M.C. Ferris, T. Munson and D. Ralph

A homotopy method for mixed complementarity problems based on the PATH solver

Abstract Mixed complementarity problems can be recast as zero finding problems for the normal map, a function that is smooth on the interior of each of the cells of a piecewise linear manifold of $\mathbb{R}^n$, called the normal manifold. We develop a predictor-corrector, or path following, homotopy method based upon using piecewise linear approximations to the piecewise smooth normal map. A description of an implementation using technology found in the PATH solver is given along with computational experience on the MCPLIB test suite.

1 Introduction

The complementarity problem arises in many different applications [14]. The original source of these problems were the optimality conditions of linear, quadratic, and nonlinear programs. Since that time, many other applications, including game theory, economic equilibria, and structure design/failure have been postulated and solved as complementarity problems. In this paper, we will focus on the mixed (nonlinear) complementarity problem of finding $z \in [l, u]$ such that for all $i = 1, \ldots, n$,

$$\begin{array}{rcl} f_i(z) & \geq & 0 \quad \text{if } z_i = l_i \\ f_i(z) & \leq & 0 \quad \text{if } z_i = u_i \\ f_i(z) & = & 0 \quad \text{if } l_i < z_i < u_i, \end{array}$$

where $f : \mathbb{R}^n \to \mathbb{R}^n$ is a continuously differentiable function, l_i and u_i are fixed, possibly infinite numbers satisfying

$$-\infty \leq l_i < u_i \leq +\infty$$

and $[l, u]$ denotes the set of $z \in \mathbb{R}^n$ such that $l_i \leq z_i \leq u_i$ for all i. We call this problem MCP(f, l, u) or MCP for short.

Many monotone equilibrium problems and convex optimization problems whose constraints C are *not* simple boxes $[l, u]$ can also be reformulated as MCPs. For example, consider the nonlinear program

$$\min \phi(z) \;\text{ subject to }\; Az = b,\; g_j(z) \leq 0 \text{ for } j = 1, \ldots, q$$

where f and each g_j are real, differentiable, convex functions on $\mathbb{R}^n$, and $A \in \mathbb{R}^{p \times n}$, $b \in \mathbb{R}^p$. If the Slater constraint qualification holds, namely there exists $\hat{z}$ such that

$A\hat{z} = b$ and $g_j(\hat{z}) < 0$ for each j, then finding a minimizer z, global or local, of this nonlinear program is equivalent to solving its stationarity or Karush-Kuhn-Tucker conditions; see [21]. The KKT conditions form an MCP in z and auxiliary variables $\mu \in \mathbb{R}^p$ and $\lambda \in \mathbb{R}^q$ called (KKT or Lagrange) multipliers. Define an MCP function $f : \mathbb{R}^{n+p+q} \to \mathbb{R}^{n+p+q}$ by

$$f(x, \mu, \lambda) \;=\; (\nabla\phi(x) \,-\, A^T\eta \,-\, \nabla g(x)^T\lambda,\; Ax - b,\; g(x))$$

where $\nabla\phi(x)$ is the gradient vector of ϕ at x, $g(x) = (g_1(x), \ldots, g_q(x))$ and $\nabla g(x)$ is the $q \times n$ Jacobian matrix of g at x. Also let the box $[l, u]$ be $\mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^q_-$. Then the associated MCP is precisely the KKT conditions of the nonlinear program, hence is equivalent to the nonlinear program.

Our method will be based on a reformulation of the mixed complementarity problem. To explain this reformulation, let us first consider a nonempty polyhedral subset of $\mathbb{R}^n$, C. Associated with each of the faces of C is a (full-dimensional) polyhedral set, and the collection of these polyhedra comprise a piecewise linear manifold of $\mathbb{R}^n$ called the *normal manifold*, denoted $\mathcal{N}_C$. These polyhedra are called the *cells* of $\mathcal{N}_C$; a full description along with important properties is given in [24] (see also [22] for further investigation). For example, when $C = \mathbb{R}^n_+$, that is $C = [l, u]$ with each $l_i = 0$ and $u_i = \infty$, the cells of $\mathcal{N}_C$ are the orthants of $\mathbb{R}^n$.

We denote by $\pi_C(\cdot)$ the Euclidean projection mapping onto the set C; thus for the box $C = [l, u]$, $\pi_C(x)$ is the vector whose ith component is l_i if $x_i \le l_i$, x_i if $l_i \le x_i \le u_i$, and u_i otherwise. The normal map [24] induced by (f, C) is the function $f_C : \mathbb{R}^n \to \mathbb{R}^n$ given by

$$f_C(x) \;=\; f(\pi_C(x)) + x - \pi_C(x).$$

It is well known that the projection map $\pi_C(x)$ and hence the normal map $f_C(x) = f(\pi_C(x)) + x - \pi_C(x)$ are smooth in the interior of each of the cells of $\mathcal{N}_C$. The key point is that for $C = [l, u]$, if x satisfies $f_C(x) = 0$, then $z = \pi_C(x)$ solves MCP. Furthermore, if z solves MCP then $x = z - f(z)$ is a zero of f_C.

We propose and implement a homotopy (or continuation) method for finding zeroes of the normal map $f_C(x)$ when $C = [\ell, u]$ based on classical predictor-corrector ideas. Previously proposed algorithms [29, 27, 30] use linear approximations to single (smooth) pieces of a piecewise smooth mapping, see also [16, 17]. The method proposed in this paper is based on piecewise linear approximations of a piecewise smooth mapping. Both our method and that of [30] are based upon the theoretical foundations developed in [1, 2, 3].

An alternative homotopy approach investigated recently [5] is to form a homotopy that is smooth except in the limit as the homotopy parameter reaches its final value. This makes application of smooth homotopy codes attractive though some care near the end of the homotopy path, when the problem is tending toward nonsmoothness, may be needed.

Section 2 gives some background material on homotopy methods. Section 3 outlines the theoretical underpinnings of our method. Section 4 describes the implementation of a large scale code based on this theory and Section 5 gives some computational results on problems from MCPLIB [7].

2 Background

We give the notation and background for homotopy methods, in particular for normal maps induced by smooth functions and polyhedral convex sets.

2.1 Homotopy mappings and paths

Continuation or homotopy methods are a technique to trace the zeroes of a homotopy mapping, starting from an easily constructed zero and moving towards the solution to the problem of interest. To describe such methods, we first need to understand the homotopy mapping.

For a fixed vector $a \in \mathbb{R}^n$, we construct a homotopy mapping $H_C^a : \mathbb{R}^{n+1} \to \mathbb{R}$ from the function $x \mapsto x - a$ to the function f_C by interpolation: for $(x,t) \in \mathbb{R}^n \times [0,1]$,

$$H_C^a(x,t) \ = \ (1-t)(x-a) + t f_C(x).$$

Note that $x = a$ is the unique solution of $H_C^a(x,0) = 0$, and $H_C^a(x,1) = 0$ if and only if x is a zero of the normal map. Hence if we can calculate the endpoint $x^a(1)$ of the homotopy path, then we have solved the original problem. The idea of a continuation method is to analytically or numerically determine a path of solutions $x^a(t)$ of the equation $H_C^a(\cdot,t) = 0$ from $t = 0$, where $x^a(0) = a$ is the unique solution, to $t = 1$, where as mentioned, $x^a(1)$ is a zero of the normal map.

A standard list of properties of the homotopy path $x^a(t)$ follows. We only state these for the case of H_C^a defined above, though the class of homotopies with these properties is considerably larger [3] as we explain after Theorem 1. Parts 1 and 2 below summarize results that are due, in essence, to [3], while part 3 is due to [27]; see the proof for details. Parts 4 and 5 seem to be new. Part 5 refers to subanalytic functions [20, 15] which we define, for completeness, following the statement of the theorem. However we refer to [11, Section 5.2] for a succinct introduction to the properties of subanalytic sets and functions, and additional references. Also, the notation $t \to 1^-$ means $t \to 1$, $t < 1$.

Theorem 1 *Let $f : \mathbb{R}^n \to \mathbb{R}^n$ be C^1 and C be a nonempty polyhedral convex set in $\mathbb{R}^n$. For almost all $a \in \mathbb{R}^n$:*

1. *There exist $T > 0$ and a piecewise smooth path $x^a : [0,T) \to \mathbb{R}^n$ such that $x^a(0) = a$, and $\{(x^a(t),t) ; t \in [0,T)\}$ is a connected component of the set $\{(x,t) : t \in [0,T),\, H_C^a(x,t) = 0\}$.*

2. *The function $x^a(\cdot)$ is nondegenerate in the sense that for each $t \in [0,T)$, first, $x^a(t)$ lies either in the interior of a unique cell of the normal manifold $\mathcal{N}_C$ or in*

the relative interior of a facet (an $(n-1)$-dimensional face of a cell) that is the intersection of two cells of $\mathcal{N}_C$; and, second, if $x^a(t)$ lies in a cell σ and g_σ denotes the smooth mapping that represents H^a_C on σ, then the Jacobian (derivative) matrix $\nabla g_\sigma(x^a(t), t)$ has full rank.

3. *If f is C^2 and there exists a point $u \in C$ and a scalar $\rho > \|u\|$ such that for each $c \in C$ with $\|c\| = \rho$ we have $\langle f(c), c-u\rangle \geq 0$, then $\{x^a(t) : t \in [0,1)\}$ is bounded and its closure contains a point $\hat{x}$ with $f_C(\hat{x}) = 0$.*

4. *If f is C^2, $T \geq 1$, and there exists a limit point $\hat{x}$ of $p(t)$ as $t \to 1^-$ such that the directional derivative $f'_C(\hat{x};\cdot)$ is invertible, then the arc length of the path $x^a : [0,1) \to \mathbb{R}^n$ is finite and $x^a(t) \to \hat{x}$ as $t \to 1^-$.*

5. *If $T \geq 1$ and $\hat{x}$ is a limit point of $x^a(t)$ as $t \to 1^-$, then $f_C(\hat{x}) = 0$. If, in addition, either $T > 1$, or f is subanalytic then $x^a(t) \to \hat{x}$ as $t \to 1^-$.*

Proof Parts 1 and 2 are essentially due to [3, Theorem 1]. The fact that the graph of x^a is a connected component of $\{(x,t) : t \in [0,T),\ H^a_C(x,t) = 0\}$ is not stated but can be easily deduced from the discussion on transversality prior to [3, Theorem 1]. See also [30, Theorem 1] and subsequent discussion of these transversality properties. Part 3 is quoted from [28, Proposition 5.1].

For part 4, suppose $\hat{x}$ is a limit point of $x^a(t)$ as $t \to 1^-$ such that $f'_C(\hat{x};\cdot)$ is invertible. The implicit function theorem of [23] says that the path $x^a(t)$ is locally uniquely defined for t near 1^-: for some $\epsilon > 0$ and all t near 1 with $t < 1$, the homotopy equation $H^a_C(x,t) = 0$ has a unique solution within distance ϵ of $\hat{x}$. It follows that the path $\{(x^a(t),t) : t \in [0,1)\}$ is bounded and the only limit point of $x^a(t)$ as $t \to 1^-$ is $\hat{x}$ itself; i.e. $x^a(t)$ converges to $\hat{x}$ as $t \to 1^-$. Now invertibility of $f'_C(\hat{x};\cdot)$ is equivalent to other properties such as strong regularity of the associated variational inequality at $\pi_C(\hat{x})$, [30, Definition 2.2], and coherent orientation of the normal map $\nabla f(\hat{x})_{K(\hat{x})}$, [24], where $K(\hat{x})$ is the critical cone to C at $\hat{x}$, and this normal map actually coincides with $f'_C(\hat{x};\cdot)$; see [27, Section 5] and [30, Section 2] for discussion and related results. Therefore the analysis of [27, Proposition 5.6] can be adapted to show finite arc-length, using boundedness of the path and the fact that $\hat{x}$ is the only limit point of $x^a(t)$ as $t \to 1^-$.

For part 5, it is clear by continuity of f_C that $f_C(\hat{x}) = 0$ if $(\hat{x}, 1)$ lies in the closure of $\{(x^a(t),t) : t \in [0,1)\}$. If $T > 1$, continuity of $x^a(\cdot)$ immediately yields convergence of $x^a(t)$ to $\hat{x}$ as $t \to 1$. Suppose instead that f is subanalytic. We will show convergence of $x^a(t)$ to $\hat{x}$ as $t \to 1^-$ by refining an argument in the proof of [11, Theorem 5.10]. All polyhedral convex sets are subanalytic so that the projection π_C mapping is also subanalytic [11, Lemma 5.9]. It follows that H^a_C is also subanalytic since the composition of subanalytic mappings is also subanalytic, hence that $(H^a_C)^{-1}(0)$ is a subanalytic set. Let $P^a = \{(x^a(t),t) : t \in [0,1)\}$; from part 1, taking $T = 1$, there is a neighborhood U of P^a such that $U \cap (H^a_C)^{-1}(0) = P^a$. Without loss of generality assume U is subanalytic; so then P^a, the intersection of subanalytic sets, is also subanalytic. This shows that $x^a : [0, T) \to \mathbb{R}^n$ is subanalytic. The lemma in the Appendix concludes the proof. $\square$

With regard to part 5: a set S in $\mathbb{R}^n$ is *subanalytic* if there exists a semianalytic set T in a higher dimensional space $\mathbb{R}^N$ such that $S = \{x \in \mathbb{R}^n : (x, y) \in T\}$. A set T in $\mathbb{R}^N$ is *semianalytic* if for each $\bar{z} \in T$ there is a neighborhood V of $\bar{z}$ such that $T \cap V$ can be written as the finite union of sets of the form

$$V \cap \{z \in \mathbb{R}^N : f_i(z) = 0, i = 1, \ldots, I; g_j(z) < 0, j = 1, \ldots, J\}$$

where I and J are nonnegative integers, and each f_i and g_j is a real analytic mapping on $\mathbb{R}^N$. A function is subanalytic if its graph is a subanalytic set. Thus the class of subanalytic functions is rather broad.

We note that Theorem 1 holds, almost in its entirety, if we replace the homotopy mapping H_C^a by one of the form

$$(1-t)\phi(a, x) + t f_C(x)$$

where $\phi : \mathbb{R}^{m+n} \to \mathbb{R}^n$ is a *sufficient* mapping [3], that is ϕ is smooth such that its partial Jacobian matrix with respect to $a \in \mathbb{R}^m$, $\nabla_a \phi(a, x)$, has full rank for all $(a, x) \in \mathbb{R}^m \times \mathbb{R}^n$. (For example, the function $\phi(a, x) = x - a$ used in the definition of H_C^a is sufficient.) To be precise, in order to generalize Theorem 1 using a homotopy with a sufficient mapping as described, we only need to replace the qualifier $a \in \mathbb{R}^n$ by $a \in \mathbb{R}^m$ and, in the statement of part 5, the condition "f is subanalytic" by "f and ϕ are subanalytic".

For future reference, we define the homotopy path as the set

$$P^a = \{(x^a(t), t)) : t \in [0, T)\}$$

where T is the maximum value in $(0, 1]$ such that part 1 of Theorem 1 holds.

2.2 Predictor-corrector methods

Predictor-corrector methods are a class of algorithms that attempt to numerically traverse the homotopy path by a sequence of predictor-corrector iterations. Suppose for the moment that H_C^a is smooth, e.g. $C = \mathbb{R}^n$ and $f_C = f$. At step k we are given an iterate (x^k, t^k) that is approximately on the path, that is $x^k \approx x^a(t^k)$. The predictor step identifies a nonzero vector $d = (d_x, d_t) \in \mathbb{R}^n \times \mathbb{R}^1$ tangent to the path at the current point and estimates a new point further along the path $(x', t') = (x^k, t^k) + hd$ for some small $h > 0$. The direction d is found as a vector in the kernel of the Jacobian matrix $\nabla H_C^a(x^k, t^k)$ that maintains the correct orientation, that is we want to move tangent to the curve in the direction that increases arc length. The corrector step then tries to identify a point $(x^a(s), s)$ on the path near to (x', t'); this is used to define the next iterate, $(x^{k+1}, t^{k+1}) = (x^a(s), s)$. The corrector step is usually carried out by a version of Newton's method that uses the Moore-Penrose inverse of the $n \times (n+1)$ matrix $\nabla H^a_C(x,t)$ starting with $(x,t) = (x', t')$ and proceeding until $H^a_C(x,t)$ is approximately zero; see [4].

When H^a_C is piecewise smooth, as it is when f is C^1 and C is polyhedral convex, it is possible to mimic the predictor and corrector steps taken in the smooth case by either staying within one cell, or identifying an adjacent cell and moving into it. This means that at each predictor step a cell σ_k containing the current iterate (x^k, t^k) is identified, and the predictor direction d is chosen using the Jacobian matrix of the smooth mapping corresponding to H^a_C on σ^k. The Moore-Penrose-Newton iteration is undertaken in a similar way. This is the line of development in [30]. We note that these algorithms, and our proposed method to follow, include some nontrivial technical details that are necessary to make it theoretically viable. In particular, it is assumed that the chosen starting point a yields a path $x^a(t)$ that is nondegenerate as described in Theorem 1, part 2. The implementation must take special care in the near-degenerate case for which the path passes through or near a point that is contained in three or more cells.

We prefer to use a piecewise linear (PL) predictor approximation, that is to generate a piecewise linear tangent path to the piecewise smooth homotopy path. An advantage of the PL predictor is that, in principle, it allows the near degenerate situation to be handled in the same way as the nondegenerate case. Another advantage of the PL tangent path approach is that we can use the code developed over several versions of the PATH solver [8] for MCP; we rely on this code to handle numerical degeneracy in the PL path for example.

Consistency would dictate that each step of the corrector also be taken with respect to a PL model of the actual path. However for simplicity of implementation we have chosen a method more like that of [30]: we pick a cell containing the current point and apply a Moore-Penrose-Newton iteration as a heuristic for decreasing the distance to the path. Details of both the predictor and corrector step are given below.

3 Homotopy method

3.1 Predictor step

To describe the predictor step, we need to define a piecewise linear path that is tangent to the piecewise smooth homotopy path P^a at a given point. To do this, we need to describe a linearization of H^a_C.

For convenience in approximating H^a_C, we write it in another way. Let $H^a : \mathbb{R}^{n+1} \to \mathbb{R}^n$ be the C^1 homotopy from the mapping $z \mapsto z - a$ to $f(z)$,

$$H^a(z,t) \;=\; (1-t)(z-a) + tf(z),$$

and observe that

$$H^a_C(x,t) \;=\; H^a(\pi_C(x),t) \;+\; x - \pi_C(x).$$

We define the linearization of the smooth mapping H^a at any point $(z,t) \in \mathbb{R}^{n+1}$ using the first two terms of the Taylor's series expansion: for $(z',t') \in \mathbb{R}^{n+1}$,

$$\mathcal{L}H^a(z,t)(z',t') \;=\; H^a(z,t) \,+\, [(1-t)I + t\nabla f(z)](z'-z) \,+\, [f(z)-(z-a)](t'-t).$$

Now we define the linearization of the piecewise smooth mapping H^a_C at (x,t) as a function of $(x',t') \in \mathbb{R}^{n+1}$,

$$\mathcal{L}H^a_C(x,t)(x',t') \;=\; \mathcal{L}H^a(\pi_C(x),t)(\pi_C(x'),t') \,+\, x' - \pi_C(x').$$

This is a piecewise linear mapping in (x',t'). It is a "point based approximation" to H^a_C at (x,t) in the terminology of [25].

By the "piecewise linear path that is tangent to the piecewise smooth homotopy path P^a at (x^k,t^k)" we mean the set

$$P^a_k \;=\; \{(x,t) \,:\, t \text{ near } t^k,\, \mathcal{L}H^a_C(x^k,t^k)(x,t)=0\}. \tag{3.1}$$

Note $(x^k,t^k) \in P^a_k$. Given a step size $h_k > 0$ our task is to partially construct P^a_k, starting from (x^k,t^k) and moving in the direction that is associated with increasing the arc-length of the actual homotopy path P^a. This "one-sided" path is denoted Q^a_k. We keep moving along Q^a_k until we determine a point (x',t') at distance h_k from (x^k,t^k). The new point (x',t') is our prediction for the next point on the path P^a. We need a corrector scheme to move from (x',t') to a nearby point in P^a; this will be discussed in Section 3.2.

For the subsequent discussion we assume for all $t \in [0,1)$ that either $x^a(t)$ lies in the interior of a cell of $\mathcal{N}_C$ or in the relative interior of one of its facets, and that the Jacobian matrices at $(x^a(t),t)$ of the smooth functions representing H^a_C on these cells have full rank. These properties hold for almost all a by Theorem 1.

We note that the full rank of the Jacobians of H^a_C near (x^k,t^k), together with nondegeneracy of x^k, is enough to show that the set P^a_k does indeed define a piecewise linear path, i.e. a one-dimensional piecewise affine manifold, for t near t^k. This PL path is tangent to the piecewise smooth homotopy path P^a at (x^k,t^k) if $(x^k,t^k) \in P^a$. It will be necessary to generate $Q^a_k \subset P^a_k$ starting from (x^k,t^k) and moving in one of the two possible directions as determined by the orientation calculation presented next. In practice we allow for (x^k,t^k) to be near rather than in P^a — c.f. the tolerance ϵ_c used in the corrector step to follow — so that P^a_k is approximately tangent to P^a. Such details, including the invariance of orientation on the homotopy path, are spelled out in [1, 2, 3, 4].

Orientation of the path

An orientation parameter $\eta = \pm 1$ is used to decide which way to go on the path P^a_k when starting at (x^k,t^k). The orientation parameter is defined prior to the first predictor step of the homotopy method as

$$\eta \;=\; \operatorname{sgn}\det \begin{bmatrix} \nabla H^a_C(a,0) \\ (d^0)^T \end{bmatrix},$$

where $d^0 = (-f(a), 1)$. Note H^a_C is differentiable at any $(x, 0)$. Initially $t^0 = 0$, we have $x^0 = a = x^a(0)$, and $\nabla H^a_C(a, 0) = [I\, f(a)]$, so that the chosen vector d^0 is in the kernel of $\nabla H^a_C(x^0, t^0)$ as in the smooth case. Therefore

$$\begin{aligned}\eta &= \operatorname{sgn}\det \begin{bmatrix} I & f(a) \\ -f(a)^T & 1 \end{bmatrix} \\ &= \operatorname{sgn}\det \left(\begin{bmatrix} I & 0 \\ -f(a)^T & 1 + \|f(a)\|^2 \end{bmatrix} \begin{bmatrix} I & f(a) \\ 0 & 1 \end{bmatrix} \right) \;=\; 1,\end{aligned}$$

where we are using the fact that the determinant of a product of two matrices is the product of the determinants of the matrices. Writing $d^0 = (d_x, d_t) \in \mathbb{R}^n \times \mathbb{R}$, we have chosen $d_t = 1 > 0$ since initially we plan to move along the path P^a by increasing t, i.e. our first prediction will be a point $(a, 0) + h_0 d^0$, for some $h_0 > 0$, which we believe will be close to a point $(x^a(t), t)$ with $t \approx h d_t$.

Let $(x^k, t^k) \in P^a$, σ be a cell containing x^k, and $\mathcal{L}H^a_C(x^k, t^k)$ be represented on σ by an affine map whose Jacobian $J \in \mathbb{R}^{n\times(n+1)}$ at (x^k, t^k) has full rank. So the kernel of J is a one-dimensional space. Let $\ker J$ be spanned by a nonzero vector $d = (d_x, d_t) \in \mathbb{R}^n \times \mathbb{R}$, where we also ensure $x^k + s d_x \in \sigma$ for small $s > 0$ by using $-d$ instead of d if necessary (this is possible since x^k is interior either to the cell σ or to the union of two cells). Now calculate the sign η_k of the determinant of the $(n + 1) \times (n + 1)$ matrix consisting of the matrix J augmented by the row d^T (see Section 4.1).

If $\eta_k = \eta$ then we generate the first part of the path Q^a_k starting at (x^k, t^k) and moving in the nonzero direction $d^0 = d$ specified by the (kernel of the) Jacobian of $\mathcal{L}H^a_C(x^k, t^k)$ in σ. Otherwise $\eta_k = -\eta$, and we move in the "opposite direction" which is either $d^0 = -d$ if $x^k \in \operatorname{int}\sigma$ or, if x^k is in the intersection of σ and another cell σ', a nonzero direction d^0 constructed with reference to the cell σ' instead of σ_k. (Invariance of orientation is demonstrated, in part, by the result that if σ^0 denotes the cell σ or σ' associated with d^0, depending on which situation occurs, and J^0 is the Jacobian at (x^k, t^k) of $\mathcal{L}H^a_C(x^k, t^k)$ restricted to σ^0, then the matrix consisting of J^0 augmented with the row $(d^0)^T$ has determinantal sign equal to η.)

Generating the path Q^a_k

The PL path $Q^a_k \subset P^a_k$ is generated, one line segment at a time, in the following way. The path generation procedure is initialized with $(\xi^0, \tau^0) = (x^k, t^k)$, the nonzero direction $d = d^0$ as described in Section 3.1, and an associated cell σ^0. In fact, at each step we are given a point $(\xi, \tau) \in P^a_k$, a cell σ of the normal manifold $\mathcal{N}_C$ that contains ξ, and a nonzero direction $d = (d_x, d_t) \in \mathbb{R}^n \times \mathbb{R}$ such that $\xi + s d_x \in \sigma$ for all small $s > 0$. The next line segment on the path is generated by finding the maximum value, possibly infinite, of $s \geq 0$ such that $\xi + s d_x \in \sigma$; denote this value by s'. If s' is finite and the path Q^a_k is nondegenerate (as it must be for almost all a) then the new point is $(\xi', \tau') = (\xi, \tau) + s'd \in Q_k^a$ and there is a cell σ' distinct from σ such that (ξ', τ') lies in the relative interior of a common facet of σ' and σ. It is now possible, as in Section 3.1, to choose a unit direction $d' = (d'_x, d'_t)$ in the kernel of the Jacobian J' of $\mathcal{L}H_C^a(x^k, t^k)$ when restricted to σ', such that $\xi' + sd'_x \in \sigma'$ for small $s \geq 0$. The remainder of the path is generated inductively.

A more formal justification of this path generating procedure can be given in terms of [10] where it is also shown how the degenerate case can be handled. It follows from this paper that Q_k^a is composed of finitely many line segments, the last of which is either a ray or contains the starting point (x^k, t^k).

The path generating procedure may be terminated in any step that finds (d_x, d_t) and s', given (ξ, τ) and a cell σ containing ξ, as follows. Recall h_k is a positive parameter bounding the distance of points on Q_k^a to the starting point (x^k, t^k). Termination occurs at the point $(x', t') = (\xi, \tau) + s(d_x, d_t)$ for the first $0 < s \leq s'$ such that any one of the following conditions holds:

(i) $\|(x', t') - (\xi^0, \tau^0)\| = h_k$

(ii) $t' = 1$

(iii) $t' = 0$

(iv) $(x', t') = (\xi^0, \tau^0)$.

Stopping condition (iii) is used since there is only one solution $x = x^0 = a$ to the equation $H_C^a(x, 0) = 0$, and we are not interested in revisiting x^0. Stopping condition (iv) is used to prevent a kind of cycling, that is traveling in an endless loop back.

3.2 Corrector step

We are given a corrector tolerance $\epsilon_c > 0$, that is we expect every iterate (x^k, t^k) to satisfy

$$\left\| H_C^a(x^k, t^k) \right\| \leq \epsilon_c.$$

Therefore if the point (x', t') is the result of a predictor step from (x^k, t^k), then the role of the corrector step is to find the next iterate (x^{k+1}, t^{k+1}) somewhat near (x', t') such that $\left\| H_C^a(x^{k+1}, t^{k+1}) \right\| \leq \epsilon_c$. If an iterative corrector scheme is unable to provide such a point within J_c iterations, where J_c is a positive integer, then the corrector step is deemed to have failed, and a new predicted point (x', t') must be provided.

A theoretically robust corrector scheme

The corrector step can be extended from the Moore-Penrose Newton method [4] for smooth operators, which we paraphrase as follows. Suppose $H : \mathbb{R}^{n+k} \to \mathbb{R}^n$ is a smooth function, and $z^0 \in \mathbb{R}^{n+k}$ is given; for continuation methods it is enough to take

$k = 1$. For each $j = 0, 1, 2, \ldots$, we assume the Jacobian matrix $\nabla H(z^j) \in \mathbb{R}^{(n+k)\times n}$ has full rank, and define z^{j+1} as the nearest point, in Euclidean distance, to z^j in $\{z : H(z^j) + \nabla H(z^j)(z - z^j) = 0\}$. This geometrically motivated algorithm is usually stated, equivalently, as $z^{j+1} = z^j - \nabla H(z^j)^\dagger H(z^j)$ where $\nabla H(z^j)^\dagger$ is the Moore-Penrose inverse of $\nabla H(z^j)$. It converges locally at a Q-quadratic rate to a zero z^* of H if $\nabla H(z^*)$ has full rank and ∇H is Lipschitz near z^*.

The paper [9] provides an extension of this algorithm, designed for solving generalized equations, that can be easily adapted to finding zeroes of the nonsmooth function H_C^a. This approach has also been studied in some detail in [19]. We give a version of the Newton method of [9] for the case of the corrector step. It uses the parameters $\epsilon_c > 0$ and $J_c \in \mathbb{N}$.

Corrector Step A, a function of (x', t').

Let $(y^0, s^0) = (x', t')$ and $j = 0$.

While $\|H_C^a(y^j, s^j)\| > \epsilon_c$ and $j < J_c$

Find a globally nearest point (y^{j+1}, s^{j+1}) to (x', t') in the set

$$\{(y, s) : \mathcal{L}H_C^a(y^j, s^j)(y, s) = 0\}. \tag{3.2}$$

Let $j = j + 1$.

(end While)

Let $(x^\dagger, t^\dagger) = (y^j, s^j)$.

Using Theorem 1 it follows that for almost every a, if (y^j, s^j) is near enough to the path P^a then the linearized set defined in (3.2) is a piecewise linear path at least in a neighborhood of (y^j, s^j). Thus determining a globally nearest point to this path (within a closed neighborhood of (y^j, s^j)) is made computationally possible by examining each of the finitely many line segments of the path.

Suppose f is a C^2 function. It is known [9, 19] that if $\mathcal{L}H_C^a$ satisfies certain regularity conditions, then the above Newton method converges Q-quadratically to a zero of H_C^a. Suppose further that the homotopy path P^a is bounded and, for some point $(\hat{x}, \hat{t})$ in its closure, the partial directional derivative of H_C^a with respect to x, $(H_C^a)'_x(\hat{x}, \hat{t}; \cdot)$, is invertible. Then we claim the following statement can be established using the Newton convergence result, Theorem 1 and compactness arguments: for almost all vectors a, there exist $h > 0$, $\epsilon_c > 0$, and $J_c \in \mathbb{N}$ such that for any $t \in [0, \hat{t})$ and x with $\|H_C(x, t)\| \leq \epsilon_c$, if the predictor step generates (x', t') from (x, t, h), then Corrector Step A terminates in one step and the distance of (y^1, s^1) to the path P^a is a small order of the distance from $(y^0, s^0) = (x', t')$ to P^a.

We comment that the choice of a viable ϵ_c seems to depend on the "how far the path is from degenerate points", hence on a, so it is not clear how to choose ϵ_c appropriately other than as a small positive number.

A heuristic corrector scheme

To simplify our implementation, that is to avoid computation of a globally nearest point to the generally nonconvex set (3.2), it is easy to extend the approach of Kojima and Shindo [18] for piecewise smooth systems in the same number of variables as equations to underdetermined piecewise smooth equations. At the jth iterate, the idea is to identify any cell σ_j containing the current point $(y^j, s^j) \in \mathbb{R}^{n+1}$, and apply one step of the Moore-Penrose Newton method to the smooth mapping representing H_C^a on σ_j.

> **Corrector Step B**, a function of (x', t').
>
> Let $(y^0, s^0) = (x', t')$ and $j = 0$.
>
> While $\|H_C^a(y^j, s^j)\| > \epsilon_c$ and $j < J_c$
>
> > Let $(y^{j+1}, s^{j+1}) = (y^j, s^j) - \nabla H_j(y^j, s^j)^\dagger H_C^a(y^j, s^j)$
> > where σ_j is a cell of $\mathcal{N}_C$ containing (y^j, s^j), and H_j is the C^1 representation of H_C^a on σ_j.
> > Let $j = j + 1$.
>
> (end While)
>
> Let $(x^\dagger, t^\dagger) = (y^j, s^j)$.

We stress that Corrector Step B is only a heuristic for piecewise smooth systems because the convergence analysis is not as strong as it is for the classical Newton's method or its extensions in [9, 19]. See [6] for a useful but limited convergence theory of this method and some generalizations.

3.3 Formal Homotopy Algorithm

As described above, if a predictor step produces a point that cannot be sufficiently corrected in J_c or fewer iterations of the corrector step, then a revised, more conservative prediction is made, and the correction step is attempted again. We define this formally using the constants $\epsilon_t \geq 0$, $\epsilon_f > 0$ and $\lambda \in (0, 1)$ at the outer level of the algorithm; $\eta \in \{\pm 1\}$ in the predictor step; and $\epsilon_c > 0$ and $J_c \in \mathbb{N}$ in the corrector step.

> **Homotopy algorithm** for solving $H_C^a(x, 1) = 0$.
>
> Let $(x^0, t^0) = (a, 0) \in P$, $k = 0$, $\hat{h} \in (0, 1]$.
>
> While $|t^k - 1| > \epsilon_t$ or $\left\|f_C(x^k)\right\| > \epsilon_f$
>
> > Let $h^k = \hat{h}$.
> >
> > Repeat:

$(x', t', h^k) \longleftarrow$ **Predictor step**(x^k, t^k, h^k).
$(x^\dagger, t^\dagger) \longleftarrow$ **Corrector step B**(x', t').
Let $h_k = \lambda h_k$.

Until $\left\|H_C(x^\dagger, t^\dagger)\right\| \leq \epsilon_c$.
Let $(x^{k+1}, t^{k+1}) = (x^\dagger, t^\dagger)$, $k = k + 1$.

(end While)

4 Implementation

The implementation of the algorithm is presented in two parts. The predictor step uses a suitable modification of the linear complementarity problem solver contained in the PATH code. The corrector uses a heuristic based on the Moore-Penrose idea. We assume hereafter that $C = [\ell, u]$.

4.1 Prediction

PATH [8] is an implementation of a nonsmooth Newton method for solving mixed complementarity problems. At each iteration, a linear complementarity problem is solved using a homotopy method, i.e. it follows the zero curve of a homotopy from $(x^k, 0)$ to $(x', 1)$. The technique employed to generate this path is the same as that developed in Section 3.1. The code has special rules to deal with the cases where the Jacobian of the affine map on σ does not have full row rank and where there is degeneracy in the path [12].

The predictor step of our PL homotopy algorithm uses the same code to generate Q_k^a for the homotopy $\mathcal{L}H_C^a(x^k, t^k)(x, t)$ starting in a cell σ. We explain the implementation of the algorithm of Section 3.1 by generating a representation of this linearized normal map, an indexing scheme that determines σ, and the Jacobian, J, of the affine map restricted to σ. We then discuss how a set of columns spanning J and the orientation is calculated. We continue with a presentation of how the cells of the normal manifold are traversed and the termination rules used by the code.

Representation of the linearized normal map

We will describe the linearized normal map $\mathcal{L}H_C^a(x^k, t^k)(\cdot, \cdot\cdot)$ by giving a formula for the affine mapping that defines it on each cell. We first decompose $x \in \mathbb{R}^n$ into a triple $(z, w, v) \in \mathbb{R}^{3n}$, then give a simple indexing scheme for representing each cell of $\mathcal{N}_C$, and finally combine these two to describe $\mathcal{L}H_C^a(x^k, t^k)(\cdot, \cdot\cdot)$ on each cell.

Represent each x as $z - w + v$ where

$$z = \pi_C(x),\ w = [\pi_C(x) - x]_+,\ v = [x - \pi_C(x)]_+ \tag{4.1}$$

and, for any vector c, c_+ is the vector whose ith component is $\max\{c_i, 0\}$. So $z \in C$ and w, v satisfy

$$w, v \geq 0 \tag{4.2}$$

and the complementarity conditions

$$\langle z - l, w\rangle = 0, \ \langle u - z, v\rangle = 0. \tag{4.3}$$

Conversely, for any $z \in C$ and nonnegative w and v satisfying (4.3), the vector given by $x = z - w + v$ is such that (4.1) holds.

Define a *representation* $\mathcal{A} = \{\mathcal{A}_i\}$ by $\mathcal{A}_i = i$, $n+i$ or $n+2i$, where $\mathcal{A}_i \neq n+i$ if $l_i = -\infty$ and $\mathcal{A}_i \neq n+2i$ if $u_i = \infty$. By associating $\mathcal{A}$ with the indices of the vector (z, w, v), we abuse notation by writing, say, $w_i \in \mathcal{A}$ to indicate that $n+i \in \mathcal{A}$. It is clear that there is a one-to-one correspondence between the cells

$$\begin{array}{rcll} \sigma_{\mathcal{A}} & = & \{x : & x_i \leq l_i \quad \text{if } w_i \in \mathcal{A} \\ & & & l_i \leq x_i \leq u_i \quad \text{if } z_i \in \mathcal{A} \\ & & & u_i \leq x_i \quad \text{if } v_i \in \mathcal{A}\} \end{array}$$

and the representations $\mathcal{A}$, and that $\mathcal{N}_C$ is precisely the family of cells indexed by representations $\mathcal{A}$.

Let $\sigma_{\mathcal{A}}$ be a cell and $\mathcal{A}$ be its representation. Then for $x \in \sigma_{\mathcal{A}}$ and (z, w, v) given by (4.1), the relationships (4.3) and (4.2) imply

$$\begin{array}{lllllll} z_i = l_i & \text{and} & w_i = l_i - x_i & \text{and} & v_i = 0 & \text{if } w_i \in \mathcal{A}, \\ z_i = x_i & \text{and} & w_i = 0 & \text{and} & v_i = 0 & \text{if } z_i \in \mathcal{A}, \\ z_i = u_i & \text{and} & w_i = 0 & \text{and} & v_i = x_i - u_i & \text{if } v_i \in \mathcal{A}. \end{array}$$

Thus as x varies within $\sigma_{\mathcal{A}}$, only the components of (z, w, v) in the representation $\mathcal{A}$ can change and, furthermore, the transformation $x \mapsto (z, w, v)$ is affine.

Given x^k let $z^k = \pi_C(x^k)$, $w^k = [\pi_C(x^k) - x^k]_+$ and $v^k = [x^k - \pi_C(x^k)]_+$. Then rewrite $\mathcal{L}H_C^a(x^k, t^k)(x, t) = 0$ as the equivalent system consisting of (4.2), (4.3) together with

$$N(z, w, v, t) + q \ = \ 0, \tag{4.4}$$

where $I \in \mathbb{R}^{n\times n}$ is the identity,

$$\begin{array}{rcll} N & = & [A \ -I \ I \ r] & \in \mathbb{R}^{(3n+1)\times n} \\ A & = & t^k\nabla f(z^k) + (1-t^k)I & \in \mathbb{R}^{n\times n} \\ r & = & f(z^k) - (z^k - a) & \in \mathbb{R}^n \\ q & = & t^k f(z^k) + (1-t^k)(z^k - a) - Mz^k - t^k r & \\ & = & t^k z^k - a - t^k\nabla f(z^k) z^k & \in \mathbb{R}^n. \end{array}$$

So a solution (z, w, v, t) of (4.2)–(4.4) yields $\mathcal{L}H_C^a(x^k, t^k)(z - w + v, t) = 0$.

Since only the components in $\mathcal{A}$ change when $x \in \sigma_{\mathcal{A}}$, it is clear that for all $x \in \sigma_{\mathcal{A}}$ and (z, w, v) representing x

$$\mathcal{L}H_C^a(x^k, t^k)(x, t) \;=\; N_{\mathcal{A}}\,(z, w, v)_{\mathcal{A}} \,+\, rt \,+\, q^{\sigma}.$$

Here $q^{\sigma} = q + N_{\bar{\mathcal{A}}}\,(z, w, v)_{\bar{\mathcal{A}}}$ where $\bar{\mathcal{A}} = \{1, \ldots, 3n\} \setminus \mathcal{A}$ so that the subvector $(z, w, v)_{\bar{\mathcal{A}}}$ is constant in σ.

The PATH code works with the triplet (z, w, v) and $N_{\mathcal{A}}$. However, in x-space the Jacobian, J, of $\mathcal{L}H_C^a(x^k, t^k)(x, t)$ restricted to $\sigma_{\mathcal{A}}$ is given by $\begin{bmatrix} J_x & J_t \end{bmatrix} = \begin{bmatrix} N_{\mathcal{A}}\Delta & r \end{bmatrix}$ where $\Delta \in \Re^{n \times n}$ is the diagonal matrix

$$\Delta_{i,i} = \begin{cases} -1 & \text{if } w_i \in \mathcal{A} \\ 1 & \text{otherwise.} \end{cases}$$

This fact will be used in the calculation of the orientation.

Bases

At the start of the predictor step at iteration k we are given $x^k \in \mathbb{R}^k$, and a representation $\mathcal{A}$ whose corresponding cell σ contains x^k. If x^k is in the intersection of two or more cells we use $\mathcal{A}$ to describe the cell we are interested in. We have seen that the linear part of $\mathcal{L}H_C^a(x^k, t^k)$ on σ can be described using the matrix $\begin{bmatrix} N_{\mathcal{A}} & r \end{bmatrix} \in \mathbb{R}^{n \times (n+1)}$. To determine our initial search direction, we need to construct a basis for the range space and kernel of this map. More formally, given a cell σ and its representation $\mathcal{A}$, define a *basis* $\mathcal{B}$ as a list of n indices taken from $\mathcal{A} \cup \{3n+1\}$ such that $N_{\mathcal{B}}$ is invertible.

The linear complementarity code is initially given a candidate basis, $\tilde{\mathcal{B}}$. It is possible that $N_{\tilde{\mathcal{B}}}$ is not invertible. In this case we attempt to uncover a basis. We apply a LU factorization to the $n \times (n + 1)$ matrix $N_{\mathcal{A} \cup \{3n+1\}}$ to determine the linearly dependent column. The remaining columns form a basis, $\mathcal{B}$.

If more that one linearly dependent column is identified, the theory breaks down and this procedure fails. If x^k is in the intersection of two or more cells, we try to choose an adjacent cell with full row rank. Specifically, the adjacent cell attempted is $\bar{\sigma}$ with representation:

$$\bar{\mathcal{A}}_i = \begin{cases} w_i & \text{if } z_i^k = l_i \\ v_i & \text{if } z_i^k = u_i \\ z_i & \text{otherwise.} \end{cases}$$

We choose a candidate basis as $\bar{\mathcal{B}} = \bar{\mathcal{A}}$. We then apply the above procedure with $\bar{\mathcal{B}}$ replacing $\tilde{\mathcal{B}}$. In the event this fails, an error is reported and the code halts.

The reason we use a candidate basis as our initial guess as to the invertible basis as opposed to always factoring the larger $n \times (n + 1)$ matrix is that the code written for

PATH only uses the $n \times n$ basis matrix for computations. We did not want to rewrite the code to use the $n \times (n+1)$ matrix. Typically the candidate basis is invertible and we save a factorization.

Orientation

We now have a initial basis $\mathcal{B}^k$ and a linearly dependent column $\beta^k = (\mathcal{A} \cup \{3n+1\}) \setminus \mathcal{B}^k$. We need to determine the initial direction of the path Q_k^a, that is whether the variable $(z, w, v, t)_{\beta^k}$ is going to initially increase or decrease. To do this, we need to find the orientation.

For the orientation, we determine the sign of the determinant of the following matrix:

$$\begin{bmatrix} J_x & J_t \\ d_x^T & d_t \end{bmatrix} = \begin{bmatrix} N_{\mathcal{A}}\Delta & r \\ d_x^T \Delta & d_t \end{bmatrix} = \begin{bmatrix} B & b \\ d_B^T & 1 \end{bmatrix} P$$

where $\begin{bmatrix} d_x^T & d_t \end{bmatrix}$ spans the kernel of $\begin{bmatrix} J_x & J_t \end{bmatrix}$, $B = \begin{bmatrix} N_{\mathcal{A}}\Delta & r \end{bmatrix}_{\mathcal{B}^k}$, $b = \begin{bmatrix} N_{\mathcal{A}}\Delta & r \end{bmatrix}_{\beta^k}$, and P is a permutation matrix. Let LU be the decomposition of B into an invertible lower triangular matrix, L, and an invertible upper triangular matrix, U. Letting $d_B^T = -U^{-1}L^{-1}b = -B^{-1}b$, we can see that $\begin{bmatrix} d_B^T & 1 \end{bmatrix}$ spans the kernel of $\begin{bmatrix} B & b \end{bmatrix}$. Combining these facts, we have

$$\begin{aligned}
\begin{bmatrix} B & b \\ d_B^T & 1 \end{bmatrix} P &= \begin{bmatrix} LU & b \\ d_B^T & 1 \end{bmatrix} P \\
&= \begin{bmatrix} L & 0 \\ d_B^T U^{-1} & 1 - d_B^T U^{-1} L^{-1} b \end{bmatrix} \begin{bmatrix} U & L^{-1}b \\ 0 & 1 \end{bmatrix} P \\
&= \begin{bmatrix} L & 0 \\ d_B^T U^{-1} & 1 + \|d_B\|^2 \end{bmatrix} \begin{bmatrix} U & L^{-1}b \\ 0 & 1 \end{bmatrix} P.
\end{aligned}$$

Therefore, the orientation is just the sign of the determinant of L times the sign of the determinant of U multiplied by the sign of determinant of P. We will initially increase the entering variable if the orientation is η and decrease it otherwise. We then traverse the cells using a complementary pivoting strategy.

Complementary Pivoting

Given the basis, $\mathcal{B}^k$, and the entering variable, β^k, we calculate a direction, $d = -N_{\mathcal{B}^k}^{-1} N_{\beta^k}$. We will either increase or decrease variable $(z, w, v, t)_{\beta^k}$ along this direction depending upon the orientation calculation above. A ratio test (on (z, w, v, t)) determines when we run into the boundary of a cell and hence gives a leaving variable and steplength. We then perform a pivot by updating $\mathcal{B}^k$ to include the entering variable and exclude the leaving variable. The new entering variable is chosen as follows:

- If z_i leaves at its lower bound, w_i is the new entering variable.
- If z_i leaves at its upper bound, v_i is the new entering variable.
- If w_i leaves at its lower bound, z_i is the new entering variable at its lower bound.
- If v_i leaves at its lower bound, z_i is the new entering variable at its upper bound.
- If t leaves, we are done.

This defines the new cell along with a corresponding basis, entering variable, and direction. The code uses a rank-1 update of the LU factorization of the old basis to find the needed decomposition of the new basis.

Termination

The four conditions for termination presented in Section 3.1 are used in the code. If t leaves the basis we terminate according to the complementary pivoting rules. This corresponds to $t = 0$ or $t = 1$. We added a termination based on the distance from the starting point. In the code, we relax the distance based termination rule to be

$$\left\|(\pi_C(x), t) - (\pi_C(x^k), t^k)\right\| \ = \ h_k$$

because the approximation error, i.e. the difference between $\mathcal{L}H_C^a(x^k, t^k)(x, t)$ and $H_C^a(x, t)$ depends more on $\pi_C(x) - \pi_C(x^k)$ than on $[x - \pi_C(x)] - [x^k - \pi_C(x^k)]$.

Special Rules

The corrector can lead us to a place where $t^k > 1$. In this case we reverse the direction we are traveling. We put a lower bound on the t variable of 1 and an upper bound on t of ∞. We then use the above direction calculation, but take the opposite direction in which to initially travel. Everything then continues as normal, with the termination being that t leaves at its lower bound or we take a step of the maximum allowed distance.

4.2 Correction

The corrector code is given an initial (x', t') from the predictor. The implementation of the corrector uses spacer steps which move the current iterate closer to P^a. They are defined as follows:

$$\hat{y}_i^0 = \begin{cases} (\pi_C(x') - [H^a(x', t')]_+)_i & \text{if variable } \pi_C(x') = l_i \\ (\pi_C(x') + [-H^a(x', t')]_+)_i & \text{if variable } \pi_C(x') = u_i \\ x_i' & \text{otherwise.} \end{cases}$$

Let $s^0 = t'$. We then calculate $\|H_C^a(y^0, s^0)\|$ and check to see if it is zero as detailed in Section 3.2. If it is not, we use the Moore-Penrose idea to move back onto the zero curve of the homotopy.

The Moore-Penrose iterate solves a minimization problem:

$$\begin{array}{ll} \min & \|(x,t) - (y^j, s^j)\|^2 \\ \text{subject to} & \begin{bmatrix} J_x & J_t \end{bmatrix} \begin{bmatrix} x \\ t \end{bmatrix} = -H_C^a(\hat{x}, \hat{t}) \end{array}$$

where $\begin{bmatrix} J_x & J_t \end{bmatrix}$ is the Jacobian of the affine map $\mathcal{L}H_C^a(y^j, s^j)(\cdot,\cdot)$ on σ, a cell containing y^j. We use the representation $\mathcal{A}'$ to determine the Jacobian as developed in the previous section.

To solve this problem, we solve the following system of equations:

$$\begin{bmatrix} I & 0 & -J_x^T \\ 0 & 1 & -J_t^T \\ J_x & J_t & 0 \end{bmatrix} \begin{bmatrix} d_x \\ d_t \\ d_u \end{bmatrix} = \begin{bmatrix} 0 \\ -H_C^a(y^j, s^j) \end{bmatrix}.$$

The direction (d_x, d_t) is nonzero since $H_C^a(y^j, s^j) \neq 0$. We move in the direction (d_x, d_t) until either the full step is taken or we encounter the boundary of σ to obtain y^{j+1} and s^{j+1} and an adjacent cell σ'. We the check for termination, and if necessary perform another iteration.

The code does not allow the residual to increase from one corrector step to the next and returns an error code if at the end the homotopy parameter is negative. If the corrector fails, we take a half predictor step and attempt the correct again. A half predictor step is quickly found using the reconstruction technique documented in [12].

4.3 Updates

The parameters in the code are updated after each major iteration (predictor, corrector sequence). We modify the maximum distance allowed to travel to become larger if we are doing well and make it smaller if we are doing poorly.

5 Results

The algorithm was implemented using the current version of the linear complementarity problem solver found in PATH. The framework reported in [12] was used so that we could easily access problems generated by the GAMS modeling language. The preprocessor for complementarity problems developed [13] is available to the code. However, for testing purposes, it was *not* used.

We ran the code on the MCPLIB [7] test suite of problems. The MPSGE [26] models in the collection were omitted due to the fact that function and Jacobian evaluations for these models have the undesirable side-effect of changing the evaluation point, thereby making corrector steps impossible to perform.

The model name, dimension of the problem, number of function evaluations (Func), and solution time (Time), for the first starting point of the remaining models in this test set is reported in Table 5.1. The test was carried out on a Sun Ultrasparc machine that has 768 megabytes of available memory.

We note that most of the academic test problems which cause problems for other algorithms (e.g. billups, dirkse1, ralph, simple-ex) are easily solved using the predictor-corrector code. Over the entire set of 77 models with 436 total starting points, we achieved a 67.7% success rate. Of the 141 failures, 83 of these occurred for starting points associated with just two problems, namely `games` and `tinsmall`.

The current implementation demonstrates that homotopy methods can be generated for large scale complementarity problems. The choice of homotopy function (in our case $x - a$) is critical for the success of the these methods. We were disappointed with the codes robustness; in particular the implementation is very sensitive to how the original problem is formulated (e.g. where $f_1(x) = 0$ or $-f_1(x) = 0$.

Three approaches may be useful in improving these results. Firstly, a different choice of homotopy could be used instead of $x - a$ to attempt to match it mored closely to the underlying problem. Secondly, the preprocessor for MPC described in [13] could be used to exploit more fully underlying problem structure. Finally, this preprocessor could be adapted to automatically identify poorly posed problems and reformulate appropriately. This is the subject of future research.

References

[1] J. C. Alexander. The topological theory of an embedding method. In H. Wacker, editor, *Continuation Methods*, pages 37–68. Academic Press, New York, 1978.

[2] J. C. Alexander, R. B. Kellog, T.-Y. Li, and J .A. Yorke. Piecewise smooth continuation. Manuscript, 1979.

[3] J. C. Alexander, T.-Y. Li, and J .A. Yorke. Piecewise smooth homotopies. In B. C. Eaves, F. J. Gould, H.-O. Peitgen, and M. J. Todd, editors, *Homotopy Methods and Global Convergence*, pages 1–14. Plenum Press, New York, 1983.

[4] E. L. Allgower and K. Georg. *Numerical Continuation Methods, An Introduction.* Springer-Verlag, Berlin, 1990.

[5] S. C. Billups, A. L. Speight, and L. T. Watson. Nonmonotone path following methods for nonsmooth equations and complementarity problems. Technical report, Department of Mathematics, University of Colorado, Denver, Colorado, 1999.

Problem	Size	Func	Time	Problem	Size	Func	Time
badfree	5	44	0.03	mr5mcf	350	-	fail
bert_oc	5000	-	fail	munson3	2	-	fail
bertsekas	15	441	0.66	munson4	1	-	fail
billups	1	32	0.02	nash	10	203	0.71
bratu	5625	181	13208.96	ne-hard	3	111	0.08
cammcf	242	-	fail	obstacle	2500	-	fail
choi	13	21	2.66	opt_cont	288	173	12.10
colvdual	20	269	0.45	opt_cont127	4096	641	1726.96
colvnlp	15	275	0.33	opt_cont255	8192	1888	12470.15
cycle	1	23	0.01	opt_cont31	1024	323	114.73
degen	2	27	0.02	opt_cont511	16384	-	fail
dirkse1	2	135	0.08	pgvon105	105	619	12.37
duopoly	63	-	fail	pgvon106	106	-	fail
eckstein	1	21	0.01	pies	42	1595	4.91
ehl_k40	41	195	6.68	powell	16	61	0.20
ehl_k60	61	233	18.26	powell_mcp	8	75	0.11
ehl_k80	81	265	38.50	qp	4	39	0.03
ehl_kost	101	291	74.67	ralph	2	33	0.02
electric	158	-	fail	runge	1	13	0.01
explcp	16	98	0.19	scarfanum	13	88	0.23
fixedpt	406	-	fail	scarfasum	14	46	0.13
forcebsm	184	-	fail	scarfbnum	39	653	2.46
forcedsa	186	-	fail	scarfbsum	40	621	3.28
freebert	15	439	0.65	shubik	30	1534	2.78
gafni	5	48	0.07	simple-ex	17	173	0.67
games	16	-	fail	simple-red	13	121	0.65
hanskoop	14	58	0.10	sppe	27	447	0.80
hansmcf	43	139	0.74	tinloi	146	54	1.16
hydroc06	29	-	fail	tinsmall	42	647	3.58
hydroc20	99	-	fail	tobin	42	305	0.79
jel	6	317	0.37	trafelas	2376	-	fail
josephy	4	43	0.03	trig	10	-	fail
kojshin	4	61	0.04	vonthmcf	125	-	fail
kyh-scale	28	-	fail	xu1	8	73	0.07
kyh	28	-	fail	xu2	8	73	0.07
lincont	419	-	fail	xu3	11	63	0.08
mathinum	3	47	0.04	xu4	11	39	0.04
mathisum	4	35	0.03	xu5	20	70	0.12
methan08	31	-	fail				

Table 5.1: Results on MCPLIB Problems

[6] X. Chen, Z. Nashed, and L. Qi. Convergence of Newton's method for singular smooth and nonsmooth equations using adaptive outer inverses. *SIAM Journal on Optimization*, 7:445–462, 1997.

[7] S. P. Dirkse and M. C. Ferris. MCPLIB: A collection of nonlinear mixed complementarity problems. *Optimization Methods and Software*, 5:319–345, 1995.

[8] S. P. Dirkse and M. C. Ferris. The PATH solver: A non-monotone stabilization scheme for mixed complementarity problems. *Optimization Methods and Software*, 5:123–156, 1995.

[9] A. L. Dontchev. Local convergence of the Newton method for generalized equations. *C.R. Acad. Sci. Paris Series I*, 332:327–331, 1996.

[10] B. C. Eaves. A short course in solving equations with PL homotopies. In R. W. Cottle and C. E. Lemke, editors, *Nonlinear Programming*, pages 73–143, Providence, RI, 1976. American Mathematical Society, SIAM–AMS Proceedings.

[11] F. Facchinei and J. S. Pang. Total stability of variational inequalities. Manuscript, Universitá di Roma "La Sapienza", Dipartimento di Informatica e Sistemistica, Roma, Italy, 1988.

[12] M. C. Ferris and T. S. Munson. Interfaces to PATH 3.0: Design, implementation and usage. *Computational Optimization and Applications*, 12:207–227, 1999.

[13] M. C. Ferris and T. S. Munson. Preprocessing complementarity problems. Mathematical Programming Technical Report 99-07, Computer Sciences Department, University of Wisconsin, Madison, Wisconsin, 1999.

[14] M. C. Ferris and J. S. Pang. Engineering and economic applications of complementarity problems. *SIAM Review*, 39:669–713, 1997.

[15] H. Hironaka. Introduction to real-analytic sets and real-analytic maps. Technical report, Instituto Matematico "L. Tonelli", Università di Pisa, Pisa, Italy, 1973.

[16] M. Kojima. Recent advances in mathematical programming. x. computation of fixed points by continuation methods. *Systems and Control*, 25:421–430, 1981.

[17] M. Kojima and R. Hirabayashi. Continuous deformation of nonlinear programs. *Mathematical Programming Study*, 21:150–198, 1984.

[18] M. Kojima and S. Shindo. Extensions of Newton and quasi-Newton methods to systems of PC^1 equations. *Journal of Operations Research Society of Japan*, 29:352–374, 1986.

[19] B. Kummer. Metric regularity: characterizations, nonsmooth variations, and successive approximation. *Journal of Mathematical Analysis and its Applications*.

[20] M. S. Lojasiewicz. *Ensembles semi-analytiques.* Institut des Hautes Etudes Scientifiques, Bures-sur-Yvette, 1964.

[21] O. L. Mangasarian. *Nonlinear Programming.* McGraw–Hill, New York, 1969. SIAM Classics in Applied Mathematics 10, SIAM, Philadelphia, 1994.

[22] D. Ralph. On branching numbers of normal manifolds. *Nonlinear Analysis, Theory, Methods and Applications*, 22:1041–1050, 1994.

[23] S. M. Robinson. An implicit–function theorem for a class of nonsmooth functions. *Mathematics of Operations Research*, 16:292–309, 1991.

[24] S. M. Robinson. Normal maps induced by linear transformations. *Mathematics of Operations Research*, 17:691–714, 1992.

[25] S. M. Robinson. Newton's method for a class of nonsmooth functions. *Set Valued Analysis*, 2:291–305, 1994.

[26] T. F. Rutherford. Applied general equilibrium modeling with MPSGE as a GAMS subsystem: An overview of the modeling framework and syntax. *Computational Economics, forthcoming*, 1999.

[27] H. Sellami. *A Continuation Method for Normal Maps.* PhD thesis, University of Wisconsin – Madison, Madison, Wisconsin, 1994.

[28] H. Sellami. A homotopy continutation method for normal maps. *Mathematical Programming*, pages 317–337, 1998.

[29] H. Sellami and S. M. Robinson. Homotopies based on nonsmooth equations for solving nonlinear variational inequalities. In G. Di Pillo and F. Giannessi, editors, *Nonlinear Optimization and Applications*, pages 327–343. Plenum Press, New York, 1996.

[30] H. Sellami and S. M. Robinson. Implementation of a continutation method for normal maps. *Mathematical Programming*, pages 563–578, 1997.

Appendix

We use a property of subanalytic sets [20, 15] in the proof the next result: If S is a subanalytic set in $\mathbb{R}^n$ then for any point $\hat{x}$ in its closure there is a continuous function $g : [0,1] \to \mathbb{R}^n$ with $g(1) = \hat{x}$ and $g(s) \in S$ for $s \in [0,1)$.

Lemma 2 *If $p : [0,1) \to \mathbb{R}^n$ is a subanalytic function then either there exists the limit $\lim_{t\to 1^-} p(t)$, or $\|p(t)\| \to \infty$ as $t \to 1^-$.*

Proof Let $p : [0,1) \to \mathbb{R}^n$ be subanalytic, and $\hat{p}$ be a limit point of $p(t)$ as $t \to 1^-$. It is sufficient to show that the limit $\lim_{t\to 1^-} p(t)$ exists and equals $\hat{p}$.

The graph of p, $P = \{(t, p(t)) : t \in [0,1)\}$, is subanalytic by definition and $(1, \hat{p})$ lies in its closure. Let $g : [0,1] \to \mathbb{R}^n$ be a continuous function with $g(1) = (1, \hat{p})$ and $g(s) \in P$ for $s \in [0,1)$. Denote the first component function of $g(s)$ by $\gamma(s) \in [0,1]$. Hence for $s \in [0,1)$, $\gamma(s) \in [0,1)$ and $g(s) = (\gamma(s), p \circ \gamma(s))$.

Let $\epsilon > 0$. We will complete the proof by providing $\delta \in (0,1]$ such that

$$\|p(t) - \hat{p}\| \ < \ \epsilon \quad \text{for all } t \in [1-\delta, 1). \tag{5.1}$$

Continuity of g at $s = 1$ yields $\delta_g \in (0,1]$ such

$$\|p \circ \gamma(s) - \hat{p}\| \ < \ \epsilon \quad \text{for all } s \in [1-\delta_s, 1).$$

Now $\gamma(1-\delta_g) < 1 = \gamma(1)$ so continuity of γ ensures (by the intermediate value theorem) that $\gamma([1-\delta_g, 1)) \supset [\gamma(1-\delta_g), 1)$. Let $\delta = 1 - \gamma(1-\delta_g) \in (0,1]$ and deduce (5.1) from the previous bound. □

An easy corollary extends a result of [11, Theorem 5.10], in the context of a homotopy method applied to variational inequalities, in which parts 2 and 3 of were shown to be equivalent.

Corollary 3 *For a subanalytic function $p : [0,1) \to \mathbb{R}^n$ the following statements are equivalent:*

1. $\liminf_{t\to 1^-} \|p(t)\| < \infty$.
2. $\limsup_{t\to 1^-} \|p(t)\| < \infty$.
3. $\lim_{t\to 1^-} p(t)$ *exists.*

Acknowledgements

We are grateful to Jeremy Glick for preliminary work on smooth homotopy methods and to Stephen M. Robinson for discussions on homotopy methods applied to variational inequalities.

The work of Ferris and Munson was based on research supported by the National Science Foundation grant CCR-9972372 and the Air Force Office of Scientific Research grant F49620-98-1-0417 and the work of Ralph was based on research partially supported by the Australian Research Council.

Michael C. Ferris, Todd S. Munson
Computer Sciences Department,
University of Wisconsin,
Madison, WI 53706, USA.

Daniel Ralph
Department of Mathematics,
University of Melbourne,
Melbourne, Australia.