Noname manuscript No.
(will be inserted by the editor)

Pathfollowing for Parametric Mathematical Programs with Complementarity Constraints

Vyacheslav Kungurtsev · Johannes Jäschke

Received: date / Accepted: date

Abstract In this paper we study procedures for pathfollowing parametric mathematical programs with complementarity constraints. We present two procedures, one based on the penalty approach to solving standalone MPCCs, and one based on tracing active set bifurcations arising from doubly-active complementarity constraints. We demonstrate the performance of these approaches on a variety of examples with different types of stationary points and also a simple engineering problem with phase changes.

Keywords Parametric optimization · Sensitivity · Mathematical Programs with Complementarity Constraints · Complementarity · Pathfollowing · Nonlinear Model Predictive Control

1 Introduction

We are interested in tracing the approximate path of solutions to a mathematical program with complementarity constraints (MPCC) subject to a parameter, as the parameter changes. Specifically, we study the problem,

$$\begin{array}{rl} \min_{x\in\mathbb{R}^{n+2n_c}} & f(x,t) \\ \text{s.t.} & g(x,t) \geq 0, \\ & h(x,t) = 0, \\ & \min\{x_j, x_{j+n_c}\} = 0, \text{ for all } j \in \{n+1,...,n+n_c\} \end{array} \tag{1}$$

where the more general case of nonlinear complementarity constraints can be covered by this case with the use of slack variables. We consider the situation wherein one has a reasonably accurate estimate of the solution at $t = 0$ and intends to find the solution at $t = 1$ using parametric sensitivity, i.e., the directional derivative of solutions to (1), or some related problem, with respect to t.

VK was supported by the OP VVV project CZ.02.1.01/0.0/0.0/16 019/0000765 "Research Center for Informatics". JJ acknowledges support from DNVGL.

V. Kungurtsev
Intelligent Data Analysis, Department of Computer Science, Faculty of Electrical Engineering, Czech Technical University in Prague. E-mail: vyacheslav.kungurtsev@fel.cvut.cz,

J. Jäschke
Department of Chemical Engineering, Norwegian University of Science and Technology (NTNU) E-mail: jaschke@ntnu.no

A primary motivation for this arises in real-time nonlinear model predictive control. There are a number of problems arising in process control engineering wherein the dynamic process has a complementarity structure (e.g., see [1,19]). Using the advanced step NMPC framework [25] or the real-time iteration [4], one can pathfollow along the solution using sensitivity updates from the predicted and measured state. For example in advanced step NMPC, at a given iteration of a model predictive control procedure, using the predicted state from the previous time step, there is an approximate solution of (1) at $t = 0$ which is used in an applied control, and meanwhile the solution of the MPCC at the next time step is solved offline. Then the state is measured at the real next time step and the sensitivity update is applied to traverse the solution from the imperfect predicted state by the offline solution (considered as $t = 0$) to the real, measured state (considered as $t = 1$). See [11, 23] for illustrations of this approach.

Considered as standalone problems, MPCCs are typically solved by procedures that approximate them by NLPs. Thus the parametric MPCC can be approximated by an appropriate parametric NLP or set of NLPs. This notion, of course, will be made precise later, and is one of the main points of this paper.

The main contribution of this paper is the first study of pathfollowing for parametric MPCCs. Currently, there is no literature presenting any algorithm to trace the solution of a parametric MPCC from an approximate solution at one value of a parameter to another. In this paper we study two appropriate methods based on NLP approximations of the underlying MPCC for each parameter. In particular we study the penalty method for solving MPCCs, and also, noting the importance of doubly-active complementarity constraints in indicating potential branching points of the parametric MPCCs, a procedure that traces these along their underlying NLPs.

We found that the guarantees and experimental performance of pathfollowing resembles solving standalone MPCCs in an analogous manner. In particular, the penalty method is able to trace, just as the penalty method is able to solve, stationary points satisfying weaker conditions in regards to the presence of descent directions. It is simple to implement, however, and for problems without spurious stationary points or many bi-active constraints, it performs just as well and reliably as pathfollowing on a standalone NLP. However, for problems in which B-stationarity, the most precise form of stationarity is sought, which is also difficult to solve for as a standalone problem, a more complicated and combinatorial approach is necessary. Given the required tracking of different branches, the gain in reliability of quality solutions is at the cost of a more difficult to implement and slower pathfollowing algorithm.

1.1 Previous Literature

There is no literature on pathfollowing of MPCC. Parametric optimization for standard NLPs has enjoyed two periods of intensive activity, around the publication of the seminal book [8] and more recently with the development of higher quality nonlinear models for process control motivating the rise of nonlinear model predictive control, e.g., [4–6, 24]. In [15] the problem with weaker constraint qualifications was studied. The procedures are based on sensitivity, in particular, directional derivatives of the solution to an NLP subject to a differential parameter change, whose properties depend on the constraint qualifications holding at the solution.

In the case of MPCCs, typically no standard constraint qualifications are satisfied (aside from Guignard's). Thus only the second algorithm in [14] for generic branching could potentially be applicable for pathfollowing MPCCs. The algorithm requires an approximation to the complete characterization of the optimal Lagrange multiplier set, and in general is not tailored for MPCCs specifically.

There are a number of algorithms for solving standalone MPCCs. The most standard ones are based on regularization or penalization [20], and have been proven to find the C-stationary points of the problem. C-stationary points have non-trivial directions of descent, however, so a filter sequential linear complementarity based approach was presented in [18] to obtain solutions satisfying B-stationarity, the most inclusive stationarity measure for which no descent direction exists.

Sensitivity of MPCCs is studied in [2, 13, 21]. In combination with algorithmic pathfollowing ideas for NLP and algorithms for solving standalone MPCCs, these form the inspiration for the procedures and analysis presented in this paper.

2 Background

In this section we first review the optimality conditions of the standalone MPCC, and then proceed to describe the general properties of the solution paths of parametric MPCCs as the value of the parameter changes, and finally review pathfollowing procedures for parametric NLPs.

2.1 Optimality Conditions and Constraint Qualifications

Following [21] we present three related NLPs to (1) based around a feasible point x^*. The first, the *tightened NLP*, will be defined as,

$$\begin{aligned} \min_{x\in\mathbb{R}^{n+2n_c}} \; & f(x,t) \\ \text{s.t. } & g(x,t) \geq 0, \\ & h(x,t) = 0, \\ & x_i = 0 \text{ for } x_i^* = 0,\ i \in \{n+1,...,n+2n_c\} \\ & x_i \geq 0 \text{ for } x_i^* > 0,\ i \in \{n+1,...,n+2n_c\}. \end{aligned} \tag{2}$$

Next, the *relaxed NLP* is defined as,

$$\begin{aligned} \min_{x\in\mathbb{R}^{n+2n_c}} \; & f(x,t) \\ \text{s.t. } & g(x,t) \geq 0, \\ & h(x,t) = 0, \\ & x_i = 0,\ x_{i+n_c} \geq 0 \text{ for } i \text{ such that } x_i^* = 0 \text{ and } x_{i+n_c}^* > 0,\ i \in \{n+1,...,n+n_c\} \\ & x_i = 0,\ x_{i-n_c} \geq 0 \text{ for } i \text{ such that } x_i^* = 0 \text{ and } x_{i-n_c}^* > 0,\ i \in \{n+n_c+1,...,n+2n_c\} \\ & x_i, x_{i+n_c} \geq 0 \text{ for } i \text{ such that } x_i^* = 0 \text{ and } x_{i+n_c}^* = 0,\ i \in \{n+1,...,n+n_c\}. \end{aligned} \tag{3}$$

It holds that if x^* is a local minimizer of the relaxed NLP then it is a local minimizer of the MPCC, and if it is a local minimizer of the MPCC then it is a local minimizer of the tightened NLP. The three are equivalent if *strict complementarity* holds, in which case $x_i^* > 0$ or $x_{i+n_c}^* > 0$ for all $i \in \{n+1,...,n+n_c\}$.

Let us denote some index sets of active complementary components associated with a feasible point x^*,

$$\begin{aligned} \mathcal{A}_1(x^*) &\triangleq \{i | i \in \{n+1,...,n+n_c\},\ x_i^* = 0,\ x_{i+n_c}^* > 0\}, \\ \mathcal{A}_2(x^*) &\triangleq \{i | i \in \{n+n_c+1,...,n+2n_c\},\ x_i^* = 0,\ x_{i-n_c}^* > 0\}, \\ \mathcal{A}_1^0(x^*) &\triangleq \{i | i \in \{n+1,...,n+n_c\},\ x_i^* = 0,\ x_{i+n_c}^* = 0\}, \\ \mathcal{A}_2^0(x^*) &\triangleq \{i | i \in \{n+n_c+1,...,n+2n_c\},\ x_i^* = 0,\ x_{i-n_c}^* = 0\} \end{aligned}$$

Denote the set

$$\mathcal{I}(x^*) = \{I \subseteq \{i | i \in \mathcal{A}_1(x^*) \cup \mathcal{A}_1^0(x^*) \cup \mathcal{A}_2(x^*) \cup \mathcal{A}_2^0(x^*)\} : \forall i \in \{n+1,...,n+n_c\}, \{i, i+n_c\} \cap I \neq \emptyset\}$$

i.e., the set of all active complementarity variables such that at least one from every complementarity pair is included, and the *Index* $I \in \mathcal{I}(x^*)$ *NLP* as,

$$\begin{array}{rl} \min_x & f(x,t), \\ \text{s.t.} & g(x,t) \geq 0, \\ & h(x,t) = 0, \\ & x_i = 0,\ i \in I, \\ & x_i \geq 0,\ i \in \{n+1,...,n+2n_c\} \setminus I \end{array} \tag{4}$$

We will denote the *MPCC Lagrangian* with multipliers $\{\lambda, \mu, \sigma\}$ as,

$$\mathcal{L}(x,\lambda,\mu,\sigma,t) = f(x,t) - \lambda^T g(x,t) - \mu^T h(x,t) - \sigma^T x,$$

where it is assumed that $\sigma_i = 0$ for $i \leq n$.

We say that *stationarity* at x^* of the parametric MPCC at t holds if there exist multipliers $\{\lambda^*, \mu^*, \sigma^*\}$ such that,

$$\begin{array}{l} \nabla f(x^*,t) - (\lambda^*)^T \nabla g(x^*,t) - (\mu^*)^T \nabla h(x^*,t) - \sigma^* = 0, \\ \lambda^* \cdot g(x^*,t) = 0,\ \lambda^* \geq 0,\ g(x^*,t) \geq 0,\ h(x^*,t) = 0 \\ \min\{x_i^*, x_{i+n_c}^*\} = 0, \text{ for all } i \in \{n+1,...,n+n_c\} \\ \sigma_i^* = 0 \text{ for } i < n+1 \\ \sigma_i^* x_i^* = 0, \text{ for } i \in \{n+1,...,n+2n_c\} \end{array} \tag{5}$$

The MPCC is *C-stationary* if also $\sigma_{i_1}\sigma_{i_2} \geq 0$ for all $i_1 \in \{n+1,...,n+n_c\}$ and $i_2 = i_1 + n_c$.

It is *M-stationary* if $\sigma_{i_1} > 0$, $\sigma_{i_2} > 0$, or $\sigma_{i_1}\sigma_{i_2} = 0$ for all $i_1 \in \{n+1,...,n+n_c\}$ and $i_2 = i_1 + n_c$.

The MPCC is *S-stationary* if $\sigma_i \geq 0$ for all $i \in \{n+1,...,n+2n_c\}$. S-stationarity is also called strong stationarity, and is the strongest notion of stationarity. It is equivalent to the KKT property of the relaxed NLP (3).

The tightest notion of stationarity, i.e., the one for which no feasible descent directions exist, is *B-stationarity*, which holds if $d = 0$ is the solution to the LPEC,

$$\begin{array}{rl} \min_{d \in \mathbb{R}^{n+2n_c}} & \nabla f(x^*,t)^T d \\ \text{s.t.} & g(x^*,t) + \nabla g(x^*,t)^T d \geq 0, \\ & h(x^*,t) + \nabla h(x^*,t)^T d = 0, \\ & 0 \leq x_i^* + d_i \perp x_{i+n_c}^* + d_{i+n_c} \geq 0,\ i \in \{n+1,...,n+n_c\}. \end{array} \tag{6}$$

This is equivalent to x^* being feasible and stationary for all Index I NLP for $I \in \mathcal{I}(x^*)$. If $\mathcal{A}_1^0(x^*) = \mathcal{A}_2^0(x^*) = \emptyset$ then B-stationarity is equivalent to S-stationarity. Otherwise, S-stationarity, as the strongest stationarity measure, implies B-stationarity (however, a local minimizer may exist that satisfies B, but not S-stationarity).

We say that x^* satisfies MPCC-LICQ, MPCC-MFCQ, or MPCC-SMFCQ if the tightened NLP at x^* satisfies the LICQ, MFCQ or SMFCQ, respectively.

There are a number of second order conditions, we first report the strongest, appearing in [21],

Theorem 1 *[21, Theorem 7]*

– *If x^* is a local minimizer of (1) at t and the MPCC-SMFCQ condition holds at x^*, then it is a strongly stationary point with unique optimal multipliers $(\lambda^*, \mu^*, \sigma^*)$ and*

$$d^T \frac{\partial^2}{\partial x^2} \mathcal{L}(x^*, \lambda^*, \mu^*, \sigma^*, t) d \geq 0$$

for any d such that d is feasible for (6) and $\nabla f(x^*,t)^T d = 0$.

– *If x^* is a strongly stationary point of* (1) *at* t*, and for every direction* d *feasible for* (6) *and* $d^T \nabla f(x^*, t) = 0$*, there exist* $(\lambda^*, \mu^*, \sigma^*)$ *satisfying strong-stationarity such that*

$$d^T \frac{\partial^2}{\partial x^2} \mathcal{L}(x^*, \lambda^*, \mu^*, \sigma^*, t) d > 0$$

There are a number of other second order optimality conditions corresponding to each of the types of critical points given. These are described in [9]. It is beyond the scope of the paper to describe each of the conditions, only to say that they correspond to standard notions of second order sufficiency with critical directions corresponding to cones natural to the specific multiplier types for M and C-stationarity.

2.2 Properties of the Parametric Solution Path

Given the combinatorial structure of MPCCs, it is natural to consider that at a solution x^* of an MPCC, for each possible $I \in \mathcal{I}(x^*)$, the corresponding index I NLP has its associated parametric sensitivity properties. This is formalized in [21, Theorem 9] in the case of strong second order sufficiency and the LICQ holding at x^* for every I, corresponding to unique NLP minimizers for perturbations of t. Of course, weaker constraint qualifications can give rise to more complicated parametric sensitivity structure for each NLP (e.g., [10]). This is a natural extension of parametric NLP in general, as an outgrowth of the combinatorial indexed NLPs corresponding elements of $\mathcal{I}(x^*)$.

If the multipliers satisfy *upper level strict complementarity* however, then the disappearance of the weakly double-active multipliers results in a unique path.

Definition 1 Upper level strict complementarity holds at $\{x^*, \lambda^*, \mu^*, \sigma^*\}$ if $\sigma_i^* > 0$ for $i \in \mathcal{A}_1^0(x^*) \cup \mathcal{A}_2^0(x^*)$.

Theorem 10 [21] states that given this condition and an LICQ condition on the gradient matrices corresponding to all combinations of $I \in \mathcal{I}(x^*)$ and inequality constraints such that $\lambda_j^* > 0$ in union with all the subsets such that $\lambda_j^* = 0$ and j is active, there exists a unique path of solutions for local perturbations of t in (1). These solutions maintain B-stationarity or C-stationarity, as the case may be, and if B-stationarity and the second order sufficiency condition hold, then they are local minimizers.

In [2] the standard critical sets of parametric optimization [8] are studied in the context of MPCCs. In this paper we restrict ourselves to parametric MPCC problems for which the MPCC-second order sufficiency conditions and MPCC-LICQ hold, leaving only the "type two" degenerate critical point. This is the case where either one λ_i^* or σ_i^* is zero for an active constraint. The first case is similar to the corresponding one in in standard nonlinear programming–there is a nonsmooth parametric path for which the branch on one side has some particular constraint feasible but inactive, and the other side the constraint stays feasible but the multiplier becomes strongly active, generically speaking (i.e., for a dense subset of the set of potential real-valued parametric optimization problems). In the case of $i \in \left(\mathcal{A}_0^1(x^*) \cup \mathcal{A}_0^2(x^*)\right) \cap \{i : \sigma_i^* = 0\}$ at $\bar{t}$, i.e., a double active complementarity constraint that has a null multiplier, generically speaking, one branch will have a variable in $\mathcal{A}_0^1(x^*) \cup \mathcal{A}_0^2(x^*)$ become inactive (i.e., positive), and σ_i^* will change sign at $\bar{t}$. It can change sign in either direction, however, thus potentially a C-stationary point could become just M-stationary, or S-stationary to just weakly stationary.

Let us define partial strict complementarity.

Definition 2 Partial strict complementarity holds at $\{x^*, \lambda^*, \mu^*, \sigma^*\}$ if for all $i \in \mathcal{A}_1^0(x^*)$, either $\sigma_i^* > 0$ or $\sigma_{i+n_c}^* > 0$.

In [13] the strong stability of C-stationary points is investigated. It is shown [13, Theorem 1] that under MPEC-LICQ, MPEC-SOSC and partial strict complementarity as well as the second order sufficiency for every Index $I \in \mathcal{I}(x^*)$ NLP, it holds that C-stationary points are strongly stable, i.e., there is a unique path of C-stationary points with respect to perturbations. Subsequently, they show that same result under MPEC-LICQ, MPEC-SOSC and upper level strict complementarity. The results hold analogously for strongly stationary points, but not M-stationary points.

2.3 Pathfollowing for Nonlinear Programming

We recall here some basic notions about pathfollowing for standard (non-complementarity) parametric nonlinear programming. Thus consider (x^*, λ^*, μ^*) satisfying stationarity (so with $n_c = 0$). Let $\Lambda(x^*, t)$ correspond to the set of all (λ^*, μ^*) satisfying the conditions for x^*.

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

The Lagrangian function associated with the NLP is $L(x, \lambda, \mu, t) = f(x, t) - g(x, t)^T \lambda - h(x, t)^T \mu$. The Hessian of the Lagrangian with respect to x is denoted by

$$H(x, \lambda, \mu, t) := \nabla^2_{xx} f(x, t) - \sum_{i=1}^{m_i} \lambda_i \nabla^2_{xx} g_i(x, t) - \sum_{i=1}^{m_e} \mu_i \nabla^2_{xx} h_i(x, t).$$

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

Definition 3 (Strong second-order sufficient conditions (SSOSC)) A primal-dual pair (x^*, y^*) satisfies the *strong second-order sufficient optimality conditions* at t if it satisfies the first-order KKT conditions and

$$d^T H(x^*, \lambda^*, \mu*, t) d > 0 \text{ for all } d \in \tilde{\mathcal{C}}(x^*, \lambda^*, \mu^*, t) \setminus \{0\}, \quad (7)$$

where $d \in \tilde{\mathcal{C}}(x^*, \lambda^*, \mu^*, t)$ if $\nabla_x g_i(x^*, t)^T d = 0$ for $i \in \mathcal{A}_+(x^*, \lambda^*, t)$ and $\nabla_x h(x^*, t)^T d = 0$.

The weaker second-order sufficiency condition is given as follows.

Definition 4 (Second-order sufficient conditions (SOSC)) A primal-dual pair (x^*, y^*) satisfies the *second-order sufficient optimality conditions* at t if it satisfies the first-order KKT conditions and

$$d^T H(x^*, \lambda^*, \mu*, t) d > 0 \text{ for all } d \in \hat{\mathcal{C}}(x^*, \lambda^*, \mu^*, t) \setminus \{0\}, \quad (8)$$

where $d \in \hat{\mathcal{C}}(x^*, \lambda^*, \mu^*, t)$ if $\nabla_x g_i(x^*, t)^T d = 0$ for $i \in \mathcal{A}_+(x^*, \lambda^*, t)$, $\nabla_x g_i(x^*, t)^T d \geq 0$ for $i \in \mathcal{A}_0(x^*, \lambda^*, t)$ and $\nabla_x h(x^*, t)^T d = 0$.

The generalized SSOSC is defined as,

Definition 5 (Generalized SSOSC (GSSOSC)) A primal vector x^* satisfies the *generalized strong second-order sufficient optimality conditions* at t if (x^*, λ^*, μ^*) satisfies the SSOSC for all $(\lambda^*, \mu^*) \in \Lambda(x^*, t)$.

Using [12], we know that a QP approximating,

$$\begin{array}{ll} \min_{\Delta x} & \frac{1}{2}\Delta x^T H(x^*,\lambda^*,\mu^*,t)\Delta x + \Delta t^T \frac{\partial^2 L(x^*,\lambda^*,\mu^*,t)}{\partial t \partial x}\Delta x \\ \text{subject to} & \nabla_t h(x^*,t)^T \Delta t + \nabla_x h(x^*,t)^T \Delta x = 0, \\ & \nabla_t g_i(x^*,t)^T \Delta t + \nabla_x g_i(x^*,t)^T \Delta x = 0, \ \text{for } i \in \mathcal{A}_+(x^*,\lambda^*,t), \\ & \nabla_t g_i(x^*,t)^T \Delta t + \nabla_x c_i(x^*,t)^T \Delta x \geq 0, \ \text{for } i \in \mathcal{A}_0(x^*,\lambda^*,t). \end{array} \tag{9}$$

corresponds to the directional derivative along the homotopy path of solutions, if the LICQ and SSOSC hold at x^* at t.

Here, and in the sequel of the paper, assume that $\nabla_{xt} f(x,t)$, $\nabla_t g(x,t)$ and $\nabla_t h(x,t)$ are all constant. By lifting the nonlinear dependence (i.e., introducing a variable z in place of t in the problem functions and adding the constraint $z = t$) any parametric NLP can be written this way.

In [14, Section 2-3] a *predictor-corrector* QP is introduced, which incorporates both the directional derivative of the solution along with a step towards the solution at the next value of the homotopy parameter, given an approximate solution at the initial value of the parameter $(\bar{x},\bar{\lambda},\bar{\mu})$. It can be seen that under affine dependence the QP is equivalent to,

$$\begin{array}{ll} \min_{\Delta x} & \nabla f(\bar{x},t+\Delta t)^T \Delta x + \frac{1}{2}\Delta x^T H(\bar{x},\bar{\lambda},\bar{\mu},t)\Delta x + \Delta t^T \frac{\partial^2 L(\bar{x},\bar{\lambda},\bar{\mu},t)}{\partial t \partial x}\Delta x \\ \text{subject to} & h(\bar{x},t) + \nabla_t h(\bar{x},t)^T \Delta t + \nabla_x h(\bar{x},t)^T \Delta x = 0, \\ & g_i(\bar{x},t) + \nabla_t g_i(\bar{x},t)^T \Delta t + \nabla_x g_i(\bar{x},t)^T \Delta x = 0, \ \text{for } i \in \mathcal{A}_+(\bar{x},\bar{\lambda},t), \\ & g_i(\bar{x},t) + \nabla_t g_i(\bar{x},t)^T \Delta t + \nabla_x c_i(\bar{x},t)^T \Delta x \geq 0, \ \text{for } i \in \mathcal{A}_0(\bar{x},\bar{\lambda},t). \end{array} \tag{10}$$

This corresponds to the standard sensitivity-based predictor-corrector approach for pathfollowing approximate solutions of NLPs, which we will use in the Algorithms discussed in this paper.

The active sets $\mathcal{A}_+(\bar{x},\bar{\lambda},t)$ and $\mathcal{A}_0(\bar{x},\bar{\lambda},t)$ are estimated based on [7], also developed and studied procedurally in [14]. In particular, for a standard NLP (i.e., $n_c = 0$), we define the optimality residual as

$$\eta(x,\lambda,\mu) = \left\| \begin{pmatrix} \nabla_x f(x,t) - \nabla h(x,t)\mu - \nabla g(x,t)\lambda \\ \min(g(x,t),\lambda) \\ h(x,t) \end{pmatrix} \right\|,$$

and estimate the active sets as follows,

$$\begin{array}{l} \mathcal{A}(x,\lambda,\mu) = \{i : \ g_i(x,t) \leq \eta(x,\lambda,\mu)^\gamma\} \\ \mathcal{A}^+(x,\lambda,\mu) = \{i \in \mathcal{A}(x,\lambda,\mu) : \ \lambda_i > \eta(x,\lambda,\mu)^\gamma\} \\ \mathcal{A}^0(x,\lambda,\mu) = \mathcal{A}(x,\lambda,\mu) \setminus \mathcal{A}^+(x,\lambda,\mu) \end{array} \tag{11}$$

where $0 < \gamma < 1$.

3 Algorithms

In this section we describe two Algorithms for pathfollowing solutions to parametric MPCCs, based on approximating the parametric MPCC by an NLP or set of NLPs and using the predictor-corrector QP (10).

3.1 Penalty Methods

In order to obtain a solution of the MPCC, penalty methods instead solve the NLP,

$$\begin{array}{rl} \min_{x\in\mathbb{R}^{n+2n_c}} & f(x,t)+\rho x_{n+1:n+n_c}^T x_{n+n_c+1:n+2n_c} \\ \text{s.t.} & g(x,t)\geq 0, \\ & h(x,t)=0, \\ & x_j \geq 0, \text{ for all } j\in\{n+1,...,n+2n_c\} \end{array} \tag{12}$$

for some penalty parameter $\rho > 0$.

There are several important results relating the stationarity properties of the original problem and the penalty relaxation.

Theorem 2 *[20, Theorem 5.1-5.2, Proposition 5.3]*

- *If x^* is strongly stationary for (1) at t then for ρ sufficiently large, x^* is stationary for (12). Furthermore, if MPEC-LICQ, MPEC-MFCQ, or MPEC-SOSC hold for (1) at t at x^* then LICQ, MFCQ and SOSC hold at (12) at t at x^*, respectively.*
- *If x^* is stationary for (12) and x^* is feasible for (1), then it is strongly stationary for (1). And furthermore if LICQ and SOSC hold at x^* for (12) at t, then MPEC-LICQ and MPEC-SOSC holds at x^* for (1) at t.*
- *If x^* is strongly stationary for (1) at t, MPEC-LICQ holds at x^*, and $\sigma_i^*+\sigma_{i+n_c}^* > 0$ for all $i\in\mathcal{A}_1^0(x^*)$, then for ρ sufficiently large, there is a neighborhood of x^* such that every stationary point of (12) is a strongly stationary point for (1).*

From the proofs in [20], the minimum ρ that is required for the first two results is equal to,

$$\bar{\rho}_1 = 1+\max\left(0, \max_{i\in\mathcal{A}_1(x^*)}\frac{\sigma_i^*}{x_{i+n_c}^*}, \max_{i\in\mathcal{A}_2(x^*)}\frac{\sigma_i^*}{x_{i-n_c}^*}\right), \tag{13}$$

and for the third result, it must satisfy,

$$\rho\xi+\max_i \sigma_i^* - \xi > 0,$$

where ξ is equal to,

$$\xi = \frac{1}{2}\min\left(\min_{i\in(\mathcal{A}_1(x^*)+n_c)\cup(\mathcal{A}_2(x^*)-n_c)} x_i^*, \min_{i:\sigma_i^*>0}\sigma_i^*\right)$$

Note, however, that this does not imply the strong second order sufficiency conditions for the penalty NLP.

One can obtain the multipliers σ^* from the solution of the penalty function minimization by taking [3],

$$\begin{array}{l} \sigma_i^* = z_i - \rho x_{i+n_c},\ i\in\{n+1,...,n+n_c\}, \\ \sigma_i^* = z_i - \rho x_{i+n_c},\ i\in\{n+n_c+1,...,n+2n_c\} \end{array} \tag{14}$$

where z^* are the multipliers for the bound constraints $x^*\geq 0$ in the penalty problem.

3.1.1 Description of the Algorithm

We write the penalty objective as,

$$f_p(x,t;\rho) = f(x,t) + \rho x_{n+1:n+n_c}^T x_{n+n_c+1:n+2n_c}$$

and correspondingly the Lagrangian of the penalty problem as,

$$L_p(x,\lambda,\mu,z,t;\rho) = f_p(x,t;\rho) - \lambda^T g(x,t) - \mu^T h(x,t) - z^T x.$$

with Hessian matrix,

$$H_p(x,\lambda,\mu,t;\rho) = \nabla^2_{xx} L_p(x,\lambda,\mu,z,t;\rho)$$

(where we drop the dependence on z since x appears linearly in L_p).

We present the Penalty pathfollowing algorithm as Algorithm 1.

Algorithm 1 Penalty Pathfollowing Method for MPCC

Input: $t = 0$, x, λ, μ, z close to $(x^*, \lambda^*, \mu^*, z^*)$ satisfying the optimality conditions of (12) and feasible for (1) at $t = 0$.

Compute the distance to solution residual $\eta(x,\lambda,\mu,z,t)$

Estimate the active set A and the strongly active set A^+ by (11).

while $t < 1$ **do**

  Solve

$$\begin{array}{ll} \min_{\Delta x} & \nabla_x f_p(x,t;\rho) + \frac{1}{2}\Delta x^T H_p(x,\lambda,\mu,t;\rho)\Delta x + \Delta p^T \frac{\partial^2 L_p(x,\lambda,\mu,t;\rho)}{\partial t \partial x}\Delta x \\ \text{subject to} & h(x,t) + \nabla_t h(x,t)^T \Delta t + \nabla_x h(x,t)^T \Delta x = 0, \\ & g_i(x,t) + \nabla_t g_i(x,t)^T \Delta t + \nabla_x g_i(x,t)^T \Delta x = 0, \text{ for } i \in A_+(x,\lambda,t), \\ & g_i(x,t) + \nabla_t g_i(x,t)^T \Delta t + \nabla_x c_i(x,t)^T \Delta x \geq 0, \text{ for } i \in A_0(x,\lambda,t), \\ & x_i = 0,\ i \in A_+(x,z), \\ & x_i \geq 0,\ i \in \{n+1,...,n+2n_c\} \setminus A_+(x,z) \end{array} \quad (15)$$

  to obtain $(\Delta x, \hat{\lambda}, \hat{\mu})$.

  If $x + \Delta x$ is infeasible for (1), then increase ρ, repeat loop.

  if $\eta(x+\Delta t, \hat{\lambda}, \hat{\mu}, z, t+\Delta t)$ is sufficiently small **then**

    Set $t = t + \Delta t$, $x = x + \Delta x$,$\lambda = \hat{\lambda}$,$\mu = \hat{\mu}$, increase Δt.

    Estimate the active set A and the strongly active set A^+.

  else

    Decrease Δt

  end if

end while

3.1.2 Convergence

The convergence tracking result associated with the penalty algorithm is given below in Theorem 3. In general, for well-behaved problems the procedure successfully traces the solutions, but has several limitations. First, only strongly stationary points are guaranteed to be traced. This does not necessarily imply that weaker notions of stationarity that are inherently spurious, like M or C-stationarity, will not be traced, and it also implies that not all B-stationary points will be traced. Furthermore, an "a posteriori" assumption must hold, that the SSOSC conditions hold for the penalty problem associated with the MPCC, in particular, The SOSC for the original MPCC is not enough to ensure the predictor-corrector QP is able to generate the solution path for the MPCC.

Theorem 3 *If there is an isolated primal solution path of strongly stationary points $x^*(t)$ for some $t \in [t_a, t_b]$, then for all t, $x^*(t)$ is a solution path for* (12) *for $t \in [t_a, t_b]$. Furthermore, if*

1. *MPEC-LICQ holds for all* $x^*(t)$,
2. *there is an initial estimate* $x(t_a)$ *sufficiently close to* $x^*(t_a)$,
3. *The SSOSC holds for* (12) *for all* $x^*(t)$ *satisfying SOSC,*
4. *The right-hand side of* (13) *is bounded by* $\bar{\rho}$ *for all primal-dual complementarity solutions* $x^*(t), \sigma^*(t)$, $t \in [t_a, t_b]$.
5. $\sigma_i^*(t) + \sigma_{i+n_c}^*(t) > 0$ *for all complementarity multipliers associated with* $x^*(t)$,

then applying Algorithm 1 with $\rho \geq \bar{\rho}$, *starting from* $x(t_a)$, *successfully traces an approximate solution path* $x(t)$, *i.e., for any* D, *there exists a* D_0 *such that if* $\|x(t_a) - x^*(t_a)\| \leq D_0$ *then* $\|x(t) - x^*(t)\| \leq D$ *for all* t.

Proof Consider $x^*(t_a)$ the closest point to $x(t_a)$ satisfying strong-stationarity. From Theorem 2 we can conclude that $x^*(t_a)$ is solution to (12) at t_a for $\rho \geq \bar{\rho}$. Furthermore, the LICQ, SOSC, and, by assumption SSOSC hold at $x^*(t_a)$ as a solution to (12). We can extend this to say that these conditions hold for all $x^*(t)$ as solutions to (12), and the final statement of Theorem 2 implies there is a one-to-one correspondence between $x^*(t)$ as a solution to (12) and for the MPCC, with a region around $x^*(t)$ for which no solutions satisfying these properties exist for either problem.

Now we can bring $x(t_a)$ as close as necessary to $x^*(t_a)$ such that the conditions of [14, Theorem 4.2] (alternatively, as these conditions imply strong regularity for the KKT system defined as a variational inequality, [6]), are satisfied as applied to (12), the predictor-corrector QP (15) associated with the appropriate active and strongly active set estimates, is able to successfully trace a primal-dual solution with $\|x(t) - x^*(t)\| \leq D$.

Note that we could implement a version for which MPEC-MFCQ, rather than the MPEC-LICQ, is necessary for convergence and tracking, given in [15], however, for ease of presentation we use the variant requiring unique multipliers.

3.2 Active Set Index Method

The next pathfollowing algorithm is intended to trace B-stationary points, which as indicated describe the tightest notion of optimality. Their importance and the combinatorial approach of the Active Set algorithm described here is inspired by the SLPEC-EQP Algorithm presented in [18] (see also [16, Chapter 7]).

In that paper, it was observed that M, C and weak stationary points still have feasible descent directions, and B-stationarity is the precise stationarity concept associated with the lack of such descent directions. S-stationarity is the strongest notion of stationarity, but seeking S-stationarity can neglect points that are B but not S stationary.

The primary driver towards global convergence in [18] is the LPEC,

$$\begin{array}{rl} \min_d & \nabla f(x,t)^T d, \\ \text{s.t.} & g(x,t) + \nabla g(x,t)^T d \geq 0, \\ & h(x,t) + \nabla h(x,t)^T d = 0, \\ & 0 \leq x_i + d_i \perp x_{i+n_c} + d_{i+n_c} \geq 0,\ i \in \{n+1, ..., n+n_c\}, \\ & \|d\| \leq \rho, \end{array} \tag{16}$$

where $\rho > 0$ is a trust-region parameter. It is clear that $d = 0$ being the solution to this LPEC corresponds to B-tationarity. The SLPEC-EQP algorithm [18] iterates by solving this LPEC, using the solution to estimate the active set, and performing an equality constrained QP step to accelerate towards convergence. The authors prove the global convergence of this procedure to a B-stationary point. Recall that a local minimizer to an LPEC being a zero step implies that the full set of index-I NLPs is optimal we focus on estimating the set $\mathcal{I}$, noting that the double active set indicates possible branching in the directional derivative of the solution.

Thus, given a point x that is close to a B-stationary point x^* at t, we seek to estimate the active set $\mathcal{A}^0(x,t) = \{i \in \{n+1,...,n+2n_c\} : x_i^* = 0\}$ of complementary variables.

However, there is no certifiably accurate method to identify the set of index I NLPs of an MPCC. Thus we must make the following assumption,

Assumption 1 *We have a reliable means of estimating $\mathcal{A}^0(x,t)$ from x, if x is sufficiently close to x^*, a B-stationary point.*

Heuristically, we consider i as active, and thus a potential member of the index I NLP when $x_i \le \epsilon_0$ for some tolerance ϵ_0. Practically, we form the set of branches based on an estimate of the doubly-active set, i.e., $\{i : x_i \le \epsilon_0,\, x_{i+n_c} \le \epsilon_0\}$. In our practical experiments, we found that a threshold of $\epsilon_0 = 10^{-3}$ worked reasonably well.

We then use this to form an estimate of the set $\mathcal{I}$ and perform a parametric pathfollowing on each NLP. Denote this estimate to be $I_0(x,t)$, i.e.,

$$I_0(x,t) = \{i \in \mathcal{A}^0(x,t) : \forall i \in \{n+1,...,n+n_c\},\ \text{either } i \in I_0(x,t) \text{ or } i+n_c \in I_0(x,t)\} \quad (17)$$

We can make one simple adjustment to save unnecessary duplicate branches: note that if a multiplier corresponding to an inequality constraint is zero at the solution to an NLP, its corresponding linearized constraint in the sensitivity QP is associated with an inequality. This makes the QP redundant to another QP for which the corresponding index is not in the index of the I-NLP if one exists. This can hold if there are more that one doubly-active constraint. as two inequalities.)

3.2.1 Description of the Algorithm

We present the Active Set MPEC pathfollowing algorithm as Algorithm 2.

First, given an approximate B-stationary solution x to the MPCC at t, we estimate and prune the set of index I NLPs. For each $I \in I_0(x,t)$ defined in (17), we form the NLP and solve it by an appropriate solver for standalone NLPs in order to obtain the multiplier estimates. If no appropriate dual variable approximate solutions (i.e., with the right signs) exist, then this I is not in $I_0(x,t)$ and we discard it. Otherwise, the predictor-corrector QP is then used to obtain the next solution. The QP we solve is the following.

$$\begin{array}{ll} \min_{\Delta x} & \nabla_x f(x,t;\rho) + \frac{1}{2}\Delta x^T H(x,\lambda,\mu,t;\rho)\Delta x + \Delta p^T \frac{\partial^2 L(x,\lambda,\mu,t;\rho)}{\partial t \partial x}\Delta x \\ \text{subject to} & h(x,t) + \nabla_t h(x,t)^T \Delta t + \nabla_x h(x,t)^T \Delta x = 0, \\ & g_i(x,t) + \nabla_t g_i(x,t)^T \Delta t + \nabla_x g_i(x,t)^T \Delta x = 0,\ \text{for } i \in A_+(x,\lambda,t), \\ & g_i(x,t) + \nabla_t g_i(x,t)^T \Delta t + \nabla_x c_i(x,t)^T \Delta x \ge 0,\ \text{for } i \in A_0(x,\lambda,t), \\ & x_i = 0,\ i \in \mathcal{A}^0(x,t), \\ & x_i \ge 0,\ i \in \{n+1,...,n+2n_c\} \setminus \mathcal{A}^0(x,t) \end{array} \quad (18)$$

We then check, after the step, if the solution still satisfies the original basic complementarity conditions. Then, we recalculate $I_0(x+\Delta x, t+\Delta t)$ if necessary and investigate if the new primal solution estimate remains stationary for every single index I NLP in $I_0(x+\Delta x, t+\Delta t)$ by obtaining the appropriate dual solution minimizing the stationarity norm,

$$\min_{\sigma_{I_+} \ge 0, \lambda \ge 0} \|\nabla f(x+\Delta x, t+\Delta t) - \lambda^T \nabla g(x+\Delta x, t+\Delta t) - \mu^T \nabla h(x+\Delta x, t+\Delta t) - \sigma\|^2 \quad (19)$$

where $I_+ \subseteq \{n+1, ..., n+2n_c\}$ such that $x_i \geq 0$ for $i \in I_+$. In particular, if the objective value is less than 10^{-3} then we consider it approximately stationary, in our Algorithm. However, this value needs to be found by considerable experimental tuning.

If the point is not feasible, or not approximately optimal for every $I \in I_0(x + \Delta x, t + \Delta t)$, or if the pathfollowing fails, then we cut this branch and proceed to starting the next branch. Finally, if it tracks the solution until the final iterate at $t = 1$, for the sake of completeness we pathfollow the other branches as well.

Algorithm 2 Active Set Pathfollowing Method for MPCC

Input: $t = 0$, x close to x^*, a B-stationary point.

Estimate the set of index I active sets $I_0(x, 0)$.
while $t < 1$ **do**
  for Each $I \in I_0(x, t)$ **do**
    For the index I NLP, find the best estimate for the multipliers (λ, μ, σ) ,
    Estimate the active and strongly active sets $\mathcal{A}(x, \lambda, \mu, \sigma), \mathcal{A}^+(x, \lambda, \mu, \sigma)$ by (11)
    Solve (18) to obtain a predictor-corrector step d_I.
    If the point $x + d_I$ is not approximately feasible for the complementarity, set $d_I = \emptyset$
    Check the objective value of (19) for every other $I \in I_0(x + d_I, t + \Delta t)$.
    If the point $x + d_I$ is KKT-stationary for every other $I \in I_0(x + d_I, t + \Delta t)$, then set $d_J = \emptyset$
  end for
  If there exists a d_I for some I, set $x = x + d_I$.
  If there is more than one such d_I, in parallel set many such new primal solution estimates.
  Re-estimate the double complementary active set $I_0(x + d_I, t + \Delta t)$ and conduct the next pathfollowing step..
end while

3.2.2 Convergence

Theorem 4 *Consider a (finite) set of branches of B-stationary points $\{x_k^*(t)\}_{k=1,...,K}$ for $x_k^*(t)$ defined in $t \in [t_{a_k}, t_{b_k}]$. Given an initial point $x(t_0)$ sufficiently close to $x^*(t_{\hat{k}})$ for some $\hat{k}$, and desired threshold D, Algorithm 2 will generate a set of paths $\{x_k(t)\}$ such that, $\|x_k(t) - x_k^*(t)\| \leq D$ for all $t \in [t_{a_k}, t_{b_k}]$.*

Proof Under Assumption 1, since B-stationarity is equivalent to a set of NLPs being optimal, $x(t_0)$ is an approximate stationary point to each Index I NLP found in $I_0(x, 0)$. If there exists a branch $x^*(t_0 + \Delta t)$ for any $\Delta t > 0$, then by [14, Theorem 4.2] the active set estimation and predictor-corrector QP will trace it to desired accuracy. By the fact that the distance to the solution of any NLP is $O(\cdot)$ of the optimality residual $\eta(\cdot)$, a true B-stationary point will satisfy a low objective value of (19) for every index-I NLP.

If during the pathfollowing another constraint becomes doubly-active at $\bar{t}$, it is added to $I_0(x(\bar{t}), \bar{t})$, and subsequently traced from $\bar{t}$ on.

4 Examples

4.1 Example 1

Consider the problem,

$$\begin{array}{c}\min_{x \in \mathbb{R}^2} \; (x_1 - t)^2 + (x_2 + t)^2 \\ \text{s.t. } 0 \leq x_1 \perp x_2 \geq 0.\end{array}$$

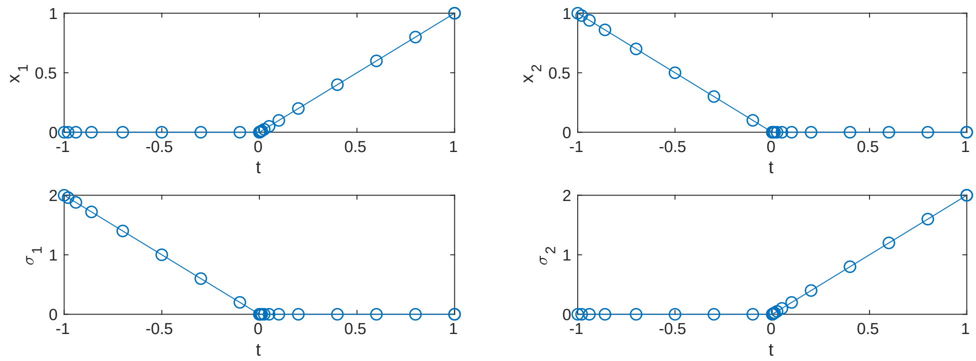

Fig. 1 Illustration of the traced solutions to Example 1 found using the penalty algorithm. The circles are the points visited by the method. As expected, since the solutions are all strongly stationary, the complementarity multipliers are zero when the corresponding variable is inactive and become positive as they become active, and vice versa.

For $t < 0$, $(0, -t)$ is a local minimizer and the origin is weakly stationary. For $t = 0$, the origin is a local minimizer. For $t > 0$, $(t, 0)$ is a local minimizer and the origin is weakly stationary.

Penalty Method

The penalty method is able to successfully trace the solution $(0, -t)$ for $t \in [-1, 0]$ through $(t, 0)$ for $t \in [0, 1]$.

The primal solutions as well as the complementarity multiplier estimates given by (14) are shown in Figure 1.

Active Set Method

The active set method is able to successfully trace the solution $(0, -t)$ for $t \in [-1, 0]$ through $(t, 0)$ for $t \in [0, 1]$. It recognizes the double-active set at $t = 0$ and discards the branch of the Index-I NLP associated with the path of solutions for $t \in [-1, 0]$, switching to the other branch thereafter.

4.2 Example 2

This problem is inspired by [18].

$$\begin{array}{c}\min_{x\in\mathbb{R}^2}\ (x_1 - t)^2 + x_2^3 + x_2^2 \\ \text{s.t. } 0 \le x_1 \perp x_2 \ge 0.\end{array}$$

For $t < 0$, the origin is the local minimizer and only stationary point. For $t > 0$, the point $(t, 0)$ is a local minimizer and the origin is M-stationary.

Penalty Method

In this case, the penalty method always traces the origin for $t \in [-1, 0]$, and the local minimizer $(t, 0)$ for $t \in [0, 1]$. Thus it does not appear to be captured by the M-stationary point. The set of points traced along the homotopy is shown in Figure 2.

Active Set Method

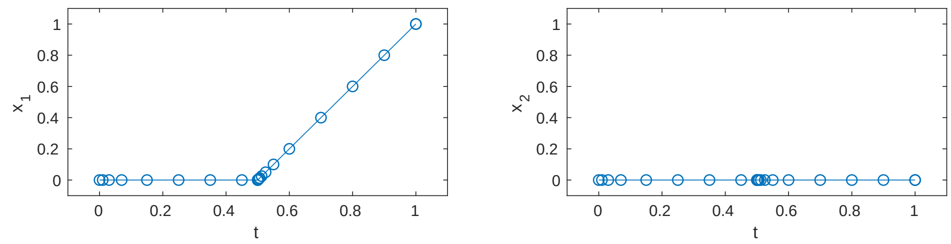

Fig. 2 Illustration of the penalty algorithm solutions for Example 2.

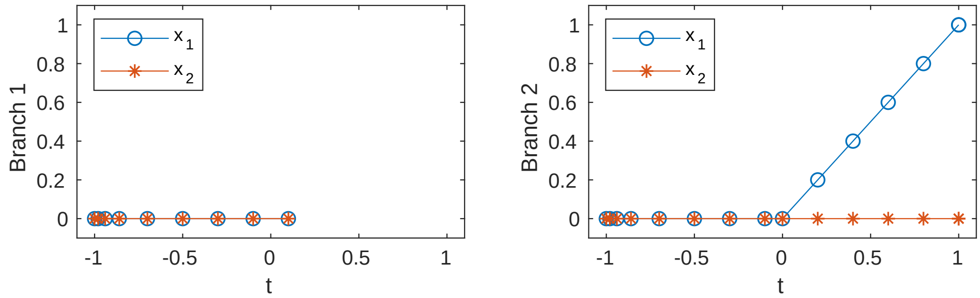

Fig. 3 Illustration of the active set algorithm solutions for Example 2.

The active set method traces two solutions along the bi-active set along $t \in [-1, 0]$. Subsequently, the one holding $x_1 = 0$ stops pathfollowing at $t = 0$ and the one holding $x_2 = 0$ pathfollows successfully along $(t, 0)$ for $t \in [0, 1]$.

The solutions are shown in Figure 3.

4.3 Example 3

This problem arises from [21]. Consider,

$$\begin{array}{c} \min_{x\in\mathbb{R}^2} (x_1 - t)^2 + (x_2 - t)^2 \\ \text{s.t. } 0 \le x_1 \perp x_2 \ge 0. \end{array}$$

For $t \le 0$, the origin is the unique minimizer, and is a B-stationary and C-stationary point. For $t > 0$, the origin is C-stationary, and $(t, 0)$ and $(0, t)$ are both local minimizers.

For any t, the Lagrangian Hessian is $\nabla^2_{xx}\mathcal{L} = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$, as this matrix is positive definite, it is also positive definite on any cone of directions.

We illustrate the problem in Figure 4.

Penalty Method

Note that for $t = 0$, the Lagrangian Hessian for (12) is $\nabla^2_{xx}\mathcal{L}_p = \begin{pmatrix} 2 & \rho \\ \rho & 2 \end{pmatrix}$. At the origin, every multiplier is null, i.e., $\sigma^*_{1,2} = z^*_{1,2} - \rho x^* = 0$, thus no constraint is strongly active and both variables are weakly active at the origin. Thus, although $\nabla^2_{xx}\mathcal{L}_p$ is positive definite along the cone of feasible directions, which are $\begin{pmatrix} 1 & 0 \end{pmatrix}^T$ and $\begin{pmatrix} 0 & 1 \end{pmatrix}^T$, it is not in the subspace corresponding

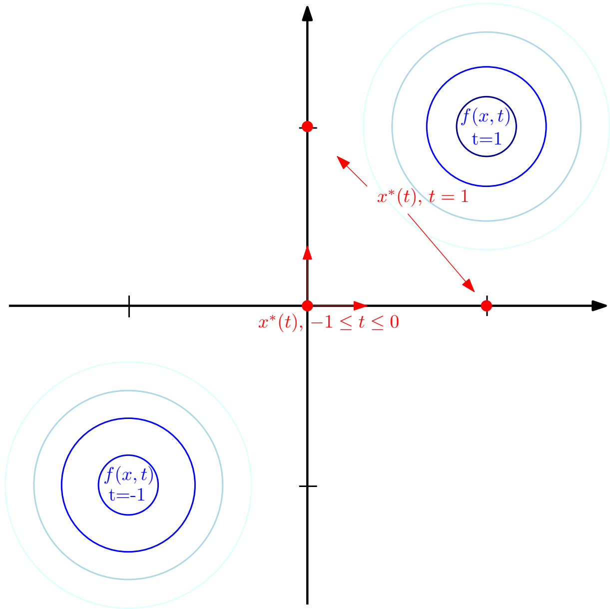

Fig. 4 Illustration of Example 2. The circular level sets of the objective function at $t=-1$ and $t=1$ are drawn as circles. As t traces from -1 to 1, the level sets move continuously from the bottom left to the top right. The local minimizers are marked as red dots. Up until $t=0$, i.e., for $t\in[-1,0]$, the unique local minimizer is the origin. For $t\in(0,1]$ there are two local minimizers, $(t,0)$ and $(0,t)$. In addition, there is a C-stationary point at the origin, for which clearly both axes are a descent direction, indicating the weakness of this stationarity condition .

to the strong second order sufficiency conditions, which includes $\begin{pmatrix}1 & -1\end{pmatrix}^T$. Thus, predictably the penalty algorithm is not able to pathfollow past $t=0$, in particular as the Hessian in the null-space of the pathfollowing QP (9) is indefinite, the corresponding QP is nonconvex. Thus, the results on NLP pathfollowing do not imply in theory, and in practice, that a standard QP solver would be able to solve the corresponding predictor-corrector QP.

The optimality conditions for the QP (10) are,

$$\begin{aligned}&-2t+2\Delta x_1+\rho\Delta x_2=0,\\&-2t+2\Delta x_2+\rho\Delta x_1=0,\\&d_1\geq 0,\; d_2\geq 0.\end{aligned}$$

From these equations, it must hold that $\Delta x_1=\Delta x_2>0$. Note that this is not a feasible direction for the original problem, and thus the QP for the pathfollowing of (12) does not correspond to pathfollowing for the original MPCC.

The penalty method is able to trace the origin solution for $t<0$, and for $t>0$ it appears to pick either local minimizer with equal probability, while never converging to the C-stationary point, and finds it impossible to pathfollow too close to $t=0$.

Active Set Method The active set method, started at $t=-1$, recognizes the presence of the doubly-active set and proceeds with two separate pathfollowings, one with fixed $x_1=0$ and the other with fixed $x_2=0$. The algorithm is able to follow the origin for $t\in[-1,0]$ and each path follows the separate local minimizers ($(t,0)$ and $(0,t)$ for $t\in[0,1]$). The two branches of solutions are shown in Figure 5.

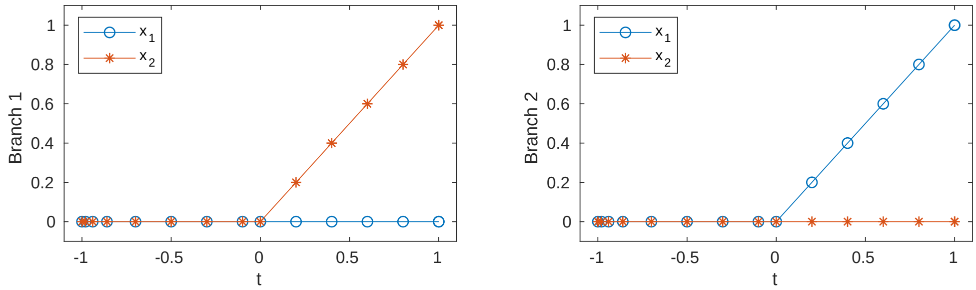

Fig. 5 Illustration of the active set algorithm solutions for Example 2.

4.4 Example 4

This problem is original to this paper.

$$\begin{array}{l} \min_{x\in\mathbb{R}^2} x_1^2 + (x_2+t)^2 \\ \quad\text{s.t. } 0 \le x_1 \perp x_2 \ge 0. \end{array}$$

For $t < 0$ the origin is C-stationary and $(0, -t)$ is a local minimizer. For $t \ge 0$, the origin is a local minimizer and only stationary point.

Penalty Method For $t \in [-1, 0]$, the local minimizer, $(0, -t)$ is traced by the penalty pathfollowing method, and for $t \in [0, 1]$, it traces the origin. If, however, instead of finding the initial solution with the standalone penalty method, the penalty pathfollowing is initialized for $t < 0$ at the origin then it continues to pathfollow the parametric C-stationary point. Thus, because the standalone penalty method seeks descent, it is likely to pick a local minimizer over a weaker stationary point, if both exist, although it still recognizes the latter as a stationary point for the penalty problem.

Active Set Method For $t \in [-1, 0]$, the local minimizer, $(0, -t)$ is traced by the active set pathfollowing method, and for $t \in [0, 1]$, it traces the origin.

4.5 Example 5

This problem is original to this paper.

$$\begin{array}{l} \min_{x\in\mathbb{R}^2} (x_1-1)^2 + (x_2+t)^2 \\ \quad\text{s.t. } x_2 - x_1 \ge 0, \\ \qquad 0 \le x_1 \perp x_2 \ge 0. \end{array}$$

For all $t \in [0, 1]$ the origin is the unique local minimizer. The stationarity conditions for the problem are,

$$\begin{array}{l} \sigma_1 - \lambda = -2, \\ \sigma_2 + \lambda = 2t, \\ \lambda \ge 0. \end{array}$$

Thus both $(\lambda^*, \sigma_1^*, \sigma_2^*) = (2, 0, 2t-2)$ and $(\lambda^*, \sigma_1^*, \sigma_2^*) = (0, -2, 2t)$ are M-stationary, and furthermore for the index $\{1\}$ NLP, the second set of multipliers is optimal, and for the index-$\{2\}$ NLP the first set is optimal, and thus the origin is also B-stationary for all $t \in [0, 1]$.

Penalty Method

The penalty method is unable to find a satisfactory initial point. Initializing from one of the stationary points does not result in any successful pathfollowing step.

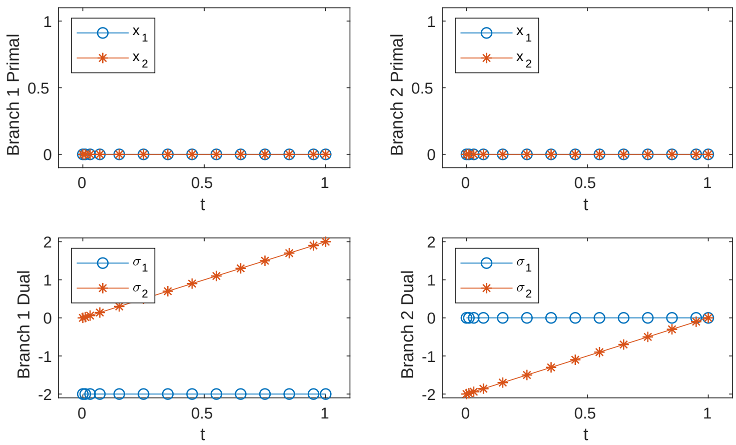

Fig. 6 Illustration of the active set algorithm solutions for Example 5.

Active Set Method

The active set method is able to trace both index I-NLPs for this set of problems, tracing the origin for $t = [0, 1]$ along the multipliers $(2, 0, 2t - 2)$ for the index-$\{1\}$-NLP and $(0, -2, 2t)$ for the index-$\{2\}$-NLP. The primal and dual solutions traced are shown in Figure 6.

4.6 Example 6

This problem arises in the MacMPEC collection of MPEC problems [17].

$$\begin{aligned} &\min_{x\in\mathbb{R}^2}\ x_1 + x_2 - (1-t)x_3 \\ &\text{s.t. } 4x_1 - x_3 \geq 0, \\ &\quad\ 4x_2 - x_3 \geq 0, \\ &\quad\ 0 \leq x_1 \perp x_2 \geq 0. \end{aligned}$$

The point $x^* = (0, 0, 0)$ is B-stationary but not M-stationary for $t \in [0, \frac{3}{4}]$ and strongly stationary for $t \in [\frac{3}{4}, 1]$. The problem does not satisfy MPEC-LICQ, and two options of multipliers are $\lambda^* = (0, 1 - t)$ and $\sigma^* = (1, -3 + 4t)$ as well as $\lambda^* = (1 - t, 0)$ and $\sigma^* = (-3 + 4t, 1)$, indicating that x^* is optimal for both the index-1 and index-2 NLP, but is otherwise only weakly stationary.

Penalty Method

For $t \in [0, \frac{3}{4}]$, the penalty method is unable to find a satisfactory initial point. Initializing from one of the stationary points does not result in any successful pathfollowing step, with the error that every predictor-corrector QP problem is non-convex. For $t \in [\frac{3}{4}, 1]$ the algorithm is able to pathfollow the solution $x^* = (0, 0, 0)$.

Active Set Method

The active set method pathfollows along the origin, doing so twice, once for each index in the double active set, for $t \in [0, 1]$.

4.7 Example Flash Calculation

This section describes a classic chemical engineering problem appearing in, e.g., [22, Chapter 7]. It models vapor/liquid equilibrium, wherein the Gibbs energy is minimized at some given temperature and pressure. In this case we consider pressure to be fixed and vary Temperature. We consider a feed stream F that is separated into vapor V and liquid L product, of three different compounds $\{y_i\}$ and $\{x_i\}$, respectively, with $\sum_i x_i = 1$ and $\sum_i y_i = 1$, and $x_i + y_i = z_i$. Using Raoult's law, we have that,

$$y_i/x_i = K_i = p_i^{sat}(T)/p$$

where p_i^{sat} is a nonlinear function of temperature,

$$p_i^{sat} = 10^{A_i - \frac{B_i}{T+C_i}}$$

The fraction $a \in [0,1] = V/F$ of gas is determined by the Rachford-Rice equation,

$$\sum_i \frac{z_i(K_i - 1)}{1 + a_t(K_i - 1)} = 0.$$

However, the bubble point temperature is $T_b = 382.64$ and the dew point $T_d = 393.30$, and thus for $T < T_b$, the solution for a_t will be negative, but clearly every substance must be liquid, and for $T > T_b$, the solution for a_t will be greater than one, but clearly every substance must be gas.

Thus the true ratio of gas a will be determined by the complementarity system,

$$\begin{aligned} &a - s_v + s_l - a_t = 0, \\ &0 \le s_l \perp L \ge 0, \\ &0 \le s_v \perp V \ge 0, \\ &0 \le a \le 1 \end{aligned}$$

The resulting problem is a nonlinear complementarity system. In addition, the nonlinear equations relating temperature to the component distributions are ill-conditioned. Thus, to define the MPEC, we,

- lift Raoult's law and the Rachford-Rice equation, making K_i, $\frac{1}{(K_i-1)}$ (which we write as k_i), $\log p_i^{sat}$ variables, alongside V, L, x_i, y_i, a, a_t, T, s_V and s_L.
- We consider T as both a parameter and a variable, and define an equality constrained $T = T_{\text{target}}$, and we trace the solution for $T_t = 380$ to 400.
- When taking into account active constraints at the solutions of the NCP, the entire system does not satisfy MPEC-LICQ, thus we do not make the vapor fraction a hard constraint, but define our objective to be,

$$f(x,t) = \frac{1}{2}(aF - V)^2$$

The constants we use in the experiments are given in Table 1.

Table 1 Constants Used in the Flash Calculation.

A_1	A_2	A_3	B_1	B_2	B_3
3.97786	4.00139	3.93002	1064.840	1170.875	1182.774
C_1	C_2	C_3	z_1	z_2	z_3
−41.136	−48.833	−52.532	0.5	0.3	0.2

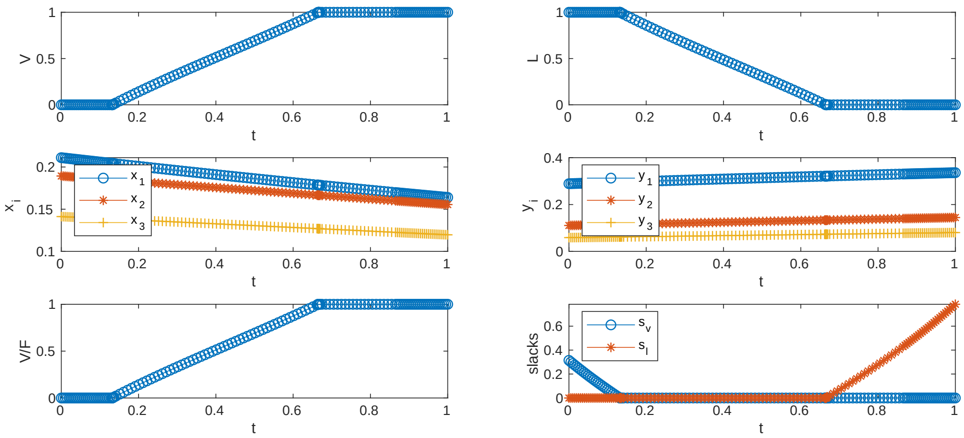

Fig. 7 Primal solutions for the Liquid-Phase flow problem .

4.7.1 Penalty Method

The penalty method accurately tracks the solution, achieving the liquid-gas balance as expected.

Every solution is a strongly stationary point, and the objective function is strictly convex, and so all the necessary conditions for the convergence of the penalty method for solving the MPEC and tracing a parametric curve are satisfied.

The solutions for all of the variables are given in Figure 7. As expected, the variables X and s_x, as well as Y and s_y are complementary, with the slacks being positive when their corresponding complementary variables are zero, and zero when they are inactive. The proportion of gas a corresponds to V. The variables k_i appear to vary the most nonlinearly, resembling an exponential function with respect to the parameter, which is equivalent to the temperature.

Note that the required step-size for pathfollowing is generally fairly small, and gets smaller during active set changes.

4.7.2 Active Set Method

The active set method also traces the set of solutions. At each active set change, a double-active set was recognized. The previous branch was eventually cut, and the other branch initiated. Although the method converged for this problem, it took considerable tuning with regards to tolerances for the double-active constraints, on the one hand, and the required tolerance for checking the other index I NLPs, while tracing a branch, on the other. In particular, numerically, a double-active constraint will stay active for a few homotopy steps before being seen as no longer bi-active, and meanwhile the steps must still be accepted as valid for the other index Is NLP.

4.8 Discussion

The two phase flow example illustrates the general efficacy of the procedures outlined above on real problems. The simpler problems indicate some subtle distinctions of the two methods. In particular, the penalty method is easier to tune and implement, traces S-stationary points, but not B-stationary, and also it can trace C-stationary points, which may not be desired by the practitioner. By contrast, the active set method is able to trace B-stationary points and does not trace and spurious stationary points. However, since it must trace every combination of selections of doubly-active constraints, it is generally slower and more difficult to tune.

5 Conclusion

In this paper we studied the parametric mathematical program with complementarity constraints, presenting the first investigation of algorithmically pathfollowing these programs. We formulated two algorithms, one based on the penalty method for standalone MPCCs, and one based on following the possible branches of NLPs suggested by double-active constraints and the B-stationarity condition. We studied their performance on a set of examples constructed to distinguish the types of parametric properties there could be, illustrating the performance on this illustrative set of problems, as well as a "real-world" problem associated with chemical process engineering.

Pathfollowing parametric optimization problems can be important in a number of applications, including nonlinear model predictive control, as well as assessing the practical sensitivity of a solution with respect to a parameter. We intend that this work presents a solid first step in the development of literature and algorithms for the pathfollowing of parametric MPCCs. In general, the tradeoffs associated with seeking tighter versus looser notions of stationarity inherent to standalone methods for solving MPCCs carried over to pathfollowing parametric ones, and in each case we expect this to enrich further understanding of MPCCs in general and algorithms for pathfollowing parametric MPCCs in particular.

Acknowledgements

We would like to thank Sven Leyffer for sharing his insight regarding the different stationary point concepts addressed in this paper and suggesting one of the examples.

References

1. Baumrucker, B., Renfro, J., Biegler, L.T.: Mpec problem formulations and solution strategies with chemical engineering applications. Computers & Chemical Engineering **32**(12), 2903–2913 (2008)
2. Bouza, G., Guddat, J., Still, G.: Critical sets in one-parametric mathematical programs with complementarity constraints. Optimization **57**(2), 319–336 (2008)
3. Coulibaly, Z., Orban, D.: An \ell_1 elastic interior-point method for mathematical programs with complementarity constraints. SIAM Journal on Optimization **22**(1), 187–211 (2012)
4. Diehl, M., Bock, H.G., Schlöder, J.P.: A real-time iteration scheme for nonlinear optimization in optimal feedback control. SIAM Journal on control and optimization **43**(5), 1714–1736 (2005)
5. Dinh, Q.T., Savorgnan, C., Diehl, M.: Adjoint-based predictor-corrector sequential convex programming for parametric nonlinear optimization. SIAM Journal on Optimization **22**(4), 1258–1284 (2012)
6. Dontchev, A.L., Krastanov, M., Rockafellar, R.T., Veliov, V.M.: An euler–newton continuation method for tracking solution trajectories of parametric variational inequalities. SIAM Journal on Control and Optimization **51**(3), 1823–1840 (2013)
7. Facchinei, F., Fischer, A., Kanzow, C.: On the accurate identification of active constraints. SIAM Journal on Optimization **9**(1), 14–32 (1998)
8. Guddat, J., Vazquez, F.G., Jongen, H.T.: Parametric optimization: singularities, pathfollowing and jumps. Springer (1990)
9. Guo, L., Lin, G.H., Ye, J.J.: Second-order optimality conditions for mathematical programs with equilibrium constraints. Journal of Optimization Theory and Applications **158**(1), 33–64 (2013)
10. Izmailov, A.: Solution sensitivity for karush–kuhn–tucker systems with non-unique lagrange multipliers. Optimization **59**(5), 747–775 (2010)
11. Jäschke, J., Yang, X., Biegler, L.T.: Fast economic model predictive control based on nlp-sensitivities. Journal of Process Control **24**(8), 1260–1272 (2014)

12. Jittorntrum, K.: Solution point differentiability without strict complementarity in nonlinear programming. In: Sensitivity, Stability and Parametric Analysis, pp. 127–138. Springer (1984)
13. Jongen, H.T., Shikhman, V., Steffensen, S.: Characterization of strong stability for c-stationary points in mpcc. Mathematical programming **132**(1-2), 295–308 (2012)
14. Kungurtsev, V., Diehl, M.: Sequential quadratic programming methods for parametric nonlinear optimization. Computational Optimization and Applications **59**(3), 475–509 (2014)
15. Kungurtsev, V., Jaschke, J.: A predictor-corrector path-following algorithm for dual-degenerate parametric optimization problems. SIAM Journal on Optimization **27**(1), 538–564 (2017)
16. Lenders, F.J.M.: Numerical methods for mixed-integer optimal control with combinatorial constraints. Ph.D. thesis (2018)
17. Leyffer, S.: Macmpec: Ampl collection of mpecs. Argonne National Laboratory (2000)
18. Leyffer, S., Munson, T.S.: A globally convergent filter method for mpecs. Preprint ANL/MCS-P1457-0907, Argonne National Laboratory, Mathematics and Computer Science Division (2007)
19. Raghunathan, A.U., Biegler, L.T.: Mathematical programs with equilibrium constraints (mpecs) in process engineering. Computers & chemical engineering **27**(10), 1381–1392 (2003)
20. Ralph*, D., Wright, S.J.: Some properties of regularization and penalization schemes for mpecs. Optimization Methods and Software **19**(5), 527–556 (2004)
21. Scheel, H., Scholtes, S.: Mathematical programs with complementarity constraints: Stationarity, optimality, and sensitivity. Mathematics of Operations Research **25**(1), 1–22 (2000)
22. Skogestad, S.: Chemical and energy process engineering. CRC press (2008)
23. Suwartadi, E., Kungurtsev, V., Jäschke, J.: Sensitivity-based economic nmpc with a path-following approach. Processes **5**(1), 8 (2017)
24. Zavala, V.M., Anitescu, M.: Real-time nonlinear optimization as a generalized equation. SIAM Journal on Control and Optimization **48**(8), 5444–5467 (2010)
25. Zavala, V.M., Biegler, L.T.: The advanced-step nmpc controller: Optimality, stability and robustness. Automatica **45**(1), 86–93 (2009)