SQP METHODS FOR PARAMETRIC NONLINEAR OPTIMIZATION

Vyacheslav Kungurtsev* Moritz Diehl †

July 2013

Abstract

Sequential quadratic programming (SQP) methods are known to be efficient for solving a series of related nonlinear optimization problems because of desirable hot and warm start properties–a solution for one problem is a good estimate of the solution of the next. However, standard SQP solvers contain elements to enforce global convergence that can interfere with the potential to take advantage of these theoretical local properties in full. We present two new predictor-corrector procedures for solving a nonlinear program given a sufficiently accurate estimate of a similar problem. The procedures attempt to trace a homotopy path between solutions of the two problems, staying within the local domain of convergence for the series of problems generated. We provide theoretical convergence and tracking results, as well as some numerical results demonstrating the robustness and performance of the methods.

Key words. Parametric nonlinear programming, Nonlinear programming, nonlinear constraints, augmented Lagrangian, sequential quadratic programming, SQP methods, stabilized SQP, regularized methods, model predictive control.

AMS subject classifications. 49J20, 49J15, 49M37, 49D37, 65F05, 65K05, 90C30

*Department of Computer Science and Optimization in Engineering Center (OPTEC), KU Leuven, Celestijnenlaan 200A, B-3001 Leuven- Heverlee, Belgium. (vyacheslav.kungurtsev@cs.kuleuven.be).

†Department of Microsystems Engineering IMTEK, University of Freiburg, Georges-Koehler-Allee 102, 79110 Freiburg, Germany and Electrical Engineering Department (ESAT-STADIUS) and Optimization in Engineering Center (OPTEC), KU Leuven, Kasteelpark Arenberg 10, B-3001 Leuven- Heverlee, Belgium. (mdiehl@esat.kuleuven.be).

1. Introduction

In many engineering applications it is common to solve a series of nonlinear programs dependent on a parameter. This can arise in a context of online tuning software for design problems, where a user slightly adjusts the objective or constraints to study how the optimal solution of a system behaves relative to certain perturbations, or in model predictive control, where a dynamically evolving process is controlled in real time. Since with each change, or each time step, the structure of the problem does not fundamentally change, it should be expected that the solution of one problem is a good initial guess for the subsequent problem. It is natural to develop theoretical methodology to make the concept of a "good" initial guess more precise, and formulate practical algorithms that take advantage of this methodology to be able to robustly and efficiently solve such a sequence of nonlinear programs.

In particular, we analyze properties of parametric exact Hessian sequential quadratic programming (SQP) methods. SQP methods are well known to have desirable "hot-start" properties, in contrast to interior point methods. This is because in a conventional SQP method, if the active set has stabilized, the algorithm behaves like Newton's method on the optimality conditions that are satisfied at the solution. However, in practice, even arbitrarily close to the solution, a QP solver may not estimate the optimal active set. Furthermore, typical NLP solvers may modify the Hessian or use a positive definite quasi-Newton variant in order to prevent the possibility of forming nonconvex QPs. Here, we use robust active set estimates and exact Hessians with each QP subproblem lying in the Newton domain of convergence in a sequence of homotopy steps.

We consider a parametric nonconvex nonlinear optimization problem of the form,

$$\begin{aligned} \underset{x\in\mathbb{R}^n}{\text{minimize}} \quad & f(x,t) \\ \text{subject to} \quad & c(x,t) \geq 0, \end{aligned} \tag{1.1}$$

where $f : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}$, $c : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}^m$ are nonlinear two times Lipschitz continuously differentiable (possibly nonconvex) functions and $t \in [0, 1]$. In particular, $t = 0$ and $t = 1$ correspond to two optimization problems of interest and f and c are homotopies with $(f(x, 0), c(x, 0))$ and $(f(x, 1), c(x, 1))$ corresponding to the two optimization problems. This problem can be embedded in a more general study of the problem,

$$\begin{aligned} \underset{x\in\mathbb{R}^n}{\text{minimize}} \quad & f(x,p) \\ \text{subject to} \quad & c(x,p) \geq 0, \end{aligned} \tag{1.2}$$

where, for ease of subsequent exposition, we relax notation in treating f and c as the same in (1.1) and (1.2). Here $p \in \mathcal{P} \subseteq \mathbb{R}^d$ is a parameter. Problem (1.1) corresponds to the analysis of problems at two $p_1, p_2 \in \mathcal{P}$ of interest, and using the solution of (1.2) at p_1 to guide an initial point for problem (1.2) at p_2 using $p(t) = tp_1 + (1 - t)p_2$. We do not include equality constraints, even though they are pervasive in engineering problems, because they do not provide additional challenge to the machinery developed, and the analysis and algorithms presented in this paper immediately extend to the case of mixed inequality and equality (or even just equality) constraints.

We recall the definition of the first-order optimality conditions,

Definition 1.1. (First-order necessary conditions) *A primal-dual pair* (x^*, y^*) *satisfies the* first-order necessary optimality conditions *for* (1.1) *if,*

$$\begin{array}{rcl} \nabla_x f(x^*,t) & = & J(x^*,t)^T y^*, \\ c(x^*,t) & \geq & 0, \\ c(x^*,t)^T y^* & = & 0, \\ y^* & \geq & 0. \end{array} \tag{1.3}$$

where $J(x,t)$ denotes the Jacobian constraint matrix for problem t at x. 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 the paper.

We will define $\mathcal{A}(x^*,t)$ to be the set of indices 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 let $\Lambda(x^*)$ be the set of y^* such that (x^*,y^*) satisfy (1.3), and define $\mathcal{A}_+(x^*,t) = \cup_{y^* \in \Lambda(x^*)} \mathcal{A}_+(x^*,y^*,t)$ and $\mathcal{A}_0(x^*,t) = \cap_{y^* \in \Lambda(x^*)} \mathcal{A}_0(x^*,y^*,t)$.

The Lagrangian function associated with (1.1) is $L(x,y,t) = f(x,t) - c(x,t)^T y$, where y is an m-vector of dual variables associated with the inequality constraints. The Hessian of the Lagrangian with respect to x is denoted by $H(x,y,t) = \nabla^2 f(x,t) - \sum_{i=1}^m y_i \nabla^2 c_i(x,t)$.

We also define the second-order sufficiency condition.

Definition 1.2. (Second-order sufficient conditions (SOSC)) *A primal-dual pair* (x^*,y^*) *satisfies the* second-order sufficient optimality conditions *if it satisfies the first-order conditions* (1.3) *and*

$$d^T H(x^*,y^*,t) d > 0 \;\; \textit{for all} \;\; d \in \mathcal{C}(x^*,y^*,t) \setminus \{0\}, \tag{1.4}$$

where $d \in \mathcal{C}(x^*,y^*,t)$ *if* $\nabla c_i(x^*,t)^T d = 0$ *for* $i \in \mathcal{A}_+(x^*,y^*,t)$ *and* $\nabla c_i(x^*,t)^T d \geq 0$ *for* $i \in \mathcal{A}_0(x^*,y^*,t)$.

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

Definition 1.3. (Strong second-order sufficient conditions (SSOSC)) *A primal-dual pair* (x^*,y^*) *satisfies the* strong second-order sufficient optimality conditions *if it satisfies the first-order conditions* (1.3) *and*

$$d^T H(x^*,y^*,t) d > 0 \;\; \textit{for all} \;\; d \in \tilde{\mathcal{C}}(x^*,y^*,t) \setminus \{0\}, \tag{1.5}$$

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

It can be seen that the cone of directions on which positive definiteness of the Hessian must hold is larger in the case of the SSOSC. If *strict complementarity*, the condition that $\mathcal{A}_0(x^*,y^*,t) = \emptyset$, holds, then the two are equivalent.

We make one assumption throughout the paper.

Assumption 1.1. *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.*

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

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

where $x^*(t)$ is the closest primal solution to (1.1) at t, and $\Lambda(x^*(t))$ is the set of Lagrange multipliers $\{y^*\}$ satisfying (1.3) together with x^*. 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

$$\eta(x,y,t) = \left\| \begin{pmatrix} \nabla f(x,t) - J(x,t)^T y \\ \min(c(x,t), y) \end{pmatrix} \right\|.$$

These two quantities bound each other under the SOSC.

Lemma 1.1. *If the second-order sufficiency condition for optimality holds at $(x^*, y^*)(t)$, 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,*

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

Alternatively, if the second-order sufficiency condition holds for all $(x^*, y^*)(t)$, with y^* in a compact subset of $\Lambda(x^*)$, and (x,y) is sufficiently close to this subset, the estimate also holds.

Proof. See, e.g., Wright [40, Theorem 3.1 and Theorem 3.2]. ∎

The rest of the paper is organized as follows. In the rest of this section we review the the previous literature on this topic. Section 2 discusses predictor-corrector methods in the context of parametric nonlinear optimization. Section 3 discusses an approach to solving the problem using conventional SQP, and Section 4 provides theoretical tracking error results. Section 5 discusses an approach to solving degenerate problems, with Section 6 providing the convergence results. We analyze some numerical results in Section 7 and finally conclude the paper and present some directions in future research in Section 8

1.1. Previous work

The local analysis of nonlinear programs depending on parameters has a rich literature. In particular, optimization theorists have long been interested in the existence and stability of solutions, error bounds, and directional differentiability of nonlinear programs subject to perturbations. For some comprehensive surveys, see [4, 24]. For some classical results, see [21, 30, 32, 33].

The study of parametric optimization from the perspective of continuation is more terse. Some seminal work includes [20] and [28], in which the form of a parametric solution is analyzed, e.g., what conditions lend themselves to the satisfaction

of the implicit function theorem, and which result in folds or other forms of bifurcations in the solution set.

The work of Guddat et al. [14] summarizes many of these results, as well as provides algorithms for tracing homotopy paths of parametric NLPs as well as jumping across singularities. Some of the ideas presented foreshadow the more rigorously treated active set estimation and locally convergent procedures used in this paper.

From the reverse perspective, homotopy methods have been used as a method for solving one-off NLPs (see, e.g., [38] for an early work). There continues a literature on this for some contexts of convex problems, but as the fundamental problem and the algorithms formulated are entirely different than the methods discussed here, so we do not discuss this line of work further.

More recently, the development of practical algorithms for solving a series of related NLPs has been spurred by the recent interest in on-line process optimization, (see, e.g., [34]). In [5], an SQP algorithm is presented for solving model predictive control problems under the general principle of not iterating each solution to optimality but solving a single subproblem at each step.

Zavala and Anitescu et al. [41] analyze the parametric NLP problem in the context of generalized equations, and provide theoretical tracking results for an augmented Lagrangian algorithm. Tran-Dinh et al. [36] formulate an adjoint-based sequential convex programming algorithm to solve parametric NLPs. Using the generalized equation framework given by Zavala and Anitescu, they prove a theoretical result on the tracking error and show numerical comparisons of their method against several other approaches for nonlinear model predictive control of a hydro power plant. The QP solver qpOASES [9–11] is designed to use a homotopy path to solve successive QP problems generated from an SQP solver applied to model predictive control. In [19] an inexact variant of this is explored, where the Hessian (approximation) used in one problem is used in the subsequent problem, with a correction term. They show that this inexact strategy can result, in some cases, in considerable savings in computation time.

Contribution: This paper is unique in providing convergence theory for an algorithm solving a sequence of related NLPs in a procedure that uses exact Hessians in a robust and reliable way, taking full advantage of the potential strengths of SQP methods for this context.

1.2. Notation

The ith component of a vector labeled with a subscript will be denoted by $[\,\cdot\,]_i$, e.g., $[v]_i$ is the ith component of the vector v. The subvector of components with indices in the index set $\mathcal{S}$ is denoted by $[\,\cdot\,]_{\mathcal{S}}$, e.g., $[v]_{\mathcal{S}}$ is the vector with components v_i for $i \in \mathcal{S}$. Similarly, if M is a symmetric matrix, then $[M]_{\mathcal{S}}$ denotes the symmetric matrix with elements m_{ij} for $i, j \in \mathcal{S}$.

Let $\{\alpha_j\}_{j\geq 0}$ be a sequence of scalars, vectors or matrices and let $\{\beta_j\}_{j\geq 0}$ be a sequence of positive scalars. If there exists a positive constant γ such that $\|\alpha_j\| \leq \gamma\beta_j$, we write $\alpha_j = O(\beta_j)$. If $\|\alpha_j\| \leq \gamma_j\beta_j$ with $\gamma_j \to 0$, we write $\alpha_j = o(\beta_j)$.

2. Predictor corrector methods for parametric nonlinear optimization

Consider the more general problem of finding a solution of $F(x, 1) = 0$, given a known $\bar{x}$ such that $F(\bar{x}, 0) = 0$.

Predictor-corrector methods are a class of methods that include a predictor step that initially attempts to estimate the solution and a corrector step to refine the estimate.

At the simplest level, for a homotopic method to be successful, i.e., reach a solution of $F(x, 1) = 0$, at the minimum it should have a sequence of Newton iterates, as described in Algorithm 2.1,

Algorithm 2.1 Newton corrector homotopy.

1: Begin with x a solution to $F(x, t) = 0$ at $t = 0$. Initialize $\Delta t = 1$.
2: Choose constants γ_1 and γ_2 satisfying $0 < \gamma_1 < 1$ and $\gamma_2 > 1$.
3: **while** $t < 1$ **do**
4: Solve $\Delta x = -F'(x, t + \Delta t)^{-1} F(x, t + \Delta t)$.
5: **if** $\|F(x + \Delta x, t + \Delta t)\|$ is sufficiently small **then**
6: Set $t = t + \Delta t$.
7: Set $x = x + \Delta x$.
8: Let $\Delta t = \min(\gamma_2 \Delta t, 1 - t)$.
9: **else**
10: Let $\Delta t = \gamma_1 \Delta t$.
11: Go to Step 4
12: **end if**
13: **end while**

In this case, there is only a corrector step. The initial point at which Newton's method is applied is the original x. If x happens to lie in the Newton domain of rapid local convergence for the problem at t_2, then the procedure would take a successful step and the condition at Step 4 in Algorithm 2.1 is satisfied. Hence, this procedure is effective at generating a point that sufficiently contracts $||F(x, t)||$ for $t_2 - t_1$ small enough that the domain of convergence relative to x does not change too much, as illustrated in Figure 1.

However, by including a predictor step, described in Algorithm 2.2, we incorporate first order information in order to estimate the location of the solution at t_2. Here, if the first-order estimate is sufficiently accurate, we may take a larger step along t and still achieve the desired level of contraction in $\|F\|$ by staying within the domain of local convergence, as illustrated in Figure 2.

A canonical predictor-corrector algorithm using one step of a linear tangential predictor and a step of Newton's method is given as Algorithm 2.2.

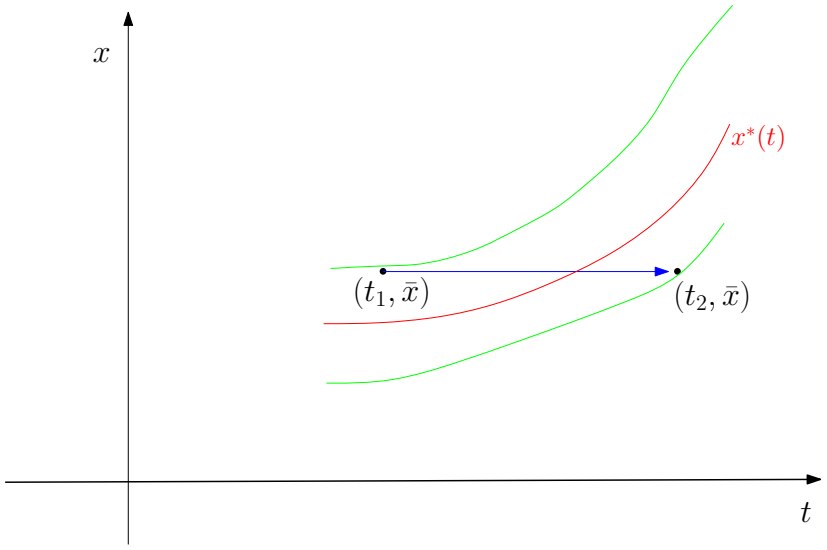


Figure 1: Using the previous $\bar{x}$ as the predictor

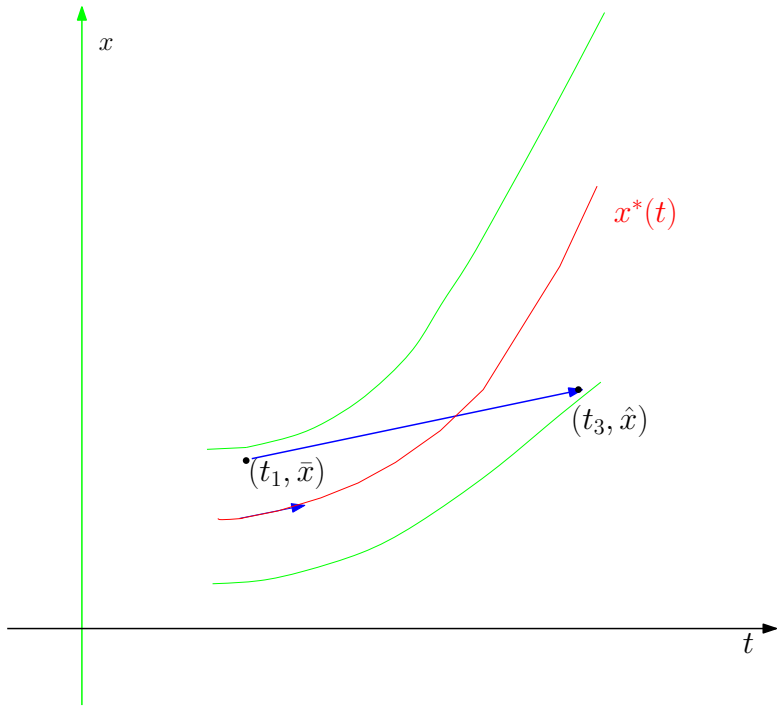


Figure 2: Using an additional tangential predictor

Algorithm 2.2 Predictor-corrector.

1: Begin with x a solution to $F(x,t)=0$ at $t=0$. Initialize $\Delta t=1$.
2: Choose constants γ_1 and γ_2 satisfying $0<\gamma_1<1$ and $\gamma_2>1$.
3: **while** $t<1$ **do**
4: Calculate $d=\frac{\partial x(t)}{\partial t}$.
5: Set $\widehat{x}=x+\Delta t d$.
6: Solve $\Delta x=-F'(\widehat{x},t+\Delta t)^{-1}F(\widehat{x},t+\Delta t)$.
7: **if** $\|F(\widehat{x}+\Delta x,t+\Delta t)\|$ is sufficiently small **then**
8: Set $t=t+\Delta t$.
9: Set $x=\widehat{x}+\Delta x$.
10: Let $\Delta t=\min(\gamma_2\Delta t,1-t)$.
11: **else**
12: Let $\Delta t=\gamma_1\Delta t$.
13: Go to Step 5
14: **end if**
15: **end while**

This is the Euler-Newton method using exact function and derivative definitions, as described in Allgower and Georg [1].

Note that d uniquely exists if the assumptions of the implicit function theorem hold. In practice, for generic cases, d is calculated as $d=\left(\frac{\partial F(x,t)}{\partial x}\right)^{-1}\frac{\partial F(x,t)}{\partial t}$.

For a couple references on continuation methods, see [1, 35].

2.1. Predictors for parametric NLP

In the context of the problem of parametric NLP (1.1), the direction of the predictor is the first-order estimate of the change in x and y that would solve the problem corresponding to a change in the parameter p. It also indicates which constraints become inactive with the parameter change.

In this paper we explore a strategy where a predictor and corrector are computed together in one QP subproblem, as well as a strategy where a predictor and corrector are computed separately. In a general sense, the predictor (as a separate step, or as an aspect of the QP) estimates the newly active or inactive constraints, and the corrector refines the step on a properly reduced space with a Newton type iteration.

The existence of a predictor, i.e., directional differentiability of the parametric NLP, has long been established to hold under strong regularity (see, e.g., Bonnans and Shapiro [3, Proposition 3.3]). The case where the LICQ does not hold is less straightforward. In Levy [25], the concept of proto-differentiability was introduced and directional differentiability in this sense was shown to exist for the primal solutions $x^*(t)$ under the Mangasarian-Fromovitz constraint qualification. In the case where the constant rank constraint qualification holds at x, Ralph and Dempe [29] demonstrated the Bouligand-differentiability of the optimal primal solution, and in particular the existence of a well-defined piecewise smooth path with respect to the parameters, as well as showing a straightforward way to calculate this direction.

Note that, in the case of nonunique multipliers, which we study in the Section 5,

what directional differentiability of the optimal dual solution corresponds to is ambiguous, since it is expected that the perturbed problem has multiple dual solutions as well.

2.1.1. Active set estimate

The directional derivative provides a one-sided estimate of the active set. In particular, for small perturbations of an NLP, the optimal inactive constraints remain inactive and a subset of the weakly active constraints may become inactive. Estimation of newly active sets is performed by the rest of the components of the procedure. Depending on whether a one or two-step procedure is used, this means that the estimation of newly active constraints is performed by linearized constraints at the QP subproblem of the predicted point, or a separate active set estimation procedure.

Facchinei et al. [7] presents an estimate of the active set,

$$\mathcal{A}_\epsilon(x, y, t) = \{i : c_i(x, t) \leq \eta(x, y, t)^\gamma\}, \tag{2.1}$$

where γ is a scalar satisfying $0 < \gamma < 1$. In the case of the separate predictor and corrector step strategy algorithm, described in Section 5, we use this estimate to determine the active set on which to perform the corrector step.

2.1.2. Tangential direction

Izmailov [17] provides a thorough study of the directional differentiability of primal-dual solution pairs. Consider a step $(\Delta t, \Delta x)$ The careful analysis is based on the splitting of the null multipliers by the index sets,

$$\begin{aligned}
\mathcal{A}_0^+ &= \{i \in \mathcal{A}_0 | \tfrac{\partial c_i}{\partial t}\Delta t + \tfrac{\partial c_i}{\partial x}\Delta x = 0,\ \Delta y_i > 0\},\\
\mathcal{A}_0^0 &= \{i \in \mathcal{A}_0 | \tfrac{\partial c_i}{\partial t}\Delta t + \tfrac{\partial c_i}{\partial x}\Delta x = 0,\ \Delta y_i = 0\},\\
\mathcal{A}_0^N &= \{i \in \mathcal{A}_0 | \tfrac{\partial c_i}{\partial t}\Delta t + \tfrac{\partial c_i}{\partial x}\Delta x > 0,\ \Delta y_i = 0\},
\end{aligned} \tag{2.2}$$

indicating null constraints such that the perturbations change the multipliers to be strongly active, to remain null, or the constraint to become inactive.

For ease of exposition, we will now ignore the multipliers corresponding to the inactive constraints, as any reasonable predictor will be expected to maintain them to be zero. Thus, in this section, y and J will correspond to the multipliers and constraint gradients of just the active constraints at the current solution.

The following result, found in several sources, an important set of equations characterizing the directional derivative of an NLP.

Theorem 2.1. *Izmailov [17, Theorem 2.1] Let $(\bar{x}, \bar{y})$ be a solution of system (1.1) with $t = \bar{t}$. Assume there exist sequences $\{x_k, y_k, t_k, s_k\} \to \{\bar{x}, \bar{y}, \bar{t}, 0\}$, where $s_k \in \mathbb{R}$, such that for each k, (x_k, y_k) solves (1.1) with $t = t_k$. Then, any limit point of*

$(\Delta t, \Delta x, \Delta y)$ of the sequence $\{t_k - \bar{t}, x_k - \bar{x}, y_k - \bar{y}\}/s_k$ satisfies,

$$\begin{aligned} \frac{\partial^2 L(\bar{x},\bar{y},\bar{t})}{\partial x \partial t}\Delta t + \frac{\partial^2 L(\bar{x},\bar{y},\bar{t})}{\partial^2 x}\Delta x + J(\bar{x},\bar{t})^T \Delta y &= 0, \\ \frac{\partial c(\bar{x},\bar{t})_+}{\partial t}\Delta t + \frac{\partial c(\bar{x},\bar{t})_+}{\partial x}\Delta x &= 0, \\ \frac{\partial c(\bar{x},\bar{t})_0}{\partial t}\Delta t + \frac{\partial c(\bar{x},\bar{t})_0}{\partial x}\Delta x &\geq 0, \\ \Delta y_0 &\geq 0, \\ \Delta y_0^T(\frac{\partial c(\bar{x},\bar{t})_0}{\partial t}\Delta t + \frac{\partial c(\bar{x},\bar{t})_0}{\partial x}\Delta x) &= 0. \end{aligned} \tag{2.3}$$

Note that the theorem does not state that this sequence exists, nor that the limit of $\{t_k - \bar{t}, x_k - \bar{x}, y_k - \bar{y}\}/s_k$ exists. The rest of [17] establish conditions under which these two properties hold. Nevertheless this set of equations form the computational basis for generating predictor steps.

2.1.3. Corrector

In standard continuation, correctors are a variation of Newton's method. SQP methods can correspond to Newton's method in a direct way. In particular, for (x_k, y_k) close to a primal dual solution (x^*, y^*), if $\mathcal{A}(x_k) = \mathcal{A}(x_{k+1}) = \mathcal{A}(x^*)$ then the QP solution solves the system,

$$\begin{pmatrix} H(x_k, y_k) & \nabla c(x_k) \\ \nabla c(x_k)^T & 0 \end{pmatrix} \begin{pmatrix} \Delta x_k \\ y_{k+1} \end{pmatrix} = \begin{pmatrix} g(x_k) - \nabla c(x_k)^T y_k \\ -c(x_k) \end{pmatrix},$$

which is the Newton system for the first-order optimality conditions. Most proofs of quadratic or superlinear local convergence of SQP methods rely on the correct estimation of the active set and thereafter this system being solved.

Correct estimation of the active set is of importance from both the theoretical and practical perspective. Even if the reduced Hessian is positive definite at the solution, the Hessian could still be indefinite, and for a QP at a base point in the neighborhood of the solution, there may be more than one solution, and even an unbounded solution. A "wrong" or unbounded solution is particularly likely to be selected by the QP solver if strict complementarity does not hold. A local convergence result can still be shown where a desired solution to a quadratic program (QP) *exists* that is part of a quadratically convergent sequence, however, in the case of multiple solutions, its selection depends on the QP solver. In order to have a well-defined algorithm that selects a unique sequence of iterates that are solutions to bounded QPs, all of the QPs we solve will be on a specific reduced space that will ensure their convexity.

3. Conventional SQP methods for Homotopy solutions

In this section we discuss a conventional SQP method for solving nonlinear programming homotopy problems. We begin with stating our assumptions for the problem.

Definition 3.1. *The* strict Mangasarian-Fromovitz constraint qualification *holds at (x^*, t) if there exists a vector y^* such that (x^*, y^*) satisfy (1.3), the matrix $[\nabla_x c(x,t)]_{\mathcal{A}_+(x^*,y^*,t)}$ has rank $|\mathcal{A}_+(x^*, y^*, t)|$, and there exists a vector d such that $\nabla[c(x^*, t)]^T_{\mathcal{A}(x^*)} d > 0$.*

The strict MFCQ is equivalent to the MFCQ holding at a point x^* satisfying the necessary optimality conditions (e.g., the existence of a Lagrange multiplier) in conjunction to the uniqueness of the Lagrange multiplier [23] (see also the note in [37]).

Assumption 3.1. *For every $t \in [0,1]$, the problem (1.1) has at least one primal-dual solution $(x^*, y^*)(t)$ to the first-order KKT conditions. Furthermore at each x^*, the SMFCQ and the SSOSC hold for $(x^*, y^*)(t)$.*

It should be emphasized that strict complementarity is not assumed and the problem would be trivially uninteresting otherwise. For an optimal solution at which strictly complementary holds, at any perturbation sufficiently small of the problem, the optimal active set is the same. Thus, if strict complementarity held at every point along the solution of the homotopy, there would be no active set changes, and the problem would be identical to one that is strictly equality-constrained. Null multipliers provide a branching point; if the constraint i is weakly active at the solution, then for a perturbation of the problem it could remain weakly active, or could become strongly active or inactive, as denoted in (2.2).

The corrector is the equivalent of a Newton step in the context of the KKT optimality conditions. This makes SQP an appropriate framework, because once the active set is correctly estimated, solving the quadratic subproblem is equivalent to a step of Newton on the equations corresponding to the first-order optimality conditions.

3.1. Properties of parametric NLP under strong regularity conditions

It shall be shown in Section 4 that Assumption 3.1 implies the existence of a unique primal-dual path of solutions $(x^*, y^*)(t)$ to (1.1). This suggests that the system (2.3) should have a unique set of solutions. Note that this system corresponds to the stationarity conditions for the QP,

$$\begin{aligned} \underset{\Delta x}{\text{minimize}} \quad & \tfrac{1}{2}\Delta x^T H(x^*, y^*, t)\Delta x + \Delta p^T \tfrac{\partial^2 L(x^*, y^*, t)}{\partial t \partial x}\Delta x \\ \text{subject to} \quad & \nabla_t c_i(x^*, t)^T \Delta t + \nabla_x c_i(x^*, t)^T \Delta x = 0, \ \text{for } i \in \mathcal{A}_+(x^*, y^*, t), \\ & \nabla_t c_i(x^*, t)^T \Delta t + \nabla_x c_i(x^*, t)^T \Delta x \geq 0, \ \text{for } i \in \mathcal{A}_0(x^*, y^*, t). \end{aligned} \tag{3.1}$$

Under Assumption 3.1, the stationary point of this QP is the local minimizer. Let a current iterate $(\bar{x}, \bar{y}, t)$ be close to the optimal solution $(x^*, y^*)(t)$. Enforcing a step towards feasibility, we consider the QP,

$$\begin{aligned} \underset{\Delta x}{\text{minimize}} \quad & \nabla f(\bar{x}, t + \Delta t)^T \Delta x + \tfrac{1}{2}\Delta x^T H(\bar{x}, \bar{y}, t + \Delta t)\Delta x \\ \text{subject to} \quad & c_i(\bar{x}, t + \Delta t) + \nabla_x c_i(\bar{x}, t + \Delta t)^T \Delta x = 0, \ \text{for } i \in \mathcal{A}_+(x^*, y^*, t), \\ & c_i(\bar{x}, t + \Delta t) + \nabla_x c_i(\bar{x}, t + \Delta t)^T \Delta x \geq 0, \ \text{for } i \in \mathcal{A}_0(x^*, y^*, t). \end{aligned} \tag{3.2}$$

It may appear as though formulating this QP requires an extra function evaluation. However, this is not the case if we reformulate the problem so that the dependence

on the parameter is linear. In particular, note that (1.2) can be written as,

$$\begin{aligned} \underset{x\in\mathbb{R}^n, z\in\mathbb{R}^d}{\text{minimize}} \quad & f(x,z) \\ \text{subject to} \quad & c(x,z) \geq 0, \\ & z = p. \end{aligned}$$

Therefore, we will assume that $\nabla f(x,t) = \nabla f(x,0)$, $\nabla c(x,t) = \nabla c(x,0)$ and $H(x,y,t) = H(x,y,0)$ for all x, y, t.

We can see now that because the solution to (3.2) corresponds to the directional derivative of (2.3), plus the step to feasibility, it is the unique predictor along the path of solutions. In addition, however, it can be expected that if the step maintains the active set, it also behaves like Newton's method on the optimality conditions, and is thus a corrector.

This immediately suggests two procedures:

- Solving (3.2) using appropriate estimates of the strongly and weakly active sets, and considering an iteration to be successful if there is quadratic contraction, or
- solving (3.2) to obtain a predictor step that is just sufficiently low in its optimality residual, and then take an additional contractive Newton step on the active set.

In this section, we present a procedure performing the first of these where the QP is both the predictor and corrector in an iterative Homotopy method. We shall consider a two-step approach in our discussion of homotopy methods for degenerate problems, in Section 5.

3.2. Statement of the algorithm

Define

$$\mathcal{A}_{+,\epsilon}(x,y,t) = \{i : y_i \geq \eta(x,y,t)^\gamma\}. \tag{3.3}$$

This will be shown to be a reliable estimate of the strongly active set at t.

As discussed in the previous section, we use this estimate to set up a reduced space in which to solve the predictor-corrector QP. In addition, we also enforce the linearized inactive inequality constraints. Thus, we solve the QP,

$$\begin{aligned} \underset{\Delta x}{\text{minimize}} \quad & \nabla f(\bar{x}, t+\Delta t)^T \Delta x + \tfrac{1}{2}\Delta x^T H(\bar{x},\bar{y},t+\Delta t)\Delta x \\ \text{subject to} \quad & c_i(\bar{x},t+\Delta t) + \nabla c_i(\bar{x},t+\Delta t)^T\Delta x = 0, \text{ for } i \in \mathcal{A}_{+,\epsilon}(\bar{x},\bar{y},t), \\ & c_i(\bar{x},t+\Delta t) + \nabla c_i(\bar{x},t+\Delta t)^T\Delta x \geq 0, \text{ for } i \in \{1,...,m\} \setminus \mathcal{A}_{+,\epsilon}(\bar{x},\bar{y},t). \end{aligned} \tag{3.4}$$

We formally state Algorithm 3.1. The procedure is relatively simple and consists of the strongly active set estimate, the QP solution, and acceptance and refusal of the step. It should be noted that in practice, parallelization will make sure that no iterations are wasted, i.e., on a machine with Q processors, we choose Q values of Δt_j, solve the corresponding QP on each processor, and pick the largest value

of Δt_j with a successful iteration to be the base Δt for the next iteration, with an equal number of Δt_j larger and smaller than Δt in the next iteration. As $Q \to \infty$, there will be no failed iterations, and the step will always be the largest such that a successful iteration can be made. Thus the worst-case computational complexity of the procedure is the number of required steps (determined by the domain of convergence of the problem) multiplied by the cost of solving the convex QP.

Algorithm 3.1 Parametric SQP Method.

1: Begin with $(\bar{x}, \bar{y})$, an approximate solution to (1.1) at $t = 0$.
2: Set constants $\gamma, \gamma_1 \in (0, 1)$, $\gamma_2 > 1$.
3: **while** $t < 1$ **do**
4: Calculate $\mathcal{A}_{+,\epsilon}(\bar{x}, \bar{y}, t)$.
5: Solve (3.4) for $(\Delta x, \Delta y)$.
6: **if** $\eta(\bar{x} + \Delta x, \bar{y} + \Delta y, t + \Delta t) \leq \eta(\bar{x}, \bar{y}, t + \Delta t)^{1+\gamma}$ **then**
7: Let $\bar{x} = \bar{x} + \Delta x$, $\bar{y} = \bar{y} + \Delta y$, and $t = t + \Delta t$.
8: Let $\Delta t = \min(1 - t, \gamma_2 \Delta t)$.
9: **else**
10: Let $\Delta t = \gamma_1 \Delta t$. Go back to Step 5
11: **end if**
12: **end while**

Note that this procedure is defined to only approximately solve the problem at $t = 1$. For a precise solution, at $t = 1$ the same QP can be successively solved, and, under the right conditions the sequence of primal-dual solutions will converge quadratically to the solution.

In addition, in a software-level implementation, we will want to avoid cycling of Δt back and forth (a successful iteration followed by a larger step in Δt, followed by failure, followed by a smaller Δt and a successful iteration, etc.) and we intend to dampen the constants in time depending on the accumulated history of successful iterations.

4. Convergence

Proposition 4.1. *For $\|p(1) - p(0)\|$ sufficiently small, there exists a continuous unique path $(x^*, y^*)(t)$ between each solution $(x^*, y^*)(0)$ of (1.1) for $t = 0$ and some $(x^*, y^*)(1)$ of (1.1) for $t = 1$.*

Proof.

Kojima [21, Theorem 7.1] implies that for each $x^*(t)$, there is a unique stationary point $x^*(t + \delta t)$ for δ sufficiently small. Furthermore, this point is isolated. This, together with Levy and Rockafellar [25, Theorem 3.1] implies that there is a continuous path from $x^*(t)$ to $x^*(t + \delta t)$, for δ sufficiently small.

We must now prove that the path of multipliers is continuous. In fact, however, we can make a stronger statement. By the strong second order sufficiency condition,

Robinson [32, Theorem 1] implies the error bound,

$$\|x^*(t) - x^*(t+\delta t)\| + \|y^*(t) - y^*(t+\delta t)\| \leq O(\|\delta t\|),$$

implying that the path is locally Lipschitz continuous.

We may continue this procedure indefinitely at $(x^*, y^*)(t+\delta t)$ until either $t = 1$ or an accumulation of open sets whose closure represents a non-KKT boundary point. ∎

Theorem 4.1. *If $(\bar{x}, \bar{y})$ is sufficiently close to some solution $(x^*, y^*)(t)$ and Δt is sufficiently small, there is a unique solution to the quadratic subproblem (3.4), and condition at step 6 in Algorithm 3.1 is satisfied.*

Proof.

We first claim that $\mathcal{A}_{+,\epsilon}(\bar{x}, \bar{y}, t) = \mathcal{A}_+(x^*(t), y^*(t), t)$. Note that the quantity $\min_i\{[y^*(t)]_i : i \in \mathcal{A}_+(x^*(t), y^*(t), t)\}$ is positive which implies the inclusion $\mathcal{A}_+(x^*(t), y^*(t), t) \subseteq \mathcal{A}_{+,\epsilon}(\bar{x}, \bar{y}, t)$ since $\eta(\bar{x}, \bar{y}, t) \to 0$ as $(\bar{x}, \bar{y}) \to (x^*, y^*)(t)$.

Assume $i \in \mathcal{A}_{+,\epsilon}(\bar{x}, \bar{y}, t)$. Recall from Lemma 1.1 that $|[\bar{y}]_i - [y^*(t)]_i| \leq O(\|\eta(\bar{x}, \bar{y}, t)\|)$. This implies that as $y_0 \to y^*(t)$, eventually $|[\bar{y}]_i - [y^*(t)]_i| < \eta(\bar{x}, \bar{y}, t)^\gamma$, which is only possible, by the definition of $\mathcal{A}_{+,\epsilon}$ (3.3), if $y^*(t)_i \neq 0$.

Proposition 4.1 implies that for Δt small enough, it holds that $\mathcal{A}_+(x^*(t), y^*(t), t) = \mathcal{A}_+(x^*(t+\Delta t), y^*(t+\Delta t), t+\Delta t)$.

The SSOSC, Kojima [21, Corollary 7.2] and Hager [15, Lemma 3] imply that (3.4) is convex and therefore has a unique solution.

Consider a standard QP centered at $(\bar{x}, \bar{y})$,

$$\begin{aligned} \underset{\Delta x}{\text{minimize}} \quad & \nabla f(\bar{x}, t+\Delta t)^T \Delta x + \tfrac{1}{2}\Delta x^T H(\bar{x}, \bar{y}, t+\Delta t)\Delta x \\ \text{subject to} \quad & c_i(\bar{x}, t+\Delta t) + \nabla c_i(\bar{x}, t+\Delta t)^T \Delta x \geq 0, \text{ for all } i. \end{aligned} \tag{4.1}$$

With $(\bar{x}, \bar{y})$ sufficiently close to $(x^*, y^*)(t)$ and Δt small enough so that $(x^*, y^*)(t)$ is sufficiently close to $(x^*, y^*)(t+\Delta t)$. we may consider this QP to be a perturbation of (1.1) at $t + \Delta t$. By Kojima [21, Corollary 7.2], there is a solution $(\widehat{x}, \widehat{y})$ of (4.1) such that the active set at $(\widehat{x}, \widehat{y})$ contains $\mathcal{A}_+(x^*(t+\Delta t), y^*(t+\Delta t, t+\Delta t)$. Since $(\widehat{x}, \widehat{y})$ also, then, solves (3.4), this is in particular the solution generated by the algorithm.

Quadratic contraction, i.e.,

$$\eta(\widehat{x} + \Delta t, \widehat{y} + \Delta t, t+\Delta t) \leq C_q(t)\eta(\widehat{x}, \widehat{y}, t+\Delta t)^2, \tag{4.2}$$

follows from the local convergence of SQP under the SMFCQ and the SOSC, see, for example, Bonnans [2], and Lemma 1.1

∎

Theorem 4.2. *If the initial point (x_0, y_0) is sufficiently close to some primal-dual solution $(x^*, y^*)(0)$, Algorithm 3.1 generates a sequence of points (x_k, y_k, t_k) satisfying either $0 = t_0 < t_1 < t_2 < ...$, or $0 = t_0 < t_1 < t_2 < ... < t_N = 1$. Furthermore, there exist $(t_1, t_2, t_3, ...)$ such that for all k if $\delta(x_0, y_0) \leq \delta_0$ for some δ_0, it holds that,*

$$\delta(x_k, y_k, t_k) \leq \delta(x_0, y_0, 0) \tag{4.3}$$

Proof. Consider (x_1, y_1, t_1). By the continuity of the primal-dual solution path, Proposition 4.1, we have that for any ϵ_1, there exists t_1 such that,

$$\delta(x_0, y_0, t_1) = \|(x_0, y_0) - (x^*, y^*)(t_1)\| \leq \epsilon_1. \tag{4.4}$$

Let $\epsilon_1 = \min(\delta(x_0, y_0, 0), C_1(t_1)/(C_2^2(t_1)C_q(t_1)))$, where C_1, C_2, and C_q are constants arising from Lemma 1.1 and (4.2).

There exists a t_1 and $\delta(x_0, y_0, 0)$ sufficiently small such that the conditions of Theorem 4.1 are satisfied for t_1,

$$\eta(x_1, y_1, t_1) \leq \eta(x_0, y_0, t_1)^2.$$

and condition 6 in Algorithm 3.1 is satisfied, implying, after applying Lemma 1.1 (4.4) and (4.2),

$$\begin{aligned}
\delta(x_1, y_1, t_1) &\leq (1/C_1(t_1))\eta(x_1, y_1, t_1) \leq (C_q(t_1)/C_1(t_1))\eta(x_0, y_0, t_1)^2 \\
&\leq (C_2^2(t_1)C_q(t_1)/C_1(t_1))\delta(x_0, y_0, t_1)^2 \\
&\leq (C_2^2(t_1)C_q(t_1)/C_1(t_1))\min(1/(C_2^2(t_1)C_q(t_1)/C_1(t_1)), \delta(x_0, y_0, 0))\delta(x_0, y_0, 0) \\
&\leq \delta(x_0, y_0, 0).
\end{aligned}$$

Choosing t_2 to be such that $\delta(x_1, y_1, t_2)$ satisfies $\delta(x_1, y_1, t_2) \leq \min(1/(C_2^2(t_2)C_q(t_2)/C_1(t_2)), \delta(x_1, y_1, t_1))$, the same argument implies that,

$$\delta(x_2, y_2, t_2) = O(\delta(x_1, y_1, t_2)^2) \leq \delta(x_0, y_0, 0),$$

and the procedure can be continued for arbitrary k, proving the estimate (4.3). ∎

There are two points that must be made about the previous result. First, note that it may happen that the Algorithm will iterate and produce a series of approximate solutions at $t = 1$ even without tracking ($\delta(x_k, y_k, t_k) \leq \delta(x_0, y_0, 0)$). As it is still along the radius of local convergence, this is a feature rather than a drawback, i.e., it takes the largest step along the homotopy it can, thus iterating across solutions faster. If needed, the time saved in solving intermediate problems can be applied to iterate the last obtained solution locally to refine the estimated solution.

Also, it could happen that the Algorithm generates an infinite sequence. If the steps Δt get repeatedly smaller, it could happen that $\lim_{N\to\infty} \sum_i^N \Delta t_i < 1$ and the algorithm never reaches $t = 1$, the ultimately desired problem. We state a theorem indicating the necessary assumptions to prevent this from happening.

Theorem 4.3. *If*

- *each path* $(x^*, y^*)(t)$ *is globally Lipschitz continuous along* $t \in [0, 1]$.
- *the domain of quadratic attraction for* (1.1)*, i.e., the requisite maximal distance of* (x_0, y_0) *to* $(x^*, y^*)(t)$ *for quadratic local convergence, is uniformly bounded from below with respect to* t*, and the constant of quadratic convergence* $C_q(t)$ *in* (4.2) *is bounded above by* $\bar{C}_q$*, and*
- *the constants* $C_1(t)$ *and* $C_2(t)$*, and their inverses in Lemma 1.1 are uniformly bounded from above by* $\bar{C}$ *for* $t \in [0, 1]$*,*

then Algorithm (3.1) *terminates after a finite number of iterations* N *and, there exists a* Δt *such that* $t_k - t_{k-1} \leq \Delta t$ *implies that,*

$$\delta(x_N, y_N, 1) \leq \delta(x_0, y_0, 0) \tag{4.5}$$

Proof. The second assumption of the Theorem implies that quadratic contraction requires, for some $C_3 \leq 1$,

$$\delta(x_k, y_k, t_{k+1}) \leq C_3.$$

Let $\delta(x_0, y_0, 0)$ and $\|p(t_k) - p(t_{k-1})\|$ be such that, with L the Lipschitz constant associated with the first assumption of the Theorem,

$$\delta(x_0, y_0, 0) + L\|p(t_k) - p(t_{k-1})\| \leq 2\delta(x_0, y_0, 0) \leq \min(C_3, 4\bar{C}^3 C_q). \tag{4.6}$$

Let it hold that,

$$\delta(x_0, y_0, t_1) = \|(x_0, y_0) - (x^*, y^*)(t_1)\| \leq 2\delta(x_0, y_0, 0). \tag{4.7}$$

and it follows that,

$$\begin{aligned} \delta(x_1, y_1, t_1) \quad & \leq \bar{C}\eta(x_1, y_1, t_1) \leq \bar{C}\bar{C}_q\eta(x_0, y_0, t_1)^2 \leq \bar{C}^3\bar{C}_q\delta(x_0, y_0, t_1)^2 \\ & \leq 4\bar{C}^3\bar{C}_q\delta(x_0, y_0, 0)^2 \leq \delta(x_0, y_0, 0). \end{aligned}$$

Likewise, for any k,

$$\delta(x_k, y_k, t_k) \leq \delta(x_0, y_0, 0).$$

Finite termination follows from the fact that Lipschitz continuity of $(x^*, y^*)(t)$ implies that there is a constant $\widetilde{t}$ such that for $t_k - t_{k-1} \geq \widetilde{t}$, $\|p(t_k) - p(t_{k-1})\|$ satisfies (4.6).

∎

Since stability is a local property, the assumptions of the previous Theorem should appear appropriate, since a priori we do not know whether there is uniform stability across the homotopic solution path. In general we expect that for well-behaved problems, these assumptions should hold. The domain of Newton-like convergence, and the constants appearing in error bounds are nuanced topics relatively unexplored in the context of nonlinear programming and so we do not, in this paper, analyze these assumptions in more detail.

5. Algorithms for degenerate problems

In this section, we analyze problems for which the SMFCQ does not hold. Furthermore we relax the condition of the SSOSC to just the SOSC. The motivation is two-fold. First, although traditionally one perspective on degenerate problems is that it is due to a faulty presentation of the nonlinear programming problem, we disagree with this perspective for three reasons: 1) in an interest to make optimization software widely applicable, understanding abstract conditions should not necessarily be in the domain of the end user 2) it can be difficult to identify, a priori, the presence of degeneracy at the solution for nonlinear constraints and 3) some problems, like MPECs, cannot be reformulated in a way satisfying the SMFCQ.

Second, even if the nonlinear program at $t = 0$ and $t = 1$ satisfy the MFCQ, that does not mean that the SMFCQ is satisfied along every point on the homotopy path between these two. Consider, for instance, the problem,

$$\begin{aligned} \underset{x\in\mathbb{R}^2}{\text{minimize}} \quad & x_2 + x_1^2 \\ \text{subject to} \quad & x_2 \geq x_1, \\ & x_2 \geq -x_1, \\ & x_2 \geq p. \end{aligned} \tag{5.1}$$

as p changes from -1 to 1. It is clear that for $p \in [-1, 0)$, $x^*(p) = (0, 0)$ and $y^* = (1/2, 1/2, 0)$, and for $p \in (0, 1]$, $x^*(p) = (0, p)$ and $y^* = (0, 0, 1)$, however, for $p = 0$, $x^*(p) = (0, 0)$ but $y^*(p) = (\widetilde{y}, \widetilde{y}, 1 - \widetilde{y})$ for all $\widetilde{y} \in [0, 0.5]$.

Algorithm 3.1 will, for $p < 0$, correctly estimate the two strongly active constraints $x_2 \geq -x_1$ and $x_2 \geq x_1$, on which there is a unique feasible point $x = (0, 0)$. At $p = 0$ and $(\bar{x}, \bar{y}) = (0, 0)$, the multipliers satisfying the solution of the QP subproblem correspond to the optimal multipliers $\Lambda(x^*(0))$ of the original problem. However, since it can be difficult to confirm that any particular multiplier is "preferable" to the QP solver, there is no particular reason to expect it to select $y^* = (0, 0, 1)$, and subsequently estimate only the third constraint as strongly active, as is necessary to continue to trace the solution path for $p > 0$.

For the remainder of this and the subsequent section, we make the following assumption about (1.1),

Assumption 5.1. *For every $t \in [0, 1]$, the problem (1.1) has at least one primal-dual solution $(x^*, y^*)(t)$ to the first-order KKT conditions, i.e., the Lagrange multiplier set for the problem is nonempty. Furthermore, for each $x^*(t)$, there exists a compact subset $\mathcal{V}(x^*, t) \subseteq \Lambda(x^*, t)$ of the set of multipliers such that for all $y^* \in \mathcal{V}(x^*, t)$ the second-order sufficiency condition holds at (x^*, y^*). Furthermore the affine hulls of $\mathcal{V}(x^*, t)$ and $\Lambda(x^*, t)$ are the same.*

Note that by Robinson [33], Assumption 5.1 implies that any primal minimizer x^* is strict. So although we allow dual nonuniqueness, we do not consider the case of locally nonunique primal solutions. This is motivated by the intended context; optimization problems arising in engineering (whether standalone or arising from optimal control) tend to have a regularization term that have the effect of distinguishing particular primal minima. So although we are aware of developments in

the formulation of algorithms suitable for primal local non-uniqueness of solutions (e.g., Facchinei et al. [6]), we do not incorporate them in our method.

Although the example of Problem (5.1) consists of degeneracy at just a single point along the homotopy, we will be interested in analyzing the most general case and allow for degeneracy along the entire solution path. Thus, in contrast to Algorithm 3.1, we solve an *equality* constrained stabilized QP, using the active set estimate (2.1). Note that the inequality constrained stabilized SQP procedure theoretically achieves superlinear convergence under weak regularity assumptions (see, e.g., [8]) however, it is shown only that there exists a sequence of solutions converging superlinearly, and since these QPs are possibly indefinite with possibly multiple or unbounded solutions, it is not a well-defined algorithm. We may, as in Section 3, assume the SSOSC and solve a stabilized QP on the strongly active constraints, and this is one plausible procedure. We are interested, however, in problems satisfying just the regular SOSC, and under Assumption 5.1, QPs at a base point arbitrarily close to the solution can result in nonconvex QPs if there are any inequality constraints. For a thorough discussion of the challenges of indefiniteness in the context of perturbed QPs, see Chapter 5 of [22].

5.1. Calculating the extreme set of multipliers

As illustrated by problem (5.1), degeneracy can correspond to discontinuities in the optimal multiplier set path, which can correspond to changes in the optimal active set. Therefore, it becomes necessary to explore the extreme points of the region $\Lambda(x^*)$, where by extreme points we refer to the optimal multipliers with varying number of weakly active components.

To begin the algorithm, we use the active set estimate (2.1) to estimate the active set at $x^*(0)$. Then we solve a series of LPs along a combinatorial expansion of the possible set of weakly active multipliers. In particular, let $\{\aleph_k\} = \mathbb{P}(\mathcal{A}_\epsilon(x_0, y_0, 0))$, where $\mathbb{P}$ indicates the power set, the set of all subsets.

Let σ be a constant such that $0 < \sigma < \gamma < 1$. We solve the series of LPs,

$$\begin{aligned} \underset{\widetilde{y}\in\mathbb{R}^m}{\text{minimize}} \quad & \kappa \\ \text{subject to} \quad & 0 \le [\widetilde{y}]_i \le \kappa \quad \text{for all } i \in \{1, ..., m\} \\ & -\eta(x_0, y_0)^{\frac{1+\sigma}{1+\gamma}} \le [\nabla f(x_0, 0) - J(x_0, 0)^T \widetilde{y}]_i \le \eta(x_0, y_0)^{\frac{1+\sigma}{1+\gamma}}, \quad \text{for all } i \in \{1, ..., n\} \\ & \widetilde{y}_i = 0, \quad \text{for } i \notin \aleph_k, \end{aligned} \tag{5.2}$$

to generate a set of acceptable extremal multipliers. After pruning for redundancies (i.e., a multiplier appearing multiple times across $\aleph_k$), we obtain a set of initial points $\{(x_0, \widetilde{y}_k)\}$ from which to calculate a predictor step. Each $\aleph_k$ serves as an estimate of $\mathcal{A}_+(x^*(0), y_k^*(0), 0)$ for some $y_k^*(0)$, with $\mathcal{A}_\epsilon(x_0, y_0) \setminus \aleph_k$ an estimate for $\mathcal{A}_0(x^*(0), y_k^*(0), 0)$.

Note that it is immediately clear that this procedure will generate $\widetilde{y} = (0, 0, 1)$ as one of the multipliers at $p = 0$ for Problem (5.1).

5.2. Calculating the solution to the predictor equations

Recall that the equations (2.3) are stationarity conditions for the QP,

$$\begin{array}{ll} \underset{\Delta x}{\text{minimize}} & \left(\frac{\partial^2 L(\bar{x},\bar{y},\bar{t})}{\partial x \partial t}\Delta t\right)^T \Delta x + \frac{1}{2}\Delta x \frac{\partial^2 L(\bar{x},\bar{y},\bar{t})}{\partial^2 x}\Delta x, \\ \text{subject to} & \frac{\partial c(\bar{x},\bar{t})_+}{\partial t}\Delta t + \frac{\partial c(\bar{x},\bar{t})_+}{\partial x}\Delta x = 0, \\ & \frac{\partial c(\bar{x},\bar{t})_0}{\partial t}\Delta t + \frac{\partial c(\bar{x},\bar{t})_0}{\partial x}\Delta x \geq 0. \end{array} \tag{5.3}$$

However, unlike in the case of strong regularity, minimizing the QP does not correspond to finding these stationary solutions. The curvature along such points is not important for calculating a predictor and so a QP solver with provisions to ensure descent in the course of solving the quadratic program may not be fully appropriate. One alternative would be to solve a series of LPs. There are a finite number of weakly active constraints, and the series of inequalities can be solved for each combination of $y_i = 0$ or $[\frac{\partial c(\bar{x},\bar{t})_0}{\partial t}\Delta t + \frac{\partial c(\bar{x},\bar{t})_0}{\partial x}\Delta x]_i = 0$ for each $i \in \mathcal{A}_0$.

In addition, there may be an infinite set of solutions to (2.3). In fact, this is the more common case if the set of multipliers satisfying the first-order optimality conditions at $\bar{x}$ is not unique, see the examples in Levy [24], for instance. Therefore, we regularize the direction in order to certify a unique solution. Considering $\widetilde{y}_k$ for a permutation $\aleph_k$, we obtain another set of permutations $\bar{\aleph}_{kj} \in \mathbb{P}(\mathcal{A}_\epsilon(\bar{x},\widetilde{y}_k,t) \setminus \aleph_k)$, and solve a series of problems of the form,

$$\begin{array}{ll} \underset{\Delta x,\Delta y}{\text{minimize}} & \|\Delta x\| + \|\Delta y\| \\ \text{subject to} & \frac{\partial^2 L(\bar{x},\widetilde{y}_k)}{\partial x \partial t}\Delta t + \frac{\partial^2 L(\bar{x},\widetilde{y}_k)}{\partial^2 x}\Delta x + J(\bar{x},\bar{t})^T\Delta y = 0, \\ & \frac{\partial c(\bar{x},\bar{t})_+}{\partial t}\Delta t + \frac{\partial c(\bar{x},\bar{t})_+}{\partial x}\Delta x = 0, \\ & \frac{\partial c(\bar{x},\bar{t})_0}{\partial t}\Delta t + \frac{\partial c(\bar{x},\bar{t})_0}{\partial x}\Delta x \geq 0, \\ & [\Delta y]_{\mathcal{A}_0} \geq 0, \\ & [\Delta y]_i = 0 \text{ for } i \in \bar{\aleph}_{kj}, \\ & [\frac{\partial c(\bar{x},\bar{t})_0}{\partial t}\Delta t + \frac{\partial c(\bar{x},\bar{t})_0}{\partial x}\Delta x]_i = 0 \text{ for } i \notin \bar{\aleph}_{kj}, \end{array} \tag{5.4}$$

for each $\bar{\aleph}_{kj}$, where all the constraints correspond to only the estimated active inequality constraints.

We may use either the ∞ or the $1-$norm in the subproblem to obtain an LP. The first encourages an "interior" solution, many multipliers change, and the latter encourages a "sparse" solution, where many multipliers are expected to stay constant.

The number of solutions to (5.4) may be larger than the number of true differentiable solution paths, and so we cut out any solutions that do not correspond to actual approximate solutions at $t + \Delta t$ by discarding all solutions $(\widehat{x},\widehat{y})_{kj}$ that do not satisfy,

$$\eta(\widehat{x}_{kj},\widehat{y}_{kj},t+\Delta t) \leq \eta(\bar{x},\widetilde{y}_k,t)^{1-\frac{\sigma}{2}} \tag{5.5}$$

for the solution $(\widehat{x}_{kj},\widehat{y}_{kj}) = (\bar{x}+\Delta x, \widetilde{y}_k + \Delta y)$ for $\bar{\aleph}_{kj}$. If the condition does not hold, we discard this solution. Note if $(\widehat{x}_{kj},\widehat{y}_{kj})$ is calculated to be the step along the

directional derivative of a solution path that is at least C^2, then condition (5.5) holds for sufficiently small Δt by the fact that Lemma 1.1 implies that $\eta(\widehat{x}_{kj}, \widehat{y}_{kj}, t+\Delta t) \leq C_2(t+\Delta t)\delta(\widehat{x}_{kj}, \widehat{y}_{kj}, t+\Delta t) = C_2(t+\Delta t)\delta(\bar{x}, \widetilde{y}_k, t) + o(\Delta t) = O(\eta(\bar{x}, \widetilde{y}_k, t)) + o(\Delta t)$.

It appears as though this procedure, in the worst case complexity, is exponential in the number of active constraints. However, it should be noted that each of the LPs in Section 5.1, as well as all of the LPs corresponding to each $\aleph_k$ can be performed entirely in parallel, and thus the extremal multiplier and predictor calculations are perfectly parallelizable.

5.2.1. Contingencies

There are several possibilities for the solution of the predictor,

1. There are no feasible points of (5.4) and $\aleph_k = \emptyset$.
2. There are no feasible points of (5.4) and $\aleph_k \neq \emptyset$.
3. There is a solution to (5.4).

Izmailov [17, Theorem 2.4] suggests that the second case is pathological. In general, however, the Theorem also implies that strictly complementary multipliers are only certifiably stable under the linear independence constraint qualification (LICQ), suggesting that in the case of degeneracy, the first case is not unlikely. However, in this case, there are expected to be no active set changes, all constraints stay strongly active, and so a primary purpose of the predictor of estimating active set changes is irrelevant. In this case, we do not attempt to form a predictor step and move on to the corrector. In particular $(\widehat{x}_{kj}, \widehat{y}_{kj}) = (\bar{x}, \widetilde{y}_k)$ for all j, again discarding if 5.5 is unsatisfied.

5.3. Corrector step

Since we are considering the context of assuming no constraint qualifications, classical results implying that subproblems with a base point sufficiently close to a primal-dual solution (see, e.g., Robinson [30]) are not directly applicable. In particular, if the properties of strict complementarity and the full rank of the constraint Jacobian matrix are not satisfied, the solution to quadratic subproblems may not necessarily have an active set corresponding to the active set at the solution. In addition, the KKT matrix used to solve the QP subproblem,

$$K = \begin{pmatrix} H(x,y,t) & J(x,t)^T \\ J(x,t) & 0 \end{pmatrix},$$

may even be singular.

We propose a strategy whereby we estimate the active set and solve a perturbed equality constrained SQP subproblem, namely the stabilized SQP subproblem. Equality-constrained stabilized SQP has been shown [18] to converge superlinearly under assumptions weaker than the ones made in this paper.

This algorithm resembles that of Wright [40], wherein an outer algorithm is performed until a point with a sufficiently low optimality residual is generated, and then the active set is estimated and the equality constrained stabilized SQP is solved with the estimated active set corresponding to the equality constraints. If the constraints and dual variables remained feasible and the optimality residual sufficiently reduced, the equality sSQP step is considered successful.

5.3.1. Stabilized SQP iteration

After calculating the predictor step, we recalculate the active set estimate, $\mathcal{A}_\epsilon(\widehat{x}_{kj}, \widehat{y}_{kj}, t+\Delta t)$, potentially picking up new active constraints.

Solve,

$$\begin{pmatrix} H(\widehat{x}_{kj}, \widehat{y}_{kj}, t+\Delta t) & J(\widehat{x}_{kj}, t+\Delta t)_A^T \\ J(\widehat{x}_{kj}, \Delta t)_A & -\eta(\widehat{x}_{kj}, \widehat{y}_{kj}, t+\Delta t)I \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta y_A \end{pmatrix} = \begin{pmatrix} -\nabla_x L(\widehat{x}_{kj}, \widehat{y}_{kj}, t+\Delta t) \\ -c(\widehat{x}_{kj}, t+\Delta t)_A \end{pmatrix}, \tag{5.6}$$

where the subscript A corresponds to the components in the respective matrix or vector in $\aleph_k \cup (\mathcal{A}_\epsilon(\bar{x}, \widetilde{y}_k, t) \setminus \bar{\aleph}_{kj})$. We set $[\Delta y]_{A^c} = 0$.

The solution $(\Delta x, \Delta y)$ is accepted if there is sufficient superlinear contraction and if it remains feasible with respect to the inactive constraints,

$$\begin{aligned} &\eta(\widehat{x}_{kj} + \Delta x_{kj}, \widehat{y}_{kj} + \Delta y_{kj}, t+\Delta t) \leq \eta(\widehat{x}_{kj}, \widehat{y}_{kj}, t+\Delta t)^{1+\gamma}. \\ &c(\widehat{x}_{kj} + \Delta x_{kj})_i \geq 0 \text{ for } i \notin \mathcal{A}_\epsilon(\widehat{x}_{kj}, \widehat{y}_{kj}, t+\Delta t) \end{aligned} \tag{5.7}$$

5.3.2. Recalculating the extremal multipliers

The corrector step cannot be limited to an iteration involving the primal-dual solution to this equality-constrained problem. In particular, the multipliers can switch sign so that the component of a resultant dual variable becomes negative and thus, non-optimal. The set of multipliers satisfying the optimality conditions for the strictly equality-constrained problem may be larger than the set of multipliers satisfying the optimality conditions for the full inequality-constrained problem. It could be that the step corresponding to the equality-constrained problem results in an inappropriate multiplier, but there exists a multiplier that results in an appropriate decrease in the optimality conditions.

Furthermore, as this is the end of the procedure, we must again recognize that the set of optimal multipliers may not be a singleton and in order to capture the active set changes at the next iteration, we will need to calculate the set of extremal multipliers.

So we perform the same procedure as described in Section 5.1. In particular, we

solve, for each member of the power set $\aleph_{kjl} \in \mathbb{P}(\mathcal{A}_\epsilon(\widehat{x}_{kj}, \widehat{y}_{kj}, t + \Delta t))$

$$
\begin{aligned}
\underset{\widetilde{y} \in \mathbb{R}^m}{\text{minimize}} \quad & \kappa \\
\text{subject to} \quad & -\kappa \leq [\widetilde{y}]_i \leq \kappa \ \text{ for all } i \in \{1, ..., m\} \\
& -\eta(\widehat{x}_{kj} + \Delta x_{kj}, \widehat{y}_{kj} + \Delta y_{kj})^{1+\sigma} \\
& \qquad \leq [\nabla f(\widehat{x}_{kj} + \Delta x_{kj}, t + \Delta t) - J(\widehat{x}_{kj} + \Delta x_{kj}, t + \Delta t)^T \widetilde{y}]_i \\
& \ \eta(\widehat{x}_{kj} + \Delta x_{kj}, \widehat{y}_{kj} + \Delta y_{kj})^{1+\sigma} \\
& \qquad \geq [\nabla f(\widehat{x}_{kj} + \Delta x_{kj}, t + \Delta t) - J(\widehat{x}_{kj} + \Delta x_{kj}, t + \Delta t)^T \widetilde{y}]_i \\
& \quad \text{for all } i \in \{1, ..., n\} \\
& 0 \leq \widetilde{y}, \\
& \widetilde{y}_i = 0, \ \text{for } i \notin \aleph_{kjl},
\end{aligned}
\tag{5.8}
$$

Note that we have a combinatorial expansion, at each iteration there are a number of initial multipliers from which we calculate a number of predictor steps, and at each we calculate a corrector step and then a set of new extremal multipliers. In the case of degeneracy, a unique solution path $x^*(t)$ for $t \in [0, 1]$ is not guaranteed. As such, depending on the desired objective, a user of the algorithm could either 1) trace all unique valid solution paths or 2) select one to follow on a homotopy path. In the next section, we present the outline of the algorithm for the second case. Obtaining all of the solution paths would correspond to a recursive tree of instances of the procedure. Since the combinatorial expansion is entirely parallelizable on each level, however, neither objective is burdensome from the perspective of computational cost.

5.4. Statement of the algorithm

Algorithm 5.1 Predictor-corrector algorithm for degenerate parametric nonlinear programming.

Begin with $\bar{x}$, a solution to (1.1) at $t = 0$ and constants $\gamma, \gamma_1 \in (0,1)$, $\gamma_2 > 1$.
Solve (5.2) for each $\aleph_k$ permutation of weakly active multipliers.
while $t < 1$ **do**
  for each k **do**
    Solve (5.4) for each $\bar{\aleph}_{kj}$ to obtain $\{\widehat{x}, \widehat{y}\}_{kj}$.
    Discard all $(\widehat{x}, \widehat{y})_{kj}$ that do not satisfy (5.5).
    if $\{(\widehat{x}, \widehat{y})_{kj}\} = \emptyset$ **then**
      Let $(\widehat{x}_{kj}, \widehat{y}_{kj}) = (\bar{x}, \widetilde{y}_k)$ for all j.
    end if
    for each (k, j) such that $(\widehat{x}, \widehat{y})_{kj}$ exists **do**
      Calculate $\mathcal{A}_\epsilon(\widehat{x}_{kj}, \widehat{y}_{kj}, t + \Delta t)$ by (2.1).
      Solve (5.6) for $(\Delta x_{kj}, \Delta y_{kj})$.
      if $(\Delta x_{kj}, \Delta y_{kj})$ satisfies (5.7) **then**
        Solve (5.8) for each $\aleph_{kjl} \in \mathbb{P}(\mathcal{A}_\epsilon(\widehat{x}_{kj}, \widehat{y}_{kj}, t + \Delta t))$ to obtain $\{\widetilde{y}_{kjl}\}$.
      end if
    end for
  end for
  if no satisfactory $(\Delta x_{kj}, \widetilde{y}_{kjl})$ is found for all l, k, j **then**
    Let $\Delta t = \gamma_1 \Delta t$, go back to Step 5.
  else
    For some k, j, for all l with a satisfactory $(\Delta x_{kj}, \widetilde{y}_{kjl})$,
      let $\bar{x} = \widehat{x}_{kj} + \Delta x_{kj}$, $\{\aleph_k\} = \{\aleph_{kjl}\}$, $\{\widetilde{y}_k\} = \{\widetilde{y}_{kjl}\}$, $t = t + \Delta t$.
    Let $\Delta t = \max(1 - t, \gamma_2 \Delta t)$.
  end if
end while

6. Convergence

We present three general assumptions that facilitate the results presented in this section. All of these assumptions concern the properties of homotopic paths for (1.1) between $t = 0$ and $t = 1$. As such, they are higher-level, a posteriori assumptions. Consider the set valued mapping,

$$S : t \to X^*(t) = \{x^*(t) : \text{ there exists a } y^* \text{ such that } (x^*(t), y^*) \text{ satisfy (1.3)}\},$$

Assumption 6.1. *The graph of S is arcwise-connected, i.e., for any $(x^*(t_0))$ and $(x^*(t_1))$ corresponding to solutions of (1.1) with $t = t_0$ and $t = t_1$, there exists a continuous curve $(x^*(t))$ connecting $x^*(t_0)$ and $x^*(t_1)$ such that $(x^*(t))$ solves (1.1) for t. Furthermore, for each t, every local minimizer $x^*(t)$ is isolated.*

The next assumption essentially implies that there are no "dead ends" in the homotopy path,

Assumption 6.2. *For every continuous path of primal solutions that includes $x^*(0)$, for every branch point $\widetilde{t}$, the number of solutions corresponding to $t > \widetilde{t}$ is always at least one.*

This effectively rules out turning points or boundary points in the continuation path. Every primal bifurcation point must be a simple branching, a pitchfork, a transcritical or otherwise a point for which there exists at least one primal solution to (1.1) for $t > \widetilde{t}$ arcwise connected to $x(\widetilde{t})$ (which, due to this assumption and Assumption 6.1, exists).

The next assumption ensures that each multiplier $\widetilde{y}_{kjl}$ the procedure finds is in the neighborhood of optimal multipliers that satisfy the second order conditions. The assumption essentially implies that all multipliers not satisfying the second order sufficiency condition are large in norm, i.e., the SOSC is only unsatisfied for "unbounded" multipliers.

Assumption 6.3. *Let $\kappa = \max_{(t\in[0,1],\aleph\in\mathbb{P}(\mathcal{A}(x^*(t))))} \min\{\|y\|_\infty : y \in \Lambda(x^*(t)), [y]_i = 0 \text{ for } i \in \aleph\}$. For every $y^* \in \Lambda(x^*(t))$ for all $x^*(t)$ and all t, with $\|y^*\|_\infty < 2\kappa$, (x^*, y^*) lies in the relative interior of $\mathcal{V}(x^*(t))$, with $\mathcal{V}(x^*(t))$ defined as in Assumption 5.1.*

Theorem 6.1. *For $(\widehat{x}, \widehat{y})$ sufficiently close to $(x^*, \text{int}(\mathcal{V}(x^*)))(t + \Delta t)$ with $\mathcal{V}(x^*)$ as in Assumption 6.1, there exists a solution to the system (5.6) satisfying the condition (5.7), and a solution to (5.8) for some $\aleph_{kjl}$*

Proof. Wright [40, Theorem 3.3] implies that $\mathcal{A}_\epsilon(\widehat{x}, \widehat{y}, t + \Delta t) = \mathcal{A}(x^*(t + \Delta t))$.

Consider the problem,

$$\min_x f(x, t+\Delta t)) \text{ subject to } c(x, t+\Delta t)_i = 0 \text{ for all } i \in A,\ c(x, t+\Delta t)_i \geq 0, \text{ otherwise,} \tag{6.1}$$

where $A = \mathcal{A}_\epsilon(\widehat{x}, \widehat{y}, t+\Delta t)$. Note that the solutions to each of these problems contain the primal-dual solutions $(x^*, \Lambda(x^*))$ of (1.1) for $t + \Delta t$.

Furthermore, Wright [40, Lemma 4.1] implies that this solution satisfies

$$\bar{\eta}(\widehat{x} + \Delta x, \widehat{y} + \Delta y, t + \Delta t) \leq \bar{\eta}(\widehat{x}, \widehat{y}, t + \Delta t)^2, \tag{6.2}$$

where $\bar{\eta}(x, y, t) = \left|\left|\begin{pmatrix} g(x,t) - J(x,t)^T y \\ c(x,t)_{\mathcal{A}_\epsilon} \end{pmatrix}\right|\right|$. Clearly by definition of η, since $[\widehat{y} + \Delta y]_{\mathcal{A}_\epsilon^c} = 0$, we have that $\eta(\widehat{x} + \Delta x, \widehat{y} + \Delta y, t + \Delta t) \leq \bar{\eta}(\widehat{x} + \Delta x, \widehat{y} + \Delta y, t + \Delta t)$. On the other hand, $\bar{\eta}(\widehat{x}, \widehat{y}, t + \Delta t) \leq \bar{\delta}(\widehat{x}, \widehat{y}, t + \Delta t)$, the distance to a solution of (6.1) by [40, Theorem 3.4] and $\bar{\delta}(\widehat{x}, \widehat{y}, t + \Delta t) \leq \delta(\widehat{x}, \widehat{y}, t + \Delta t)$, since $x^*(t + \Delta t)$ is isolated and the primal-dual solution set corresponding to $x^*(t + \Delta t)$ is no smaller for problem (6.1) than problem (1.1). Together, these facts, along with another application of Lemma 1.1, imply that and solution $(\Delta x, \Delta y)$ satisfies,

$$\eta(\widehat{x} + \Delta x, \widehat{y} + \Delta y, t + \Delta t) \leq \eta(\widehat{x}, \widehat{y}, t + \Delta t)^{1+\gamma}, \tag{6.3}$$

Izmailov and Solodov [18, Lemma 2] implies that the matrix in (5.6) is nonsingular and therefore there exists a unique solution to the system.

It holds that, for sufficiently small Δt and by Lemma 1.1, for $i \notin \mathcal{A}_\epsilon(\widehat{x}, \widehat{y}, t+\Delta t)$,

$$[c(\widehat{x}+\Delta x, t+\Delta t)]_i \geq [c(\widehat{x}, t+\Delta t)]_i - L\|\Delta x\| \geq \eta(\widehat{x}, \widehat{y}, t+\Delta)^\gamma - L\delta(\widehat{x}, \widehat{y}, t+\Delta t) > 0, \tag{6.4}$$

where we also use Wright [40, Lemma 4.1] for the estimate $\|\Delta x\| = O(\delta(\widehat{x}, \widehat{y}, t+\Delta t))$.

Now we claim that there exists a $\widetilde{y}$ such that,

$$\eta(\widehat{x} + \Delta x, \widetilde{y}, t + \Delta t) \leq \eta(\widehat{x}, \widehat{y}, t + \Delta t)^{1+\sigma}.$$

Consider $\widetilde{y} = \min_{y \in \Lambda(x^*, t+\Delta t)} \|\widehat{y} + \Delta y - y\|$. Note that $\|\widehat{y} + \Delta y - \widetilde{y}\| \leq \delta(\widehat{x} + \Delta x, \widehat{y} + \Delta y, t + \Delta t)$.

The optimality residual at the new primal-dual point satisfies, $\eta(x + \Delta x, \widetilde{y}, t + \Delta t) \leq 2\|\nabla f(\widehat{x} + \Delta x) - J(\widehat{x}, t + \Delta t)^T\widetilde{y}\| + 2\|\min(\widetilde{y}, c(\widehat{x} + \Delta x, t + \Delta t))\|$. We have already shown above that

$$\begin{aligned}
&\|\min(\widetilde{y}, c(\widehat{x} + \Delta x))\| \leq \|c(\widehat{x} + \Delta x, t + \Delta t)\| \leq \bar{\eta}(\widehat{x} + \Delta x, \widehat{y} + \Delta y, t + \Delta t) \\
&\quad = O(\eta(\widehat{x} + \Delta x, \widehat{y} + \Delta y, t + \Delta t)) = O(\eta(\widehat{x}, \widehat{y}, t + \Delta t)^{1+\gamma}) = o(\eta(\widehat{x}, \widehat{y}, t + \Delta t)^{1+\sigma}).
\end{aligned}$$

It holds that $\nabla f(x^*, t+\Delta t) - J(x^*, t+\Delta t)^T\widetilde{y} = 0$. Therefore, by Assumption 1.1 and Lemma 1.1

$$\begin{aligned}
&\|\nabla f(\widehat{x} + \Delta x, t + \Delta t) - J(\widehat{x}, t + \Delta t)^T\widetilde{y}\| \\
&\quad \leq \|\nabla f(\widehat{x} + \Delta x, t + \Delta t) - \nabla f(x^*, t + \Delta t) - (J(\widehat{x} + \Delta x, t + \Delta t) - J(x^*, t + \Delta t))^T\widetilde{y}\| \\
&\quad \leq L\delta(\widehat{x} + \Delta x, t + \Delta t) + L\|\widetilde{y}\|\delta(\widehat{x} + \Delta x, t + \Delta t) \\
&\quad \leq O(\eta(\widehat{x}, \widehat{y}, t + \Delta t)^{1+\gamma}) = o(\eta(\widehat{x}, \widehat{y}, t + \Delta t)^{1+\sigma}).
\end{aligned}$$

∎

The conditions implying a successful series of steps of Algorithm 5.1 are described in the next Theorem.

Theorem 6.2. *Assume for all $x^*(t)$ primal solutions of (1.1), it holds that there exists $y^* \in \Lambda(x^*(t))$ such that y^* satisfies $y^* \in \text{Argmin}\{\|y\|_\infty :\in \Lambda(x^*), [y]_i = 0 \text{ for } i \in \aleph\}$ for all $\aleph \in \mathbb{P}(\mathcal{A}(x^*(t)))$ and for at least one such y^* there exists a continuous path $y^*(s)$ for $s \in [t, t+\delta_s)$ for some δ_s with $y^*(t) = y^*$ and $(x^*(s), y^*(s))$ solving (1.1).*

Then, if the initial point (x_0, y_0) is sufficiently close to $(x^*, \mathcal{V}(x^*))(0)$, Algorithm 3.1 generates a sequence of points (x_k, y_k, t_k) satisfying either $0 = t_0 < t_1 < t_2 < ...$, or $0 = t_0 < t_1 < t_2 < ... < t_N = 1$. Furthermore, for any k, there exists $\{t_1, t_2 - t_1, t_3 - t_2, ...\}$ such that for sufficiently small initial $\delta(x_0, y_0, 0)$, it holds that,

$$\delta(x_k, y_k, t_k) \leq \delta(x_0, y_0, 0) \tag{6.5}$$

Proof. For (x_0, y_0) sufficiently close to $(x^*, \mathcal{V}(x^*))(0)$, it holds that $\mathcal{A}(x^*, 0) = \mathcal{A}_\epsilon(x_0, y_0, 0)$ and by a similar argument as in Theorem 6.1, there exists a $\widetilde{y}_k$ for some $\aleph_k \in \mathbb{P}(\mathcal{A}_\epsilon(x_0, y_0, 0))$ that solves (5.2).

By the stability of linear equations [31], we have that $\|\widetilde{y}_k - y_k^*\| = O(\delta(x_0, y_0, 0))$ where $y_k^* \in \text{Argmin}\{\|y\|_\infty : \in \Lambda(x^*), [y]_i = 0 \text{ for } i \in \aleph_k \in \Lambda(x^*)\}$. By the assumptions of the Theorem there exists $(x^*(s), y^*(s))$ solving (1.1) close to (x^*, y_k^*).

If $y^*(s)$ is also Lipschitz continuous, this implies the existence of a solution to (5.4) by Izmailov [17, Theorem 2.1]. The presence or absence of this case does not affect the argument, however.

For sufficiently small t_1, for some $\aleph_{kj}$ condition (5.5) implies that, $\mathcal{A}_\epsilon(\widehat{x}, \widehat{y}, t + \Delta t) = \aleph_k \cup (\mathcal{A}_\epsilon(\bar{x}, \widetilde{y}_k, t) \setminus \bar{\aleph}_{kj})$ for some k, j and by Wright [40, Theorem 3.3], $\mathcal{A}_\epsilon(\bar{x}_{kj}, \bar{y}_{kj}, t_1) = \mathcal{A}(x^*(t_1), y^*(t_1), t_1)$.

For sufficiently small t_1, by Assumption 6.3 and continuity of $y^*(s)$, $y^*(t_1)$ must lie in the relative interior of $\Lambda(x^*(t_1))$ and the conditions of Theorem 6.1 are satisfied at $(\widehat{x}, \widehat{y})(t_1)$ for sufficiently small t_1 and the procedure generates a point (x_1, y_1) such that, $\eta(x_1, y_1, t_1) \leq \eta(\widehat{x}, \widehat{y}, t_1)^{1+\sigma}$. The rest of the argument follows as in Theorem 4.2 and the fact that $\frac{1+\sigma}{1-\sigma/2} > 1$ implying superlinear contraction of $\eta(x, y, t)$ by (5.5) and (6.3). ∎

The existence of a path satisfying the conditions of a tangential predictor (2.3) is a nuanced subject with assumptions that are, for the purposes of this exposition, esoteric, and the interested reader can consult Izmailov [17]. We also do not include a proof of finite termination, as the number of assumptions required becomes excessive. This does suggest that successful homotopies may only be performed under small perturbations for degenerate problems, but similarly as in the conventional case, if the multiplier expansion/predictor successfully predicts newly inactive constraints, and the analogous conditions as in Theorem 4.3 along each segment of the homotopy path, and the domains of superlinear convergence are bounded from below across the entire path, then it can be seen that finite termination of Algorithm 5.1 should hold.

7. Numerical results

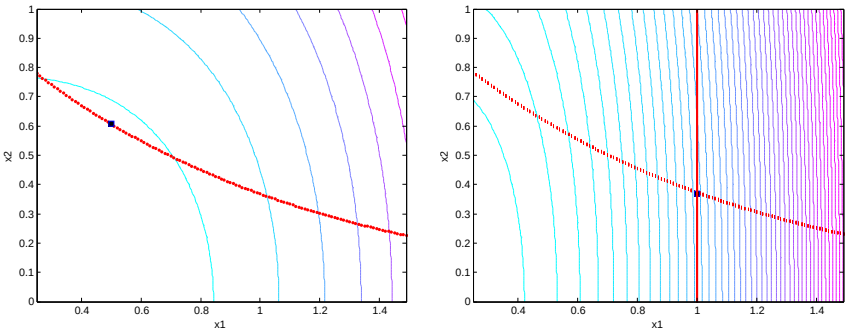
In this section we look at the performance of a MATLAB implementation of the algorithms formulated in this paper for several toy problems as well as a homotopy solution to a subset of the CUTEst test set. The first few problems will be illustrative of the properties and the needed conditions of the general algorithm, and the performance on the larger test set should give an overview of the overall performance of the algorithm. These results look encouraging, and with additional tuning, suggest that the algorithms presented in this paper can prove effective in solving the problems from the intended applications.

7.1. Problem 1: Active set change

For the first problem we illustrate Algorithm 3.1 on a simple two-dimensional nonlinear program.

$$\begin{aligned} \underset{x \in \mathbb{R}^2}{\text{minimize}} \quad & p_1 x_1^3 + x_2^2 \\ \text{subject to} \quad & x_2 - e^{-x_1} \geq 0, \\ & x_1 \geq p_2. \end{aligned} \tag{7.1}$$

Figure 3: Plot of the Problem 1 for $t = 0$ and $t = 1$.



We look at the problem starting at the approximate solution $(x_0, y_0) = ((0.5, 0.6), 1.2)$ to (7.1) with $p = (1, -4)$ tracing a path to generate an approximate solution for $p = (8, 1)$. The contour plots and constraints for the two variations of the problem are given in Figure 3.

Figure 4 plots x_1 with respect to t. Notice the steep linear change as the bound constraint becomes active. The procedure starts by reducing Δt, then achieves a contracted step, then again reduces Δt to absorb the new constraint. The problem at $t = 1$ is solved after 28 iterations in total. Algorithm 5.1 performs similarly, and since the problem is fairly simple, we do not investigate its specific properties. Although the problem is simple, the change in the problem is quite substantial in terms of both the nature of the active constraints and the grade of the objective function.

7.2. Problem 2: A degenerate problem

We look at the problem described by Wright [39], which we write as,

$$\begin{aligned} \underset{x \in \mathbb{R}^2}{\text{minimize}} \quad & x_1 \\ \text{subject to} \quad & (x_1 - 2)^2 + x_2^2 - p^2 \leq 0, \\ & (x_1 - 4)^2 + x_2^2 - (p + 2)^2 \leq 0. \end{aligned} \tag{7.2}$$

We analyze shifts in the problem close to $p = 2$. The problem is drawn in Figure 5.

The solution is at the edge of both of the circles, and increasing the value of p increases the size of the circles, while the solution remains on the leftmost edge of the circles, where they touch. The active constraint gradients are rank deficient at the solution for any p, and the SMFCQ condition does not hold. In this case, however, Algorithm 3.1 converges. It estimates one constraint as strongly active and proceeds to a solution.

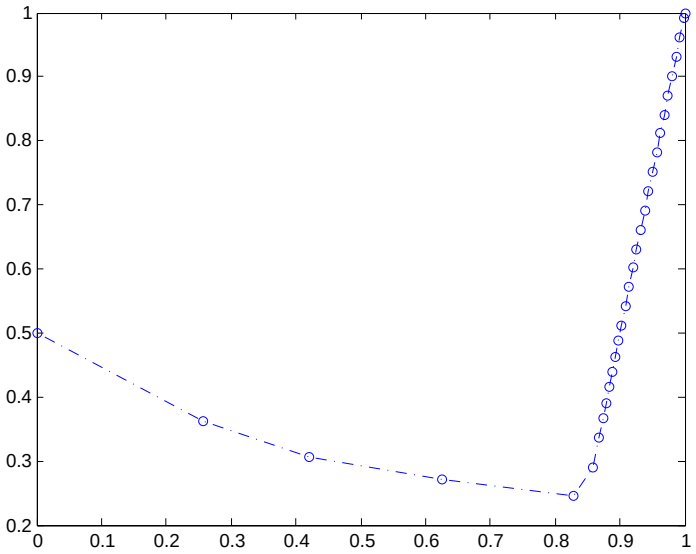


Figure 4: Problem 1: x_1 as a function of t.

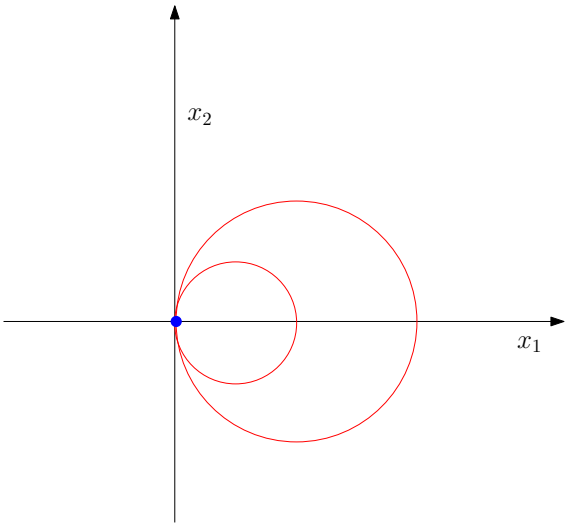
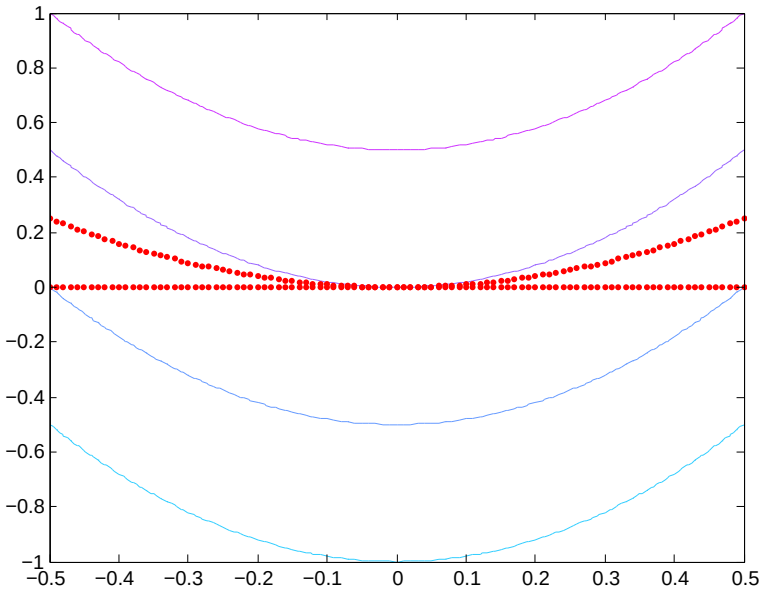


Figure 5: Problem 3.

Figure 6: Plot of the Problem 4 for $p = 0$.



We modify the problem as done in Mostafa et al. [27], to be of the form,

$$\begin{array}{ll} \underset{x\in\mathbb{R}^2}{\text{minimize}} & x_1 \\ \text{subject to} & (x_1-2)^2 + x_2^2 - p^2 \leq 0, \\ & -(x_1-4)^2 - x_2^2 + (p+2)^2 \leq 0. \end{array} \tag{7.3}$$

In this case, the MFCQ condition does not hold for any p, the set of multipliers is unbounded, with

$$y^*(p) = (y, y/2 - 1/(4p)), \text{ for any } y \geq 0.$$

Algorithm 3.1 does not converge, whereas the stabilized algorithm, Algorithm 5.1 described in Section 5 converges in 252 iterations. The sequence of points steadily converges, however, only with a small step-size, indicating a small radius of convergence. For the conventional algorithm, by contrast, degeneracy prevents a sufficiently contractive step from being generated by the subproblem.

7.3. Sensitivity to second order conditions

Consider the problem,

$$\begin{array}{ll} \underset{x\in\mathbb{R}^2}{\text{minimize}} & -2x_1^2 + x_2 \\ \text{subject to} & -x_1^2 + x_2 \geq p, \\ & x_2 \leq 0. \end{array} \tag{7.4}$$

We plot the problem in Figure 6. Consider attempting to solve a homotopy path

from $p = 0$ to $p = -1$. At $p = 0$, the optimal solution is $x^* = (0, 0)$ and $y^* = (1+\bar{y}, \bar{y})$ for $\bar{y} \geq 0$, and in fact $(0, 0)$ is the only feasible point for problem (7.4) at $p = 0$. The Hessian matrix is,

$$H(x^*, y^*) = \begin{pmatrix} -2 + 2\bar{y} & 0 \\ 0 & 0 \end{pmatrix},$$

which is negative definite for $\bar{y} \in [0, 1)$, and the Jacobian is $J(x^*) = \begin{pmatrix} 0 & 1 \\ 0 & -1 \end{pmatrix}$ for which the null-space is $v_n = \begin{pmatrix} \widetilde{v} \\ 0 \end{pmatrix}$ for all $\widetilde{v} \in \mathbb{R}$, implying that the second order sufficient conditions (both strong and regular)are not satisfied. It can be shown that the second order Guignard constraint qualification does not hold as well, implying that the second order necessary conditions are not necessary at a local minimizer.

As theoretically expected, for an initial point close to $x^* = (0, 0)$ and $y^* = (1, 0)$, Algorithm 3.1 attempts to solve an unbounded quadratic program, and the stabilized variation, Algorithm 5.1 fails to generate a step with the desired contraction rate, and neither achieve a successful iteration.

7.4. Tests on the CUTEr test set

We performed a series of tests on the CUTEr test set [13]. We selected a subset of the constrained "HS" problems [16], and, in order, for nonconvex problems, to attempt to find solutions satisfying the second order sufficient conditions, we used an algorithm known to converge to points satisfying the second order necessary conditions [12] to find solutions. We discarded three problems for which an optimal solution was not found, and two for which the second order sufficiency condition did not hold, leaving 108 problems for the test set. We then perturbed the primal-dual solution by some factor ξ with $\xi < |\bar{\xi}| < 1$, e.g., $(x_0, y_0) = (1+\xi)(x^*, y^*)$ and formed a homotopy problem. The homotopy was formed such that (x_0, y_0) is a solution for $t = 0$ and (x^*, y^*) is a solution for $t = 1$. This was done by changing the problem,

$$\begin{aligned} \underset{x \in \mathbb{R}^n}{\text{minimize}} \quad & f(x) - (1-t)(\nabla f(x_0) - J(x_0)^T y_0)^T x \\ \text{subject to} \quad & c(x) \geq (1-t)\min(0, c(x_0)), \end{aligned} \tag{7.5}$$

where min is component-wise.

We report the number of successful path-following to an approximate solution of (7.5) at $t = 1$ in Table 1 and the mean number of iterations for problems for which both algorithms were successful in Table 2. The median number of iterations was always 1, so the average iterations are the unstable mean of a highly skewed distribution. It appears that the stabilized variation is overall less robust than standard parametric SQP but generally requires fewer major iterations. Typical failed problems produce a large number successful tracking steps, after driving Δt to be very small. It should be noted that the nature of each iteration in the stabilized variation is different, however, and so the difference in the number of iterations does not necessarily indicate improved computational performance in all cases.

Since neither algorithm is supposed to be a standalone NLP solver, the results must be interpreted carefully. For many problems, the procedure is expected to

Table 1: Number of Successful Path-following homotopies to an approximate primal-dual solution (out of 108 problems).

Algorithm	$\bar{\xi} = 1$	$\bar{\xi} = 10^{-4}$	10^{-8}
Parametric SQP	58	86	94
Stabilized PSQP	33	76	86

Table 2: Average Number of Iterations for Successful Path-following Homotopies to an Approximate Primal-dual Solution, for Problems for Which Both Algorithms Converge.

Algorithm	$\bar{\xi} = 10^{-4}$	10^{-8}
Parametric SQP	6.38	22.96
Stabilized PSQP	6.62	12.28

yield a quadratic contraction in one inexpensive step, and sometimes a few, in a relatively simple algorithm. Thus, in a procedure to solve a series of related NLPs, the homotopy approach can be attempted as a first choice, and if it appears that the problems happen to be so far apart as to require an excessive number of steps, one can revert to some reliable NLP solver.

In Figure 7, we show a performance profile of parametric SQP, stabilized parametric SQP and SNOPT for $\bar{\xi} = 10^{-4}$. A performance profile provides an "at a glance" comparison of a set of algorithms on a large test set. Let $r_{p,s}$ be the ratio of major iterations for solver s as compared to the best-performing solver on problem p. Let $|\mathcal{A}|$ denote the size of a set. The function,

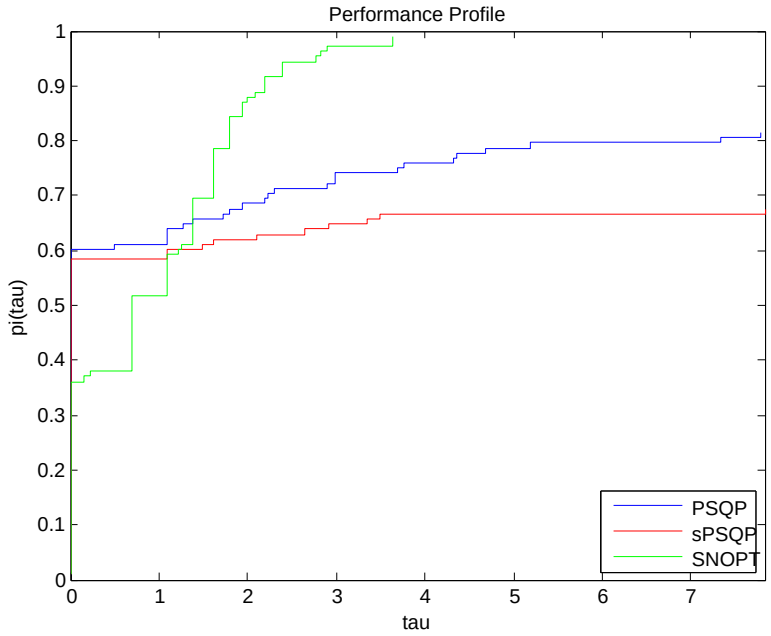
$$\pi_s(\tau) = \frac{1}{|\mathcal{P}|}|\{p \in \mathcal{P} : \log_2(r_{p,s}) \leq \tau\}|,$$

expresses the proportion of problems that are solved in at worst 2^τ iterations times the number of iterations the best solver takes. The performance profile plots $\pi_s(\tau)$ as a function of τ for the different solvers. A solver starting and initially staying at a comparatively a high-value on the vertical axis indicates a fast solver, and a solver that has a comparatively high value on the vertical axis for larger values on the horizontal axis represents a reliable solver.

In this case we measure the performance with respect to the number of major iterations. Since the task of parametric SQP is to generate an approximate solution of (1.1) at $t = 1$, we set the optimality tolerance for SNOPT to be the best optimality level reached by either PSQP or sPSQP, or if neither converged, we set it to 10^{-8}.

SNOPT solves all of the HS problems. Both parametric SQP and the stabilized variation, however, are faster, as they are the fastest performers on the vast majority of the problems. This suggests that as a strategy to solve a problem given

Figure 7: Performance profile of parametric SQP, stabilized parametric SQP and SNOPT for $\bar{\xi} = 10^{-4}$.



a solution to a similar problem, the homotopy procedures described in this paper are expected to exhibit good performance, relative to the conventional approach of using conventional SQP solvers for each nonlinear program in the sequence.

8. Conclusion and future work

Thus far, parametric nonlinear programming has been primarily a theoretical study of the local properties of solutions subject to a parameter, or heuristics for solving a series of related NLPs. Detailed, thorough study of the properties of nonlinear programs dependent on a parameter for the purpose of algorithm development is a largely unexplored subject in the context of SQP methods.

We developed and analyzed an SQP-based algorithm for the topic. It is well known that SQP is effective for solving a series of related NLPs due to its "hot/warm-start" advantages, and the ideas developed here largely formalized these notions. Exact Hessians were used throughout, because since we are able to estimate the active set at the optimal solution, indefiniteness is no longer a concern.

We noted the assumptions necessary to show solvability and finite termination of any homotopy-SQP procedure. Some of these are non-standard in the optimization literature, and thus point to a large number of additional future theoretical and applied research topics. In particular, these include studying the domain of convergence for (1.1) as a function of t. The Assumptions for the degenerate case, Assumptions 6.1 and 6.2 are relatively "high-level" in that they are about the solu-

tion path themselves. A study of the regularity properties needed to ensure these properties hold for all p in a particular range coulbe be a fruitful source of some insightful stability analysis.

From the more applied perspective, there is also the analysis of problems from engineering design and model predictive control. We are interested in exploring the presentation of the requisite properties of the problems that must hold in order for the conditions of the solution path, or the regularity assumptions necessary for these properties, to be maintained.

In addition, we can extend the analysis to problems with multiple solutions, offering the possibility of tracing one or multiple solution paths. This can include a procedure of singularity detection and pseudo-arclength continuation methods, as in [26].

Finally, we are very interested in the tuning and performance of these algorithms for practical problems arising in model predictive control, and developing easy-to-use software for a flexible NLP solver. We believe that incorporating these in a general procedure that adaptively chooses the correct algorithm by the presence or lack of degeneracy, switching to a regular NLP solver when the step-size is driven too small, could be more efficient than solving a series of NLPs by standard solvers, and we intend to implement these procedures and investigate this conjecture.

Acknowledgements

We would like to thank Greg Horn for insightful discussions in regards to some of the details of the numerical software implementation of the algorithms, in particular with regards to parallelization.

This research was supported by Research Council KUL: PFV/10/002 Optimization in Engineering Center OPTEC, GOA/10/09 MaNet and GOA/10/11 Global real- time optimal control of autonomous robots and mechatronic systems. Flemish Government: IOF/KP/SCORES4CHEM, FWO: PhD/postdoc grants and projects: G.0320.08 (convex MPC), G.0377.09 (Mechatronics MPC); IWT: PhD Grants, projects: SBO LeCoPro; Belgian Federal Science Policy Office: IUAP P7 (DYSCO, Dynamical systems, control and optimization, 2012-2017); EU: FP7-EMBOCON (ICT-248940), FP7-SADCO (MC ITN-264735), FP7-TEMPO, ERC ST HIGHWIND (259 166), Eurostars SMART, ACCM.

References

[1] E. L. Allgower and K. Georg. *Introduction to Numerical Continuation Methods*. Colorado State University Press, 1990. 8

[2] J. Bonnans. Local Analysis of Newton-Type Methods for Variational Inequalities and Nonlinear Programming. *Appl. Math. Optim*, 29:161–186, 1994. 14

[3] J. F. Bonnans and A. Shapiro. Optimization problems with perturbations: A guided tour. *SIAM Review*, 40:228–264, 1998. 8

[4] J. F. Bonnans and A. Shapiro. *Perturbation Analysis of Optimization Problems*. Springer, 2000. 4

[5] M. Diehl. *Real-Time Optimization for Large Scale Nonlinear Processes*. PhD thesis, Universität Heidelberg, 2001. 5

[6] F. Facchinei, A. Fischer, and M. Herrich. A family of newton methods for nonsmooth constrained systems with nonisolated solutions. *Mathematical Methods of Operations Research*, pages 1–11, 2011. 18

[7] F. Facchinei, A. Fischer, and C. Kanzow. On the accurate identification of active constraints. *SIAM Journal on Optimization*, 9:14–32, 1998. 9

[8] D. Fernández and M. Solodov. Stabilized sequential quadratic programming for optimization and a stabilized newton-type method for variational problems. *Mathematical programming*, 125:47–73, 2010. 18

[9] H. Ferreau. An Online Active Set Strategy for Fast Solution of Parametric Quadratic Programs with Applications to Predictive Engine Control. Master's thesis, University of Heidelberg, 2006. 5

[10] H. Ferreau. qpOASES – An Open-Source Implementation of the Online Active Set Strategy for Fast Model Predictive Control. In *Proceedings of the Workshop on Nonlinear Model Based Control – Software and Applications, Loughborough*, pages 29–30, 2007. 5

[11] H. Ferreau, C. Kirches, A. Potschka, H. Bock, and M. Diehl. qpOASES: A parametric active-set algorithm for quadratic programming. *Mathematical Programming Computation*, 2013. (under review). 5

[12] P. Gill, V. Kungurtsev, and D. Robinson. A regularized sqp method convergent to second order optimal points. *Submitted to SIAM Journal of Optimization*, 2013. 30

[13] N. Gould, D. Orban, and P. L. Toint. Cutest: a constrained and unconstrained testing environment with safe threads. *Cahier du GERAD G*, 2013:27, 2013. 30

[14] J. Guddat, F. G. Vasquez, and H. Jongen. *Parametric Optimization: Singularities, Pathfollowing and Jumps*. Teubner, Stuttgart, 1990. 5

[15] W. W. Hager and M. S. Gowda. Stability in the presence of degeneracy and error estimation. *Mathematical Programming*, 85(1):181–192, 1999. 14

[16] S. K. Hock W. Test examples for nonlinear programming codes. In *Lecture Notes in Economics and Mathematical Systems, Vol. 187*. Springer, 1981. 30

[17] A. Izmailov. Solution sensitivity for karush-kuhn-tucker systems with non-unique lagrange multipliers. *Optimization*, 59(5):747–775, 2010. 9, 10, 20, 26

[18] A. F. Izmailov and M. V. Solodov. Stabilized sqp revisited. *Mathematical programming*, 133:93–120, 2012. 20, 24

[19] T. C. Johnson, C. Kirches, and A. Wächter. An active-set quadratic programming method based on sequential hot-starts. page 2013. Available at optimization online. 5

[20] H. T. Jongen and G. W. Weber. On parametric nonlinear programming. *Annals of Operations Research*, 27:253–283, 1990. 4

[21] M. Kojima. Strongly stable stationary solutions in nonlinear programs. *Analysis and Computation of Fixed Points*, 1980. 4, 13, 14

[22] V. Kungurtsev. *Second Derivative SQP Methods*. PhD thesis, UC-San Diego, 2013. 18

[23] J. Kyparisis. On uniqueness of kuhn-tucker multipliers in nonlinear programming. *Mathematical Programming*, 32(2):242–246, 1985. 11

[24] A. B. Levy. Solution sensitivity from general principles. *Siam J. Control Optim.*, 40:1–38, 2001. 4, 19

[25] A. B. Levy and R. Rockafellar. *Advances in Nonsmooth Optimization*, chapter Sensitivity of Solutions in Nonlinear Programs with Nonunique Multipliers, pages 215–223. World Scientific Publishing, 1995. 8, 13

[26] B. N. Lundberg and A. B. Poore. Numerical continuation and singularity detection methods for parametric nonlinear programming. *SIAM Journal on Optimization*, 3:134–154, 1993. 33

[27] E.-S. M. Mostafa, L. N. Vicente, and S. J. Wright. *Global Optimization and Constraint Satisfaction*, chapter Numerical behavior of a stabilized SQP method for degenerate NLP problems, pages 123–141. Springer, 2003. 29

[28] A. Poore and C. Tiahrt. Bifurcation problems in nonlinear parametric programming. *Mathematical Programming*, 39:189–205, 1987. 4

[29] D. RALPH and S. DEMPE. Directional derivatives of the solution of a parametric nonlinear program. *Mathematical programming*, 70:159–172, 1995. 8

[30] S. Robinson. Perturbed Kuhn-Tucker points and rates of convergence for a class of nonlinear programming algorithms. *Mathematical Programming*, 7:1–16, 1974. 4, 20

[31] S. M. Robinson. Stability theory for systems of inequalities. part i: Linear systems. *SIAM Journal on Numerical Analysis*, 12:754–769, 1975. 26

[32] S. M. Robinson. Stability theory for systems of inequalities, Part II: Differentiable nonlinear systems. *SIAM Journal on Numerical Analysis*, 13(4):497–513, 1976. 4, 14

[33] S. M. Robinson. Generalized equations and their solutions, part ii: Applications to nonlinear programming. *Optimality and Stability in Mathematical Programming*, pages 200–221, 1982. 4, 17

[34] S. Sequeira, M. Graellis, and L. Puigjaner. Real-time evolution for on-line optimization of continuous processes. *Ind. Eng. Chem. Res.*, 41:1815–1825, 2002. 5

[35] R. Seydel. *Practical Bifurcation and Stability Analysis*. Springer Science and Business Media, 2010. 8

[36] Q. Tran-Dinh, C. Savorgnan, and M. Diehl. Adjoint-based Predictor-Corrector Sequential Convex Programming for Parametric Nonlinear Optimization. *SIAM J. Optimization*, 22(4):12581284, 2012. 5

[37] G. Wachsmuth. On licq and the uniqueness of lagrange multipliers. *Operations Research Letters*, 2012. 11

[38] L. T. Watson. Solving the nonlinear complementarity problem by a homotopy method. *SIAM J. Control Optim*, 17:36–46, 1979. 5

[39] S. J. Wright. Superlinear convergence of a stabilized sqp method to a degenerate solution. *Computational Optimization and Applications*, 11:253–275, 1998. 27

[40] S. J. Wright. An algorithm for degenerate nonlinear programming with rapid local convergence. *SIAM Journal on Optimization*, 15(3):673–696, 2005. 4, 21, 24, 25, 26

[41] V. Zavala and M. Anitescu. Real-Time Nonlinear Optimization as a Generalized Equation. *SIAM J. Control Optim.*, 48(8):5444–5467, 2010. 5