^*_{t_0}, t_0)] = \psi^{-1}[0], \quad w^*_t \in \psi^{-1}[r(w^*_t, t)].$$

Definition 2.1. *(Strong Regularity [15]). It is said that $w^*_{t_0}$ is a strongly regular solution of the LGE* (2.11) *if there exists a neighborhood $V_W \subseteq W$ of $w^*_{t_0}$ and a neighborhood $V_Z \subseteq Z$ of $r(w^*_{t_0}, t_0) = 0$, such that for every $r \in V_Z$,* (2.11) *has a unique solution $w = \psi^{-1}[r] \in V_W$, and the inverse mapping $\psi^{-1} : V_Z \to V_W$ is Lipschitz with constant L_ψ. That is, for any $r_1, r_2 \in V_Z$,*

$$\|\psi^{-1}[r_1] - \psi^{-1}[r_2]\| \leq L_\psi \|r_1 - r_2\|.$$

Establishing conditions for strong regularity consists of seeking properties of the derivative matrix $F_w(w^*_{t_0}, t_0)$ guaranteeing that ψ^{-1} becomes a single-valued function. To explain this, we consider the case $K = \Re^n_+$. At a given solution $w^*_{t_0}$, system (2.11) will have three different components,

$$(\mathbf{M}w^*_{t_0} + \mathbf{b})_j = 0,\ (w^*_{t_0})_j > 0, \quad j = 1 : n_a, \tag{2.13a}$$
$$(\mathbf{M}w^*_{t_0} + \mathbf{b})_j = 0,\ (w^*_{t_0})_j = 0, \quad j = n_a + 1 : n_s + n_a, \tag{2.13b}$$
$$(\mathbf{M}w^*_{t_0} + \mathbf{b})_j > 0,\ (w^*_{t_0})_j = 0, \quad j = n_s + n_a + 1 : n, \tag{2.13c}$$

where $n = n_a + n_s + n_i$, $\mathbf{M} := F_w(w^*_{t_0}, t_0)$, and $\mathbf{b} := F(w^*_{t_0}, t_0) - \mathbf{M}w^*_{t_0}$. By eliminating the last n_i inactive components from the system, $\mathbf{M}$ can be reduced to

$$\hat{\mathbf{M}} = \begin{bmatrix} \mathbf{M}_{11} & \mathbf{M}_{12} \\ \mathbf{M}_{21} & \mathbf{M}_{22} \end{bmatrix}, \tag{2.14}$$

and (2.11) can be expressed in the reduced form

$$r \in \hat{\mathbf{M}}y + \hat{\mathbf{b}} + \mathcal{N}_{\Re^{n_a} \times \Re^{n_s}_+}(y), \tag{2.15}$$

where $y \in \Re^{na+n_s}$ and $\hat{\mathbf{b}}^T = [\mathbf{b}_1^T \ \mathbf{b}_2^T]$.

THEOREM 2.2. *(Theorem 3.1 in [15]). Consider the system* (2.15) *and the operator*

$$\psi\, y := \hat{\mathbf{M}}y + \mathbf{b} + \mathcal{N}_{\Re^{n_a} \times \Re^{n_s}_+}(y).$$

Necessary and sufficient conditions for ψ^{-1} to be Lipschitzian are (i) $\mathbf{M}_{11}$ is nonsingular and (ii) $\mathbf{M}_{22} - \mathbf{M}_{21}\mathbf{M}_{11}^{-1}\mathbf{M}_{12}$ have positive principal minors. In Section 4, we will interpret these conditions in the context of parametric NLP. Using this basic set of tools, we now establish results that will allow us to construct algorithms able to track the solution manifold of (2.1).

3. Approximation Results. In our discussion, we will make the following blanket assumption.

ASSUMPTION 1. *The mapping $F_w(w,t)$ is Lipschitz in both arguments with constant L_{F_w}, $\forall w \in \mathcal{W}, t \in \mathcal{T}$.*

THEOREM 3.1. *Assume $w^*_{t_0}$ is strongly regular. Then, there exist neighborhoods V_W and V_T and a unique and Lipschitz continuous solution $w^*_t \in V_W$ of the GE* (2.1) *that satisfies, for each $t = t_0 + \Delta t \in V_T$,*

$$\textit{(i)} \quad \|w^*_t - w^*_{t_0}\| \le L_w \Delta t \tag{3.1}$$

with $L_w > 0$. In addition, consider the approximate solution $\bar{w}_t$ computed from the perturbed LGE (2.11) *with $r = F(w^*_{t_0}, t_0) - F(w^*_{t_0}, t)$. We have that $\bar{w}_t$ satisfies*

$$\textit{(ii)} \quad \|w^*_t - \bar{w}_t\| = o(\Delta t),$$

and, if Assumption 1 holds,

$$\textit{(iii)} \quad \|w^*_t - \bar{w}_t\| = O(\Delta t^2).$$

Proof. Result (i) follows from strong regularity (Def. 2.1) and Lipschitz continuity of ψ^{-1}. This can be established under a fixed-point argument for sufficiently small Δt, as shown in Theorem 2.1 in [15] and Theorem 5.13 in [4]. Result (ii) follows from strong regularity (Def. 2.1) and from the definition of the residual (2.10) for w^*_t,

$$\begin{aligned}
\|w^*_t - \bar{w}_t\| &\le L_\psi \|r(w^*_t, t) - r\| \\
&\le L_\psi \| \left(F(w^*_{t_0}, t_0) + F_w(w^*_{t_0}, t_0)(w^*_t - w^*_{t_0}) - F(w^*_t, t)\right) - \left(F(w^*_{t_0}, t_0) - F(w^*_{t_0}, t)\right) \| \\
&\le L_\psi \|F_w(w^*_{t_0}, t_0)(w^*_t - w^*_{t_0}) - F(w^*_t, t) + F(w^*_{t_0}, t)\|.
\end{aligned}$$

From the integral mean value theorem we have

$$F(w^*_t, t) - F(w^*_{t_0}, t) = \int_0^1 F_w(w^*_{t_0} + \chi(w^*_t - w^*_{t_0}), t)(w^*_t - w^*_{t_0}) d\chi, \tag{3.2}$$

so we obtain, after replacing (3.2) in the preceding equation, that

$$\begin{aligned}
\|w^*_t - \bar{w}_t\| &\le L_\psi \|w^*_t - w^*_{t_0}\| \left\| \int_0^1 \left(F_w(w^*_{t_0}, t_0) - F_w(w^*_{t_0} + \chi(w^*_t - w^*_{t_0}), t)\right) d\chi \right\| \\
&\overset{(i)}{\le} L_\psi L_w \Delta t \left\| \int_0^1 \left(F_w(w^*_{t_0}, t_0) - F_w(w^*_{t_0} + \chi(w^*_t - w^*_{t_0}), t)\right) d\chi \right\| = o(\Delta t).
\end{aligned}$$

Result (iii) is a consequence of the Lipschitz continuity of F_w,

$$\begin{aligned}\|w_t^* - \bar{w}_t\| &\leq L_\psi L_w \Delta t \int_0^1 \left\|\left(F_w(w_{t_0}^*, t_0) - F_w(w_{t_0}^* + \chi(w_t^* - w_{t_0}^*), t)\right)\right\| d\chi \\ &\leq L_\psi L_w \Delta t \int_0^1 L_{F_w}(\chi \left\|(w_t^* - w_{t_0}^*)\right\| + \Delta t) d\chi \\ &\overset{(i)}{\leq} L_\psi L_w \Delta t L_{F_w} \left(\frac{1}{2} L_w \Delta t + \Delta t\right) \leq L_\psi L_w L_{F_w} \left(\frac{1}{2} L_w + 1\right) \Delta t^2 = o(\Delta t^2).\end{aligned}$$

The proof is complete. ▯

Having a reference solution $w_{t_0}^*$, we can compute the approximate solution $\bar{w}_t$ by solving the LCP (2.8) or the QP (2.9) with $r = F(w_{t_0}^*, t_0) - F(w_{t_0}^*, t)$. From Theorem 3.1, we see that this approximation can be expected to be close to the optimal solution w_t^* even in the presence of active-set changes. In our approximate algorithm, however, we relax the requirement that $w_{t_0}^*$ be available. Instead, we consider a linearization point $\bar{w}_{t_0}$ located in the neighborhood of $w_{t_0}^*$. In addition, we assume that the LCP is not solved exactly. In other words, $\bar{w}_t$ is the solution of the LGE,

$$r_\epsilon \in F(\bar{w}_{t_0}, t) + F_w(\bar{w}_{t_0}, t_0)(w - \bar{w}_{t_0}) + \mathcal{N}_K(w), \tag{3.3}$$

where $r_\epsilon \in \Re^n$ represents a given solution error. This system can be posed in form (2.11) by the following definition:

$$r = r_\epsilon + F(w_{t_0}^*, t_0) + F_w(w_{t_0}^*, t_0)(w - w_{t_0}^*) - F(\bar{w}_{t_0}, t) - F_w(\bar{w}_{t_0}, t_0)(w - \bar{w}_{t_0}). \tag{3.4}$$

Note that, in this case, the perturbation r is an implicit function of the solution $w = \bar{w}_t$. However, note that (3.4) is used only as an analytical tool. In practice, it is not solved as an implicit LGE. In the following theorem we establish stability conditions for the tracking error $\|\bar{w}_t - w_t^*\|$.

Theorem 3.2. *(Stability of Tracking Error). Assume $w_{t_0}^*$ is a strongly regular solution of (2.11). Define $\bar{w}_t$ as the solution of the perturbed LGE (3.3) where $\bar{w}_{t_0}$ is a point in the neighborhood V_W of $w_{t_0}^*$. The associated residual $r(\bar{w}_{t_0}, t_0)$ is assumed to satisfy*

$$\|r(\bar{w}_{t_0}, t_0) - r(w_{t_0}^*, t_0)\| \leq \delta_r,$$

with $\delta_r > 0$. Assume there exists $\delta_\epsilon > 0$ such that $\|r_\epsilon\| \leq \delta_\epsilon$. If there exists $\kappa > 0$ and if Δt is chosen sufficiently small such that

$$L_{F_w} L_\psi \left(L_w + 1\right) \Delta t \, \delta_r \leq \kappa \Delta t^2 \tag{3.5a}$$

$$\left(L_w L_{F_w} \left(\frac{1}{2} L_w + 1\right) + \kappa\right) \Delta t^2 + \delta_\epsilon \leq \delta_r \left(1 - \frac{3}{2} L_{F_w} L_\psi^2 \delta_r\right), \tag{3.5b}$$

then the tracking error remains stable:

$$\|\bar{w}_{t_0} - w_{t_0}^*\| \leq L_\psi \delta_r \quad \Rightarrow \quad \|\bar{w}_t - w_t^*\| \leq L_\psi \delta_r.$$

Proof. From strong regularity (Def 2.1) and the assumed initial residual we have

$$\|\bar{w}_{t_0} - w_{t_0}^*\| \leq L_\psi \|r(\bar{w}_{t_0}, t_0) - r(w_{t_0}^*, t_0)\| \leq L_\psi \delta_r. \tag{3.6}$$

To bound $\|\bar{w}_t - w_t^*\|$ we need to bound the distance between the associated residuals. From (3.4) and (2.10) we have

$$
\begin{aligned}
&r - r(w_t^*, t)\\
&= r_\epsilon + F(w_{t_0}^*, t_0) + F_w(w_{t_0}^*, t_0)(\bar{w}_t - w_{t_0}^*) - F(\bar{w}_{t_0}, t) - F_w(\bar{w}_{t_0}, t_0)(\bar{w}_t - \bar{w}_{t_0})\\
&\quad - F(w_{t_0}^*, t_0) - F_w(w_{t_0}^*, t_0)(w_t^* - w_{t_0}^*) + F(w_t^*, t)\\
&= r_\epsilon + F_w(w_{t_0}^*, t_0)(\bar{w}_t - w_{t_0}^*) - F(\bar{w}_{t_0}, t) - F_w(\bar{w}_{t_0}, t_0)(\bar{w}_t - \bar{w}_{t_0})\\
&\quad - F_w(w_{t_0}^*, t_0)(w_t^* - w_{t_0}^*) + F(w_t^*, t)\\
&= r_\epsilon + F(w_t^*, t) - F_w(w_{t_0}^*, t_0)(w_t^* - w_{t_0}^*) - F(w_{t_0}^*, t)\\
&\quad + F(w_{t_0}^*, t) - F_w(\bar{w}_{t_0}, t_0)(w_{t_0}^* - \bar{w}_{t_0}) - F(\bar{w}_{t_0}, t)\\
&\quad + F_w(w_{t_0}^*, t_0)(\bar{w}_t - w_t^* + w_t^* - w_{t_0}^*) - F_w(\bar{w}_{t_0}, t_0)(\bar{w}_t - w_t^* + w_t^* - w_{t_0}^*).
\end{aligned}
$$

As in Theorem 3.1, we use the mean value theorem to compute the bounds:

$$
\begin{aligned}
\|F(w_t^*, t) - F_w(w_{t_0}^*, t_0)(w_t^* - w_{t_0}^*) - F(w_{t_0}^*, t)\| &\leq L_w L_{F_w} \left(\frac{1}{2} L_w + 1\right) \Delta t^2\\
\|F(w_{t_0}^*, t) - F_w(\bar{w}_{t_0}, t_0)(w_{t_0}^* - \bar{w}_{t_0}) - F(\bar{w}_{t_0}, t)\| &\leq L_{F_w} \left(\frac{1}{2} L_\psi^2 \delta_r^2 + L_\psi \delta_r \Delta t\right).
\end{aligned}
$$

We also have $\|r_\epsilon\| \leq \delta_\epsilon$. The remaining terms can be bounded as follows:

$$
\begin{aligned}
&\|F_w(w_{t_0}^*, t_0)(\bar{w}_t - w_t^* + w_t^* - w_{t_0}^*) - F_w(\bar{w}_{t_0}, t_0)(\bar{w}_t - w_t^* + w_t^* - w_{t_0}^*)\|\\
&\quad \leq L_{F_w} \|w_{t_0}^* - \bar{w}_{t_0}\| (\|\bar{w}_t - w_t^*\| + \|w_t^* - w_{t_0}^*\|)\\
&\quad \leq L_{F_w} L_\psi \delta_r \|\bar{w}_t - w_t^*\| + L_{F_w} L_w L_\psi \delta_r \Delta t.
\end{aligned}
$$

Merging terms, and moving all terms containing $\|\bar{w}_t - w_t^*\|$ to the left, we obtain

$$
\begin{aligned}
&\|\bar{w}_t - w_t^*\| \leq L_\psi \|r - r(w_t^*, t)\|\\
&\leq L_\psi \delta_\epsilon + L_\psi L_w L_{F_w} \left(\frac{1}{2} L_w + 1\right) \Delta t^2 + L_\psi L_{F_w} \left(\frac{1}{2} L_\psi^2 \delta_r^2 + L_\psi \delta_r \Delta t\right)\\
&\qquad\qquad + L_\psi L_{F_w} L_\psi \delta_r \|\bar{w}_t - w_t^*\| + L_\psi L_{F_w} L_w L_\psi \delta_r \Delta t \Longrightarrow\\
&\|\bar{w}_t - w_t^*\|\\
&\leq \frac{L_\psi \delta_\epsilon + L_\psi L_w L_{F_w} (\frac{1}{2} L_w + 1) \Delta t^2 + L_\psi L_{F_w} (\frac{1}{2} L_\psi^2 \delta_r^2 + L_\psi \delta_r \Delta t) + L_\psi L_{F_w} L_w L_\psi \delta_r \Delta t}{1 - L_{F_w} L_\psi^2 \delta_r}.
\end{aligned}
$$

To establish stability, we need to find conditions for Δt such that $\|\bar{w}_t - w_t^*\| \leq L_\psi \delta_r$. This implies,

$$
L_\psi \delta_r \geq \frac{L_\psi \delta_\epsilon + L_\psi L_w L_{F_w} (\frac{1}{2} L_w + 1) \Delta t^2 + L_\psi L_{F_w} (\frac{1}{2} L_\psi^2 \delta_r^2 + L_\psi \delta_r \Delta t) + L_\psi L_{F_w} L_w L_\psi \delta_r \Delta t}{1 - L_{F_w} L_\psi^2 \delta_r}.
$$

Dividing through by L_ψ, multipliying with the denominator, and simplifying we have

$$
\delta_r - \frac{3}{2} L_{F_w} L_\psi^2 \delta_r^2 \geq \delta_\epsilon + L_w L_{F_w} \left(\frac{1}{2} L_w + 1\right) \Delta t^2 + L_{F_w} L_\psi \left(L_w + 1\right) \Delta t \delta_r.
$$

This condition is satisfied if (3.5a)-(3.5b) hold. The proof is complete. ▯

COROLLARY 3.3. *Assume conditions of Theorem 3.2 hold* $\forall\, t \in [t_0, t_f]$. *Then,*

$$\|\bar{w}_{t_{k+1}} - w^*_{t_{k+1}}\| \leq L_\psi \delta_r,\ t_{k+1} = t_k + k \cdot \Delta t,\ \forall\, k \leq \frac{t_f - t_0}{\Delta t}.$$

Proof. Follows by induction. ▯

Discussion of Theorem 3.2. At every t_k, $\Delta\bar{w}_{t_k}$ is obtained by solving the LGE (3.3). The approximation of $w^*_{t_{k+1}}$ is obtained from $\bar{w}_{t_{k+1}} = \bar{w}_{t_k} + \Delta\bar{w}_{t_k}$. We have thus created an algorithm that tracks the solution manifold of the parametric GE (2.1) by solving inexactly (within δ_ϵ) one LCP at every step. This allows us to use an iterative algorithm that can be terminated early.

From Theorem 3.2, a condition for (3.5a)-(3.5b) to hold is that

$$\|r(\bar{w}_{t_k}, t_k) - r(w^*_{t_k}, t_k)\| = \|F(w^*_{t_k}, t_k) + F(w^*_{t_k}, t_k)(\bar{w}_{t_k} - w^*_{t_k}) - F(\bar{w}_{t_k}, t_k)\| \leq \delta_r, \quad (3.7)$$

where $r(w^*_{t_k}, t_k) = 0$. This condition gives a guideline for monitoring the progress of the algorithm. Condition (3.5a) can be satisfied easily for $\delta_r = o(\Delta t), O(\Delta t^2)$. Condition (3.5b) is stricter. If $\delta_r = o(\Delta t)$, this condition states that the solution error should be at least $\delta_\epsilon = o(\Delta t)$. The first term on the left-hand side represents the tracking error of $\bar{w}_t$ if $w^*_{t_0}$ is used as linearization point. If we choose $\delta_r = O(\Delta t^2)$ at the initial point, and $\delta_\epsilon = O(\Delta t^2)$ at all subsequent iterations, there will exist κ such that for all Δt sufficiently small the tracking error is $O(\Delta t^2)$ as stated in Theorem 3.1. Note that a small L_ψ is beneficial because it relaxes both (3.5a) and (3.5b). As seen in Theorem 2.2, this Lipschitz constant can be related to the conditioning of the derivative matrix F_w.

We also note that the technique of proof for Theorem 3.2 is similar to the one concerning the geometrical infeasibility of a time-stepping method [1] for differential variational inequalities (DVI) [13]. Indeed, one can prove that the parameteric solution w^*_t satisfies a DVI. Nevertheless, the fact that the problem has no dynamics makes it easy to solve directly rather than casting it as a DVI.

4. On-Line Nonlinear Programming. We now specialize the results of the previous sections to parametric NLP problems of the form

$$\min \quad f(x,t), \quad \text{s.t.} \quad c(x,t) = 0, \quad x \geq 0. \quad (4.1)$$

Here, $x \in \Re^n$ and the mappings $f : \Omega \times T \to \Re$, $c : \Omega \times T \to \Re^m$ are assumed to be continuously differentiable from the open sets $\Omega \subseteq \Re^n$ and $T \subseteq \Re$. To simplify our discussion and without loss of generality, we consider only the case where all components of x are subject to inequality constraints. The first-order optimality conditions of this problem are

$$\nabla_x \mathcal{L}(w,t) - \nu = 0,\ c(x,t) = 0,\ x^T \nu = 0,\ x \geq 0,\ \nu \geq 0. \quad (4.2)$$

The Lagrange function is defined as

$$\mathcal{L}(w,t) = f(x,t) + \lambda^T c(x,t), \quad (4.3)$$

where $\lambda \in \Re^m$ are Lagrange multipliers and $w^T = [x^T, \lambda^T]$. Note that (4.2) can be formulated without introducing the extra variables $\nu \in \Re^n$. These multipliers are introduced only for clarity in the presentation. The optimality conditions can be posed as a GE of the form,

$$0 \in \nabla_x \mathcal{L}(w,t) - \nu + \mathcal{N}_{\Re^n}(x) \quad (4.4a)$$

$$0 \in c(x,t) + \mathcal{N}_{\Re^m}(\lambda) \quad (4.4b)$$

$$0 \in x + \mathcal{N}_{\Re^n_+}(\nu). \quad (4.4c)$$

If we linearize the optimality conditions around a given solution $w^*_{t_0}$ we get,

$$0 \in \begin{bmatrix} H_{xx}(w^*_{t_0}, t_0) & J_x^T(x^*_{t_0}, t_0) & -\mathbb{I}_n \\ J_x(x^*_{t_0}, t_0) & & \\ \mathbb{I}_n & & \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta \lambda \\ \Delta \nu \end{bmatrix} + \begin{bmatrix} \nabla_x \mathcal{L}(w^*_{t_0}, t_0) - \nu^*_{t_0} \\ c(x^*_{t_0}, t_0) \\ x^*_{t_0} \end{bmatrix} + \begin{bmatrix} \mathcal{N}_{\Re^n}(x) \\ \mathcal{N}_{\Re^m}(\lambda) \\ \mathcal{N}_{\Re^n_+}(\nu) \end{bmatrix}. \tag{4.5}$$

Here, $\Delta x := x - x^*_{t_0}$, $\Delta\lambda := \lambda - \lambda^*_{t_0}$, $\Delta\nu := \nu - \nu^*_{t_0}$, $J_x(x^*_{t_0}, t_0) := \nabla_x c(x^*_{t_0}, t_0)$, and $H_{xx}(w^*_{t_0}, t_0) := \nabla_{xx}\mathcal{L}(w^*_{t_0}, t_0)$. As shown in Section 2, to establish conditions for strong regularity, we eliminate the n_i components corresponding to the pair $(x^*_{t_0})_j > 0, (\nu^*_{t_0})_j = 0$. This gives a reduced matrix of the form

$$\left[\begin{array}{c|c} \mathbf{K}(w^*_{t_0}, t_0) & -\mathbb{E} \\ \hline \mathbb{E}^T & \end{array}\right] = \left[\begin{array}{ccc|c} H_{xx}(w^*_{t_0}, t_0) & J_x^T(x^*_{t_0}, t_0) & -\mathbb{I}_{n_a} & -\mathbb{I}_{n_s} \\ J_x(x^*_{t_0}, t_0) & & & \\ \mathbb{I}_{n_a} & & & \\ \hline \mathbb{I}_{n_s} & & & \end{array}\right], \tag{4.6}$$

where $\mathbb{E} = [\mathbb{I}_{n_s} \,|\, 0 \,|\, 0]$.

THEOREM 4.1. *(Strong Regularity of NLP). Let $f(x, \cdot)$ and $c(x, \cdot)$ be functions from the open set $\Omega \in \Re^n$ into $\Re, \Re^m$ that are at least twice differentiable at a point $x^*_{t_0} \in \Omega$. Suppose that $w^*_{t_0}$ solves* (4.4). *If, (i) for every nonzero vector $w \in \Re^n$ satisfying $J_x(x^*_{t_0}, t_0)w = 0$, $\mathbb{I}_{n_a} w = 0$, one has $w^T H_{xx}(w^*_{t_0}, t_0)w > 0$, and (ii) $[J_x^T(x^*_{t_0}, t_0) \,|\, \mathbb{I}_{n_a} \,|\, \mathbb{I}_{n_s}]$ is full rank, then* (4.5) *is strongly regular at this point.*

Proof. From Theorem 2.2 we have that it suffices $\mathbf{K}(w^*_{t_0}, t_0)$ to be nonsingular and the Schur complement matrix $\mathbb{E}^T \mathbf{K}(w^*_{t_0}, t_0)^{-1}\mathbb{E}$ to be positive definite. As shown in Theorem 4.1 in [15], this is consequence of conditions (i) and (ii). ☐

The conditions of Theorem 4.1 are the strong second-order conditions and the linear independence constraint qualification (LICQ) (Chapter 12 in [12]). As seen in Section 2, strong regularity guarantees that there exist nonempty neighborhoods where the solution of the linearized GE is a Lipschitz continuous function of the problem data. A similar result has been obtained in [8] without resorting to GE results. In [16] it is shown that by weakening LICQ to the Mangasarian-Fromovitz constraint qualification (MFCQ), the Lipschitz continuity properties of the solution are lost (see discussion after Corollary 4.3). The reason is that LICQ guarantees that the multifunction (4.4) becomes a single-valued function on a neighborhood of the solution (i.e., the multipliers are unique). Nevertheless, boundedness results still hold under MFCQ. We emphasize that strict complementarity slackness is not necessary to guarantee strong regularity. This property is crucial since, as t varies and the active-set change, points at which complementarity slackness does not hold will be encountered.

Consider the perturbed QP problem formed at $\bar{w}^T_{t_0} = [\bar{x}^T_{t_0}, \bar{\lambda}^T_{t_0}]$ in the neighborhood of $w^*_{t_0}$,

$$\min_{\Delta x \geq -\bar{x}_{t_0}} \quad \nabla_x f(\bar{x}_{t_0}, t)^T \Delta x + \frac{1}{2}\Delta x^T H_{xx}(\bar{w}_{t_0}, t_0)\Delta x \tag{4.7a}$$

$$\text{s.t.} \quad c(\bar{x}_{t_0}, t) + J_x(\bar{x}_{t_0}, t_0)\Delta x = 0, \tag{4.7b}$$

where $\Delta x = x - \bar{x}_{t_0}$. Note the perturbation $t_0 \leftarrow t$ in the equality constraints and in the gradient of the objective function. The solution of this problem is given by the step $\Delta\bar{w}_t$ toward

w_t^*. The optimality conditions of this QP formulate an LGE of the form (3.3). Therefore, the results of Theorem 3.2 apply directly.

5. Augmented Lagrangean Strategy. The approximation results of the previous sections can be used to derive algorithms to track of the solution manifold of the NLP (4.1). For instance, as we have seen, solving a single QP (4.7) at each time step is sufficient. In our context, however, we assume that the QPs are large-scale and may contain many degrees of freedom and active bounds. Therefore, it is crucial to have a fast solution strategy for the QP itself in order to keep Δt as small as possible. Here, we propose to reformulate the NLP using an augmented Lagrangean function and solve the underlying QP using a PSOR strategy. The justification of this approach is provided at the end of this section. To derive our strategy, we define the augmented Lagrangean function,

$$\mathcal{L}_A(x,\bar{\lambda},t,\rho) = f(x,t) + \bar{\lambda}^T c(x,t) + \frac{\rho}{2}\|c(x,t)\|^2. \tag{5.1}$$

A strategy to solve the original NLP (4.1) consists of computing solutions of the augmented Lagrangean subproblem

$$\min_{x\geq 0} \quad \mathcal{L}_A(x,\bar{\lambda},t,\rho) \tag{5.2}$$

for a sequence of increasing ρ. In the following, we assume that the penalty parameter ρ is not updated but remains fixed to a sufficiently large value. Consequently, we drop from the notation any dependencies on this parameter. Note that the multipliers $\bar{\lambda}$ act as parameters of the augmented Lagrangean subproblem. The solution of the subproblem is defined as $x^*(\bar{\lambda},t)$. The multipliers can be updated externally as

$$\bar{\lambda} \leftarrow \bar{\lambda} + \rho\, c(x^*(\bar{\lambda},t),t). \tag{5.3}$$

We thus define the solution pair $x^*(\bar{\lambda},t)$, $\Lambda^*(\bar{\lambda},t) = \bar{\lambda} + \rho\, c(x^*(\bar{\lambda},t),t)$. The first-order conditions of (5.2) can be posed as a GE of the form

$$0 \in \nabla_x \mathcal{L}_A(x,\bar{\lambda},t) + \mathcal{N}_{\Re^n_+}(x), \tag{5.4}$$

where

$$\nabla_x \mathcal{L}_A(x,\bar{\lambda},t) = \nabla_x f(x,t) + (\bar{\lambda} + \rho\, c(x,t))^T \nabla_x c(x,t).$$

The linearized version of (5.4) defined at the NLP solution $x^*_{t_0}$, $\bar{\lambda} = \lambda^*_{t_0}$ is given by

$$r \in \nabla_x \mathcal{L}_A(x^*_{t_0},\lambda^*_{t_0},t_0) + \nabla_{xx}\mathcal{L}_A(x^*_{t_0},\lambda^*_{t_0},t_0)(x - x^*_{t_0}) + \mathcal{N}_{\Re^n_+}(x) \tag{5.5}$$

for $r = 0$. To establish perturbation results for the augmented Lagrangean LGE in connection with those of the original NLP (4.1), we consider the following *equivalent* formulation of (5.4), proposed in [3]:

$$0 \in F(w,p(\bar{\lambda}),t) + \mathcal{N}_{\Re^n_+ \times \Re^m}(w), \tag{5.6}$$

where

$$F(w,p(\bar{\lambda}),t) = \begin{bmatrix} \nabla_x f(x,t) + \Lambda^T \nabla_x c(x,t) \\ c(x,t) + p(\bar{\lambda}) + \frac{1}{\rho}(\lambda^*_{t_0} - \Lambda) \end{bmatrix}, \tag{5.7}$$

$w^T = [x^T\, \Lambda^T]$, and

$$p(\bar{\lambda}) = \frac{1}{\rho}(\bar{\lambda} - \lambda^*_{t_0}). \tag{5.8}$$

For $t = t_0$ and $\bar{\lambda} = \lambda^*_{t_0}$, we have $p(\bar{\lambda}) = 0$, $x^*(p(\bar{\lambda}), t) = x^*_{t_0}$, and $\Lambda^*(p(\bar{\lambda}), t) = \lambda^*_{t_0}$. The solution of GE (5.6) is denoted as $w^*(p(\bar{\lambda}), t)$. The linearized version of (5.6) at $w^*_{t_0}$ is

$$r \in F(w^*_{t_0}, 0, t_0) + F_w(w^*_{t_0}, 0, t_0)(w - w^*_{t_0}) + \mathcal{N}_{\Re^n_+ \times \Re^m}(w), \tag{5.9}$$

where

$$F_w(w^*_{t_0}, 0, t_0) = \begin{bmatrix} \nabla_{xx}\mathcal{L}(w^*_{t_0}, t_0) & \nabla_x c(x^*_{t_0}, t_0) \\ \nabla_x^T c(x^*_{t_0}, t_0) & -\frac{1}{\rho}\mathbb{I}_m \end{bmatrix}. \tag{5.10}$$

After applying the reduction procedure of Section 2 to the derivative matrix (5.10) we can show that, for sufficiently large ρ, the reduced matrix satisfies conditions of Theorem 2.2 at a strongly regular solution $w^*_{t_0}$. The proof of this assertion is long and will be omitted here. It follows along the lines of the results of Section 4 and uses the results of Proposition 2.4 in [3]. In particular, one needs to show that the negative diagonal matrix in the bottom right-hand corner of (5.10) does not affect significantly the conditioning of the derivative matrix for sufficiently large ρ. Because of the equivalence between (5.4) and (5.6), the same can be argued for the Hessian matrix $\nabla_{xx}\mathcal{L}_A(x^*_{t_0}, \lambda^*_{t_0}, t_0)$. We emphasize that the reformulation (5.6) is considered only for theoretical purposes. In practice, (5.4) is solved.

We now establish the following approximation results in the context of the augmented Lagrangean framework.

LEMMA 5.1. *Assume $w^*_{t_0}$ is a strongly regular solution of* (5.5). *Then, there exist neighborhoods V_W,V_T, and V_p where the solution of the augmented Lagrangean subproblem* (5.2) *satisfies, for each $t = t_0 + \Delta t \in V_T$, $p(\bar{\lambda}) \in V_p$,*

$$(i) \quad \|w^*(\bar{\lambda}, t) - w^*_{t_0}\| \leq \frac{L_w}{\rho}\|\bar{\lambda} - \lambda^*_{t_0}\| + L_w \Delta t. \tag{5.11}$$

Furthermore, consider the approximate solution $\bar{x}(\bar{\lambda}, t)$ obtained from the perturbed LGE (5.5) *with*

$$r = \nabla_x \mathcal{L}_A(x^*_{t_0}, \lambda^*_{t_0}, t_0) - \nabla_x \mathcal{L}_A(x^*_{t_0}, \bar{\lambda}, t), \tag{5.12}$$

and associated multiplier $\bar{\Lambda}(\bar{\lambda}, t) = \bar{\lambda} + \rho\, c(\bar{x}(\bar{\lambda}, t), t)$. The pair, denoted by $\bar{w}(\bar{\lambda}, t)$, satisfies

$$(ii) \quad \|\bar{w}(\bar{\lambda}, t) - w^*(\bar{\lambda}, t)\| = O\left(\left(\Delta t + \frac{1}{\rho}\|\bar{\lambda} - \lambda^*_{t_0}\|\right)^2\right). \tag{5.13}$$

Proof. The result follows from the equivalence between (5.4) and (5.6), by recalling that $p(\lambda^*_{t_0}) = 0$, $p(\bar{\lambda}) = \frac{1}{\rho}\|\bar{\lambda} - \lambda^*_{t_0}\|$, and by applying Theorem 3.1. ☐

This result states that the solution of a perturbed augmented Lagrangean LGE formed at $w^*_{t_0}$ provides a second-order approximation of the subproblem solution $w^*(\bar{\lambda}, t)$. The impact of the multiplier error can be made arbitrarily small by fixing ρ to a sufficiently large value. Stability of the tracking error is established in the following theorem. Here, we relax the requirement of the availability of $w^*_{t_0}$. In addition, we establish conditions for the step size Δt and the penalty parameter ρ guaranteeing that, by solving a single augmented Lagrangean LGE per time step, the tracking error remains stable.

THEOREM 5.2. *(Stability of Tracking Error for Augmented Lagrangean). Assume $w^*_{t_0}$ is a strongly regular solution of* (5.5). *Define $\bar{x}(\bar{\lambda}, t)$ as the solution of the LGE,*

$$r_\epsilon \in \nabla_x \mathcal{L}_A(\bar{x}_{t_0}, \bar{\lambda}, t) + \nabla_{xx}\mathcal{L}_A(\bar{x}_{t_0}, \bar{\lambda}, t_0)(x - \bar{x}_{t_0}) + \mathcal{N}_{\Re^n_+}(x), \tag{5.14}$$

with associated multiplier update $\bar{\Lambda}(\bar{\lambda},t)=\bar{\lambda}+\rho\, c(\bar{x}(\bar{\lambda},t),t)$. The pair is denoted by $\bar{w}(\bar{\lambda},t)$. The reference linearization point $\bar{w}_{t_0}^T=[\bar{x}_{t_0}^T,\bar{\Lambda}_{t_0}^T]$ with $\bar{\Lambda}_{t_0}=\bar{\lambda}+\rho\, c(\bar{x}_{t_0},t_0)$ is assumed to exist in the neighborhood V_W of $w_{t_0}^*$. The associated residual $r(\bar{w}_{t_0},t_0)$ is assumed to satisfy $\|r(\bar{w}_{t_0},t_0)-r(w_{t_0}^*,t_0)\|\leq\delta_r$ with $\delta_r>0$. Furthermore, assume there exists $\delta_\epsilon>0$ such that $\|r_\epsilon\|\leq\delta_\epsilon$. If there exists $\kappa>0$, Δt is chosen sufficiently small and ρ sufficiently large such that

$$L_{F_w}L_\psi\left(L_w\left(1+\frac{L_\psi}{\rho}\right)+1\right)\Delta t\delta_r+\frac{L_w}{\rho}\left(\delta_r+\frac{L_w}{L_\psi}\Delta t\right)\leq\kappa\left(\Delta t+\frac{L_\psi\delta_r}{\rho}\right)^2 \tag{5.15a}$$

$$\left(L_wL_{F_w}\left(\frac{1}{2}L_w+1\right)+\kappa\right)\left(\Delta t+\frac{L_\psi\delta_r}{\rho}\right)^2+\delta_\epsilon\leq\delta_r\left(1-\left(\frac{3}{2}+\frac{L_\psi}{\rho}\right)L_{F_w}L_\psi^2\delta_r\right). \tag{5.15b}$$

Then, the tracking error remains stable:

$$\|\bar{w}_{t_0}-w_{t_0}^*\|\leq L_\psi\delta_r \quad\Rightarrow\quad \|\bar{w}(\bar{\lambda},t)-w_t^*\|\leq L_\psi\delta_r.$$

Proof. Using the equivalence between (5.4) and (5.6), we have $\bar{w}(\bar{\lambda},t)=\bar{w}(p(\bar{\lambda}),t)$. Consequently, we need to bound

$$\begin{aligned}\|\bar{w}(p(\bar{\lambda}),t)-w_t^*\|&=\|\bar{w}(p(\bar{\lambda}),t)-w^*(p(\bar{\lambda}),t)+w^*(p(\bar{\lambda}),t)-w_t^*\|\\&\leq\|\bar{w}(p(\bar{\lambda}),t)-w^*(p(\bar{\lambda}),t)\|+\|w^*(p(\bar{\lambda}),t)-w_t^*\|.\end{aligned} \tag{5.16}$$

The second term, the distance between the solution of the augmented Lagrangean subproblem $w^*(p(\bar{\lambda}),t)$ and the NLP solution w_t^*, can be bounded by using the Lipschitz continuity property,

$$\begin{aligned}\|w^*(p(\bar{\lambda}),t)-w_t^*\|&=\|w^*(p(\bar{\lambda}),t)-w^*(p(\lambda_t^*),t)\|\\&\leq L_w\|p(\bar{\lambda})-p(\lambda_t^*)\|\\&\leq L_w\|p(\bar{\lambda})\|+L_w\|p(\lambda_t^*)\|.\end{aligned} \tag{5.17}$$

The distance between $w^*(p(\bar{\lambda}),t)$ and the approximate solution of the LGE (5.14) follows from the definition of strong regularity.

$$\|\bar{w}(p(\bar{\lambda}),t)-w^*(p(\bar{\lambda}),t)\|\leq L_\psi\|r-r(w^*(p(\bar{\lambda}),t),t)\|.$$

From the equivalence between (5.4) and (5.6) we have that solving (5.14) is equivalent to solving

$$r_\epsilon\in F(\bar{w}_{t_0},p(\bar{\lambda}),t)+F_w(\bar{w}_{t_0},p(\bar{\lambda}),t_0)(w-\bar{w}_{t_0})+\mathcal{N}_{\Re_+^n\times\Re^m}(w).$$

Consequently, the perturbation r associated to $\bar{w}(p(\bar{\lambda}),t)$ is given by

$$\begin{aligned}r=r_\epsilon+F(w_{t_0}^*,0,t_0)+F_w(w_{t_0}^*,0,t_0)(w-w_{t_0}^*)-F(\bar{w}_{t_0},p(\bar{\lambda}),t)&\\-F_w(\bar{w}_{t_0},p(\bar{\lambda}),t_0)(w-\bar{w}_{t_0}),&\end{aligned}$$

with $w = \bar{w}(p(\bar{\lambda}), t)$. The residual $r(w^*(p(\bar{\lambda}), t), t)$ is obtained from (2.10). We have

$$\begin{aligned}
\mathbf{A} &= r - r(w^*(p(\bar{\lambda}), t)) \\
&= r_\epsilon + F(w^*_{t_0}, 0, t_0) + F_w(w^*_{t_0}, 0, t_0)(\bar{w}(p(\bar{\lambda}), t) - w^*_{t_0}) \\
&\quad - F(\bar{w}_{t_0}, p(\bar{\lambda}), t) - F_w(\bar{w}_{t_0}, p(\bar{\lambda}), t_0)(\bar{w}(p(\bar{\lambda}), t) - \bar{w}_{t_0}) \\
&\quad - F(w^*_{t_0}, 0, t_0) - F_w(w^*_{t_0}, 0, t_0)(w^*(p(\bar{\lambda}), t) - w^*_{t_0}) + F(w^*(p(\bar{\lambda}), t), p(\bar{\lambda}), t) \\
&= r_\epsilon + F_w(w^*_{t_0}, 0, t_0)(\bar{w}(p(\bar{\lambda}), t) - w^*_{t_0}) + F(w^*_{t_0}, p(\bar{\lambda}), t) \\
&\quad - F(\bar{w}_{t_0}, p(\bar{\lambda}), t) - F_w(\bar{w}_{t_0}, p(\bar{\lambda}), t_0)(\bar{w}(p(\bar{\lambda}), t) - \bar{w}_{t_0}) \\
&\quad - F_w(w^*_{t_0}, 0, t_0)(w^*(p(\bar{\lambda}), t) - w^*_{t_0}) + F(w^*(p(\bar{\lambda}), t), p(\bar{\lambda}), t) - F(w^*_{t_0}, p(\bar{\lambda}), t) \\
&= r_\epsilon + F(w^*(p(\bar{\lambda}), t), p(\bar{\lambda}), t) - F_w(w^*_{t_0}, 0, t_0)(w^*(p(\bar{\lambda}), t) - w^*_{t_0}) - F(w^*_{t_0}, p(\bar{\lambda}), t) \\
&\quad + F(w^*_{t_0}, p(\bar{\lambda}), t) - F_w(\bar{w}_{t_0}, p(\bar{\lambda}), t_0)(w^*_{t_0} - \bar{w}_{t_0}) - F(\bar{w}_{t_0}, p(\bar{\lambda}), t) \\
&\quad + F_w(w^*_{t_0}, 0, t_0)(\bar{w}(p(\bar{\lambda}), t) - w^*_t + w^*_t - w^*_{t_0}) \\
&\quad - F_w(\bar{w}_{t_0}, p(\bar{\lambda}), t_0)(\bar{w}(p(\bar{\lambda}), t) - w^*_t + w^*_t - w^*_{t_0}).
\end{aligned}$$

We use the mean-value theorem,

$$\begin{aligned}
&F(w^*(p(\bar{\lambda}), t), p(\bar{\lambda}), t) - F(w^*_{t_0}, p(\bar{\lambda}), t) \\
&\qquad = \int_0^1 F_w(w^*_{t_0} + \chi(w^*(p(\bar{\lambda}), t) - w^*_{t_0}), p(\bar{\lambda}), t)(w^*(p(\bar{\lambda}), t) - w^*_{t_0}) d\chi,
\end{aligned}$$

to compute the following bound:

$$\begin{aligned}
\mathbf{B} &= \|F(w^*(p(\bar{\lambda}), t), p(\bar{\lambda}), t) - F_w(w^*_{t_0}, 0, t_0)(w^*(p(\bar{\lambda}), t) - w^*_{t_0}) - F(w^*_{t_0}, p(\bar{\lambda}), t)\| \\
&\leq \int_0^1 \left\| \left(F_w(w^*_{t_0} + \chi(w^*(p(\bar{\lambda}), t) - w^*_{t_0}), p(\bar{\lambda}), t) - F_w(w^*_{t_0}, 0, t_0)\right)(w^*(p(\bar{\lambda}), t) - w^*_{t_0}) \right\| d\chi \\
&\leq \|w^*(p(\bar{\lambda}), t) - w^*_{t_0}\| \int_0^1 L_{F_w} \left(\chi \|w^*(p(\bar{\lambda}), t) - w^*_{t_0}\| + \|p(\bar{\lambda})\| + \Delta t\right) d\chi \\
&\leq \frac{1}{2} L_{F_w} \|w^*(p(\bar{\lambda}), t) - w^*_{t_0}\|^2 + L_{F_w} \|w^*(p(\bar{\lambda}), t) - w^*_{t_0}\| (\|p(\bar{\lambda})\| + \Delta t) \\
&\leq \frac{1}{2} L_{F_w} L_w^2 \left(\|p(\bar{\lambda})\| + \Delta t\right)^2 + L_{F_w} L_w \left(\|p(\bar{\lambda})\| + \Delta t\right) \left(\|p(\bar{\lambda})\| + \Delta t\right) \\
&\leq L_w L_{F_w} \left(\frac{1}{2} L_w + 1\right) \left(\Delta t + \|p(\bar{\lambda})\|\right)^2.
\end{aligned}$$

Similarly,

$$\begin{aligned}
\mathbf{C} &= \|F(w^*_{t_0}, p(\bar{\lambda}), t) - F_w(\bar{w}_{t_0}, p(\bar{\lambda}), t_0)(w^*_{t_0} - \bar{w}_{t_0}) - F(\bar{w}_{t_0}, p(\bar{\lambda}), t)\| \\
&\leq \int_0^1 \|F_w(\bar{w}_{t_0} + \chi(\bar{w}_{t_0} - w^*_{t_0}), p(\bar{\lambda}), t) - F_w(\bar{w}_{t_0}, p(\bar{\lambda}), t_0)\| \|\bar{w}_{t_0} - w^*_{t_0}\| d\chi \\
&\leq \|\bar{w}_{t_0} - w^*_{t_0}\| \int_0^1 L_{F_w} \left(\chi \left|\left|\bar{w}_{t_0} - w^*_{t_0}\right|\right| + \Delta t\right) d\chi \\
&\leq \frac{1}{2} L_{F_w} \|\bar{w}_{t_0} - w^*_{t_0}\|^2 + L_{F_w} \|\bar{w}_{t_0} - w^*_{t_0}\| \Delta t \\
&\leq \frac{1}{2} L_{F_w} L_\psi^2 \delta_r^2 + L_{F_w} L_\psi \delta_r \Delta t.
\end{aligned}$$

The remaining terms can be bounded as

$$\begin{aligned}\mathbf{D} &= \|F_w(w^*_{t_0},0,t_0)(\bar{w}(p(\bar{\lambda}),t) - w^*_t + w^*_t - w^*_{t_0}) - F_w(\bar{w}_{t_0},p(\bar{\lambda}),t_0)(\bar{w}(p(\bar{\lambda}),t) - w^*_t + w^*_t - w^*_{t_0})\| \\ &\leq \|F_w(w^*_{t_0},0,t_0) - F_w(\bar{w}_{t_0},p(\bar{\lambda}),t_0)\|\|\bar{w}(p(\bar{\lambda}),t) - w^*_t + w^*_t - w^*_{t_0}\| \\ &\leq L_{F_w}\left(\|w^*_{t_0} - \bar{w}_{t_0}\| + \|p(\bar{\lambda})\|\right)\left(\|\bar{w}(p(\bar{\lambda}),t) - w^*_t\| + \|w^*_t - w^*_{t_0}\|\right) \\ &\leq L_{F_w}\left(L_\psi \delta_r + \|p(\bar{\lambda})\|\right)\left(\|\bar{w}(p(\bar{\lambda}),t) - w^*_t\| + L_w \Delta t\right).\end{aligned}$$

Using $\|r_\epsilon\| \leq \delta_\epsilon$ and merging terms **B**, **C**, and **D** into **A**, we obtain

$$\begin{aligned}\|\bar{w}(p(\bar{\lambda}),t) - w^*(p(\bar{\lambda}),t)\| \leq\ & L_\psi \delta_\epsilon + L_\psi L_w L_{F_w}\left(\frac{1}{2}L_w + 1\right)\left(\Delta t + \|p(\bar{\lambda})\|\right)^2 \\ & + L_\psi \frac{1}{2} L_{F_w} L_\psi^2 \delta_r^2 + L_\psi L_{F_w} L_\psi \delta_r \Delta t \\ & + L_\psi L_{F_w}\left(L_\psi \delta_r + L_\psi \|p(\bar{\lambda})\|\right)\left(\|\bar{w}(p(\bar{\lambda}),t) - w^*_t\| + L_w \Delta t\right).\end{aligned} \tag{5.18}$$

We substitute (5.17) and (5.18) in (5.16) and apply,

$$\begin{aligned}\|p(\bar{\lambda})\| &\leq \frac{1}{\rho}\|\bar{\lambda} - \lambda^*_{t_0}\| \leq \frac{L_\psi}{\rho}\|r(\bar{w}_{t_0},t_0) - r(w^*_{t_0},t_0)\| \leq \frac{L_\psi}{\rho}\delta_r \\ \|p(\lambda^*_t)\| &\leq \frac{1}{\rho}\|\lambda^*_t - \lambda^*_{t_0}\| \leq \frac{L_w}{\rho}\Delta t\end{aligned} \tag{5.19}$$

to obtain

$$\begin{aligned}&\|\bar{w}(p(\bar{\lambda}),t) - w^*_t\| \\ &\quad \leq L_\psi \delta_\epsilon + L_\psi L_w L_{F_w}\left(\frac{1}{2}L_w + 1\right)\left(\Delta t + \|p(\bar{\lambda})\|\right)^2 \\ &\qquad + L_\psi \frac{1}{2} L_{F_w} L_\psi^2 \delta_r^2 + L_\psi L_{F_w} L_\psi \delta_r \Delta t \\ &\qquad + L_\psi L_{F_w}\left(L_\psi \delta_r + L_\psi \|p(\bar{\lambda})\|\right)\left(\|\bar{w}(p(\bar{\lambda}),t) - w^*_t\| + L_w \Delta t\right) \\ &\qquad + L_w \|p(\bar{\lambda})\| + L_w \|p(\lambda^*_t)\| \\ &\quad \leq L_\psi \delta_\epsilon + L_\psi L_w L_{F_w}\left(\frac{1}{2}L_w + 1\right)\left(\Delta t + \frac{L_\psi \delta_r}{\rho}\right)^2 \\ &\qquad + \frac{1}{2} L_\psi L_{F_w} L_\psi^2 \delta_r^2 + L_\psi L_{F_w} L_\psi \delta_r \Delta t \\ &\qquad + L_{F_w} L_\psi^2 \delta_r \left(1 + \frac{L_\psi}{\rho}\right)\left(\|\bar{w}(p(\bar{\lambda}),t) - w^*_t\| + L_w \Delta t\right) \\ &\qquad + L_w \frac{L_\psi \delta_r}{\rho} + L_w^2 \frac{\Delta t}{\rho}.\end{aligned}$$

For stability we require $\|\bar{w}(p(\bar{\lambda}),t) - w^*_t\| \leq L_\psi \delta_r$. This implies

$$\begin{aligned}L_\psi \delta_r \geq\ & \frac{L_\psi \delta_\epsilon + \frac{L_w}{\rho}\left(L_\psi^2 \delta_r + L_w \Delta t\right) + L_\psi L_w L_{F_w}\left(\frac{1}{2}L_w + 1\right)\left(\Delta t + \frac{L_\psi \delta_r}{\rho}\right)^2}{1 - L_{F_w} L_\psi^2 \delta_r \left(1 + \frac{L_\psi}{\rho}\right)} \\ & + \frac{L_\psi L_{F_w}(\frac{1}{2}L_\psi^2 \delta_r^2 + L_\psi \delta_r \Delta t) + L_{F_w} L_\psi^2 \delta_r \left(1 + \frac{L_\psi}{\rho}\right) L_w \Delta t}{1 - L_{F_w} L_\psi^2 \delta_r \left(1 + \frac{L_\psi}{\rho}\right)}.\end{aligned}$$

Dividing through by L_ψ and rearranging we have,

$$
\begin{aligned}
&\delta_r - L_{F_w} L_\psi^2 \delta_r^2 \left(1 + \frac{L_\psi}{\rho}\right) \\
&\quad \geq \delta_\epsilon + \frac{L_w}{\rho}\left(\delta_r + \frac{L_w}{L_\psi}\Delta t\right) + L_w L_{F_w}\left(\frac{1}{2}L_w + 1\right)\left(\Delta t + \frac{L_\psi \delta_r}{\rho}\right)^2 \\
&\qquad + L_{F_w}(\frac{1}{2}L_\psi^2\delta_r^2 + L_\psi \delta_r \Delta t) + L_{F_w} L_\psi L_w \delta_r \left(1 + \frac{L_\psi}{\rho}\right)\Delta t
\end{aligned}
$$

$$
\begin{aligned}
&\delta_r - \left(\frac{3}{2} + \frac{L_\psi}{\rho}\right) L_{F_w} L_\psi^2 \delta_r^2 \\
&\quad \geq \delta_\epsilon + L_w L_{F_w}\left(\frac{1}{2}L_w + 1\right)\left(\Delta t + \frac{L_\psi \delta_r}{\rho}\right)^2 \\
&\qquad + L_{F_w} L_\psi \left(L_w\left(1 + \frac{L_\psi}{\rho}\right) + 1\right)\delta_r \Delta t + \frac{L_w}{\rho}\left(\delta_r + \frac{L_w}{L_\psi}\Delta t\right).
\end{aligned}
$$

This last condition is satisfied if (5.15a)-(5.15b) hold. The proof is complete . ◻

Discussion of Theorem 5.2. The recursive stability result of Corollary 3.3 also applies in this context. Conditions (5.15a)-(5.15b) reduce to (3.5a)-(3.5b) for $\rho \to \infty$. Therefore, similar order results to those of Theorem 3.2 can be expected for sufficiently large ρ. Note also that the initial multiplier error (bounded by δ_r) always appears divided by ρ. This indicates that relatively large initial multiplier errors can be tolerated by increasing ρ. Nevertheless, note that the second term on the left hand side of (5.15a) remains $o(\Delta t)$ even if $\delta_r = O(\Delta t^2)$. In other words, this condition is more restrictive than (3.5a). This term arises from the application of the Lipschitz property to bound $\|\lambda_t^* - \lambda_{t_0}^*\|$ One could try the alternative approach of using the residuals to obtain less conservative bounds for $\|w^*(p(\bar{\lambda}), t) - w^*(p(\lambda_t^*), t)\|$ in (5.17). Even with this approach, however, we have not been able to do so. This difficulty seems to be related to the fact that the multiplier update is only first-order [3], so, at least in the form presented here, we find it unlikely that the method would succeed in converging as $\Delta t \to 0$ without letting $\rho \to \infty$.

As a final remark, we point out that the stability conditions can be satisfied for fixed and sufficiently large κ as long as $\rho = O\left(\frac{1}{\Delta t}\right)$ and $\delta_r = O(\Delta t^2)$. This has the side effect of having ρ effectively as a penalty parameter, a situation that resembles the use of a smoothing barrier function and that may raise stability problems. While both penalizations arise in different contexts, an important question is whether the augmented Lagrangean penalization is more stable than that obtained by using smoothing penalty functions. In our scheme, the penalty parameter is *finite* for every fixed Δt, and the scheme is guaranteed to be stable. For continuation schemes, however, stability results incorporating smoothing functions are currently lacking. A simple numerical comparison will be presented in the next section. A more rigorous stability analysis is a left as an interesting topic for future research.

In order to solve the QP associated to the LGE (5.14), we follow a PSOR approach. The QP has the form,

$$
\min_{z \geq \alpha} \quad \frac{1}{2} z^T \mathbf{M} z + \mathbf{b}^T z. \tag{5.20}
$$

Any solution of this QP solves the LCP,

$$
\mathbf{M}z + \mathbf{b} \geq 0, \quad z - \alpha \geq 0, \quad (z - \alpha)^T(\mathbf{M}z + \mathbf{b}) = 0. \tag{5.21}
$$

Consider the following PSOR algorithm adapted from [10, 11]:

PSOR Algorithm.
Given $z^0 \geq \alpha$, compute for $k = 0, 1, ..., n_{iter}$,

$$z_i^{k+1} = (1-\omega) z_i^k - \frac{\omega}{\mathbf{M}_{ii}} \left(\sum_{j<i} \mathbf{M}_{ij} z_j^{k+1} + \sum_{j>i} \mathbf{M}_{ij} z_j^k - \mathbf{b}_i \right)$$
$$z_i^{k+1} = \max\left(z_i^{k+1}, \alpha_i\right), \quad i = 1, ..., n, \tag{5.22}$$

where ω is the relaxation factor.

THEOREM 5.3. *(Theorem 2.1 in [11]). Let* $\mathbf{M}$ *be symmetric positive definite. Then, each accumulation point of the sequence* $\{z^k\}$ *generated by* (5.22) *converges to a solution of the LCP* (2.8)*. The rate of convergence is R-linear.*

A suitable measure of progress of the PSOR algorithm is the projected gradient (or residual) $P_K(\mathbf{M}\, z + \mathbf{b})$, where $K := \{z \,|\, z \geq \alpha\}$ and

$$(P_K(g))_j = \begin{cases} \min\{0, g_j\} & \text{if } z_j = \alpha_j \\ g_j & \text{if } z_j > \alpha_j. \end{cases} \tag{5.23}$$

This is based on the fact that a solution of (5.20) satisfies $P_K(\mathbf{M}\, z + \mathbf{b}) = 0$. Similarly, the progress of the algorithm can be monitored by using the projected gradient of the augmented Lagrangean function $P_{\Re^n_+}\left(\nabla_x \mathcal{L}_A(\bar{x}_{t_k}, \bar{\lambda}_{t_k}, t_k, \rho)\right)$. This is a more direct convergence check of (2.1), as opposed to (3.7). The computational complexity of PSOR is at most $O(n^2)$ where n is the dimension of x. We can now establish our tracking algorithm (4.1), which we refer to as AugLag:

AugLag Tracking Algorithm.
Given $\bar{x}_{t_0}, \bar{\lambda}_{t_0}$, Δt, ρ, n_{iter},

1. Evaluate $\nabla_x \mathcal{L}_A(\bar{x}_{t_k}, \bar{\lambda}_{t_k}, t_{k+1}, \rho)$ and $\nabla_{xx} \mathcal{L}_A(\bar{x}_{t_k}, \bar{\lambda}_{t_k}, t_k, \rho)$.
2. Compute step $\Delta \bar{x}_{t_{k+1}}$ by applying n_{iter} PSOR iterations to (5.20) with $\mathbf{M} = \nabla_{xx} \mathcal{L}_A(\bar{x}_{t_k}, \bar{\lambda}_{t_k}, t_k, \rho)$, $\mathbf{b} = \nabla_x \mathcal{L}_A(\bar{x}_{t_k}, \bar{\lambda}_{t_k}, t_{k+1}, \rho)$.
3. Update primal variables $\bar{x}_{t_{k+1}} = \bar{x}_{t_k} + \Delta \bar{x}_{t_{k+1}}$ and multipliers $\bar{\lambda}_{t_{k+1}} = \bar{\lambda}_{t_k} + \rho\, c(\bar{x}_{t_{k+1}}, t_{k+1})$.
4. Set $k \leftarrow k + 1$.

Justification of Augmented Lagrangean Framework. To justify the choice of the AugLag framework from a computational point of view we make the following observations. If the QP (4.7) is sparse, full-space active-set and interior-point solvers are the most efficient alternatives [2, 18]. In on-line NLP applications, active-set strategies have been preferred because warm-start information can be used to reduce the number of iterations, as opposed to interior-point methods. However, the time per iteration in an interior-point solver tends to be smaller because the linear algebra can be done more efficiently. The reason is that the structure of the Karush-Kuhn-Tucker matrix is fixed and, consequently, symbolic factorizations need to be applied only once. In addition, high-level structures can be exploited. In most active-set and interior-point implementations, direct indefinite linear solvers are used to compute the search step. The accuracy of these steps is high. However, the computational overhead of a single factorization can be very high as well. As an alternative, one could consider the use of iterative linear solvers such as QMR, GMRES, or PCG in an interior-point framework [5]. A problem with this approach is that multiple linear systems still need to be solved since active-set identification (i.e., barrier parameter update) is performed externally. This situation could be avoided by fixing the barrier parameter to a given value. However, as we will see in the next section, this smoothing approach is not very robust. Based on these observations, we argue that the AugLag strategy is attractive because: (i) the iteration matrix (Hessian of the augmented Lagrangean) remains at least positive semi-definite close to the solution manifold, (ii) it performs linear algebra and active-set identification tasks simultaneously, (iii) it can exploit warm-start information, and (iv) it has a favorable computational complexity. We emphasize that achieving a high accuracy with PSOR might require a very large number of iterations. As demonstrated by Theorem 5.2, however, this does not represent a limitation in an on-line setting. Nevertheless, it does limits the attractiveness of PSOR in a more general NLP context.

6. Numerical Example. To illustrate the developments, we consider the receding-horizon control of a nonlinear CSTR [9]. The optimal control formulation is given by

$$\begin{aligned}
&\min_{u(\tau)} \int_t^{t+T} \left(w_T(z_T - z_T^{sp})^2 + w_C(z_C - z_C^{sp})^2 + w_u(u - u^{sp})^2\right) d\tau \\
&\text{s.t. } \frac{dz_C}{d\tau} = \frac{z_C - 1}{\theta} + k_0 z_C \exp\left[\frac{-E_a}{z_T}\right], \quad z_C(0) = z_C(t) \\
&\frac{dz_T}{d\tau} = \frac{z_T - z_T^f}{\theta} - k_0 z_C \exp\left[\frac{-E_a}{z_T}\right] + \alpha u(z_T - z_T^{cw}), \quad z_T(0) = z_T(t) \\
&z_C^{min} \le z_C \le z_C^{max}, \quad z_T^{min} \le z_T \le z_T^{max}, \quad u^{min} \le u \le u^{max}.
\end{aligned}$$

The system involves two states, $z(\tau) = [z_C(\tau),\, z_T(\tau)]$, corresponding to dimensionless concentration and temperature, and one control, $u(\tau)$, corresponding to the cooling water flow rate. The *model* time dimension is denoted by τ, and the *real* time dimension is denoted by t. Accordingly, the receding-horizon is defined as $\tau \in [t,\, t+T]$, and the initial conditions are $z_T(t)$ and $z_C(t)$. The model parameters are $z_T^{cw} = 0.38$, $z_T^f = 0.395$, $E_a = 5$, $\alpha = 1.95 \times 10^4$, $\theta = 20$, $k_0 = 300$, $w_C = 1 \times 10^6$, $w_T = 1 \times 10^3$, and $w_u = 1 \times 10^{-3}$. The bounds are set to $z_C^{min} = 0$, $z_C^{max} = 0.5$, $z_T^{min} = 0.5$, $z_T^{max} = 1.0$, $u^{min} = 0.25$, and $u^{max} = 0.45$. The set-points are denoted by the superscript *sp*. For implementation, the optimal control problem is converted into an NLP of the form in (4.1) by applying an implicit Euler discretization scheme with $N = 100$ grid points and $\Delta\tau = 0.25$. The NLP is parametric in the initial states, which are implicit functions of t. To apply the AugLag tracking algorithm, we define a simulation horizon $t \in [t_0, t_f]$ which is divided into N_s points with states $z(t_k),\ k = 0, ..., N_s$ and $\Delta t = t_{k+1} - t_k$. We set the augmented Lagrangean penalty parameter to $\rho = 100$. To solve the augmented Lagrangean QP at each step, we fix the number of PSOR iterations to 25. To illustrate the importance of handling non smoothness effects in a consistent manner, we compare the performance AugLag with two continuation algorithms incorporating different smoothing barrier functions. The first algorithm (Log Barrier) smooths out the inequality constraints by using terms of the form $\mu \cdot \log(x - x^{min}) + \mu \cdot \log(x^{max} - x),\ \mu = 1.0$ [17, 18]. The second algorithm (Sqrt Barrier) incorporates terms of the form $\mu \cdot \mathrm{sqrt}(x - x^{min}) + \mu \cdot \mathrm{sqrt}(x^{max} - x),\ \mu = 100$ [14, 6]. To prevent indefiniteness of the barrier functions near the boundaries of the feasible region, we incorporate a fraction to the boundary rule of the form,

$$x = \min(\max(x, x^{min} + \epsilon), x^{max} - \epsilon), \quad \epsilon = 1 \times 10^{-3}.$$

We initialize the algorithms by perturbing an initial solution $w_{t_0}^*$ as $\bar{w}_{t_0} \leftarrow w_{t_0}^* \cdot \delta_w$ where $\delta_w > 0$ is a perturbation parameter. This perturbation generates the initial residual $r(\bar{w}_{t_0}, t_0)$

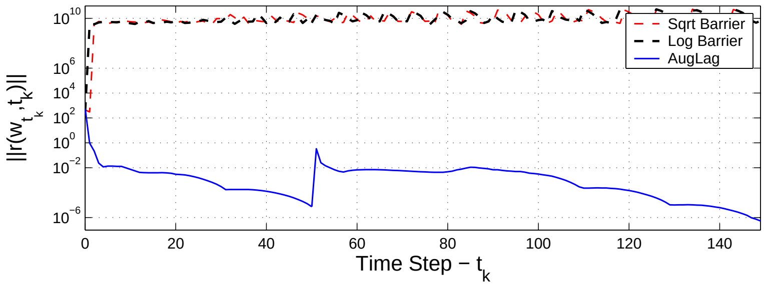

FIG. 6.1. *Residual trajectories for Barrier and AugLag continuation algorithms with* $\delta_w = 1.25$.

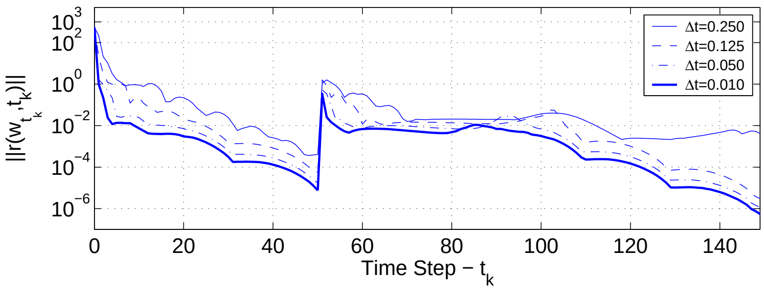

FIG. 6.2. *Residual trajectories for AugLag with increasing* Δt.

. An additional perturbation, in the form of a set-point change, is introduced at $t_k = 50$. The residuals along the manifold $r(\bar{w}_{t_k}, t_k)$ are computed from (3.7). In our numerical experiments, all the algorithms are able to tolerate relatively large initial perturbations, maintaining the residuals stable along the entire time horizon. Log Barrier destabilizes at $\delta_w = 1.20$ while Sqrt Barrier destabilizes at $\delta_w = 1.25$. AugLag remains stable in both cases, tolerating perturbations as large as $\delta_w = 5.0$. In Figure 6.1, we present the norm of the residuals along the simulation horizon with $\Delta t = 0.025$ and for an initial perturbation of $\delta_w = 1.25$. As can be seen, even if the initial residual is large, $O(10^2)$, AugLag remains stable. In addition, the use of smoothing functions introduces numerical instability. We now illustrate the effect of Δt on the residual of AugLag. Here, the initial residual is generated by using $\delta_w = 5.0$ and can go as high as $O(10^3)$. In Figure 6.2, note that the residual levels remain stable, implying that δ_r is at least $O(10^3)$. The set-point change generates a residual that is only $O(10^0)$ and can be tolerated with no problems. The PSOR residuals r_ϵ at the beginning of the horizon and at $t_k = 50$ are $O(10^{-1})$ and go down to $O(10^{-6})$ when the system reaches the set-points. In Figure 6.3, we present control and temperature profiles for $\Delta t = 0.25$ and $\Delta t = 0.01$. As expected, the tracking error decreases with the step size. We note that the PSOR strategy does a good job at identifying the active-set changes in subsequent steps. At a single step, up to 100 changes were observed. For the larger step size, note that even if the active-sets do not match, the residuals remain bounded and the system eventually converges to the optimal trajectories.

7. Conclusions and Future Work. We have presented a framework for the analysis of parametric nonlinear programming (NLP) problems based on generalized equation concepts. The framework allows us to derive approximate algorithms for on-line NLP. We demonstrate that if points along a solution manifold are consistently strongly regular, it is possible to

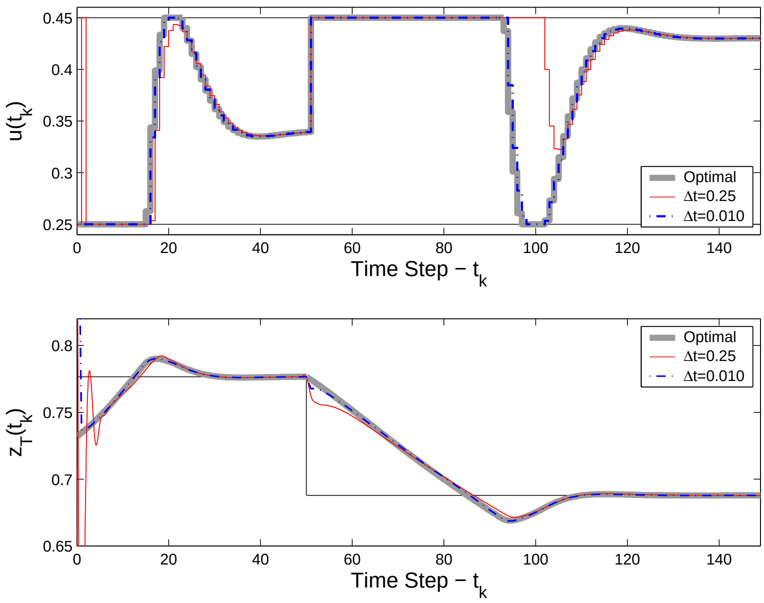

FIG. 6.3. *AugLag and optimal trajectories for the control (top) and temperature (bottom).*

track the manifold approximately by solving a single linear complementarity problem (LCP) per time step. We established sufficient conditions that guarantee that the tracking error remains bounded to second order with the size of the time step, even if the LCP is solved only to first-order accuracy. We present a tracking algorithm based on an augmented Lagrangean reformulation and a projected successive overrelaxation strategy to solve the LCPs. We demonstrate that the algorithm is able to identify multiple active-set changes and reduce the tracking errors efficiently. As part of our future work, we will establish a more rigorous comparison between the stability properties of the augmented Lagrangean penalization and of smoothing approaches. In addition, we will study the possibility of proving convergence, without requiring the augmented Lagrangean penalty parameter to go to ∞. We will establish a more efficient implementation of the algorithm and will perform a detailed computational analysis. This strategy has the potential to solve large-scale NLPs in real-time environments. In addition, we are interested in exploring a strategy able to adapt the number of PSOR iterations (and thus the step size) along the manifold. Such a strategy could reduce the tracking errors and improve the robustness of the algorithm.

Acknowledgments. This work was supported by the U.S. Department of Energy through contract DE-AC02-06CH11357.

REFERENCES

[1] M. ANITESCU AND G. HART, *A constraint-stabilized time-stepping approach for rigid multibody dynamics with joints, contact and friction*, International Journal for Numerical Methods In Engineering, 60, 2335–2371, 2004.

[2] R. A. BARTLETT, A. WACHTER AND L. T. BIEGLER. *Active-set vs. interior point strategies for model predictive control.* In Proceedings of Americal Control Conference. 4229–4233, 2000.

[3] D. .BERTSEKAS. *Constrained Optimization and Lagrange Multiplier Methods.* New York, Academic Press, 1982.

[4] F. Bonnans and A. Shapiro. *Perturbation Analysis of Optimization Problems*. New York, Springer-Verlag, 2000.
[5] R. H. Byrd, J. C. Gilbert and J. Nocedal. *A trust-region method based on interior-point techniques for nonlinear programming. Math. Programm.* 89, 149–185, 2000.
[6] M. Diehl, H. J. Ferreau and N. Haverbeke. *Efficient numerical methods for nonlinear MPC and moving horizon estimation*. In *Nonlinear Model Predictive Control* (pp. 391-417). Berlin, Springer-Verlag, 2009.
[7] M. Diehl, H.G. Bock, and J.P. Schlöder. *A real-time iteration scheme for nonlinear optimization in optimal feedback control*. SIAM J. Control and Optim. 43(5),1714–1736, 2005.
[8] W. W. Hager. *Lipschitz continuity for constrained processes.* SIAM J. Control Optim., 17(3), 321–338, 1979.
[9] G. A. Hicks and W.H.Ray, *Approximation methods for optimal control synthesis*. Can. J. Chem. Eng., 49, 522–529, 1971.
[10] Kočvara, M. and Jochem Zowe. *An iterative two-step algorithm for linear complementarity problems.* Numerische Mathematik, 68, 95–106, 1994.
[11] O. L. Mangasarian. *Solution of symmetric linear complementarity problems by iterative methods.* JOTA, 22(4), 465–485, 1977.
[12] J. Nocedal and S. J. Wright. *Numerical Optimization*. New York, Springer-Verlag, 1999.
[13] J. Pang and D. Stewart, *Differential variational inequalities*, Mathematical Programming, 113, 345–424, 2008.
[14] T. Ohtsuka. *A continuation/GMRES method for fast computation of non-linear receding horizon control.* Automatica., 40, 563–574, 2004.
[15] S. M. Robinson. *Strongly Regular Generalized Equations.* Mathematics of Operations Research, 5(1), 43–61, 1980.
[16] S. M. Robinson. *Generalized Equations and their Solutions. II. Applications to Nonlinear Programming.* Math. Programming Stud., 19, 200–221, 1982.
[17] A. G. Wills and W. P. Heath, *Barrier function based model predictive control*, Automatica, 40, 1415–1422, 2004.
[18] V. M. Zavala and L. T. Biegler. Nonlinear Programming Strategies for State Estimation and Model Predictive Control. In *Nonlinear Model Predictive Control*, (pp. 419-432). Berlin, Springer-Verlag, 2009.