TECHNICAL REPORT, DEPARTMENT OF CHEMICAL ENGINEERING, NTNU
23.03.2016

CONVERGENCE OF A PREDICTOR-CORRECTOR SENSITIVITY BASED PARAMETRIC NLP SOLVER FOR DUAL-DEGENERATE PROBLEMS

VYACHESLAV KUNGURTSEV[†], JOHANNES JÄSCHKE[‡]

Abstract. Parametric optimization, tracking the solution of an optimization problem dependant on a parameter as the parameter changes, is an important research area for a number of practical problems. Most path-following algorithms for tracing a solution path of a parametric nonlinear optimization problem are only certifiably convergent under strong regularity assumptions about the problem functions, in particular, the linear independence of the constraint gradients at the solutions, which implies a unique multiplier solution for each nonlinear program. In this paper we propose and prove convergence results for a procedure designed to solve problems satisfying a weaker set of conditions, allowing for non-unique (but bounded) multipliers, applying known sensitivity results for this class of problems. Each iteration along the path consists of three parts: (1) a Newton corrector step for the primal and dual variables, this is obtained by solving a linear system of equations. (2) a tangential predictor step for the primal and dual variables, which is found as the solution of a quadratic programming problem, and (3) a jump step for the dual variables, which is found as the solution of a linear programming problem. We present a convergence proof, and demonstrate the robustness of the algorithm numerically on a couple of illustrative examples.

1. Introduction. We consider the parametric optimization problem, with $f : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}$ and $c : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}^m$,

$$\begin{aligned} \min_{x\in\mathbb{R}^n} \quad & f(x,t) \\ \text{subject to} \quad & c_i(x,t) = 0,\ i \in \mathcal{E}, \\ & c_i(x,t) \geq 0,\ i \in \mathcal{I}. \end{aligned} \tag{1.1}$$

where $\mathcal{E} = \{1, ..., m_e\}$ and $\mathcal{I} = \{m_e + 1, ..., m\}$ and we seek to trace the solution path along a parameter change from $t = 0$ to $t = 1$.

We assume that $\nabla_x c(x, t_1) = \nabla_x c(x, t_2)$ and $\nabla_{xx} f(x, t_1) = \nabla_{xx} f(x, t_2)$ for all t_1, t_2 as well as $\nabla_{xt} f(x_1, t) = \nabla_{xt} f(x_2, t)$ and $\nabla_t c(x_1, t) = \nabla_t c(x_2, t)$ for any two x_1 and x_2. In particular, these conditions imply that $f(x, t)$ and $c(x, t)$ are of the form $f(x, t) = f_0(x) + (a_f^T x)t$ and $c(x, t) = c^0(x) + a_c t$ where $a_f \in \mathbb{R}^n$ and $a_c \in \mathbb{R}^m$. This places the optimization problem under the standard notion of *canonical perturbations* [17].

Note that the problem class in problem (1.1) is not as restrictive as it may seem, as a more generic parametric optimization problem,

$$\begin{aligned} \min_{x\in\mathbb{R}^n} \quad & \tilde{f}(x,s) \\ \text{subject to} \quad & \tilde{c}_i(x,s) = 0,\ i \in \{1, ..., \tilde{m}_e\}, \\ & \tilde{c}_i(x,s) \geq 0,\ i \in \{\tilde{m}_e + 1, ..., \tilde{m}\}, \end{aligned}$$

[†]Agent Technology Center, Department of Computer Science, Faculty of Electrical Engineering, Czech Technical University in Prague. Research supported by the European social fund within the framework of realizing the project Support of inter-sectoral mobility and quality enhancement of research teams at the Czech Technical University in Prague, CZ.1.07/2.3.00/30.0034 and the Cisco-Czech Technical University Sponsored Research Agreement project WP5 vyacheslav.kungurtsev@fel.cvut.cz

[‡]Department of Chemical Engineering, Norwegian University of Science and Technology (NTNU) jaschke@ntnu.no

where $s \in \mathbb{R}^p$ is a vector and the solution is traced from s_0 to s_f, and $\tilde{f}(x,s)$ and $\tilde{c}(x,s)$ have more arbitrary, potentially nonlinear dependence on s can be rewritten in the form (1.1) by the incorporation of another variable z, writing $\tilde{f}$ and $\tilde{c}$ as $\tilde{f}(x,z)$ and $\tilde{c}(x,z)$ and adding the equality constraint $c_{\tilde{m}+1} = z - (1-t)s_0 + ts_f = 0$.

Parametric problems such as (1.1) occur in many applications such as model predictive control [4, 13, 18], stochastic optimization [6], global [26] and bilevel optimization [2]. Fast and accurate performance is especially demanded for real-time model predictive control, for which procedures that perform with sufficient speed for linear problems are abundant [24], but the implementation for nonlinear models has generally been more challenging

In this paper we present an algorithm for parametric optimization algorithms that applies sensitivity theory of optimization problems subject to perturbations. A novel feature of this algorithm is that it will be shown that it is convergent for problems that are dual-degenerate, i.e., do not have a unique multiplier for the optimization problem at every value of the parameter. In particular we shall rely on the sensitivity results presented in [19], which shows B-differentiability of the primal solution path under a set of weak regularity conditions.

1.1. Notation. Given vectors a and b with the same dimension, the vector with ith component $a_i b_i$ is denoted by $a \cdot b$. Similarly, $\min(a,b)$ is the vector with components $\min(a_i, b_i)$. The vectors e and e_j denote, respectively, the column vector of ones and the jth column of the identity matrix I. The dimensions of e, e_i and I are defined by the context. The ith component of a vector labeled with a subscript will be denoted by $[v_\alpha]_i$. Similarly, if $\mathcal{K}$ is an index set, $v_\mathcal{K}$ and $[v_\alpha]_\mathcal{K}$ indicate the vector with $|\mathcal{K}|$ components composed of the entries of v and v_α, respectively, corresponding to those indices in $\mathcal{K}$. If there exists a positive constant γ such that $\|\alpha_j\| \leq \gamma\beta_j$, we write $\alpha_j = O(\beta_j)$. If there exists a sequence $\gamma_j \to 0$ such that $\|\alpha_j\| \leq \gamma_j\beta_j$, we say that $\alpha_j = o(\beta_j)$.

1.2. Background. We recall the definition of the first-order optimality conditions,

DEFINITION 1.1 (First-order necessary conditions). *A vector $x^* \in \mathbb{R}^n$ satisfies the* first-order necessary optimality conditions *for* (1.1) *at t if there exists a $y^* \in \mathbb{R}^m$ such that,*

$$\begin{aligned} \nabla_x f(x^*,t) &= \nabla_x c(x^*,t) y^*, \\ c_i(x^*,t) &= 0,\ i \in \mathcal{E} \\ c_i(x^*,t) &\geq 0,\ i \in \mathcal{I} \\ c(x^*,t)^T y^* &= 0, \\ y_i^* &\geq 0,\ i \in \mathcal{I}. \end{aligned} \tag{1.2}$$

We denote $\Lambda(x^*,t)$ as the set of dual vectors y^* corresponding to x^* such that (x^*,y^*) satisfy the first order necessary conditions at t.

Denoting the normal cone $\mathcal{N}(y)$ to be,

$$\mathcal{N}(y) = \begin{cases} \{x | x \geq 0 \text{ and } x^T y = 0\} & \text{if } y \geq 0, \\ \emptyset & \text{otherwise} \end{cases}$$

an alternative formulation of (1.2) is given as,

$$\begin{aligned} \nabla_x f(x^*,t) &= \nabla_x c(x^*,t) y^*, \\ c_\mathcal{E}(x^*,t) &= 0, \\ c_\mathcal{I}(x^*,t) &\in \mathcal{N}(y_\mathcal{I}). \end{aligned}$$

These conditions necessarily hold for any local minimizer x^* which satisfies a constraint qualification, a geometric regularity condition involving the local properties of the feasible region. There are a number of constraint qualifications of varying restrictiveness, a few of which we will mention later in this section.

We will define $\mathcal{A}(x^*,t)$ to be the set of inequality constraint indices $i \in \mathcal{I}$ such that for $i \in \mathcal{A}(x^*,t)$, $c_i(x^*,t) = 0$, $\mathcal{A}_0(x^*,y^*,t) \subseteq \mathcal{A}(x^*,t)$ to be the set such that $i \in \mathcal{A}_0(x^*,y^*,t)$ implies that $[y^*]_i = 0$ and $\mathcal{A}_+(x^*,y^*,t) \subseteq \mathcal{A}(x^*,t)$ to be the set such that $i \in \mathcal{A}_+(x^*,y^*,t)$ implies that $[y^*]_i > 0$. We define $\mathcal{A}_+(x^*,t) = \cup_{y^* \in \Lambda(x^*,t)} \mathcal{A}_+(x^*,y^*,t)$ and $\mathcal{A}_0(x^*,t) = \cap_{y^* \in \Lambda(x^*,t)} \mathcal{A}_0(x^*,y^*,t)$.

The Lagrangian function associated with (1.1) is $L(x,y,t) = f(x,t) - c(x,t)^T y$. The Hessian of the Lagrangian with respect to x is denoted by

$$H(x,y,t) := \nabla^2_{xx} f(x,t) - \sum_{i=1}^{m} y_i \nabla^2_{xx} c_i(x,t).$$

The strong form of the second-order sufficiency condition is defined as follows.

DEFINITION 1.2 (Strong second-order sufficient conditions (SSOSC)).

A primal-dual pair (x^*,y^*) *satisfies the* strong second-order sufficient optimality conditions *at* t *if it satisfies the first-order conditions* (1.2) *and*

$$d^T H(x^*,y^*,t) d > 0 \text{ for all } d \in \tilde{\mathcal{C}}(x^*,y^*,t) \setminus \{0\}, \tag{1.3}$$

where $d \in \tilde{\mathcal{C}}(x^*,y^*,t)$ *if* $\nabla_x c_i(x^*,t)^T d = 0$ *for* $i \in \mathcal{A}_+(x^*,y^*,t) \cup \mathcal{E}$.

The generalized SSOSC is defined as,

DEFINITION 1.3 (Generalized SSOSC (GSSOSC)). *A primal vector* x^* *satisfies the* generalized strong second-order sufficient optimality conditions *at* t *if* (x^*,y^*) *satisfies the SSOSC for all* $y^* \in \Lambda(x^*,t)$.

We shall now define a few constraint qualifications relevant for this paper.

DEFINITION 1.4. *The* Linear Independence Constraint Qualification (LICQ) *holds for* (1.1) *at* t *for a feasible point* x *if the set of vectors* $\{\nabla_x c_i(x,t)\}_{i \in \mathcal{E} \cup \mathcal{A}(x,t)}$ *is linearly independent.*

DEFINITION 1.5. *The* Mangasarian-Fromovitz Constraint Qualification (MFCQ) *holds for* (1.1) *at* t *for a feasible point* x *if,*

1. $\{\nabla_x c_i(x,t)\}_{i \in \mathcal{E}}$ *is linearly independent, and*
2. *There exists a* p *such that* $\nabla_x c_i(x,t)^T p > 0$ *for all* $i \in \mathcal{A}(x,t)$.

Equivalently, by the theorem of the alternative [22], the MFCQ holds if there is no set of scalars $\{\alpha_i\}_{i \in \{1,\ldots,m\}}$ *such that*

1. *For* $i \in \mathcal{I}$, $\alpha_i \geq 0$,
2. *Either there exists* $i \in \mathcal{E}$ *such that* $\alpha_i \neq 0$ *or* $\sum_{i \in \mathcal{I}} \alpha_i > 0$ *and,*
3. $$\sum_{i \in \mathcal{E} \cup \mathcal{A}(x,t)} \alpha_i \nabla_x c_i(x,t) = 0.$$

DEFINITION 1.6. *The* Constant Rank Constraint Qualification (CRCQ) *holds for* (1.1) *at* t *for a feasible point* x *if there exists a neighborhood* $\mathcal{N}$ *of* x *such that for all subsets* $\mathcal{U} \subseteq \mathcal{E} \cup \mathcal{A}(x,t)$, *the rank of* $\{\nabla_x c_i(x,t)\}_{i \in \mathcal{U}}$ *is equal to the rank of* $\{\nabla_x c_i(\bar{x},t)\}_{i \in \mathcal{U}}$ *for all* $\bar{x} \in \mathcal{N}$.

Next, we need the notion of an outer graphical derivative and proto-differentiability [21].

DEFINITION 1.7. *For a multifunction $S : \mathbb{R}^p \rightrightarrows \mathbb{R}^d$, the* outer graphical derivative *of S at $\bar{w}$ for $\bar{v} \in S(\bar{w})$ is denoted by $DS(\bar{w}|\bar{v}) : \mathbb{R}^p \rightrightarrows \mathbb{R}^d$ and defined as,*

$$DS(\bar{w}|\bar{v})w' = \{v' | \exists w'_\nu \to w', \tau_\nu \downarrow 0 \text{ with } (\bar{v}_\nu - \bar{v})/\tau_\nu \to v' \text{ for some } \bar{v}_\nu \in S(\bar{w} + \tau_\nu w'_\nu)\}.$$

S is said to be proto-differentiable *at $\bar{w}$ for $\bar{v}$ if every vector $(w', v') \in graph(DS(\bar{w}|\bar{v})w')$ is equal to the following limit,*

$$(w', v') = \lim_{s \downarrow 0} \frac{(w(s), v(s)) - (\bar{w}, \bar{v})}{s}, \tag{1.4}$$

for some selection mapping $s \to (w(s), v(s)) : [0, \epsilon] \to graph(S)$ for some small ϵ.

In particular, we will be interested in the multi-function corresponding to a local primal-dual solution set for the NLP subject to a parameter,

$$K(t) := \{x^*(t), \Lambda(x^*(t), t)\}.$$

We define the distance of a point to the nearest primal-dual solution by $\delta(x, y, t)$,

$$\delta(x, y, t) = \sqrt{\|x - x^*(t)\|^2 + \text{dist}(y, \Lambda(x^*(t), t))^2},$$

where $x^*(t)$ is the closest primal solution to (1.1) at t. We will also sometimes write $\delta(x, t)$ to denote $\delta(x, t) = \|x - x^*(t)\|$.

The optimality residual $\eta(x, y, t)$ is defined as

$$\eta(x, y, t) = \left\| \begin{pmatrix} \nabla_x f(x, t) - \nabla_x c(x, t) y \\ c(x, t)_{\mathcal{E}} \\ [\min(c(x, t), y)]_{\mathcal{I}} \end{pmatrix} \right\|_\infty .$$

2. Background.

2.1. Previous results on parametric optimization and our contribution. In this section we review the literature on parametric optimization, highlighting several important features that have proven crucial for the development of fast and reliable algorithms.

A general pathfollowing procedure for nonlinear parametric equations includes a predictor and a corrector, where a predictor uses the tangent to the solution path to estimate the solution at a subsequent parameter value, and a corrector modifies the predictor step by incorporating additional information to take a step closer to the solution path (for an overview, see [1]). Predictor steps are first-order approximations and corrector steps use some form of Newton iteration. See, for instance [3, Chapter 5] for a discussion of Newton's method for continuation methods. Of course, optimization presents a generalized equation, instead of a standard nonlinear equation, and thus extensions of standard pathfollowing concepts need to be developed.

The seminal book [10] considered the parametric optimization problem in considerable detail. The authors classified all of the possible forms of primal or dual bifurcation of the parametric optimization problem, and the degeneracy conditions that are associated with these bifurcations. Based on this classification and the sensitivity theory available at the time, the authors formulated plausible procedures for tracing the solution path along a homotopy. New sensitivity theory has been developed after this book appeared, however, particularly for degenerate problems (see, e.g., [12]).

New results on the sensitivity of nonlinear programs subject to a parameter have motivated some predictor methods [13, 28]. However, the directional derivative (predictor) is just a first-order

estimate. If the problem is highly nonlinear, and there is notable curvature in the solution path, over the iterations, the path can slowly diverge from the true solution path.

Sequential Quadratic Programming (SQP) methods, wherein an nonlinear program (NLP) is solved by solving a sequence of quadratic programs approximating it (QP), were presented as a reasonable choice for solving parametric problems [4] because of their desirable warm-start properties. For an overview of SQP methods, see [9]. In particular, one solves the problem at each parameter with a few steps of an SQP algorithm, using the solution at the previous value of the parameter as an initial guess for the subsequent one. With a correctly estimated active set, an SQP iteration can be equivalent to a Newton step, and thus corresponds precisely to a corrector.

One major difficulty with practical implementation of SQP methods, however, is that whereas at the solution satisfying the SSOSC, the reduced Hessian (Lagrangian Hessian in the particular subspace) is positive definite, the full Hessian may not be, and the resulting subproblems may be nonconvex, with possibly multiple or unbounded solutions (for a detailed discussion of these challenges, see [14, Chapter 5]). There are a number of strategies devised to deal with these issues. but they typically require the use of inexact Hessian information, and thus lose the Newton/corrector-like behavior of the algorithm and as such the major potential benefit of using SQP for parametric optimization.

In addition, a strongly desirable property of a parametric optimization algorithm, in particular one applicable to optimal control problems is the capacity to handle degeneracy. In this paper we will assume that the MFCQ and the CRCQ hold for every t across the solution path, but not necessarily the LICQ. This conditions are typical for optimal control problems [23] because the dynamics of the system appear discretized as equality constraints in the optimization problem, and aside from certain classes of singular ODEs/DAEs, these are expected to have linearly independent gradients if they are well formulated. However, the controls and states are often subject to bound constraints or simple linear constraints, and when many of these constraints are active, the entire set of equality plus active bound constraints becomes overdetermined, and the set gradients corresponding to all active constraints become linearly dependent. However, bound constraints are "nice" in the sense of providing a strictly feasible direction (to satisfy the MFCQ) and are constant, and thus trivially have constant rank gradients across the entire primal space (thus satisfy the CRCQ).

The literature on proofs of convergence of parametric optimization algorithms, with the one exception of [15], has always assumed strong regularity, which requires linear independence of the constraint gradients at the solution, or at least the uniqueness of the optimal multiplier at the solution for every parameter [5, 27]. The paper [15] considers the case of general degeneracy, including problems satisfying no constraint qualification (but the existence of a Lagrange multiplier), but the procedure suggested requires solving a number of LPs to find possible multipliers to branch from, and thus relies on too weak assumptions for our purposes.

The development of an appropriate algorithm that follows the predictor-corrector framework in a reliable and efficient way is challenging. The predictor has to use the appropriate sensitivity results, the corrector needs to use a well-defined Newton iteration despite dual degeneracy, and there must be a procedure to handle possible discontinuous jumps along the optimal multiplier across a change in the parameter wherein there is a change in the optimal active set. The algorithm presented here satisfies all of these features.

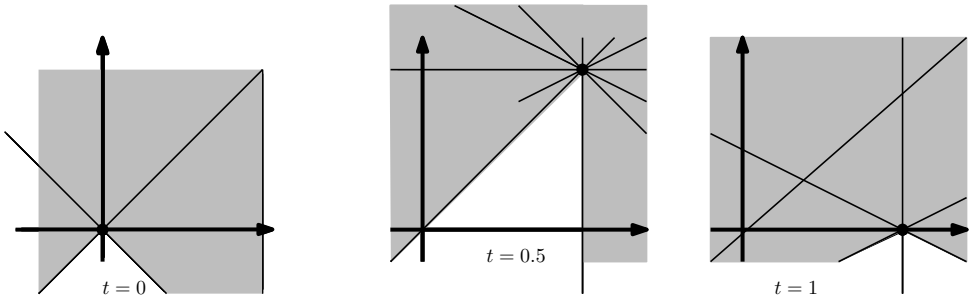

FIG. 1. *Problem* (2.1) *illustrated. The axes are for* x_1 *and* x_2*, the straight lines are the constraints, the black dot is the optimal point, and the grey area indicates the infeasible region.*

2.2. Illustrative Example. We present the following problem,

$$\begin{array}{rrl} \min_x & -e^{x_2} + \frac{1}{2}(x_1 - x_3)^2 & \\ \text{subject to} & x_3 - 10t & = 0 \\ & x_1 - x_2 & \geq 0 \\ & 10t - x_2 & \geq 0 \\ & -x_1 - x_2 + 20t & \geq 0 \\ & 5 - x_1 & \geq 0 \\ & \frac{1}{2}x_1 - x_2 + \frac{25}{2} - 10t & \geq 0 \\ & -\frac{1}{2}x_1 - x_2 + \frac{15}{2} - 10t & \geq 0 \end{array} \tag{2.1}$$

It can be verified that the solution is $x^*(t) = (10t, 10t, 10t)$ for $t \in [0, \frac{1}{2}]$ and $x^*(p) = (5, 10-10t, 10t)$ for $t \in [5, 10]$. The MFCQ, but not the LICQ holds for all $x^*(t)$, $t \in [0, 1]$ (and indeed for all t). Since all of the constraints are linear, the CRCQ also holds. The constraint gradients with respect to t are,

$$\nabla_t c(x,t) = \begin{pmatrix} 10 \\ 0 \\ 10 \\ 20 \\ 0 \\ -10 \\ -10 \end{pmatrix}$$

The first three inequality constraints are active for $t \in [0, \frac{1}{2})$, all constraints are active at $t = 0.5$ and the last three constraints are active for $t \in (\frac{1}{2}, 1]$. Noting that x_3 is just a placeholder for nonlinear dependence of the objective function with respect to the parameter, i.e., setting $x_3 = 10t$, we illustrate the problem with respect to x_1 and x_2 in Figure 1 for $t = 0, \frac{1}{2}, 1$, noting by a circular dot where the primal solution is, the lines the constraints, and the grey area corresponds to the infeasible region.

The set of optimal multipliers are,

$$\Lambda(x^*(0)) = \{(0, y_1, y_2, y_1, 0, 0, 0),\ 2y_1 + y_2 = 1,\ y_1, y_2 \geq 0\},$$
$$\Lambda(x^*(0.5)) = \{(0, y_1, y_2, y_3, y_4, y_5, y_6),\ y_1 + y_2 + y_3 + y_5 + y_6 = e^5,\ y_1 - y_3 - y_4 + \tfrac{1}{2}y_5 - \tfrac{1}{2}y_6 = 0,$$
$$y_i \geq 0,\ \text{for all } i\}$$
$$\Lambda(x^*(1)) = \{(5, 0, 0, 0, y_1, y_2, y_3),\ -2y_1 + y_2 - y_3 = -5,\ y_2 + y_3 = 1\}.$$

Notice that it holds that the last three multipliers are always zero for $t \in [0, \frac{1}{2})$, then at least two components are strictly positive for $t \in (\frac{1}{2}, 1]$. Therefore, it is desired for a parametric optimization algorithm to be able to formulate a multiplier that jumps discontinuously across $t < 0.5$ to $t > 0.5$.

3. Algorithm.

3.1. Overview. In this section we describe our new algorithm. To illustrate the procedure, we recall the notions of a pathfollowing as given in, for example, [1], and consider Figure 2. Let the current point be x_k, the solution at t_k and $t_k + \Delta t$ be $x^*(t_k)$ and $x^*(t_k + \Delta t)$, respectively. We wish to take a step from x_k that approximates $x^*(t_k + \Delta t)$. A predictor step uses information about the tangent of the solution path, using the slightly inaccurate information at x_k (as in, one can take an approximation of the tangent at $x^*(t_k)$ using problem information at x_k). Taking a pure predictor step would result in moving to x_3. A pure corrector takes a step towards a more accurate solution at a given point, which corresponds to x_2. Combining the two, if we do not re-evaluate the function or its derivatives at x_2 or x_3, a predictor-corrector would result in a step to x_4.

Alternatively, one can re-evaluate the problem functions at x_2 or x_3 yielding a more accurate pathfollowing procedure at the expense of additional computational cost. In our algorithm, however, we take a corrector step, then, without performing any additional function evaluations, use the corrector step and updated function estimates arising from linearizations, to generate a slightly more accurate predictor, and so obtain the estimate x_5 as the approximation to $x^*(t_k + \Delta t)$.

3.2. Algorithm Description. Let γ be a constant satisfying $0 < \gamma < 1$ and an estimate of the active set be [7],

$$\mathcal{A}_\gamma(x, y, t) = \{i \in \mathcal{I} : c_i(x, t) \leq \eta(x, y, t)^\gamma\}, \qquad (3.1)$$

We begin each iteration with a point x_k and a multiplier y_k such that $\{\nabla_x c_i(x_k)\}_{\mathcal{E} \cup \{i \in \mathcal{I} : y_i > 0\}}$ is linearly independent. For the initial point, we may solve the problem approximately using a globalized NLP solver, then use the procedure outlined at the end of this section to obtain this y_k. We then form an estimate of the active set A and strongly active set A_+ using (3.1) and $A_+ = \{i : y_i > 0\} \cup \mathcal{E}$.

Given a step Δt, the algorithm, at a given (x_k, y_k, t) solves the linear system to obtain the following corrector step,

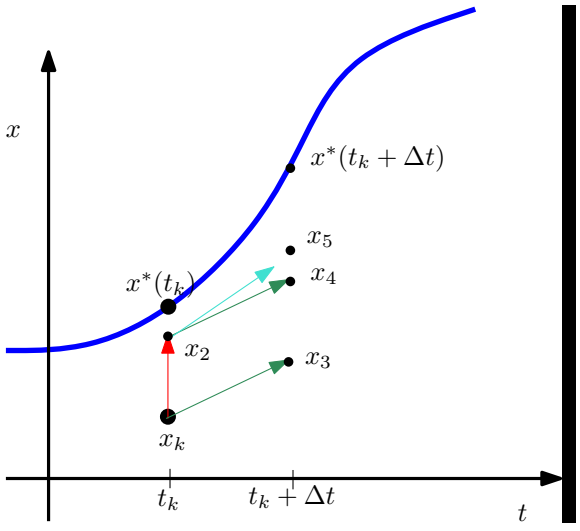

FIG. 2. *Predictor, corrector, predictor-corrector, and corrected predictor steps*

$$\begin{pmatrix} H(x_k,y_k,t) & -\nabla_x c_{A_{+,k}}(x_k,t) \\ \nabla_x c_{A_{+,k}}(x_k,t)^T & 0 \end{pmatrix} \begin{pmatrix} \Delta_c x \\ \Delta_+ y \end{pmatrix} = - \begin{pmatrix} \nabla_x f(x_k,t) - \nabla_x c(x_k,t) y_k \\ c_{A_{+,k}}(x_k,t) \end{pmatrix} \blacksquare$$

(**CorrectStep**)

to obtain $(\Delta_c x, \Delta_+ y)$. If there exists an $i \notin \mathcal{E}$ such that $[y_{A_{+,k}} + \Delta_+ y]_i < 0$ then this suggests that we are not sufficiently close to a primal-dual solution, and so we revert to an external independent globalized optimization software to find a new primal-dual point (x_k, y_k) closer to $(x^*(t), \Lambda(x^*(t)))$ and satisfying the needed linear independence conditions. Otherwise, we let $\Delta_c y \in \mathbb{R}^m$ be such that $[\Delta_c y]_{A_{+,k}} = \Delta_+ y$ and $[\Delta_c y]_{\{1,\ldots,m\} \setminus A_{+,k}} = 0$.

Given the correct strongly active set and a good starting point, at a given t, it should hold that, $(\|x_k + \Delta_c x - x^*(t)\| + \mathrm{dist}(y_k + \Delta_c y, \Lambda(x^*(t),t))) \leq C(\|x_k - x^*(t)\| + \mathrm{dist}(y_k, \Lambda(x^*(t),t)))^2$, and the corrector step results in an iterate that is closer to the primal-dual solution set.

We then solve the corrected predictor QP subproblem,

$$\begin{array}{rl} \min_{\Delta_p x} & (\nabla_x f(x_k, t+\Delta t) - \nabla_x f(x_k,t))^T \Delta_p x + \frac{1}{2} \Delta_p x^T H(x_k, y_k, t+\Delta t) \Delta_p x, \\ \text{subject to} & \nabla_t c_i(x_k,t)\Delta t + (\nabla_x c_i(x_k, t+\Delta t) + \nabla^2_{xx} c_i(x_k, t+\Delta t)\Delta_c x)^T \Delta_p x = 0, \\ & \qquad\qquad i \in A_{+,k}, \\ & \nabla_t c_i(x_k,t)\Delta t + (\nabla_x c_i(x_k, t+\Delta t) + \nabla^2_{xx} c_i(x_k, t+\Delta t)\Delta_c x)^T \Delta_p x \geq 0, \\ & \qquad\qquad i \in A_k \setminus A_{+,k} \end{array}$$

(**QPPredict**)

Denote the primal-dual solution of this subproblem as $(\Delta_p x, \Delta_p y)$. Note that by the properties of f and c as a function of t, no new function evaluations need to be performed between solving (**CorrectStep**) and (**QPPredict**).

Let $(\Delta x, \Delta y) = (\Delta_c x + \Delta_p x, \Delta_c y + \Delta_p y)$. We then check if any new constraints have become violated by too high a tolerance, or that the step was not a sufficiently accurate estimate of $(x^*(t+\Delta t), \Lambda(x^*(t+\Delta t), t+\Delta t))$. Specifically, if,

$$\eta(x_k + \Delta x, y_k + \Delta y, t + \Delta t) > \max(\eta(x_k, y_k, t), \eta(x_k, y_k, t+\Delta t))^{1+\gamma}. \tag{3.2}$$

we decrease Δt and solve (**QPPredict**) again. Otherwise, we obtain a new estimate for the active set by $A_{k+1} = \mathcal{A}_\gamma(x_k + \Delta x, y_k + \Delta y, t + \Delta t) \cup \mathcal{E}$.

In effect, the corrector QP (**CorrectStep**) produces an iterate closer to the primal-dual solution set at the given t, and without the additional term involving $\nabla^2_{xx} c_i(x_k, t+\Delta t)\Delta_c x$, (**QPPredict**) would be an estimate for the tangent of the pathfollowing solution curve, with the estimate derived from information at (x_k, y_k). Including this extra term uses the information from the corrector, without requiring any new function evaluations, to improve upon and generate a more accurate prediction.

We next solve for the new multiplier,

$$\begin{array}{rl} \min_y & y^T \nabla_t c(x + \Delta x, t + \Delta t)\Delta t \\ \text{subject to} & -|\nabla L(x_k + \Delta x, y_k + \Delta y, t + \Delta t)| \\ & \quad \leq \nabla_x f(x_k + \Delta x, t + \Delta t) + \sum_{i \in A_{k+1}} \nabla_x c_i(x_k + \Delta x, t + \Delta t) y_i \\ & \quad \leq |\nabla L(x_k + \Delta x, y_k + \Delta y, t + \Delta t)| \\ & y_{\mathcal{I}} \geq 0, \\ & y_{i \notin A_{k+1}} = 0, \end{array} \tag{LPPC}$$

where the absolute value is performed component-wise. We define the solution of this LP as $y_{k+1}^{(0)}$.

Finally we need to generate a new multiplier where the constraint gradients corresponding to the positive components are linearly independent, in order for the system defining (**QPPredict**) at the subsequent iteration to be nonsingular. We perform the procedure outlined on page 493 of [25]. In particular, if $\tilde{A}_+ = \{i : [y_{k+1}^{(0)}]_i > 0\} \cup \mathcal{E}$ is such that $\sum_{i\in\tilde{A}_+} \nabla_x c_i(x_k + \Delta x, t + \Delta t)w = 0$, then we know from dual form of the MFCQ, in Definition 1.5, that there is a w that has a component $j \in \mathcal{I}$ such that $w_j < 0$. We know that

1. $\sum_{i\in\tilde{A}_+} \nabla_x c(x_k + \Delta x, t + \Delta t)(y_i + \alpha w_i)$ is constant with respect to α for any α with $\alpha \geq 0$,
2. $[y_{k+1}^{(0)}]_i + \alpha w_i \geq 0$ for $i \in \tilde{A}_+ \cap \mathcal{I}$ for α sufficiently small, and
3. $([y_{k+1}^{(0)}]_{i\in\tilde{A}_+} + \alpha w)^T[\nabla_t c(x + \Delta x, t + \Delta t)\Delta t]_{i\in\tilde{A}_+}$ is also constant for any $\alpha \geq 0$, because otherwise (**LPPC**) would be able to find a solution with a lower objective, due to the two properties above suggesting feasibility of $[y_{k+1}^{(0)}]_i + \alpha w_i$ for (**LPPC**).

We let $[y_{k+1}^{(0)}]_{i\in\tilde{A}_+} = [y_{k+1}^{(0)}]_{i\in\tilde{A}_+} + \hat{\alpha}w$ with $\hat{\alpha} = \max\{\alpha : [y_{k+1}^{(0)}]_i + \hat{\alpha}w_i \geq 0 \text{ for } i \in \tilde{A}_+ \cap \mathcal{I}\}$. We then repeat the procedure until we find a satisfactory $y_{k+1}^{(0)}$ (i.e., satisfying the linearly independent gradients condition), which we now set as y_{k+1}.

We redefine $A_{+,k} = \{i : [y_{k+1}]_i > 0\} \cup \mathcal{E}$. We then set $x_{k+1} = x_k + \Delta x$, and iterate $k = k + 1$.

The procedure is summarized as Algorithm 1

4. Convergence of the predictor-corrector path-following algorithm.

4.1. Preliminaries. We shall use the following results throughout the convergence theory.

LEMMA 4.1. *If the SSOSC and the MFCQ hold at (x^*, y^*), then for (x, y) sufficiently close to (x^*, y^*), there exist constants $C_1(t) > 0$ and $C_2(t) > 0$ such that it holds that,*

$$C_1(t)\delta(x, y, t) \leq \eta(x, y, t) \leq C_2(t)\delta(x, y, t).$$

Proof. See, e.g., Wright [25, Theorem A.1]. □

The next lemma verifies that the active set estimate is accurate.

LEMMA 4.2. *[7, Theorem 3.7] For all x, y such that $\delta(x, y, t)$ is sufficiently small, $\mathcal{A}_\gamma(x, y, t) = \mathcal{A}(x^*(t))$.*

LEMMA 4.3. *[11, Lemma 3] Given matrices Q^* and P^*, where Q^* is symmetric, suppose that,*

$$w^T Q^* w \geq \alpha\|w\|^2, \text{ whenever } P^* w = 0,\ w \in \mathbb{R}^n.$$

Then given any $\delta > 0$, there exists $\sigma > 0$ and neighborhoods $\mathcal{P}$ of P^* and $\mathcal{Q}$ of Q^* such that,

$$v^T(Q + \frac{1}{\rho}P^T P)v \geq (\alpha - \delta)\|v\|^2,$$

for all $v \in \mathbb{R}^n$, $0 < \rho \leq \sigma$, $P \in \mathcal{P}$, and $Q \in \mathcal{Q}$.

Algorithm 1 Predictor Corrector Sensitivity Pathfollowing Method

Input: t, x, y close to $(x^*(t), \hat{Y}(t))$ such that $\{\nabla_x c_i(x,t)\}_{\{i\in\mathcal{I}:y_i>0\}\cup\mathcal{E}}$ is linearly independent.

1: Define parameters γ satisfying $0 < \gamma < 1$ and $\iota_{\max} \in \mathbb{N}$.
2: Estimate A using (3.1) evaluated at $(x, y, 0)$ and define A_+
3: Set $\iota = 1$
4: **while** $t < 1$ **do**
5: Solve (**CorrectStep**) for $(\Delta_c x, \Delta_+ y)$.
6: **if** For some $i \notin \mathcal{E}$, $[y_{A_+} + \Delta_+ y]_i < 0$ **then**
7: Solve (1.1) at t for an approximate solution (x, y).
8: Set $\iota = 1$.
9: **end if**
10: Define $\Delta_c y \in \mathbb{R}^m$ such that $[\Delta_c y]_{A_+} = \Delta_+ y$ and zero otherwise.
11: Solve (**QPPredict**) for $(\Delta_p x, \Delta_p y)$.
12: Set $(\Delta x, \Delta y) = (\Delta_p x, \Delta_p y) + (\Delta_c x, \Delta_c y)$
13: **if** (3.2) does not hold **then**
14: **if** $\iota \geq \iota_{\max}$ **then**
15: Solve (1.1) at t for an approximate solution (x, y).
16: Set $\iota = 1$.
17: Go to Line 5
18: **else**
19: Decrease Δt. Set $\iota = \iota + 1$.
20: Go to Line 11
21: **end if**
22: **end if**
23: Compute $A = \mathcal{A}_\gamma(x + \Delta x, y + \Delta y, t + \Delta t) \cup \mathcal{E}$.
24: Solve (**LPPC**) for $y^{(0)}$.
25: Define $\tilde{A}_+ = \{i : [y^{(0)}]_i > 0\} \cup \mathcal{E}$
26: **while** $\{\nabla_x c_i(x + \Delta x, t + \Delta t)\}_{i\in\tilde{A}_+}$ has rank less than $|\tilde{A}_+|$ **do**
27: Solve for w satisfying $\sum_{i\in\tilde{A}_+} \nabla_x c_i(x_k + \Delta x, t + \Delta t)w = 0$ and $\exists i$, $w_i < 0$.
28: Set $\hat{\alpha} = \max\{\alpha : [y^{(0)}]_{i\in\tilde{A}_+\cap\mathcal{I}} + \alpha w \geq 0\}$.
29: Let $[y^{(0)}]_{i\in\tilde{A}_+} = [y^{(0)}]_{i\in\tilde{A}_+} + \hat{\alpha} w$
30: **end while**
31: Set $y = y^{(0)}$.
32: Let $A_+ = \{i : [y]_i > 0\} \cup \mathcal{E}$.
33: Set $t = t + \Delta t$, $x = x + \Delta x$.
34: **end while**

The next Theorem summarizes the sensitivity results that hold under the MFCQ and the CRCQ that are the theoretical foundation for the predictor-corrector algorithm we have formulated.

THEOREM 4.4. *Let f and c be twice continuously differentiable in t and x near $(x^*(t_0), t_0)$, and let the MFCQ and the SSOSC hold at $x^*(t_0)$.*

1. *The solution $x^*(t)$ is continuous in a neighborhood of $x^*(t_0)$ and the solution function $x^*(t)$ is directionally differentiable, i.e.,*

$$x^*(t_0 + \Delta t) = x^*(t_0) + \delta x^*(x^*(t_0), t_0, \Delta t) + o(|\Delta t|), \tag{4.1}$$

where $\delta x^*(x^*(t_0), t_0, \Delta t)$ is the directional derivative of $x^*(t)$ with respect to t at t_0 scaled by Δt, i.e.,

$$\delta x^*(x^*(t_0), t_0, \Delta t) = \lim_{\epsilon \downarrow 0} \frac{x^*(t_0 + \epsilon) - x^*(t_0)}{\epsilon} \Delta t.$$

2. *Moreover, for each t in a neighborhood of t_0 and direction Δt, there exists a multiplier $\hat{y} \in \Lambda(x^*, t_0)$ such that the directional derivative uniquely solves the following convex quadratic program,*

$$\begin{array}{rl} \min_{\Delta x} & \Delta x^T \nabla_{xt} L(x^*(t_0), \hat{y}, t_0) \Delta t + \frac{1}{2} \Delta x^T H(x^*(t_0), \hat{y}, t_0) \Delta x, \\ \textit{subject to} & \nabla_x c_i(x^*(t_0), t_0)^T \Delta x + \nabla_t c_i(x^*(t_0), t_0) \Delta t = 0,\ i \in \mathcal{A}_+(x^*(t_0), \hat{y}, t_0) \cup \mathcal{E}, \\ & \nabla_x c_i(x^*(t_0), t_0)^T \Delta x + \nabla_t c_i(x^*(t_0), t_0) \Delta t \geq 0,\ i \in \mathcal{A}_0(x^*(t_0), \hat{y}, t_0). \end{array} \quad (\textbf{SensitivityQP})$$

   We denote the solution set of this program as $(\delta x^*(x^*(t_0), t_0, \Delta t), \delta Y^*(x^*(t_0), \hat{y}, t_0, \Delta t))$ (where $\delta x^*(x^*(t_0), t_0, \Delta t)$ is a singleton). Note that $\delta x^*(x^*(t_0), t_0, \Delta t)$ does not depend on $\hat{y}$ (if there are multiple $\hat{y}$ satisfying the conditions of this part of the Theorem), but $\delta Y^*(x^*(t_0), \hat{y}, t_0, \Delta t)$ may.
3. *If, in addition, the CRCQ holds, then the multiplier values $\hat{y}$ at which the QP* (**SensitivityQP**) *must be evaluated can be found as a solution of the following linear program,*

$$\begin{array}{rl} \min_y & y^T \nabla_t c(x^*(t_0), t_0) \Delta t, \\ \textit{subject to} & \nabla_x f(x^*(t_0)) - \nabla_x c(x^*(t_0), t_0) y = 0, \\ & y_{\mathcal{I}} \geq 0, \\ & c_i(x^*(t_0), t_0) y_i \, \forall i \in \{1, ..., m\}. \end{array} \quad (\textbf{SensitivityLP})$$

   We denote its solution set as $\hat{Y}(x^*(t_0), t_0)$. Note that this set is independent of Δt as long as $\Delta t > 0$.
4. *Moreover, the set $\{\nabla_x c_i(x^*(t_0), t_0)\}_{i \in \mathcal{A}_+(x^*(t_0), \hat{y}, t_0) \cup \mathcal{E}}$ is linearly independent for some $\hat{y} \in \hat{Y}(x^*(t_0), t_0)$.*
5. *The primal-dual solution set $K(t_0)$ is proto-differentiable at $(x^*(t_0), \hat{y})$ for any $\hat{y} \in \hat{Y}(x^*(t_0), t_0)$, and the outer graphical derivative is the solution set of* (**SensitivityQP**)*, i.e.,*

$$DK(t_0 | x^*(t_0), \hat{y}) \Delta t = (\delta x^*(x^*(t_0), t_0, \Delta t), \delta Y^*(x^*(t_0), \hat{y}, t_0, \Delta t))$$

   This implies, in particular, that, for any $\delta y^* \in \delta Y^*(x^*(t_0), \hat{y}, t_0, \Delta t)$

$$dist(\hat{y} + \delta y^*, \Lambda(x^*(t_0 + \Delta t), t_0 + \Delta t)) = o(|\Delta t|). \quad (4.2)$$

Proof. Parts 1-3 appear as [13, Theorem 5] and [19, Theorems 1-2].

Part 4 follows from the proof of [16, Theorem 2.2]

Part 5 follows from [17, Proposition 2.5.1], where it can be seen that [17, (2.34)] are the optimality conditions of (**SensitivityQP**), if we consider that the problem is independent of any nonlinear perturbation w, and let $v_1' = \nabla_{xt} f(x, t) \Delta t$ and $v_2' = \nabla_t c(x, t) \Delta t$. The implication (4.2) follows from the definition of proto-differentiability, e.g., in the notation of (1.4), let $v' = 1$, reparametrize

$v(s)$ to be $t_0 + \Delta t$, then since $(\delta x^*(x^*(t_0), t_0, \Delta t), \delta Y^*(x^*(t_0), \hat{y}, t_0, \Delta t))$ is the directional derivative of $K(x, y, t)$, it holds that, for any $\delta y^* \in \delta Y^*(x^*(t_0), \hat{y}, t_0, \Delta t)$, and for $\delta x^*(x^*(t_0), t_0, \Delta t)$, the corresponding selection $w(s) = (x^*(s), y^*(s))$ satisfies,

$$w(s) - w(0) = (\delta x^*(x^*(t_0), t_0, \Delta t)), \delta y^*) + \Delta t \alpha(\Delta t),$$

where $\alpha(\Delta t) \to 0$ as $\Delta t \to 0$.
□

4.2. Convergence of Algorithm 1. We make the following assumptions,

ASSUMPTION 1. *The functions $f(x,t)$ and $c(x,t)$ are two times Lipschitz continuously differentiable for all x and $t \in [0,1]$ with respect to both x and t.*

ASSUMPTION 2. *There exists a continuous primal solution path $x^*(t)$ to (1.1) for $t \in [0,1]$.* Hereafter, every result, unless otherwise noted, is with respect to a particular continuous such path $x^*(t)$.

ASSUMPTION 3. *The CRCQ, the MFCQ, and the GSSOSC hold for all $x^*(t)$, $t \in [0,1]$.*

LEMMA 4.5. *[8] $\Lambda(x^*(t), t)$ is bounded for all t.*

Note that since the KKT conditions are linear in y, this implies that $\Lambda(x^*(t), t)$ is a closed convex polytope for any given $x^*(t)$ and t.

LEMMA 4.6. *There exists a B such that for every $\hat{y} \in \hat{Y}(x^*(t), t)$ for all t, $\|\hat{y}\| \leq B$.*

Proof. Suppose there is a sequence $(\hat{y}(t_k), t_k)$ with $\hat{y}(t_k) \in \hat{Y}(x^*(t), t)$ such that $\|\hat{y}(t_k)\| \geq k$. But since t is in a compact set, there exists a convergent subsequence and a cluster point t^*. However, this implies that $\|\hat{y}(t^*)\| = \infty$, which is impossible by Lemma 4.5. □

LEMMA 4.7. *Consider the QP* (**SensitivityQP**) *evaluated at a $\hat{y}(t) \in \hat{Y}(x^*(t), t)$ satisfying Part 2 of Theorem 4.4. There exists a Δt such that for all $t \in [0,1]$, there exists a B_2 such that any multiplier $\delta y^*(t) \in \delta Y^*(x^*(t), \hat{y}(t), t)$ of the solution to* (**SensitivityQP**) *is bounded by B_2.*

Proof. It can be seen that the MFCQ holds for (**SensitivityQP**) at $(x^*(t), \hat{y}(t))$ by the definition of $\hat{y}(t)$ and the MFCQ holding for (1.1) at $x^*(t)$. Suppose that there exists (t_k) and $\delta y^*(t_k) \in \delta Y^*(x^*(t_k), \hat{y}(t), t_k)$ such that $\|\delta y^*(t_k)\| \geq k$. But this implies that there exists a cluster point t^* such that $\|\delta y(t^*)^*\| = \infty$, and this is impossible. □

We are now ready to present the main result with regards to the predictor corrector step.

THEOREM 4.8. *If $\delta(x_k, y_k, t_k)$, $\|y_k - \hat{y}\|$ and Δt are sufficiently small for some $\hat{y} \in \hat{Y}(x^*(t), t_k)$ satisfying the condition in Part 4 of Theorem 4.4, and $A_{+,k} = \mathcal{A}_+(x^*(t_k), \hat{y}, t_k) \cup \mathcal{E}$, then consider the point $(x_k + \Delta x, y_k + \Delta y)$, where $(\Delta x, \Delta y) = (\Delta_p x + \Delta_c x, \Delta_p y + \Delta_c y)$ is defined as follows:*

- *$(\Delta_c x, \Delta_+ y)$ solves* (**CorrectStep**),
- *$\Delta_c y \in \mathbb{R}^m$ is such that $[\Delta_c y]_{A_{+,k}} = \Delta_+ y$ and $[\Delta_c y]_i = 0$ for $i \in \{1, ..., m\} \setminus A_{+,k}$, and*
- *$\Delta_p x$ solves* (**QPPredict**) *with any associated dual solution $\Delta_p y$.*

This primal-dual point $(x_k + \Delta x, y_k + \Delta y)$ satisfies $\eta(x_k + \Delta x, y_k + \Delta y, t_k + \Delta t) \leq \eta(x_k, y_k, t_k)^{1+\gamma}$.

The next result states that given a primal-dual point (x_k, y_k) sufficiently close to the primal-dual solution set $(x^*(t), \Lambda(x^*(t), t)$, the solution of (**LPPC**) yields a good multiplier approximation to from which to calculate the next predictor step.

THEOREM 4.9. *For all ϵ and Δt, there exists a δ such that if dist$((x, y), (x^*(t), \Lambda(x^*(t), t))) \leq \delta$, then the solution $\bar{y}$ to*

$$\begin{array}{rl} \min_{\bar{y}} & \bar{y} \nabla_t c(x,t) \Delta t \\ \textit{subject to} & -|\nabla L(x,y,t)| \\ & \quad \leq \nabla f(x,t) + \sum_{i \in \mathcal{A}(x^*(t)) \cup \mathcal{E}} \nabla_x c_i(x,t) \bar{y}_i \\ & \quad \leq |\nabla L(x,y,t)| \end{array} \tag{4.3}$$

satisfies $dist(\bar{y}, \hat{Y}(x^*(t), t) \leq \epsilon$.

The first theorem guarantees that the primal-dual solution for the corrector-predictor problems result in a step that satisfies the conditions required by Algorithm 1 to accept the step, i.e., (3.2) holds. It implies that the primal dual point is within the ball of local superlinear contraction and the corrector step results in a point much closer to the solution of the problem for the current value of the parameter. The predictor, (**QPPredict**) is a first-order estimate of the primal-dual solution path, and by coupling it with the corrector (**CorrectStep**), we can ensure that the solution estimate tracks the solution, in the sense of being able to produce a point that is an accurate enough approximation of the appropriate primal-dual solution such that all subsequent estimates are as close as desired by some predetermined amount.

The second Theorem states that, if the active set is estimated correctly, which follows from Lemma 4.2, if the point at which the LP (**LPPC**) is evaluated sufficiently close to $(x^*(t + \Delta t), \Lambda(x^*(t + \Delta t), t + \Delta t)$ then the multiplier solution can be made arbitrarily close to the set $\hat{Y}(x^*(t + \Delta t))$, . By assumption, we can make the starting point sufficiently close to the primal dual solution $(x^*(t), \hat{y})$, so as to (possibly also by decreasing Δt) to make the solution to the predictor-corrector subproblems arbitrarily close to $(x^*(t+\Delta t), \Lambda(x^*(t+\Delta t), t+\Delta t))$. Theorem 4.9 implies that the final multiplier estimate can get within any desired distance to $\hat{Y}(x^*(t+\Delta t), t+\Delta t)$. This suggests that the new $x_k + \Delta x$ can be in the neighborhood for which the CRCQ applies at $t+\Delta t$, and the procedure used to find a multiplier in the corner of the polyhedron, i.e., one for which the constraint gradients are linearly independent, can be successfully applied to yield a primal-dual point that satisfies the conditions of Theorem 4.8 for the problem at $t + \Delta t$. The argument then repeats itself at $t + \Delta t$, and so on, until reaching $t = 1$. For any t to which we apply the argument, we can make the solution estimate at the initial $t = 0$ close enough to yield all the desired results across all the subsequent $(0, t]$ up until the current value of the parameter.

Note that all of this is in terms of $\epsilon - \delta$ reasoning, i.e., all primal-dual points can be arbitrarily close to the desired points at the subsequent iteration, but nothing stronger than that can be claimed. In particular, unless the radius of the Newton-Kantarovich ball of quadratic convergence is uniformly bounded from above and below across t, the neighborhood for which the CRCQ applies is also bounded across t, and other such conditions, finite termination without occasional reliance on a globalized algorithm cannot be ensured. However, we expect that for many well-formulated problems, it can be expected that the required conditions do hold. Since little is known about regions of applicability of constraint qualifications and local convergence for Newton methods for nonlinear programs, however, no precise statements can be made to that effect.

4.3. Proofs of results. In this section we prove the main two results of this paper.

Proof. of Theorem 4.8

The system (**CorrectStep**) is the Newton-Lagrange system for solving the NLP,

$$\begin{array}{rl} \min_x & f(x, t_k) \\ \text{subject to} & c_i(x, t_k) = 0,\ i \in \mathcal{A}_+(x^*(t_k), \hat{y}, t_k) \cup \mathcal{E}. \end{array} \tag{4.4}$$

for which $x^*(t_k)$ is a stationary solution. By the linear independence of $\{\nabla c_i(x^*(t_k), t_k)\}_{i \in \mathcal{A}_+(x^*(t_k), \hat{y}, t_k) \cup \mathcal{E}}$ (thus the MFCQ holds for (4.4)) and the SSOSC, we can invoke [20, Theorem 2.2] to conclude that the solution is isolated. Thus the Newton-Lagrange step (**CorrectStep**) is associated with local convergence to $(x^*(t_k), \hat{y})$ uniquely.

Denoting the (unique) solution to (**CorrectStep**) as $(\Delta_c x, \Delta_c y)$, we have, from the usual

Newton local convergence properties,

$$\|(x_k + \Delta_c x - x^*(t_k), y_k + \Delta_c y - \hat{y})\| = O(\|(x_k - x^*(t_k), y_k - \hat{y})\|^2). \tag{4.5}$$

and,

$$\|(\Delta_c x, \Delta_c y)\| = O(\|(x_k - x^*(t_k), y_k - \hat{y})\|). \tag{4.6}$$

By the GSSOSC, subproblem (**SensitivityQP**) is strongly convex and has a unique solution. By (4.6) and Lemma 4.3 it holds that (**QPPredict**) is also strongly convex and has a unique primal solution for $\|(x_k - x^*(t_k), y_k - \hat{y})\|$ sufficiently small.

We can now write the optimality conditions of (**QPPredict**) as, using $\nabla_x c(x, t_k) = \nabla_x c(x, t_k + \Delta t)$ and $H(x, y, t) = H(x, y, t + \Delta t)$,

$$\begin{aligned}
&(\nabla_x f(x_k, t_k + \Delta t) - \nabla_x c(x_k, t_k + \Delta t) y_k - \nabla_x f(x_k, t_k) + \nabla_x c(x_k, t_k) y_k) \\
&\qquad + H(x_k, y_k, t_k + \Delta t)\Delta_p x - (\nabla_x c(x_k, t_k) + \nabla^2_{xx} c(x_k, t_k)\Delta_c x)\Delta_p y = 0, \\
&(\nabla_x c_i(x_k, t_k + \Delta t) + \nabla^2_{xx} c_i(x_k, t_k)\Delta_c x)^T \Delta_p x + \nabla_t c_i(x_k, t_k)\Delta t = 0,\ i \in A_{+,k}, \\
&(\nabla_x c_i(x_k, t_k + \Delta t) + \nabla^2_{xx} c_i(x_k, t_k)\Delta_c x)^T \Delta_p x + \nabla_t c_i(x_k, t_k)\Delta t \in \mathcal{N}([\Delta_p y]_i),\ i \in A_k \setminus A_{+,k}
\end{aligned} \tag{4.7}$$

Now, using the properties of f and c as functions with respect to t,

$$(\nabla_x f(x_k, t_k + \Delta t) - \nabla_x c(x_k, t_k + \Delta t) y_k - \nabla_x f(x_k, t_k) + \nabla_x c(x_k, t_k) y_k) = \nabla_{xt} L(x_k, y_k, t_k)\Delta t$$

we can rewrite (4.7) as,

$$\begin{aligned}
&\nabla_{xt} L(x_k, y_k, t_k)\Delta t + H(x_k, y_k, t_k + \Delta t)\Delta_p x - (\nabla_x c(x_k, t_k) + \nabla^2_{xx} c(x_k, t_k)\Delta_c x)\Delta_p y = 0, \\
&(\nabla_x c_i(x_k, t_k) + \nabla^2_{xx} c_i(x_k, t_k)\Delta_c x)^T \Delta_p x + \nabla_t c_i(x_k, t_k)\Delta t = 0,\ i \in A_{+,k}, \\
&(\nabla_x c_i(x_k, t_k) + \nabla^2_{xx} c_i(x_k, t_k)\Delta_c x)^T \Delta_p x + \nabla_t c_i(x_k, t_k)\Delta t \in \mathcal{N}([\Delta y]_i),\ i \in A_k \setminus A_{+,k}
\end{aligned} \tag{4.8}$$

We may consider this problem as a perturbation of the optimality conditions of (**SensitivityQP**). Consider any $\overline{\delta x}$ and an associated dual $\overline{\delta y}$ solution to (4.8). We shall apply the Upper Lipschitz continuity of solutions subject to perturbations given for a QP satisfying the SSOSC and the MFCQ in [20, Theorem 4.2]. In particular, in the notation of the Theorem, for the base perturbation γ_0, $\nabla_{\delta x}\tilde{f}(\delta x; \gamma_0) = \nabla_{xt} L(x^*(t), \hat{y}, t_k)\Delta t + H(x^*(t), \hat{y}, t_k + \Delta t)\delta x$ and for the current point, considered at γ_1, $\nabla_{\delta x}\tilde{f}(\delta x; \gamma_1) = \nabla_{xt} L(x_k, y_k, t_k)\Delta t + H(x_k, y_k, t_k + \Delta t)\delta x$, and the constraint $\tilde{c}$ as $\tilde{c}(\delta x; \gamma_0) = (\nabla_x c_i(x^*(t), t_k))^T \delta x + \nabla_t c_i(x^*(t), t_k)\Delta t$ and $\tilde{c}(\delta x; \gamma_1) = (\nabla_x c_i(x_k, t_k) + \nabla^2_{xx} c_i(x_k, t_k)\Delta_c x)^T \delta x + \nabla_t c_i(x_k, t_k)\Delta t$.

We see from the conclusion of [20, Theorem 4.2] that the solution $(\overline{\delta x}, \overline{\delta y})$ to (**QPPredict**) satisfies,

$$\begin{aligned}
&\|\overline{\delta x} - \delta x^*\| + \operatorname{dist}(\overline{\delta y}, \delta Y^*(x^*(t_k), \hat{y}, t_k)) \leq \\
&\left\| \begin{pmatrix}
\nabla_{xt} L(x_k, y_k, t_k)\Delta t - \nabla_{xt} L(x^*(t_k), \hat{y}, t_k)\Delta t \\
+(H(x_k, y_k, t_k + \Delta t) - H(x^*(t_k), \hat{y}(t), t_k + \Delta t))\overline{\delta x} \\
-(\nabla_x c(x^*(t_k), t_k) - \nabla_x c(x_k, t_k) - \nabla^2_{xx} c(x_k, t_k)\Delta_c x)\overline{\delta y} \\
((\nabla_x c(x_k, t_k) + \nabla^2_{xx} c(x_k, t_k)\Delta_c x)^T - \nabla_x c(x^*(t_k), t_k)^T)\overline{\delta x} \\
+\nabla_t c(x_k, t_k)\Delta t - \nabla_t c(x^*(t_k), t_k)\Delta t
\end{pmatrix} \right\|,
\end{aligned} \tag{4.9}$$

By the fact that $\nabla_t c(x, t)$ and $\nabla_{xt} f(x, t)$ is a constant, we have that,

$$\begin{aligned}
&\nabla_{xt} L(x_k, y_k, t_k)\Delta t - \nabla_{xt} L(x^*(t_k), \hat{y}, t_k)\Delta t = 0, \text{ and} \\
&\nabla_t c(x_k, t_k)\Delta t - \nabla_t c(x^*(t_k), t_k)\Delta t = 0
\end{aligned}$$

Given the two-times Lipschitz continuity of c and f we can obtain,

$$\begin{aligned}\|(H(x_k,y_k,t+\Delta t)-H(x^*(t_k),\hat{y},t_k+\Delta t))\overline{\delta x}\| &= \big\|((\nabla^2_{xx}f(x_k,t_k+\Delta t)-\nabla^2_{xx}f(x^*(t_k),t+\Delta t)\\ &\quad -\textstyle\sum[y_k]_i(\nabla^2_{xx}c_i(x_k)-\nabla^2_{xx}c_i(x^*(t_k)))-\sum([y_k]_i-[\hat{y}]_i)\nabla^2_{xx}c_i(x^*(t_k)))\overline{\delta x}\big\|\\ &\leq ((C_L+B+\|y_k-\hat{y}\|)\|x_k-x^*(t_k)\|+B\|y_k-\hat{y}\|)\|\overline{\delta x}\|\end{aligned} \tag{4.10}$$

where C_L is an upper bound for the Lipschitz constant for function and the first and second derivatives of f and c and B is an upper bound on $\hat{y}$ by Lemma 4.6. Using Taylor's Theorem and (4.5), (4.6),

$$\begin{aligned}&\|(\nabla_x c(x^*(t_k),t_k)-\nabla_x c(x_k,t_k)-\nabla^2_{xx}c(x_k,t_k)\Delta_c x)\overline{\delta y}\|\\ &=\|(\nabla_x c(x^*(t_k),t_k)-\nabla_x c(x_k+\Delta_c x,t_k))\overline{\delta y}+O(\|\Delta_c x\|^2\|\overline{\delta y}\|)\\ &=O(\|(x_k-x^*(t_k),y_k-\hat{y})\|^2\|\overline{\delta y}\|)\end{aligned} \tag{4.11}$$

and similarly,

$$\|(\nabla_x c(x_k,t_k)^T+\nabla^2_{xx}c(x_k,t_k)\Delta_c x-\nabla_x c(x^*(t_k),t_k)^T)\overline{\delta x}\|=O(\|(x_k-x^*(t_k))\|^2\|\overline{\delta x}\|). \tag{4.12}$$

Let δy^* satisfy,

$$\|\overline{\delta y}-\delta y^*\|=\operatorname{dist}(\overline{\delta y},\delta Y^*(x^*(t_k),\hat{y},t_k)).$$

This is unique by the fact that $\delta Y^*(x^*(t_k),\hat{y},t_k)$ is a convex polyhedral set.

From applying (4.10), (4.11), and (4.12) to (4.9), and then applying the triangle inequality to write $\|\overline{\delta x}\|\leq\|\overline{\delta x}+\delta x^*\|+\|\delta x^*\|$ and $\|\overline{\delta y}\|\leq\|\overline{\delta y}+\delta y^*\|+\|\delta y^*\|$, we can deduce,

$$\begin{aligned}&\|\overline{\delta x}-\delta x^*(x^*(t_k),t_k)\|+\|\overline{\delta y}-\delta y^*\|\\ &\quad=O(\|(x_k,y_k)-(x^*(t_k),\hat{y})\|)(\|\overline{\delta x}-\delta x^*\|+\|\delta x^*\|)\\ &\qquad+O(\|(x_k,y_k)-(x^*(t_k),\hat{y})\|^2)(\|\delta x^*\|+\|\overline{\delta y}-\delta y^*\|+\|\delta y^*\|)).\end{aligned} \tag{4.13}$$

By taking $\|(x_k,y_k)-(x^*(t_k),\hat{y}(t_k))\|$ sufficiently small, we can ensure that all terms above of the form $O(\|(x_k,y_k)-(x^*(t_k),\hat{y}(t_k))\|)$ are less than one half. We can then subtract, $\frac{1}{2}\|\overline{\delta x}-\delta x^*\|+\frac{1}{2}\|\overline{\delta y}-\delta y^*\|$ from both sides of (4.13), then double both sides of the resulting equation, to get,

$$\|\overline{\delta x}-\delta x^*\|+\|\overline{\delta y}-\delta y^*\|=O(\|(x_k,y_k)-(x^*(t_k),\hat{y})\|^2)(\|\delta x^*\|+\|\delta y^*\|)+O(\|(x_k,y_k)-(x^*(t_k),\hat{y})\|)\|\delta x^*\| \tag{4.14}$$

Theorem 4.4 Part 1 and 5 implies that,

$$\|x^*(t_k)+\delta x^*(x^*(t_k),t_k)-x^*(t_k+\Delta t)\|+\|\hat{y}+\delta y^*-y^*(t_k+\Delta t)\|=o(\Delta t). \tag{4.15}$$

where $y^*(t_k+\Delta t)$ satisfies,

$$\|\hat{y}+\delta y^*-y^*(t_k+\Delta t)\|=\operatorname{dist}(\hat{y}+\delta y^*,\Lambda(x^*(t_k+\Delta t),t_k+\Delta t)).$$

Furthermore, it holds that as $\Delta t\to 0$, $\delta x^*(x^*(t_k),t_k,\Delta t)\to 0$ and by Lemma 4.7 it holds that $\|\delta y^*\|\leq B_2$. Let $E(\Delta t)=\|\delta x^*(x^*(t_k),t_k,\Delta t)\|$.

Finally, using Lemma 4.1, we get,

$$\begin{aligned}
&\eta(x_k+\Delta x, y_k+\Delta y, t+\Delta t)\\
&\le C_2(t_k+\Delta t)\left(\|x_k+\Delta_c x+\overline{\delta x}-x^*(t+\Delta t)\|+\operatorname{dist}(x_k+\Delta_c y+\overline{\delta y},\Lambda(x^*(t_k+\Delta t),t_k))\right)\\
&\le C_2(t_k+\Delta t)\left(\|x_k+\Delta_c x-x^*(t_k)\|+\|y_k+\Delta_c y-\hat{y}\|\right.\\
&\qquad +\|x^*(t_k)+\delta x^*(x^*(t_k),t_k)-x^*(t_k+\Delta t)\|+\|\hat{y}+\delta y^*-y^*(t_k+\Delta t)\|\\
&\qquad\left. +\|\overline{\delta x}-\delta x^*(x^*(t_k),t_k)\|+\|\overline{\delta y}-\delta y^*\|\right)\\
&= O(\|(x_k-x^*(t_k),y_k-\hat{y})\|^2)+o(\Delta t)+O(\|(x_k,y_k)-(x^*(t_k),\hat{y})\|^2)\\
&\qquad +O(\|(x_k,y_k)-(x^*(t_k),\hat{y})\|)E(\Delta t)\\
&= O(\|(x_k,y_k)-(x^*(t_k),\hat{y})\|^2)+o(\Delta t)+O(\|(x_k,y_k)-(x^*(t_k),\hat{y})\|)E(\Delta t)\\
&= O(\eta(x_k,y_k,t)^2)+o(\Delta t)+O(\eta(x_k,y_k,t))E(\Delta t)
\end{aligned} \tag{4.16}$$

Thus, by making Δt sufficiently small, we get the desired result for sufficiently small $\|(x_k,y_k)-(x^*(t_k),\hat{y})\|$. □

Proof. of Theorem 4.9

Let the constant vector r be defined to be $r \equiv \nabla_t c(x,t)\Delta t$.

Recall that by the MFCQ and Gauvin [8], it holds that the set $\Lambda(x^*(t),t)$ is bounded, and since the KKT conditions are linear with respect to y, it is also closed and convex, and thus compact, and is defined as a polytope.

Let $\{\tilde{y}_j\}_{j\in\{1,\dots,J\}} = \tilde{Y}(x^*(t),t)$ be the set of extreme points of this polytope. It holds that for each $\tilde{y}_j$, for the set $I(\tilde{y}_j)\subseteq\mathcal{I}$ such that $[\tilde{y}_j]_{I(\tilde{y}_j)}>0$, $\{\nabla c_i(x^*(t),t)\}_{i\in I(\tilde{y}_j)\cup\mathcal{E}}$ is linearly independent and there exists a δ_1 such that $\{\nabla c_i(\tilde{x},t)\}_{i\in I(\tilde{y}_j)\cup\mathcal{E}}$ is linearly independent for $\tilde{x}$ such that $\|\tilde{x}-x^*(t)\|\le\delta_1$ by the CRCQ condition. By the implicit function theorem, for any $\tilde{\epsilon}>0$ there exists ϵ_1 such that if J_ρ and g_ρ are perturbations of the objective and constraint gradients satisfying $\|J_\rho-\nabla_x c(x^*(t),t)\|\le\epsilon_1$ and $\|g_\rho-\nabla_x f(x^*(t),t)\|\le\epsilon_1$, it holds that $[J_\rho]_{I(\tilde{y}_j)\cup\mathcal{E}}[\tilde{y}_j(\rho)]_{I(\tilde{y}_j)\cup\mathcal{E}}=g_\rho$ with $[\tilde{y}_j(\rho)]_{\mathcal{I}}\ge 0$ and $\|\tilde{y}_j(\rho)-\tilde{y}_j\|\le\tilde{\epsilon}$.

By Part 3 of Theorem 4.4 we have that $r^T\hat{y}<r^Ty^*$ for all $y^*\in\Lambda(x^*(t),t)$ with $y^*\notin\hat{Y}(x^*(t),t)$, $\hat{y}\in\hat{Y}(x^*(t),t)$. Therefore for all ϵ_2, there exists some $\hat{\epsilon}_2$ such that for perturbations ρ of the extremal multipliers $\tilde{y}_j$ satisfying $\|\tilde{y}_j(\rho)-\tilde{y}_j\|\le\hat{\epsilon}_2$, it holds that

$$r^T\tilde{y}_{\hat{j}}(\rho)<r^T\tilde{y}_{\check{j}}(\rho)-\epsilon_2 \tag{4.17}$$

for all $\hat{j}$ such that $\tilde{y}_{\hat{j}}\in\hat{Y}(x^*(t),t)\cap\tilde{Y}(x^*(t),t)$ and $\check{j}$ such that $\tilde{y}_{\check{j}}\in\tilde{Y}(x^*(t),t)\setminus\hat{Y}(x^*(t),t)$.

Since $\nabla_x L(x,y)$, $\nabla_x f(x,t)-\nabla_x f(x^*(t),t)$ and $\nabla_x c(x,t)-\nabla_x c(x^*(t),t)$ are all $O(\delta(x,y,t))$,the constraints of (4.3) correspond to a set of perturbation of the stationarity conditions. Let us say they are all bounded by $C\delta(x,y,t)$.

Finally, choose ϵ as given in the statement of Theorem 4.9, i.e., satisfying the desired estimate $\operatorname{dist}(\bar{y},\hat{Y}(x^*(t),t)\le\epsilon$. Let $\epsilon_2\ge\epsilon$, and a corresponding $\hat{\epsilon}_2\le\frac{1}{4}\epsilon^2\left(2\max\{1,\|r\|_\infty\}\max\{\|y\|_\infty:\,y\in\Lambda(x^*(t),t)\}\right)^{-1}$ such that (4.17) holds for $\|\tilde{y}_j(\rho)-\tilde{y}_j\|\le\hat{\epsilon}_2$.

Now choose $\hat{\epsilon}_1$ and then take some appropriate x and y as stated in the conditions of the Theorem, such that for $C\delta(x,y,t)\le\min(\delta_1,\hat{\epsilon}_1)$, for all y^α satisfying

$$-|\nabla_x L(x,y,t)|\le\nabla_x f(x,t)-\nabla_x c(x,t)y^\alpha\le|\nabla_x L(x,y,t)|,\quad y^\alpha\ge 0,$$

it holds that $\|y^\alpha-\sum_{j\in J}\alpha_j\tilde{y}_j\|\le\min(\hat{\epsilon}_2,\frac{\epsilon}{2})$ for some α_j with $\alpha_j\ge 0$ for $j\in\mathcal{I}$, $\sum_j\alpha_j=1$.

Now let y^α be a solution to the (4.3), i.e., y^α is feasible and $r^T y^\alpha \leq r^T y^\beta$ for all feasible y^β. Let α_j be defined as above, with

$$\|y^\alpha - \sum_{j\in J} \alpha_j \tilde{y}_j\| \leq \min(\hat{\epsilon}_2, \frac{\epsilon}{2}) \tag{4.18}$$

Consider any feasible y^β satisfying $\|y^\beta - \tilde{y}_{\hat{k}}\| \leq \hat{\epsilon}_2$ for some $\tilde{y}_{\hat{j}} \in \hat{Y}(x^*(t),t) \cap \tilde{Y}(x^*(t),t)$.

Let $\hat{J}$ be such that if $\hat{j} \in \hat{J}$ then $\tilde{y}_{\hat{j}} \in \hat{Y}(x^*(t),t) \cap \tilde{Y}(x^*(t),t)$.

It holds that,

$$\begin{array}{l}
r^T y^\alpha \leq r^T y^\beta \\
(4.18) \Longrightarrow r^T \sum_{j\in J} \alpha_j \tilde{y}_j - \|r\|_\infty \hat{\epsilon}_2 < r^T y^\beta \\
\Longrightarrow r^T \sum_{j\in J} \alpha_j \tilde{y}_j - 2\|r\|_\infty \hat{\epsilon}_2 < r^T \tilde{y}_{\hat{j}} \\
\Longrightarrow r^T \sum_{j\in J} \alpha_j \tilde{y}_j - \frac{\epsilon^2}{2} \left(\max\{\|y\|_\infty : y \in \Lambda(x^*(t),t)\}\right)^{-1} < r^T \tilde{y}_{\hat{j}} \\
\Longrightarrow r^T \sum_{j\in \hat{J}} \alpha_j \tilde{y}_j + r^T \sum_{j\in J\setminus \hat{J}} \alpha_j \tilde{y}_j - \frac{\epsilon^2}{2} \left(\max\{\|y\|_\infty : y \in \Lambda(x^*(t),t)\}\right)^{-1} < r^T \tilde{y}_{\hat{j}} \\
\Longrightarrow \epsilon_2 \sum_{j\in J\setminus \hat{J}} \alpha_j - \frac{\epsilon^2}{2} \left(\max\{\|y\|_\infty : y \in \Lambda(x^*(t),t)\}\right)^{-1} < 0 \\
\Longrightarrow \epsilon_2 \sum_{j\in J\setminus \hat{J}} \alpha_j \left(\max\{\|y\|_\infty : y \in \Lambda(x^*(t),t)\}\right) < \frac{\epsilon^2}{2} \\
\Longrightarrow \sum_{j\in J\setminus \hat{J}} \alpha_j \left(\max\{\|y\|_\infty : y \in \Lambda(x^*(t),t)\}\right) < \frac{\epsilon}{2}
\end{array}$$

which together with (4.18) implies that $\|y^\alpha - \sum_{j\in\hat{J}} \alpha_j \tilde{y}_j\| \leq \epsilon$, proving the Theorem.
□

5. Numerical Results.

5.1. Problem with degenerate constraints throughout and an active set change. We consider problem (2.1) for $t \in [0,1]$. Notice at the point wherein there is an active set change, $t = 0.5$, the vector $\nabla_t c(x,t)\Delta t$ is,

$$\begin{pmatrix} -10 \\ 0 \\ 10 \\ 20 \\ 0 \\ -10 \\ -10 \end{pmatrix} |\Delta t|,$$

and thus the solution to (**SensitivityLP**), recalling that,

$$\Lambda(x^*(5)) = (0, y_1, y_2, y_3, y_4, y_5, y_6),\ y_1 + y_2 + y_3 + y_5 + y_6 = e^5,\ y_1 - y_3 - y_4 + \frac{1}{2}y_5 - \frac{1}{2}y_6 = 0,\ y_i \geq 0,$$

is $(0,0,0,0,0,\frac{1}{2}e^5,\frac{1}{2}e^5)$, indicating that the last two constraints should be strongly active, and for $t \in [0.5,1]$, the solution should trace along these constraints.

Indeed we find that the algorithm successfully traces the solution and (**LPPC**) performs the jump in the multipliers at $t = 0.5$. If, in the implementation of the algorithm, we turn (**LPPC**) off, then the algorithm gets stuck at $t = 0.5$ and proceeds no further along the homotopy. We show the plots of the primal and dual variables in Figure 3.

Note that x follows the true solution closely, y_2 and y_4 are always effectively zero, and there is a discontinuous jump halfway along the homotopy path where y_3 jumps from being positive to zero and y_5, y_6 jump from zero to positive.

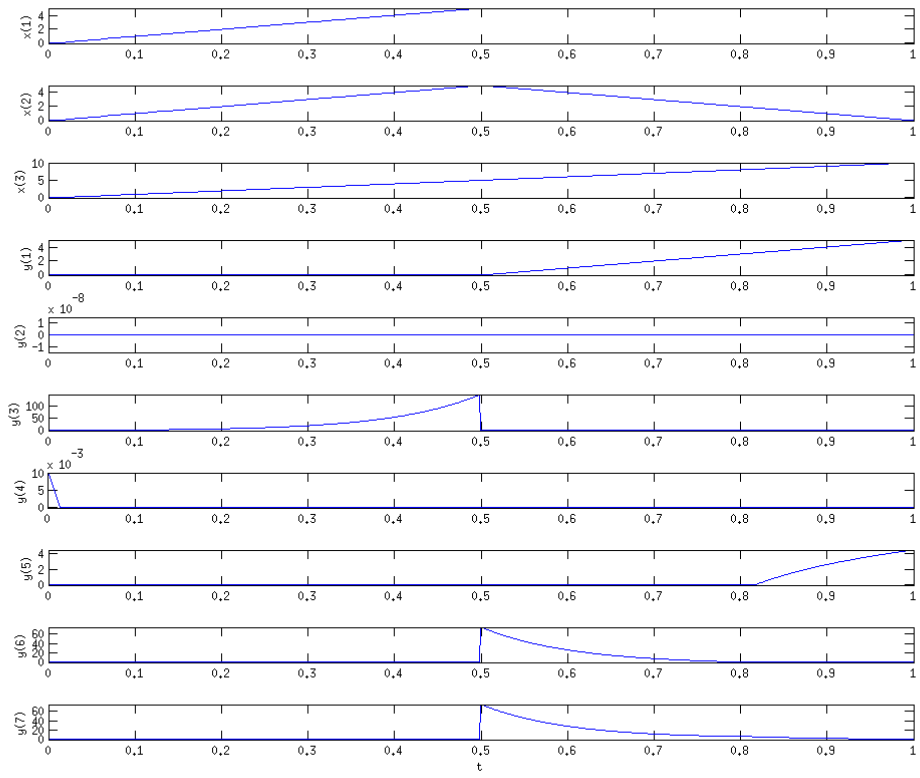

FIG. 3. *Path of primal and dual variables tracing the approximate solution to* (2.1) *across* $t = [0, 1]$

5.2. Degenerate Nonlinear Problem. We now consider a problem with nonlinear constraints. In particular, we consider the problem,

$$
\begin{aligned}
&\min_x \quad -x_2 \\
&\text{subject to:} \\
&c_1(x) := x_3 = 1 + 9t, \\
&c_2(x) := x_1 \geq 0, \\
&c_3(x) := -x_2^3 - x_1 x_2 - x_1^2 + x_3^2 \geq 0, \\
&c_4(x) := -e^{x_1} - e^{x_2} + e^{x_3} + 1 \geq 0, \\
&c_5(x) := -x_1^2 - x_1 x_2 + (x_2 - (2.5 + 0.5x_3))^2 - (2.5 + 0.5x_3)^4 x_1 - 100(x_2 - (2.5 + 0.5x_3)) \geq 0, \\
&c_6(x) := -x_1^2 + x_1 x_2 + (x_2 - (2.5 + 0.5x_3))^2 + (2.5 + 0.5x_3)^4 x_1 - 100(x_2 - (2.5 + 0.5x_3)) \geq 0,
\end{aligned}
\tag{5.1}
$$

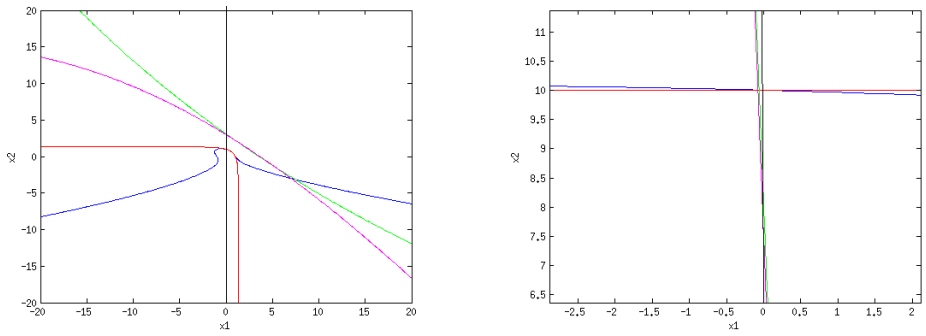

FIG. 4. *Five nonlinear constraints for p=1, 10*

For this problem $x^*(t) = (0, 1 + 9t)$ for $t \in [0, \frac{4}{9}]$, and $x^*(t) = (0, 3 + 4.5t)$ for $t \in [\frac{4}{9}, 1]$. For $t \in [0, \frac{4}{9})$ constraints 1, 2, 3, and 4 are active (including one equality constraint), and for $t \in (\frac{4}{9}, 1]$, constraints 1, 2, 5 and 6 are active. At $t = \frac{4}{9}$ all constraints are active. Since there are always at least 4 active constraints and n is three, the Jacobian is trivially rank deficient. However, for all t, $(0, 1)$ is a strictly feasible direction at $x^*(t)$, and so the MFCQ holds. Furthermore, it can be seen that the CRCQ holds at $x^*(p)$ for all p.

The constraints are shown in Figure 4. Notice that x_3 is just a placeholder for the nonlinear parameter dependence, so we can plot the constraints entirely in x_1 and x_2.

The results are given in Figure 5. Note the discontinuous jump of some of the multipliers for $t = 4/9$.

6. Conclusion. In this paper we investigated the properties of a predictor-corrector pathfollowing algorithm for parametric optimization. The algorithm consists of solving a linear system that corresponds to a corrector step, a QP that corresponds to a corrected predictor, an LP used to jump over discontinuities in the optimal Lagrange multiplier, and a procedure to obtain a new multiplier estimate corresponding to an extreme point in the approximate solution multiplier polytope. The procedure exhibits several desirable properties for an appropriate algorithm for the problems of interest, and in particular we have proven its convergence properties without assuming the LICQ holds at any of the primal solutions along the path.

In future research we will formulate this procedure in a suitable way to be applied to problems arising in NMPC, and test it on some real-world fast real-time control examples. From our numerical experience with toy examples, we believe that there is a possibility that this algorithm could be the basis of promising software for solving NMPC problems.

REFERENCES

[1] Eugene L Allgower and Kurt Georg. Numerical path following. *Handbook of numerical analysis*, 5:3–207, 1997.
[2] Stephan Dempe. *Foundations of bilevel programming*. Springer Science & Business Media, 2002.
[3] Peter Deuflhard. *Newton methods for nonlinear problems: affine invariance and adaptive algorithms*, volume 35. Springer Science & Business Media, 2011.
[4] Moritz Diehl, H Georg Bock, Johannes P Schlöder, Rolf Findeisen, Zoltan Nagy, and Frank Allgöwer. Real-

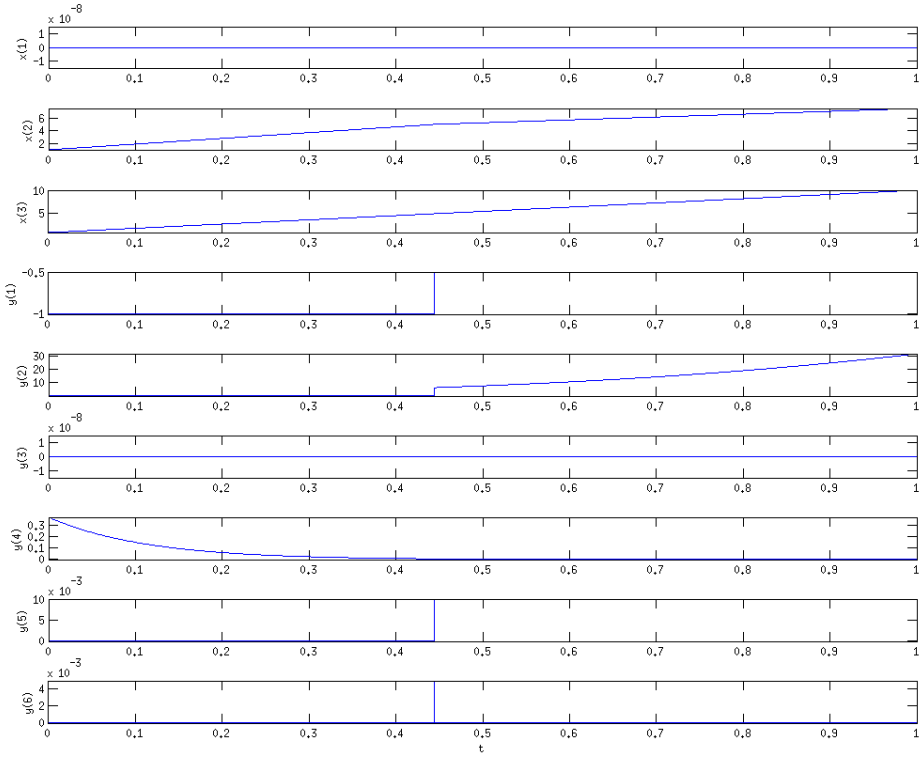

FIG. 5. *Primal-dual approximate solution trajectories for Algorithm 1 applied to* (5.1)

time optimization and nonlinear model predictive control of processes governed by differential-algebraic equations. *Journal of Process Control*, 12(4):577–585, 2002.

[5] Quoc Tran Dinh, Carlo Savorgnan, and Moritz Diehl. Adjoint-based predictor-corrector sequential convex programming for parametric nonlinear optimization. *SIAM Journal on Optimization*, 22(4):1258–1284, 2012.

[6] Jitka Dupačová. Stability and sensitivity-analysis for stochastic programming. *Annals of operations research*, 27(1):115–142, 1990.

[7] Francisco Facchinei, Andreas Fischer, and Christian Kanzow. On the accurate identification of active constraints. *SIAM Journal on Optimization*, 9(1):14–32, 1998.

[8] Jacques Gauvin. A necessary and sufficient regularity condition to have bounded multipliers in nonconvex programming. *Mathematical Programming*, 12(1):136–138, 1977.

[9] Philip E Gill and Elizabeth Wong. Sequential quadratic programming methods. In *Mixed integer nonlinear programming*, pages 147–224. Springer, 2012.

[10] Jürgen Guddat, F Guerra Vazquez, and Hubertus Th Jongen. *Parametric optimization: singularities, pathfollowing and jumps*. Springer, 1990.

[11] William W Hager. Stabilized sequential quadratic programming. *Computational optimization and Applications*, 12(1-3):253–273, 1999.

[12] AF Izmailov. Solution sensitivity for karush–kuhn–tucker systems with non-unique lagrange multipliers. *Optimization*, 59(5):747–775, 2010.

[13] Johannes Jäschke, Xue Yang, and Lorenz T Biegler. Fast economic model predictive control based on nlp-sensitivities. *Journal of Process Control*, 24(8):1260–1272, 2014.

[14] Vyacheslav Kungurtsev. *Second-derivative sequential quadratic programming methods for nonlinear optimization*. PhD thesis, UC-San Diego, 2013.

[15] Vyacheslav Kungurtsev and Moritz Diehl. Sequential quadratic programming methods for parametric nonlinear optimization. *Computational Optimization and Applications*, 59(3):475–509, 2014.

[16] Jerzy Kyparisis. Sensitivity analysis for nonlinear programs and variational inequalities with nonunique multipliers. *Mathematics of operations research*, 15(2):286–298, 1990.

[17] Adam B Levy. Solution sensitivity from general principles. *SIAM journal on control and optimization*, 40(1):1–38, 2001.

[18] Toshiyuki Ohtsuka. A continuation/gmres method for fast computation of nonlinear receding horizon control. *Automatica*, 40(4):563 – 574, 2004.

[19] Daniel Ralph and Stephan Dempe. Directional derivatives of the solution of a parametric nonlinear program. *Mathematical Programming*, 70(1-3):159–172, 1995.

[20] Stephen M Robinson. Generalized equations and their solutions, part ii: Applications to nonlinear programming. *Optimality and Stability in Mathematical Programming*, pages 200–221, 1982.

[21] R Tyrrell Rockafellar and Roger J-B Wets. *Variational analysis*, volume 317. Springer Science & Business Media, 2009.

[22] Ralph Tyrell Rockafellar. *Convex analysis*. Princeton university press, 2015.

[23] Luís N Vicente and Stephen J Wright. Local convergence of a primal-dual method for degenerate nonlinear programming. *Computational Optimization and Applications*, 22(3):311–328, 2002.

[24] Yang Wang and Stephen Boyd. Fast model predictive control using online optimization. *Control Systems Technology, IEEE Transactions on*, 18(2):267–278, 2010.

[25] Stephen J Wright. Modifying sqp for degenerate problems. *SIAM Journal on Optimization*, 13(2):470–497, 2002.

[26] Patrice Ogou Yapo, Hoshin Vijai Gupta, and Soroosh Sorooshian. Multi-objective global optimization for hydrologic models. *Journal of hydrology*, 204(1):83–97, 1998.

[27] Victor M Zavala and Mihai Anitescu. Real-time nonlinear optimization as a generalized equation. *SIAM Journal on Control and Optimization*, 48(8):5444–5467, 2010.

[28] Victor M Zavala and Lorenz T Biegler. The advanced-step nmpc controller: Optimality, stability and robustness. *Automatica*, 45(1):86–93, 2009.