Solving piecewise linear equations in abs-normal form

Andreas Griewank, Jens-Uwe Bernt, Manuel Radons and Tom Streubel

Department of Mathematics, Humboldt-Universität zu Berlin
e-mail: surname@math.hu-berlin.de

Keywords:
Switching depth, Sign real spectral radius, Coherent orientation, Generalized Jacobian, Semismooth Newton, Unfolded system, Linear complementarity
2010 MSC: 90C33, 90C56, 49J52, 65F99

Abstract: *With the ultimate goal of iteratively solving piecewise smooth (PS) systems, we consider the solution of piecewise linear (PL) equations. As shown in [Gri13] PL models can be derived in the fashion of automatic or algorithmic differentiation as local approximations of PS functions with a second order error in the distance to a given reference point. The resulting PL functions are obtained quite naturally in what we call the abs-normal form, a variant of the state representation proposed by Bokhoven in his dissertation [vB81]. Apart from the tradition of PL modeling by electrical engineers, which dates back to the Master thesis of Thomas Stern [Ste56] in 1956, we take into account more recent results on linear complementarity problems and semi-smooth equations originating in the optimization community [CPS92, Sch12, FP03]. We analyze simultaneously the original PL problem (OPL) in abs-normal form and a corresponding unfolded system (UPL), which is closely related to a linear complementarity problem (LCP). We show that the UPL, like KKT conditions and other singly switched systems, cannot be open without being injective. Hence we find that some of the intriguing PL structure described by Scholtes [Sch12] is lost in the unfolding from OPL to UPL. To both problems one may apply Newton variants with appropriate generalized Jacobians directly computable from the abs-normal representation. Alternatively, the ULP can be solved by Bokhoven's modulus method and another fixed point solver that is asymptotically faster but requires a slightly stronger convergence condition. We compile the properties of the various schemes and highlight the connection to the properties of the Schur complement matrix, in particular its signed real spectral radius as analyzed by Rump in [Rum97]. Numerical experiments and suitable combinations of the fixed point solvers and stabilized generalized Newton variants remain to be done.*

1. Introduction and Motivation

In many applications one encounters piecewise smooth (PS) functions that can be locally approximated with second order error by piecewise linear (PL) functions. In this paper we will assume throughout that all functions are continuous and thus, in fact, Lipschitz continuous. However, an extension to piecewise linear but possibly discontinuous problems should be in the back of our minds before we settle on data structures and interfaces. Discontinuous solution operators may arise for example if one considers least squares problems defined by piecewise linear systems of equations.

The process of piecewise linearization of a piecewise smooth function $F : \mathcal{D} \subset \mathbb{R}^n \mapsto \mathbb{R}^m$ given by an evaluation procedure was described in [Gri13]. The key assumption is that all nonsmoothness can be cast in terms of the absolute value function $|\cdot|$. Then piecewise linearization can be achieved in the style of algorithmic differentiation [GW08] by simply replacing all smooth elemental functions by their tangent line or plane (in case of binary operations or special functions) and the absolute value function by itself.

In contrast to conventional notions of differentiation one does not obtain a collection of derivative vectors or matrices at a given reference point $\mathring{x}$. Rather one arrives at a procedure for evaluating the unique incremental PL function $\Delta F(\mathring{x}, \Delta x) : \mathcal{D} \times \mathbb{R}^n \mapsto \mathbb{R}^m$ for which

$$F(\mathring{x} + \Delta x) \; = \; F(\mathring{x}) + \Delta F(\mathring{x}, \Delta x) + O(\|\Delta x\|^2)$$

Here the error term $\|\Delta x\|^2$ is uniform on compact subsets of $\mathcal{D} - \mathring{x}$. This means that $\Delta F(\mathring{x}, \Delta x)$ is a candidate for a nonsingular uniform Newton approximation in the sense of [FP03], although the local homeomorphism property is by no means guaranteed.

Throughout this paper we will only be concerned with the properties of the piecewise linearized function. We will also drop the decomposition into $F(\mathring{x})$ and the increment $\Delta F(\mathring{x}, \Delta x)$ and thus simply consider a globally defined piecewise linear continuous (PL) mapping

$$F(x) \; : \; \mathbb{R}^n \mapsto \mathbb{R}^m$$

Like for the (possibly) underlying nonsmooth mapping, our ultimate purpose is to solve certain basic numerical tasks, in particular (un)constrained optimization, equation solving, and the numerical integration of dynamical systems. Here we will consider, for $m = n$, the problem of solving the formally well determined system of equations

$$F(x) \; = \; 0 \in \mathbb{R}^n, \quad \text{for} \quad x \in \mathbb{R}^n \tag{1}$$

The paper is organized as follows. In Section 2 we introduce PL functions F in *abs-normal* form, a term that was apparently introduced by Barton and Khan in a more general nonlinear setting [KB12]. In Section 3 we describe the resulting polyhedral structure and give an explicit procedure for calculating generalized Jacobians of F, which were shown in [KB12, KB13] and [Gri13] to be conically active limiting Jacobians of the underlying piecewise smooth function, whenever F was obtained as its piecewise linearization. In Section 4 we examine the relation between the global properties of bijectivity and coherent orientation, which coincide under certain rather generic conditions. Section 5 discusses sufficient conditions for the global convergence of the generalized Newton method, which is often referred to as semi-smooth Newton. In Section 6 we unfold the system by elevating the intermediate switching variables to the status of full variables. As is the case for the unfolding of smooth singular equations [MD85], in this process some regularity is gained, but some information is also lost. The unfolded system UPL is always simply switched and as shown in Section 7, it can be solved by two different fixed point methods and several variants of generalized Newton. Finally, the unfolded system can also be rewritten as a linear complementarity problem [CPS92] with coherent orientation being equivalent to the P-matrix property. The final Section 8 summarizes our results and provides an outlook to further developments.

2. The abs-normal form

As also observed by Scholtes in [Sch12] any piecewise linear scalar function $f : \mathbb{R}^n \mapsto \mathbb{R}$ has a so-called **max-min** representation

$$f(x) \;=\; \max_{1\leq i\leq l}\, \min_{j\in M_i} a_j^\top x + b_j$$

where the l index sets M_i are contained in $\{1, 2 \ldots k\}$ for some $k \in N$ and the $a_i \in \mathbb{R}^n$, $b_i \in \mathbb{R}$ are constant coefficients. For a PL vector function $F : \mathbb{R}^n \mapsto \mathbb{R}^m$ each one of the m component functions can be represented in the same way. Moroever, using the equivalences

$$\max(u, w) = \tfrac{1}{2}(u + w + |u - w|) \quad \text{and} \quad \min(u, w) = \tfrac{1}{2}(u + w - |u - w|)$$

one can express all **min** and **max** expressions in terms of $s \geq 0$ absolute value functions $|z_i|$, whose arguments z_i are called *switching variables.*

Observing that each z_i is an affine function of absolute values $|z_j|$ with $j < i$ and the independents x_k for $k \leq n$, one arrives at an **abs-normal** representation

$$\begin{bmatrix} z \\ y \end{bmatrix} \;=\; \begin{bmatrix} c \\ b \end{bmatrix} + \begin{bmatrix} Z & L \\ J & Y \end{bmatrix} \begin{bmatrix} x \\ |z| \end{bmatrix} \tag{2}$$

Here the two vectors and four matrices specifying the function F have the formats

$$c \in \mathbb{R}^s,\ Z \in \mathbb{R}^{s\times n},\ L \in \mathbb{R}^{s\times s},\ b \in \mathbb{R}^m,\ J \in \mathbb{R}^{m\times n},\ Y \in \mathbb{R}^{m\times s}.$$

The matrix L is strictly lower triangular so that for given x the components of $z = z(x)$ and thus $|z|$ can be unambiguously computed one by one. Specifically, we have $L_{i,j} \neq 0$ exactly if z_i depends directly on $|z_j|$ so that there is an edge between the nodes j and i in the corresponding data dependency graph. This graph is always acyclic and the components of x, y and z represent its roots, leaves and internal vertices, respectively.

Of course, the representation (2) is by no means unique for a given mapping F. One would naturally strive to make the representation as concise as possible in some sense. Excluding incidental cancellations, we find that the smallest integer $\nu \leq s$ for which

$$L^\nu \ = \ 0$$

corresponds to the maximal number of internal nodes in any chain in the data dependency graph. We will call this the **switching depth** and consider it as key measure of the combinatorial difficulty of the function F. In this terminology, F is fully linear exactly if $\nu = 0$ with $s = 0$ and z, Z, and L are empty. We will refer to this limiting situation as the **smooth case.** We will call F **simply switched** if $\nu = 1$, a situation that arises for example in complementarity problems, where none of the nonsmooth elements are superimposed. We conjecture that, for any PL mapping $F : \mathbb{R}^n \mapsto \mathbb{R}^m$, there is an abs-normal representation with a switching depth $\nu \leq \bar{\nu}(n) = 2\,n - 1$.

Formulations similar to our abs-normal form have been used for a long time in the engineering literature. In [VBL99] several classes of PL models are compared, **Chua1** has switching depth 1 and **Grü** as well as **Bokh2** are limited to switching depth 2. It is shown there that all of them are specializations of the model **Bokh1**, which is a priori implicit in that evaluating y for given x requires the solution of an LCP with a system matrix D. However, if D is lower triangular solving the LCP requires simply a forward substitution. Moreover, one can then easily rewrite the **Bokh1** system in abs-normal form with $L = (I - D)^{-1}(I + D)$ the Möbis transform of D. Obviously L is triangular if and only if this is true for D. As we have noticed, the abs-normal form is general enough to represent all continuous PL function, so we will not use the even greater generality of the implicit **Bokh1** model. In contrast to the specification by linear pieces on polyhedra defined by systems of linear inequalities, often used in the more mathematical literature, the abs-normal form is stable and free of redundency. In particular, any perturbation of the four matrices Z, L, J and Y that preserves the strict lower triangularity of L again defines unambiguously a continuous PL function $y = F(x)$.

In the simply switched case we have $z = c + Zx$, which means that potential kinks occur at the union of the s hyperplanes $z_i(x) = c_i + e_i^\top Zx = 0$ for $i = 1 \ldots n$. We will then say that the kinks satisfy the **linear independence kink qualification LIKQ** if the normals of the hyperplanes intersecting at some point x are always linearly independent. This implies in particular that the vector $z = c + Zx$ can never have more than n vanishing components. LIKQ is implied by all square submatrices of $[c, Z] \in \mathbb{R}^{n\times(1+s)}$ of order $\min(s, n+1)$ being nonsingular. That slightly stronger condition is for example satisfied if $c = \mathbf{1}$ is the vector of ones and $Z = (\lambda_i^j)_{j=1\ldots n}^{i=1\ldots s}$ is a Vandermonde matrix at distinct abscissas λ_i for $i = 1 \ldots s$. Consequently, the polynomial $P(c, Z)$ formed by the product of the determinants of all maximal square submatrices $[c, Z]$ does not vanish at the Vandermonde choice and the same is true for almost all matrices $[c, Z] \in \mathbb{R}^{s\times(1+s)}$. In other words, LIKQ is a generic property, like linear independence of active constraints in linear optimization (LOP). If c has no zero components LIKQ is equivalent to the rows of Z being *affinely independent*.

The Rosette example

To highlight the possible properties of PL functions we take a look at the following class of examples. Positively homogeneous functions in two variables are uniquely defined by their values on the unit circle, which must be 2π periodic functions of the polar angle $\varphi(x) = \arctan(x_1, x_2)$. More specifically, we assume that we have a monotonically growing sequence of angles

$$0 = \varphi_0 < \varphi_1 < \ldots < \varphi_{n-1} < \varphi_n = 2\pi$$

and corresponding values

$$(\psi_i)_{i=0\ldots n} \text{ with } \psi_n - \psi_0 = 2p\pi \text{ for } p \in \mathbb{N}$$

By suitable subdivisions we can ensure that the increments $\varphi_i - \varphi_{i-1}$ and $|\psi_i - \psi_{i-1}|$ are all less than π. Then there exists exactly one homogenous piecewise linear function $F : \mathbb{R}^2 \mapsto \mathbb{R}^2$ such that

$$F(\cos\varphi_i, \sin\varphi_i) \;=\; (\cos\psi_i, \sin\psi_i) \quad \text{for} \quad i = 0 \ldots n$$

As shown in Figure 1 the function F can be visualized as a mapping between the triangles $(0, 0), (\cos\varphi_{i-1}, \sin\varphi_{i-1}), (\cos\varphi_i, \sin\varphi_i)$ in the domain and the triangles $(0, 0), (\cos\psi_{i-1}, \sin\psi_{i-1}), (\cos\psi_i, \sin\psi_i)$ in the range. By imposing certain conditions on the angles ψ_i we can ensure certain properties of the resulting F. More specifically, the following implications hold true

$$\begin{array}{lcl}\psi_i \text{ strictly monotone and } p = 1 & \Longrightarrow & F \text{ injective} \\ \psi_i \text{ strictly monotone and } p > 1 & \Longrightarrow & F \text{ not injective but open} \\ \psi_i \text{ are not monotone but } p > 0 & \Longrightarrow & F \text{ not open but surjective}\end{array}$$

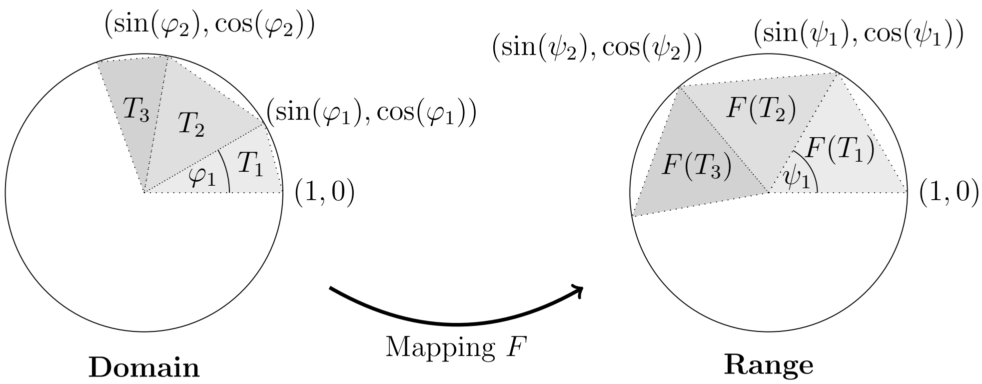

Figure 1: Piecewise linear Rosette Example on $\mathbb{R}^2$

In other words, we have a simple class of examples, which demonstrate that the well known chain of implications [Sch12]

$$\textbf{bijective} \iff \textbf{injective} \Longrightarrow \textbf{open} \Longrightarrow \textbf{surjective} \qquad (3)$$

for general PL functions cannot be strengthened. Here openness means that all images $y = F(x)$ are in the interior of $F(B_r(x))$ for any ball $B_r(x)$ about any preimage x of y. Moreover, in the PL case openness is equivalent to coherent orientation, i.e., the property that the determinants of all linear pieces have the same nonzero determinant sign. In the context of the abs-normal form we can verify this important property more or less explicitly as follows.

3. Polyhedral structure and limiting Jacobians

As in [Gri13] we define the signature vector and matrix by

$$\sigma \equiv \sigma(x) \equiv \textbf{sign}(z(x)) \in \{-1, 0, 1\}^s \quad \text{and} \quad \Sigma \equiv \Sigma(x) \equiv \textbf{diag}(\sigma) \ \in \{-1, 0, 1\}^{s \times s}$$

This vector maps $\mathbb{R}^n$ into $\{-1, 0, 1\}^s$ and represents the control flow in our calculation. As in [Gri13] one can check very easily that the sets

$$P_\sigma \ \equiv \ \{x \in \mathbb{R}^n : \sigma(x) = \sigma\}$$

are relatively open and convex polyhedra in $\mathbb{R}^n$. Being inverse images they are mutually disjoint and span the whole domain $\mathbb{R}^n$. By continuity it follows that P_σ must be open (possibly empty) if σ is definite in that all its components are nonzero. In degenerate situations there may be some indefinite σ that are nevertheless **open** in that P_σ is open.

The limiting Jacobian $\partial^L F(x)$ at some $x \in \mathbb{R}^n$, i.e., the limits of all proper Fréchet derivatives in its neighborhood, is in the PL case simply the finite set

$$\partial^L F(x) \;=\; \{J_\sigma : x \in \overline{P}_\sigma \text{ with } \sigma \text{ open}\}$$

The Clarke generalized Jacobian is the convex hull $\partial F(x) = \mathbf{conv}\left(\partial^L F(x)\right)$. In general it will be quite difficult to calculate all elements of the generating set $\partial^L F(x)$ and we will usually shy away from that combinatorial effort.

Explicit Jacobian representation

On all open σ we find that $|z| = \Sigma z$, so that the first equation in (2) yields

$$(I - L\Sigma)z = c + Zx \quad \text{and} \quad z \;=\; (I - L\Sigma)^{-1}(c + Zx)$$

Notice that due to the strict triangularity of $L\,\Sigma$ the inverse of $(I - L\Sigma)$ is well defined and polynomial in the entries of L. Moreover, due to the structural nilpotency degree ν of L we obtain the Neuman expansion

$$(I - L\Sigma)^{-1} \;=\; I + L\Sigma + (L\Sigma)^2 + \cdots + (L\Sigma)^{(\nu-1)}\,. \tag{4}$$

In the simply switched case $\nu = 1$ we have $L = 0$ and thus the expansion reduces to $I^{-1} = I$. When $\nu = 2$, we have the linear inverse $(I - L\Sigma)^{-1} = I + L\Sigma$. Substituting this expression into the second part of (2) we obtain the local representation:

Proposition 3.1. *On all open P_σ the dependents y can be directly expressed in terms of x, namely as*

$$y \;=\; b + Y\Sigma(I - L\Sigma)^{-1}c + J_\sigma\, x \quad \textit{with} \quad J_\sigma = J + Y\Sigma(I - L\Sigma)^{-1}Z \tag{5}$$

Here J_σ is the Jacobian of F restricted to P_σ. It reduces to $J_\sigma = J + Y\Sigma Z$ for simply switched problems and to J for smooth problems.

Polynomial escape

Computing generalized Jacobians J_σ according to (5) is quite simple, once an open signature σ and thus the corresponding diagonal Σ are known. To find, for a given x, some open σ with the closure $\bar{P}_\sigma$ containing x one may use the following trick, which we like to call **polynomial escape**. Due to piecewise linearity the complement $\mathcal{C}$ of all open P_σ is contained in the union of finitely many hypersurfaces. Hence, no polynomial path of the form

$$x(t) \;\equiv\; x + \sum_{i=1}^{n} e_i t^i \quad \text{with} \quad \det[e_1, e_2, \ldots, e_n] \neq 0, \quad \text{for} \quad e_i \in \mathbb{R}^n$$

can be contained in $\mathcal{C}$. In other words, we find for some σ and $\bar{t} > 0$ that $x(t) \in P_\sigma$ for all $t \in (0, \bar{t})$. The corresponding σ can be computed by lexicographic differentiation as introduced by Nesterov [Nes05] and described in a little more detail in [Gri13]. There it is also shown that any such J_σ is in fact a generalized Jacobian of the underlying nonlinear function if F was obtained by piecewise linearization. Finally, by suitably selecting $e_1 = d \neq 0$, one can make sure that the generalized Jacobian obtained is active in a cone containing the given direction d at least in its closure.

4. Coherent Orientation and Injectivity

As in the smooth case, the determinants of the Jacobains J_σ are of crucial importance for the properties of the PL function $F : \mathbb{R}^n \to \mathbb{R}^n$. It is called coherently oriented if all its Jacobians have the same nonzero determinant sign. As stated for example in [Sch12], the central property openness in the chain (3) is, for PL functions, equivalent to coherent orientation. For simply switched F, like for example all KKT systems of QOPs, we have essentially the same situation as in the affine case, namely bijectivity follows already from coherent orientation and LIKQ.

Proposition 4.1. *If F is simply switched in that $L = 0$ and its kinks satisfy LIKQ then F is bijective if and only if it is coherently oriented.*

Proof. On the basis of the mean value theorem cited above Clarke showed that F has an inverse function near some point x if all elements of the generalized Jacobian $\partial F(x)$ are nonsingular. At all points where F is differentiable that follows from the assumed coherent orientation. At all other points a certain number of $m \leq s$ components of σ vanish, which means in the simply switched case that the $s-$vector $c + Zx$ has m zero components. In fact, it may contain at most $m \leq n$ zeros since otherwise a corresponding $(n+1) \times (n+1)$ submatrix of $[c, Z]$ would have the nonzero null vector $(1, x^\top)^\top \in \mathbb{R}^{n+1}$. Without loss of generality we may assume that exactly the first $m \leq n$ components of $z = z(x)$ vanish. The remaining ones will keep their sign in a sufficiently small neighborhood of x. Due to the linear independence of the first m rows of Z we can find arbitrarily small perturbations $\Delta x \in \mathbb{R}^n$ such that the first m components of $c + Z(x + \Delta x)$ have any one of 2^m sign patterns. Correspondingly, the first m components of the signature vector $\sigma \in \mathbb{R}^s$ attain any $\{-1, 1\}$ pattern on some open domain whose closure contains the given points x. Hence, $\partial F(x)$ contains all matrices $J_\sigma = J + Y\Sigma Z$ where the last $s-m$ components of Σ are fixed and the first m may be $+1$ or -1. By assumption, all these J_σ have the same determinant sign. Changing just one $\sigma_i \in \{-1, 0, 1\}$ of the first m components continuously from -1 to $+1$ corresponds to a rank one change in the corresponding matrix J_σ, whose determinate varies linearly with respect to σ_i and therefore cannot change sign in between. Thus the J_σ along all edges have the same determinant signs, which are inherited by the ones on the face and so on. Therefore, we have shown that all generalized Jacobians are nonsingular so that F is everywhere locally injective and also globally injective. □

Any F satisfying the assumptions of the proposition is *stably coherently oriented* in that all modifications generated by small perturbation of $[c, Z]$ are also coherently oriented. To see this, one only has to note that, if one of the possible 3^s polyhedra was open at points arbitrarily close to x but had a lower dimension at x itself, the LIKQ would have to be violated. Thus, the polyhedral decomposition of all neighboring F is the same and since the respective Jacobians and their determinants are also continuous functions of the perturbation, coherent orientation is maintained. The converse is not true, since one may modify any F with an unstable decomposition at x into one that is stably coherently oriented by adding a suitable multiple α of the identity so that $F(x)$ becomes $F(x) + \alpha x$. This modification does not effect z and thus the lack of LIKQ.

Just assuming stable coherent orientation, we find that all the small perturbations satisfying LIKQ are injective and F, as the limit of such bijective perturbations, inherits this property by the following proposition.

Proposition 4.2. *Let $\mathcal{D} \subseteq \mathbb{R}^n$ be open, and let (F_k) be a sequence of continuous injective maps $F_k : \mathcal{D} \to \mathbb{R}^n$ which converges uniformly on compact sets to $F : \mathcal{D} \to \mathbb{R}^n$. Assume that the preimage $F^{-1}(y) \subseteq \mathcal{D}$ is discrete for every $y \in \operatorname{im}(F)$. Then F is injective. The proof relies on the following lemma.*

Lemma 4.3. *For every $x_0 \in \mathcal{D}$ and every $\varepsilon > 0$ with $B_\varepsilon(x_0) \subseteq \mathcal{D}$ there exists k_0 such that $F(x_0) \in F_k(B_\varepsilon(x_0))$ for all $k \geq k_0$.*

Proof (Lemma). Let $y_0 := F(x_0)$. Since $F^{-1}(y_0)$ is discrete, we can choose $r > 0$ such that $B_{2r}(x_0) \subseteq \mathcal{D}$ and $B_{2r}(x_0) \cap F^{-1}(y_0) = \{x_0\}$. After decreasing r if necessary, we can assume $r \leq \varepsilon$ for the given ε. Write $\Omega := B_r(x_0)$. Then $y_0 \notin F(\partial\Omega)$, hence, $\mathrm{dist}(y_0, F(\partial\Omega))/2 =: \delta > 0$ (note that $F(\partial\Omega)$ is compact since F is again continuous). Choose k' such that $y_k := F_k(x_0) \in B_\delta(y_0)$ for all $k \geq k'$. Choose $k_0 \geq k'$ such that $\|(F - F_k)|_{\partial\Omega}\|_\infty < \delta$ for all $k \geq k_0$. Then, for each of these k, we have

$$\mathrm{dist}(y_0, F_k(\partial\Omega)) \geq \mathrm{dist}(y_0, F(\partial\Omega)) - \|(F - F_k)|_{\partial\Omega}\|_\infty > 2\delta - \delta = \delta$$

and, consequently, $B_\delta(y_0) \subseteq \mathbb{R}^n \setminus F_k(\partial\Omega)$. Because of $y_k \in B_\delta(y_0)$, the points y_0 and y_k lie in the same connected component of $\mathbb{R}^n \setminus F_k(\partial\Omega)$. Therefore we have

$$d(F_k, \Omega, y_0) = d(F_k, \Omega, y_k)$$

where d denotes the Brouwer degree (see e.g., [Ruz04]). The right-hand side of this equation is ± 1 because $F_k|_{\overline{\Omega}}$ is an injective continuous map from a compact set to a Hausdorff space, hence, a homeomorphism onto its image. Thus, $d(F_k, \Omega, y_0) = \pm 1 \neq 0$ and, therefore, $y_0 \in F_k(\Omega)$ for all $k \geq k_0$. The statement of the Lemma now follows from $\Omega = B_r(x_0) \subseteq B_\varepsilon(x_0)$. □

Proof. (Proposition) The Proposition follows immediately by contradiction. Suppose there were $x_1 \neq x_2$ in $\mathcal{D}$ with $F(x_1) = F(x_2) =: y_0$. Choose $\varepsilon > 0$ small enough such that $B_\varepsilon(x_1)$ and $B_\varepsilon(x_2)$ are disjoint subsets of $\mathcal{D}$. Let k_1, k_2 be as in the Lemma, that is, such that $y_0 \in F_k(B_\varepsilon(x_i))$ for all $k \geq k_i$, $i = 1, 2$. Then $y_0 \in F_k(B_\varepsilon(x_1)) \cap F_k(B_\varepsilon(x_2))$ for every $k \geq \max\{k_1, k_2\}$, contradicting injectivity of the F_k. □

Hence we obtain the following strengthening of Proposition 4.1

Corollary 4.4. *If F is simply switched and stably coherently oriented in that all small perturbations have this property, then it is bijective.*

The simply switched one-dimensional example $F(x) = x - |x - \zeta| + |x + \zeta|$ is monotonically growing and thus coherently oriented if $\zeta \leq 0$ but for $\zeta > 0$ it has a slope of -1 in a small interval about the origin. Hence, for the limiting case $\zeta = 0$, where $F(x) \equiv x$, we have coherent orientation, but that property is lost for arbitrarily small $\zeta > 0$. Nevertheless, the function is of course injective so that one might conjecture that for simply switched PL functions openness already implies injectvity.

However, that is not the case as one can see from the following instance of the Rosette example.

$$F(x) \equiv \begin{bmatrix} |x_1| - |x_2| \\ \frac{1}{2}|x_1 + x_2| - \frac{1}{2}|x_1 - x_2| \end{bmatrix} \tag{6}$$

It is simply switched and coherently oriented, but not injective since F is even, so that $F(-x) = F(x)$.

The LIKQ is violated since the four kinks $\{x_1 = 0\}, \{x_2 = 0\}, \{x_1 = x_2\}$ and $\{x_1 = -x_2\}$ all intersect at the origin. Moreover, one can see that the perturbations

$$F_\varepsilon(x) \;\equiv\; \begin{bmatrix} |x_1 + \varepsilon| - |x_2 + \varepsilon| \\ \frac{1}{2}|x_1 + x_2| - \frac{1}{2}|x_1 - x_2| \end{bmatrix}$$

are no longer coherently oriented for $\varepsilon \neq 0$. More specifically, for $\varepsilon > 0$ we have the Jacobian

$$F'_\varepsilon \;=\; \begin{bmatrix} 1 & -1 \\ 0 & -1 \end{bmatrix} \quad \text{at} \quad x \;=\; \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \;=\; \begin{bmatrix} -\varepsilon/2 \\ -\varepsilon/4 \end{bmatrix}$$

whose determinant is -1 so that we do not have stable coherent orientation.

5. Generalized Newton Variants

If all elements of $\partial^L F(x_*)$ are nonsingular at some root $x_* \in F^{-1}(0)$, it follows from the celebrated theorem of Qi and Sun [QS93] that the full step Newton iteration

$$x_+ \;=\; x - A^{-1}F(x), \quad \text{with} \quad A \in \partial^L F(x_k) \tag{7}$$

converges from all x_0 sufficiently close to x_*. In fact, this result holds here trivially, since the iteration converges in one step from all points in the open neighborhood

$$\Omega(x_*) \equiv \{P_\sigma : x_* \in \overline{P}_\sigma\}^\circ$$

Of course, this means that all the combinatorial issues have already been resolved by the choice of x_0.

Much more interesting is the question under which conditions the full step Newton method (7) or a damped variant of it converge globally, i.e., from all initial points x_0. Using the mean value theorem of Clarke stated for example as Prop.7.1.16 in [FP03], one can establish by standard arguments the following global convergence result.

Proposition 5.1 (Full step convergence)**.**
Given $x_* \in F^{-1}(0)$ the full step Newton method converges from all x_0 in finitely many steps to x_* if, with respect to so some induced matrix norm, either of the following contractivity assumptions is satisfied.

$$\|I - J_\sigma^{-1} J_{\tilde{\sigma}}\| < 1, \quad \textit{for} \quad \sigma, \tilde{\sigma}\, \textbf{open}, \tag{8}$$

or

$$\|I - J_\sigma J_{\tilde{\sigma}}^{-1}\| < 1, \quad \textit{for} \quad \sigma, \tilde{\sigma}\, \textbf{open}. \tag{9}$$

In either case the root $\{x_*\} = F^{-1}(0)$ is unique.

Proof. The assumptions ensure by an application of the mean value theorem of Clarke that the error norm $\|x^+ - x_*\|$ or the residual norm $\|F(x_*)\|$, respectively, are reduced by a fixed factor at each iteration. Finite convergence from within $\Omega(x_*)$ follows as above. □

The Proposition deals with a special case of the general theory on *nonsingular uniform Newton approximations* in the sense of [FP03]. Now we will look for sufficient conditions for the contractivity properties (8) or (9) and thus global convergence of full step Newton and injectivity of F in terms of the abs-normal representation. To obtain an explicit expression for the inverses J_σ^{-1} we will assume that the matrix $J \in \mathbb{R}^{n\times n}$ representing the smooth part of our function is nonsingular. Should that a priori not be the case we can use the trivial identity

$$v \;=\; ||v| + v| - |v|, \quad \text{for} \quad v \in \mathbb{R} \tag{10}$$

to shift terms between the smooth and nonsmooth parts without changing the mapping F. However, for each modified entry we introduce two new switching variables. Now we obtain the result.

Proposition 5.2.
Assume that the abs-normal form of F has an invertible smooth part J and that

$$\hat{\rho} \equiv \|J^{-1}Y\|_p \|Z\|_p < 1 - \|L\|_p$$

Then generalized Newton converges in finitely many iterations from any x_0 to the then unique solution x_* if

$$\frac{\hat{\rho}}{(1 - \hat{\rho} - \|L\|_p)(1 - \|L\|_p)} \;<\; \frac{1}{2} \tag{11}$$

Moreover, both the solution error and the residual are monotonically reduced w.r.t the p-norm.

Proof. It follows from (5) that

$$\|I - J^{-1}J_\sigma\|_p \le \|J^{-1}Y\|_p\|(I - L\Sigma)^{-1}\|_p\|Z\|_p \;\le\; \hat\rho\big/\,(1 - \|L\|_p).$$

Hence we have by the Banach Pertubation Lemma that

$$\|J_\sigma^{-1}\|_p = \|[I - (I - J^{-1}J_\sigma]^{-1}\|_p \le 1\,/[1 - \hat\rho\,/(1 - \|L\|_p)]\ ,$$

which immediately yields for any pair of open signatures $\sigma, \tilde\sigma$

$$\|J_\sigma^{-1}J\|_p \;\le\; \frac{1 - \|L\|_p}{1 - \hat\rho - \|L\|_p} \tag{12}$$

Furthermore we derive from (5) that

$$\begin{aligned} J^{-1}\left[J_{\tilde\sigma} - J_\sigma\right] &= J^{-1}Y\left[\tilde\Sigma(I - L\tilde\Sigma)^{-1} - \Sigma(I - L\Sigma)^{-1}\right]Z \\ &= J^{-1}Y\left[(I - \tilde\Sigma L)^{-1}\tilde\Sigma - \Sigma(I - L\Sigma)^{-1}\right]Z \\ &= J^{-1}Y(I - \tilde\Sigma L)^{-1}\left[\tilde\Sigma(I - L\Sigma) - (I - \tilde\Sigma L)\Sigma\right](I - L\Sigma)^{-1}Z \\ &= J^{-1}Y(I - \tilde\Sigma L)^{-1}\left[\tilde\Sigma - \Sigma\right](I - L\Sigma)^{-1}Z \end{aligned}$$

Now taking again norms and applying standard inequalities we find

$$\|J^{-1}\left(J_{\tilde\sigma} - J_\sigma\right)\|_p \;\le\; \frac{2\hat\rho}{(1 - \|L\|_p)^2}\ . \tag{13}$$

By multiplication of (12) and (13), the last inequality ensures that both (8) and (9) are satisfied. □

Piecewise Newton

The conditions for the global convergence of full step Newton derived above are certainly rather strong and various globalizations like Ralph's path search have been proposed. On the other hand, it was observed in [Gri13] that coherent orientation implies that the fibres

$$[x_0] \;\equiv\; \{x \in \mathbb{R}^n : F(x) = \lambda F(x_0), 0 < \lambda \in \mathbb{R}\} \tag{14}$$

are, for almost all $x_0 \in \mathbb{R}^n$, bifurcation-free piecewise linear paths whose closure contains a root of F. The other, singular fibres may have bifurcations, but there is always a possibility to further reduce the residual towards a solution.

The question how this piecewise Newton method is best implemented needs further investigation, but numerical experiments are certainly encouraging [PGar]. There is a key difference between this piecewise Newton and damped Newton in that piecewise Newton is not based on just any limiting Jacobian at the current iterate, but on one that is indeed valid along the direction being taken. It cannot be guaranteed in the usual paradigm that an oracle evaluates at any x the residual $F(x)$ and some limiting Jacobian $\partial^L F(x)$.

We may summarize the results of this fourth and fifth section in the following graph of implications.

Contractivity $\Rightarrow$ Bijectivity $\Longrightarrow$ Openness $\Rightarrow$ Surjectivity
($\Longleftarrow$ if simply switched+stably coherently oriented)

The fact that the last two implications are not reversible in general was already demonstrated in Section 2 on the Rosette example, which is not simply switched. The possibility of failure for full step Newton on bijective problems can be seen in the Rosette example (6). With a right-hand side $(1, -1)$ and a starting point $(2, 1)$, Newton's method begins to cycle immediately.

6. Schur complement and the unfolded system

It turns out that we can eliminate x when the smooth part J is nonsingular .

Lemma 6.1. *Provided that* $\det(J) \neq 0$*, we have the Schur complement*

$$S \equiv L - ZJ^{-1}Y \ \in \ \mathbb{R}^{s\times s}$$

and in P_σ *it holds that*

$$\det(J_\sigma) \ = \ \det(J)\, \det(I - S\,\Sigma)$$

Moroever, if this determinant is nonzero, the inverse of J_σ *is given by*

$$J_\sigma^{-1} \ = \ J^{-1} - J^{-1}\, Y\, \Sigma\, (I - S\,\Sigma)^{-1} Z\, J^{-1} \tag{15}$$

Proof. According to Sylvester's determinant theorem we have

$$\begin{aligned}\det(J_\sigma)/\det(J) &= \det\left[I + J^{-1}Y\Sigma(I - L\Sigma)^{-1}Z\right] = \det\left[I + ZJ^{-1}\,Y\,\Sigma\,(I - L\Sigma)^{-1}\right]\\ &= \det(I - L\Sigma)^{-1}\,\det\left(I - L\,\Sigma + Z\,J^{-1}\,Y\,\Sigma\right) = \det(I - S\,\Sigma)\end{aligned}$$

where we have used that the unitary lower triangular matrix $I - L\Sigma$ has determinant 1. □

Whenever J dominates the other three submatrices, things are not too difficult, as we will see below. Notice that nonsingular linear transformations on the independents x and/or the dependents y leave the Schur complement completely unchanged. At least for (generalized) Newton variants we could therefore assume without loss of generality that $J = I$, although that does not seem to help all that much.

Rescaling the switching variables z by a positive diagonal matrix D would modify Z to DZ, Y to YD^{-1} and replace L by the similarity transformation DLD^{-1}, which is still strictly lower triangular. One can choose D such that the transformed DLD^{-1} is arbitrarily small in any one of the standard norms that are monotonic in the coordinates, but that may require a pretty wild scaling. More important is the Schur complement S, which would also be replaced by its similarity transformation DSD^{-1}.

Conditions for coherent orientation

The condition that $\det(I - S\Sigma)$ be positive for all switching matrices Σ is sufficient for coherent orientation of F. In Theorem 2.3 of [Rum97] Rump gave several equivalent properties, one of which is that the *sign real spectral radius*

$$\rho_0^s(S) \;\equiv\; \max\left\{\rho_0(\Sigma\, S) : \Sigma \in \mathbf{diag}(-1,1)^n\right\}$$

is less than 1. Here $\rho_0(S) \le \rho(S)$ denotes the *real spectral radius* of a square matrix, i.e., the largest modulus of any real eigenvalue of $S \in \mathbb{R}^{n\times n}$. The complex eigenvalues are ignored in this maximization, which makes $\rho_0(S)$ highly discontinuous with respect to S. Remarkably, $\rho_0^s(S)$ is again continuous in the entries of S and it vanishes exactly when S is permuted strictly triangular. This is true for the leading part L of our Schur complement so that we must have $\rho_0^s(S) < 1$ when the additional term $YJ^{-1}Z$ is sufficiently small. In general, deciding whether $\rho_0^s(S)$ lies below a given bound is an NP hard problem. Rump also showed that the following property is sufficient, but not necessary for $\rho_0^s(S) < 1$ and, thus, coherent orientation.

Definition 6.2. *An abs-normal form of F is called smoothly dominant if*

$$\rho \equiv \|DSD^{-1}\|_p < 1$$

for some p-matrix norm and some positive diagonal scaling D.

This condition was already used by Bokhoven in his dissertation [vB81]. Now we will show that coherent orientation can be present even when all p-norms are substantially greater than 1, i.e., when the PL system is far from being smoothly dominant.

Lemma 6.3. *There are matrices $S_n \in \mathbb{R}^{n\times n}$ with signed real spectral radius $\rho_0^s(S) \leq 0.9$ for which all p norms $\|S_n\|_p$ are greater than 1 and $\lim_n \|S_n\|_p = \infty$.*

Proof. We consider the following matrix $S = S_n$, which is just a scaling of the one given in [Rum97]

$$S = \tfrac{9}{10} \cdot (\mathbf{sign}(j-i))_{i,j=1\ldots n} \in \mathbb{R}^{n\times n} \tag{16}$$

Its L_1 and L_∞ norm are obviously equal to $\|S\|_1 = 0.9\,(n-1) = \|S\|_\infty$. Moreover, multiplying S by the first Cartesian basis vector $e_1 = (1, 0, 0, \ldots, 0) \in \mathbb{R}^n$, we find for any other $p \in [1, \infty]$

$$\|S\|_p = \max_{x\neq 0} \frac{\|Sx\|_p}{\|x\|_p} \geq \frac{\|Se_1\|_p}{\|e_1\|_p} = \|Se_1\|_p = 0.9\,(n-1)^{1/p} \xrightarrow{n\to\infty} \infty$$

Hence, we cannot achieve smooth dominance for any $p \geq 1$. However, from [Rum97] we know that the sign real spectral radius satisfies $\rho_0^s(S) = 0.9 < 1$ so that we have coherent orientation of F. $\square$

To see that smooth dominance can also arise when $\rho(|S|) > 1$ let us consider the 2×2 matrix

$$S \;=\; R(\tfrac{\pi}{2}) \;=\; \frac{0.9}{\sqrt{2}} \begin{pmatrix} 1, & -1 \\ 1, & 1 \end{pmatrix}$$

It represents a rotation by $\pi/2$ followed by a contraction by 0.9. Then we have

$$\|S\|_2 = 0.9 \;<\; 1 \;<\; 0.9\sqrt{2} \;=\; \rho(|S|)$$

As a more interesting example for smooth dominance let us consider a problem

$$Tx + \max(x, 0) = b, \quad \text{where} \quad T \;\succ\; 0$$

is symmetric positive definite, the stronger assumption used in [BC08]. Rewriting this problem in abs-normal form using $\max(x, 0) \equiv (x + |x|)/2$ we obtain

$$z = x \quad \text{and} \quad y = -b + (T + I/2)x + |z|/2$$

This corresponds to $c = 0, Z = I, L = 0, J = T + I/2, Y = I/2$ and yields the Schur complement $S \;=\; 0 - (T + I/2)^{-1}/2 \;=\; -(I + 2\,T)^{-1} \prec 0$. It is negative definite with spectral radius below 1. Hence, we have smooth dominance as $\|DSD^{-1}\|_2 \;<\; 1$ for $D = I$. We have verified that the fixed point iteration suggested in 20 below converges when T is the usual second order divided difference stencil. However, it does so very slowly and applying the generalized Newton iteration (7) and equivalently (21), also advocated by in [BC08] turns out to be much more effective.

An even stronger condition for smooth dominance and thus coherent orientation follows from the well known result of Perron-Frobenius.

Lemma 6.4. *Perron-Frobenius scaling*
Suppose the componentwise modulus $|S|$ is not permuted block-triangular. Then its spectral radius $\rho(|S|)$ is positive and the corresponding eigenvector $d \in \mathbb{R}^n$ is strictly positive such that, with $D = \mathbf{diag}(d)$ where $e \in \mathbb{R}^s$ is the vector of ones

$$D^{-1}Sd \;\equiv\; D^{-1}SD\,e = \rho(|S|)e \;\;\Longrightarrow\;\; \|D^{-1}SD\|_\infty \;=\; \rho(|S|)$$

If $\rho(|S|) = 0$, the norm $\|D^{-1}SD\|_\infty$ can be made arbitrarily small.

Proof. It is well known that all components of the eigenvector d are positive if the corresponding eigenvalue $\rho(|S|)$ is nonzero. Then we find immediately that e is the eigenvector associated with the largest eigenvalue of $|\tilde{S}|$ for $\tilde{S} = D^{-1}SD$, which in turn shows that $\|\tilde{S}\|_\infty = \||\tilde{S}|\|_\infty$ has the same value. If $\rho(|S|) = 0$, we can add $\varepsilon\, e\, e^\top$ to $|S|$ and apply the first observation to establish the second. $\square$

According to the Lemma, *absolute contractility* i.e., $\rho(|S|) < 1$, implies smooth dominance in the infinity norm. Moreover, we may always similarity transform S by some diagonal $D > 0$ such that all rows of $\tilde{S} \equiv D^{-1}SD$ have the same l_1 norm equaling $\rho(|S|) = \rho(|\tilde{S}|)$. We will call this process *equilibration.* This may not work if S is reducible in that it is permuted block triangular, which can for example be tested by the algorithm given in [DER86]. In the reducible case the unfolded system discussed below can be decomposed into several subsystems, to which our solution techniques can be applied successively. Consequently, we may assume from now on without loss of generality that the sparsity pattern of S is irreducible, which also implies $\rho_0^s(S) > 0$. Alternatively, we can scale by the left Perron-Frobenius vector $\tilde{d}$ of $|S|$ to achieve $\|\tilde{D}^{-1}S\tilde{D}\|_1 \;=\; \rho(|S|)$ for $\tilde{D} = \mathbf{diag}(\tilde{d})$, but that appears to be of little help here.

The unfolded system

We will assume throughout that J is nonsingular, hence, that S is well defined and that a suitable scaling was applied to make some norm $\|S\|_p$ small, if not necessarily less than one. So far we have looked at (2) as a system that defines a unique $z \in \mathbb{R}^s$ and thus a corresponding y for each $x \in \mathbb{R}^n$ via the first set of s triangular equations. Now suppose we have given a fixed target value y, which we can subsume into b, and compute for each z the corresponding value

$$x \;=\; x(z) \;\equiv -J^{-1}(b + Y|z|) \tag{17}$$

Substituting this result into the first equation we obtain for z the piecewise linear system

$$H(z) \;\equiv\; z - L|z| + ZJ^{-1}Y|z| \;=\; (I - S\Sigma)z \;=\; \hat{c} \;\equiv\; c - ZJ^{-1}b \tag{18}$$

Note that the generalized Jacobians $(I - S\,\Sigma)$ of the unfolded vector function $H(z)$ all have the same determinant sign exactly when $\rho_0^s(S) < 1$, which we encountered as sufficient condition for the coherent orientation of F. Generally, $F(z)$ must be coherently oriented if this is true for $H(z)$, but the converse implication is usually not true. The reason is that while all possible sign combination of z arise in the domain $\mathbb{R}^s$ of z, the switching variables $z = z(x)$ are typically restricted to a Lipschitzian submanifold in $\mathbb{R}^s$ as x ranges over $\mathbb{R}^n$.

Conversely, for any given z solving the lower part of (2) for x yields the corresponding value

$$z \;=\; z(x) \;\equiv G^{-1}(c + Zx) \quad \text{with} \quad G(z) \;\equiv\; z - L|z| \tag{19}$$

As stated by Lemma 6.4 we can make any p-norm $\|L\|_P$ of the strictly lower triangular matrix L as small as possible and in particular smaller than 1. Then the existence of G^{-1} follows not only from the triangularity of L but also the Banach fixed point theorem. Now we can observe that solutions of the original problem OPL and the unfolded problem UPL correspond to each other.

Lemma 6.5 (One-to-One solution correspondence).
Under our general assumptions with $\det(J) \neq 0$ *a point* $x_* \in \mathbb{R}^n$ *is a solution of the OPL* $F(x) = 0$ *if and only if it is a fixed point of* $x(z(x))$*, which is in turn equivalent to* $z_* = z(x_*)$ *being a fixed point of* $z(x(z))$ *and equivalently a solution of the UPL* $H(z) = \hat{c}$ *.*

Proof. We have the equivalences $F(x) = 0$

$$\begin{aligned}
&\Longleftrightarrow \quad x = -J^{-1}[b + Y|z|] \quad \text{with} \quad z = c + Zx + L|z| \\
&\Longleftrightarrow \quad x = -J^{-1}[b + Y|z|] \quad \text{with} \quad G(z) = c + Zx \\
&\Longleftrightarrow \quad x = -J^{-1}[b + Y\left|G^{-1}(c + Zx)\right|] \\
&\Longleftrightarrow \quad x = x(z(x)) \quad \Longleftrightarrow \quad z = z(x(z)) \\
&\Longleftrightarrow \quad z = G^{-1}(c + Zx) \quad \text{with} \quad x = -J^{-1}(b + Y|z|) \\
&\Longleftrightarrow \quad z = G^{-1}(c - ZJ^{-1}(b + Y|z|)) \\
&\Longleftrightarrow \quad G(z) = c - ZJ^{-1}(b + Y|z|) \\
&\Longleftrightarrow \quad z - L|z| = c - ZJ^{-1}(b + Y|z|)
\end{aligned}$$

which is equivalent to $H(z) = \hat{c}$ defined in (18) as asserted. □

We may interpret $H(z)$ as a simply switched PL function in absnormal form with $z \equiv x, Z = I = J, L = 0$, and $Y = -S$. The Schur complement is then again $0 - (-S)S\,I^{-1}I = S$, which was to be expected. Since the LIKQ condition is satisfied, the unfolded function $H(z)$ is always bijective if and only if it is open, which happens exactly when $\rho_0^s(S) < 1$.

7. Solving the unfolded system UPL

In view of Lemma 6.5 we can hope that the largely equivalent fixed point iterations $x^+ = x(z(x))$ and $z^+ = z(x(z))$ defined by (19) and (17) lead to convergence. As it turns out it is a little easier to establish convergence of the coupled iteration with respect to the z-component and the x component must then converge to its own fixed point by continuity .

Proposition 7.1. *If in some p-norm*

$$\|S - L\|_p + \|L\|_p \; < \; 1$$

then the iteration $z^+ = z(x(z))$ converges from all z_0 to the unique fixed point z_*. Moreover, the corresponding $x_* = -J^{-1}(b+Y|z_*|)$ is the unique root of $F(x) = 0$.

Proof. Since for any pair $z, y \in \mathbb{R}^s$ by the inverse triangle inequality

$$\|G(z) - G(y)\|_p \; = \; \|(z - y) - L(|z| - |y|)\|_p \; \geq \|z - y\|_p(1 - \|L\|_p|)$$

the inverse G^{-1} has the Lipschitz constant $1/(1 - \|L\|_p|)$. The Lipschitz constant of the map $R(z) \equiv \hat{c} - ZJ^{-1}Y|z|$ is simply $\|Z\,J^{-1}\,Y\|_p$, which can be expressed in terms of the Schur complement as $\|S - L\|_p$. Using the multiplicativity of Lipschitz constants we derive for the fixed point iteration $z(x(z)) = G^{-1} \circ R(z)$

$$\sup_{z \neq y} \frac{\|G^{-1} \circ R(z) - G^{-1} \circ R(y)\|_p}{\|y - z\|_p} \; \leq \; \frac{\|Z\,J^{-1}\,Y\|_p}{1 - \|L\|_p} \; = \; \frac{\|S - L\|_p}{1 - \|L\|_p}$$

Since the last upper bound is less than 1 exactly when the assumption of the proposition is satisfied, convergence follows again by Banach's fixed point theorem. The last assertion holds by substitution of (x_*, z_*) into (2). □

Modulus Algorithm

It follows immediately from the triangle inequality that the fixed point iteration is only guaranteed to converge when the problem is smooth dominant in that $\|S\|_p < 1$. Under that somewhat weaker condition one may apply the simpler fixed point iteration

$$z^+ \; = \; \hat{H}(z) \; \equiv \hat{c} + S|z| \tag{20}$$

Here no triangular substitution process is needed and S may or may not be formed explicitly. If not, we have to just solve one linear system in J at each iteration and multiply vectors by the matrices Y, Z and L. A lack of smooth dominance may then only be discovered by nonconvergence. The simple fixed point iteration was described as *modulus algorithm* in Theorem 10 on page 72 of [vB81]. We restate the basic convergence result.

Proposition 7.2. *If the abs-normal form of F is smoothly dominant in that $\rho = \|S\|_p < 1$, then the iteration* (21) *converges for all $\hat{c}$ from any z_0 to the unique solution $z_* = H^{-1}(0)$.*

Proof. To prove contractivity of $\hat{H}$ on $\mathbb{R}^s$ we note that

$$\|\hat{H}(z)-\hat{H}(\tilde{z})\|_p \;=\; \|S(|z|-|\tilde{z}|)\|_p \;\leq\; \|S\|_p\||z|-|\tilde{z}|\|_p \;\leq\; \rho\||z-\tilde{z}|\|_p \;=\; \rho\|z-\tilde{z}\|_p$$

Thus, the Banach fixed point theorem ensures linear convergence to a unique root with monotonically declining error norms $\|z_k - z_*\|_p$. □

To verify that coherent orientation is not sufficient for the fixed point iteration to converge we applied it to the example $S = S_n$ from (16) for $n = 1000$ with $\mathbf{c} = (\sin(i))_{i=1\ldots n}$ and $z_0 = \mathbf{0} \in \mathbb{R}^n$. Then $z^+ = \hat{c}+S|z|$ diverges immediately. Whether there can be convergence of the fixed point iteration from generic starting points without smooth dominance is not yet clear.

Generalized Newton on ULP

The convergence of the fixed point iterations is quite reliable, but may be asymptomatically rather slow. In particular, neither fixed point iteration promises finite convergence, so we wish to again examine Newton variants. Applying the generalized Newton method to $H(z) = \hat{c}$ we obtain the recurrence

$$z_+ \;= z - A^{-1}(H(z) - \hat{c}), \quad \text{with} \quad A \in \partial^L H(z) \tag{21}$$

Since all A now have the simple form $I - S\,\Sigma$, we obtain as a specialization of Proposition 5.1

Proposition 7.3. *If the abs-normal form of F is smoothly dominant such that $\rho = \|S\|_p < 1/3$, then the iteration* (21) *converges for all $\hat{c}$ in finitely many iterations from any z_0 to the unique solution $z_* = H^{-1}(0)$. Moreover, the p-norms of both $z - z_*$ as well as $H(z) - \hat{c}$ are monotonically reduced*

Proof. This time we simply need to bound

$$\|(I - S\Sigma)^{-1}S(\Sigma - \tilde{\Sigma})\|_p \leq \|(I - S\Sigma)^{-1}\|_p\|S\|_p\|\Sigma - \tilde{\Sigma}\|_p \leq 2\rho/(1 - \rho) < 1$$

everything else follows as above. □

Again the requirement that $\rho = \|S\|_p < 1/3$ seems rather strong. To develop an alternative condition we note that the Newton iterate z^+ reached from a current point z with $\sigma = \mathbf{sign}(z)$ and $\Sigma = \mathbf{diag}(\sigma)$ is simply given by

$$z^+ \;=\; (I - S\Sigma)^{-1}c$$

In other words, the signature vector $\sigma = \mathbf{sign}(z)$ uniquely determines the next iterate $\sigma^+ = \sigma(z^+)$, provided σ is definite, i.e., contains no zeros. For generic $\hat{c}$ the 2^n possible images σ^+ defined by definite σ will also be definite. Then we may interpret the Newton iteration as a finite automaton with the transition function $\sigma^+ = N(\sigma)$ on the state set $\mathcal{V} \equiv \{-1, 1\}^n \equiv \{-, +\}^n$. We can also regard the $\mathcal{V}$ as the vertex set of a directed graph with the edges $(\sigma, N(\sigma))$. An example with $n = 3$ depicting the counter example of Proposition 7.7. is shown in Fig. 2

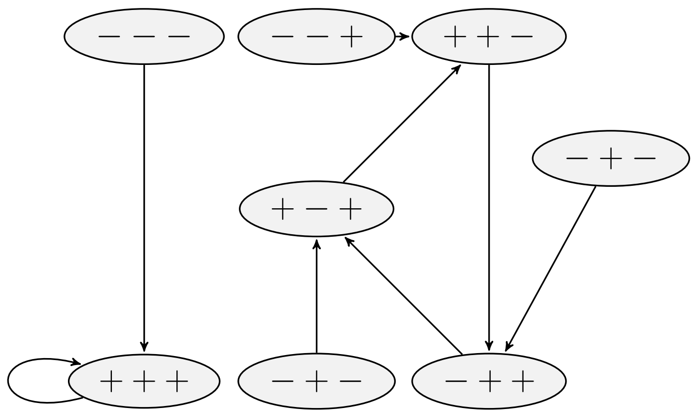

Figure 2: Transition Graph of Newton's method on unfolded system

Since all vertices σ in the directed graph $\mathcal{G}$ have a unique outgoing arc $(\sigma, N(\sigma))$ its structure is rather simple.

Proposition 7.4. *Each connected component of the transition graph $\mathcal{G}$ contains either one fixed point or a cycle of length greater than 1.*

Proof. From any initial z_0 the sequence of iterations $z_k = N^k(z_0)$ stays in the connected component of z_0 and must reach a fixed point or begin to cycle. Let $C(z_0)$ denote the set of vertices that are touched infinitely often by this sequence. Let $Prec(C(z_0))$ denote the set of all $z \in \mathcal{G}$ with $C(z) = C(z_0)$. Suppose the connected subgraph $Prec(C(z_0))$ had outgoing or incoming edges. There can be no incoming edges because repeatadly applying N to their origins would also lead to $C(z_0)$. Also there can be no outgoing edges because their origins would lead to a cycle or fixpoint outside $Prec(C(z_0))$. This completes the proof. □

To avoid cycling, we may impose the condition $\rho(|S|) \leq 1/2$, which allows us to prove finite termination and even limit the computational effort to $n^3/3$ fused multiply adds. In fact, the method simply amounts to a sign controlled version of Gaussian elimination.

Proposition 7.5. *Suppose the Schur complement $|S|$ is irreducible and absolutely contractive with $\rho = \rho(|S|) \leq 1/2$. Then the iteration (21) converges for all $\hat{c}$ in at most s iterations from any z_0 to the unique solution $z_* = H^{-1}(0)$.*

Proof. After equilibration by the Frobenius-Perron vector we may assume without loss of generality that $\rho = \rho(|S|) = \|S\|_\infty \leq 1/2$. Using the Neumann expansion we then find that for any given signature Σ

$$A_\sigma \equiv (I - S\,\Sigma)^{-1} - I = \sum_{i=1}^{\infty} (S\,\Sigma)^i \implies \|A_\sigma\|_\infty \leq \sum_{i=1}^{\infty} \|S\,\Sigma\|_\infty^i \leq \frac{\rho}{1-\rho} \leq 1$$

Dropping the counter k and the hat of $\hat{c}$ we find that the Newton iterate z^+ reached form a current point z with $\sigma = \mathbf{sign}(z)$ is simply given by

$$z^+ \;=\; (I - S\Sigma)^{-1} c \;=\; (I + A_\sigma)\, c.$$

Now we perform symmetric pivoting by reordering the equations and the components of z such that the first component c_1 of the permuted vector c is its largest, i.e., $|c_1| = \|c\|_\infty$. Then we find that

$$|z_1^+ - c_1| = |e_1^\top A_\sigma\, c| \leq \|e_1^\top A_\sigma\|_1 \|c\|_\infty \leq |c_1|\rho/(1-\rho) \;\leq |c_1|$$

Hence, the sign of the first component of z^+ is the same as $\sigma_1 \equiv \mathbf{sign}(c_1)$ or possibly zero. In either case we have $|z_1| = \sigma_1 z_1$ for sure. This will remain true over all subsequent iterations. Note that the reorderings of the equations do not affect the Newton iterates. Knowing the sign of z_1 allows us to rewrite the first equation and express it as a linear combination of the other $|z_j|$, namely

$$z_1(1 - \sigma_1 s_{11}) \;=\; c_1 + \sum_{j=2}^{s} s_{1j}|z_j| \quad\implies\quad z_1 = \frac{c_1}{1 - \sigma_1 s_{11}} + \sum_{j=2}^{s} \frac{s_{1j}|z_j|}{(1 - \bar{\sigma}_1 s_{11})}$$

Substituting this relation into the other equations, which corresponds to one step of Gaussian elimination, we obtain for $i = 2 \ldots s$

$$z_i \;=\; c_i + c_1/(1 - \sigma_1 s_{11}) + \sum_{j=2}^{s} [s_{ij} + (s_{i1}s_{1j})/(1 - \bar{\sigma}_1 s_{11})]|z_j| \;\equiv\; \tilde{c}_i + \sum_{j=2}^{s} \tilde{s}_{ij}|z_j|$$

Hence, we have reduced the piecewise linear system in s switching variables to one in $s-1$ variables. The new matrix $\tilde{S} \equiv (\tilde{s}_{ij})_{j=2\ldots s}^{i=2\ldots s}$ satisfies $\|\tilde{S}\|_\infty \leq \rho = \|S\|_\infty$ since, for each $i > 1$,

$$\sum_{j=2}^{s} |\tilde{s}_{ij}| \leq \sum_{j=2}^{s} |s_{ij}| + \frac{|s_{i1}|}{(1 - \bar{\sigma}_1 s_{11})} \sum_{j=2}^{s} |s_{1j}| \leq \rho - |s_{i1}| + \frac{|s_{i1}|(\rho - |s_{11}|)}{(1 - \bar{\sigma}_1 s_{11})} < \rho$$

Now we can apply the argument recursively to see that on each Newton iteration one additional component of z has its sign fixed for good. □

Signed Gaussian Elimination

The reduction argument in the proof can be applied directly to generate a Gaussian elimination procedure, where first the sign σ_i of the z_i associated with the currently largest constant term c_i is determined, then z_i is eliminated and so on until we reach and compute the last component, say z_j. Then one can perform a backward substitution sweep to actually compute the values of the z_i. In contrast to an LU factorization, we must repeat the process for any new right-hand side $\hat{c}$ since the pivoting order has to be adjusted.

Corollary 7.6. *Any unfolded system $H(z) = \hat{c}$ with $\rho(|S|) \leq \frac{1}{2}$ has a unique solution which can be computed by sign controlled Gaussian elimination in at most $s^3/3$ fused multiply add operations plus $O(s)$ divisions.*

Propositions 7.5 and 7.6 as well as Corollary 7.6 ensure the finite convergence of the generalized Newton method under the conditions $\rho = \|S\|_p < 1/3$ and $\rho = \rho(|S|) \leq 1/2$, respectively. Obviously, the second condition does not imply the former, but the converse is also true so that there are problems where only one but not both theorems apply. To demonstrate this we consider the example

$$S = 0.3\,[I - ee^\top/9] \;\in\; \mathbb{R}^9 \quad \text{with} \quad e = (1)_{1\ldots 9}$$

Here S is a scaled elementary reflector so that $\rho = \|S\|_2 = 0.3 \cdot 1 < 1/3$. However, one can easily check that $\rho(|S|) = 0.3 \cdot 16/9 = 1.6/3 > 0.5$ so that Proposition 7.5 applies, but neither 7.6 nor its Corollary 7.7.

Divergence on Cyclic Example

Another question that arises is whether the bound $1/2$ imposed on $\rho = \rho(|S|)$ in Proposition 7.6 and its corollary could not be weakened. The answer is that for n of any significant size the bound may only be raised a minute amount above $1/2$ without opening the possibility of divergence. More specifically, we have the following family of counter examples, whose instance for $n = 3$ was already depicted in Figure 2.

Proposition 7.7. *For $n > 2$ set $\hat{c} = (1)_{1\ldots n}$ and define $S \in \mathbb{R}^{n\times n}$ as the cyclic Töplitz matrix*

$$S = \begin{bmatrix} \mathbf{0} & a \\ a\, I_{n-1} & \mathbf{0} \end{bmatrix}.$$

Then, if $a \in \mathbb{R}$ satisfies

$$\frac{1}{2} + \frac{1}{2^n} \;\leq\; a \;\leq\; \frac{1}{\sqrt{2}},$$

the generalized Newton method cycles between n distinct and definite points when the initial z contains exactly one negative component and no zeros.

Proof. Suppose the current approximation $z = (z_i)_{i=1}^n$ consist of only positive components except for one, say $z_i < 0$. Then we will show that the next iterate z^+ has only positive iterates except for $0 > z^+_{i^+}$ with $i^+ \equiv 1 + (i \mod n)$. This relation obviously establishes the assertion, since the single negative sign will cycle infinitely often. Due to the symmetry of the situation we may assume w.l.o.g. that the last component of the current iterate z is negative. Hence where we have $\Sigma(z) = \text{diag}(1, \ldots, 1, -1)$ and the next iterate $\zeta = z^+$ is then the solution of the system of linear equations,

$$\begin{bmatrix} 1 & 0 & \ldots & a \\ -a & 1 & \ldots & 0 \\ & \ddots & & \\ & & -a & 1 \end{bmatrix} \begin{bmatrix} \zeta_1 \\ \vdots \\ \\ \zeta_n \end{bmatrix} = \begin{bmatrix} 1 \\ \vdots \\ \\ 1 \end{bmatrix}$$

Thus in terms of ζ_1 the other components ζ_i for $i = 2, \ldots, n$ are given by

$$\zeta_i = 1 + a\,\zeta_{i-1} = \left(\frac{1 - a^{i-1}}{1 - a}\right) + a^{i-1}\zeta_1$$

Substituting these expressions into the first line of the system we find

$$\zeta_1 + a\,\zeta_n = 1 \implies \zeta_1 + a\frac{(1 - a^{n-1})}{(1 - a)} + a^n\,\zeta_1 = 1 \implies \zeta_1 = \frac{1 - 2\,a + a^n}{(1 + a^n)(1 - a)}$$

Now we want to achieve a shift of the negative entry from the last to the first position during the iteration from z to z^+. So ζ_1 should become negative and ζ_2 has to stay positive. In other words, we have to impose the two conditions $\zeta_1 < 0$ and $\zeta_2 > 0$. From the first one it follows that

$$0 > 1 - a\frac{(1 - a^{n-1})}{(1 - a)} \iff \sum_{i=0}^{n-1} a^i > 2$$

and the second one is equivalent to

$$0 < \zeta_2 = 1 + a\,\zeta_1 = 1 + \frac{a}{(1 + a^n)}\left[1 - a\frac{(1 - a^{n-1})}{(1 - a)})\right] \iff 1 + a^n > 2a^2.$$

The last condition is certainly met by all $a \leq 1/\sqrt{2} < 1$. To ensure the first condition $\zeta_1 < 0$ we substitute $a = \frac{1}{2}(1 + \Delta a)$ for some $\Delta a \in (0, \sqrt{2} - 1)$. Clearly, the first condition is monotonic in a and Δa so that, if it holds for the particular $\Delta a = 2^{1-n}$, it must also hold for all greater values of that problem parameter. Now we obtain after some elementary manipulations

$$\sum_{i=0}^{n-1} \left[\frac{1}{2}(1 + \Delta a)\right]^i > 2 \iff \frac{1 - \left[\frac{1}{2}(1 + \Delta a)\right]^n}{1 - \frac{1}{2}(1 + \Delta a)} > 2 \quad \iff \Delta a < 2\sqrt[n]{\Delta a} - 1.$$

The only thing that remains to be shown is that the last inequality holds for $\Delta a \equiv 2^{1-n}$. For $n = 3$ this is easily verified by direct calculation. For all $n \geq 4$ we obtain the condition

$$2\sqrt[n]{\Delta a} - 1 \;=\; 2^{1/n} - 1 \;\geq\; 1/(2n)$$

Here, the last inequality holds for $n \geq 2$ since the function $2^{1/n} - 1 - 1/(2n)$ of n is positive for $n = 2$ and one can easily check by differentiation that it grows monotonically beyond. Now all that remains to be shown is that $2^{n-2} \;<\; 1/n$, which one can check quite easily to be indeed satisfied for all $n > 3$. This completes the proof. □

The proposition demonstrates that at least without additional structural information on S we cannot deduce the convergence of full step generalized Newton when $\rho(|S|) \in [1/2 + 1/2^n, \sqrt{2}]$. However, the following divide and conquer approach still looks promising.

Discussion of variants

If we only have $\rho(|S|) = \|S\|_\infty < 1$, it is still clear that there is a unique solution. We may then start as above by guessing that the sign of the largest $\hat{c}_i$ component determines that of the corresponding z_i and eliminate it tentatively from the system. The reduced system will then have a Schur complement $\tilde{S}$ with $\rho(|\tilde{S}|) \leq \rho(|S|) < 1$. Mostly we can expect a reduction in ρ and may then perform a reequilibration by the Perron-Frobenius vector. Descending in a depth first fashion we will eventually reach a scalar problem that can be easily solved. Handing the reduced solutions back up, we will find that the resulting value for the variable whose sign was guessed is either right, in which case we can proceed further up the decision tree, or wrong, in which case we have to switch the sign and descent for a second time. If, at first, we pick the wrong sign every time, all 2^s possibilities must be tried. This is of course highly unlikely, especially when the top level ρ is already significantly smaller than 1. Naturally, we never need to explicitly equilibrate the system matrices, but just update the Perron Frobenius eigenvector d, which then allows the determination of the proper next pivot by maximizing $d_i\hat{c}_i$. Alternatively one may try piecewise Newton, which also might need 2^s tries as far as we can see.

Because our fixed point iteration and the modulus method normally yield only linear convergence, it becomes immediately clear that it does not reduce to semi-smooth Newton. Under the assumption of smooth dominance the local convergence result of Qi et all applies and we must have finite convergence on PL problems whenever convergence occurs at all. Of course, evaluating $\hat{H}(z)$ is a lot cheaper than solving a system in the Jacobian $J_\sigma = J + Y\Sigma(I - L\Sigma)^{-1}Z$ with $\sigma = \sigma(x)$ and thus $\Sigma = \Sigma(x)$, changing from iterate to iterate. While the iteration function G is Lipschitzian, the not always unique generalized Newton steps $-J_{\sigma(x)}^{-1}F(x)$ may jump discontinuously as a function of x. Nevertheless, it might be worthwhile to switch to Newton once the signature vector σ has been stable for a few iterations. It is not yet clear whether Newton converges globally under the assumption of smooth dominance.

It is not too hard to see that (at least when full steps are taken) the generalized Newton iteration on $H(z)$ is equivalent to that applied to the partitioned equation (2) for fixed y. The key numerical effort is solving a linear system in $I - S\,\Sigma$, which is also the key effort in applying the inverse Jacobians J_σ^{-1} to any vector. In either case we first need to form the Schur complement S, which, at least formally, involves the inverse of the smooth part J. If the number s of switching variables is much smaller than n, the number of independents, we can of course compute $J^{-1}Y$ or ZJ^{-1} by solving s linear systems in J, possibly based on its LU factorization.

When $H(z)$ is injective the fibres (14) have no bifurcations at all, so tracing them in a piecewise Newton fashion seems a very promising approach. Naturally, the number of steps is not a priori bounded in any way. To see that this is not equivalent to applying piecewise Newton to the original system $F(x) = 0$ we note that in the latter case, until the final step, there will always be a nontrivial residual on the lower equation of (2), whereas the upper block will be exactly satisfied. Conversely, applying piecewise Newton to $H(z) = \hat{c}$ means that there will be a residual in the upper block but the lower equation will remain exactly satisfied. Of course, one could also try a mixture just starting from $(x, z) = (0, 0)$ so that all subsequent residuals would be multiples of (c, b). The advantages and disadvantaged of these approaches deserve to be explored in detail.

Reduction to an LCP

Decomposing $z = u - w$ with $u \perp w$ in that $u \geq 0 \leq w$ and $u^\top w = 0$, we obtain $|z| = u + w$. Substituting this into our basic equation for fixed y we obtain

$$\begin{bmatrix} u - w \\ y \end{bmatrix} = \begin{bmatrix} c \\ b \end{bmatrix} + \begin{bmatrix} Z & L \\ J & Y \end{bmatrix} \begin{bmatrix} x \\ u + w \end{bmatrix} \quad \text{with} \quad 0 \leq u \perp w \geq 0 \tag{22}$$

Assuming again that the smooth part J is nonsingular we can eliminate x and obtain with S the Schur complement as above and $\hat{c} \equiv c + YJ^{-1}(y - b)$

$$u - w \;=\; \hat{c} + S(u + w) \quad \text{with} \quad 0 \leq u \perp w \geq 0$$

Assuming furthermore that $I - S$ is nonsingular, which is certainly implied by smooth dominance, we may solve for u and obtain

$$0 \leq u \;\equiv\; q + M\,w \;\;\perp\;\; w \geq 0 \tag{23}$$

where

$$q \;\equiv (I-S)^{-1}\hat{c} \quad \text{and} \quad M \;\equiv\; (I-S)^{-1}(I+S) \tag{24}$$

This is a linear complementarity problem in standard form. Of course, in this transformation some sparsity and structure of the original piecewise equation may be lost. Nevertheless, we should keep in mind that, when the smooth Jacobian J is invertible and the Schur complement S does not have the eigenvalue 1, then we are essentially solving a complementarity problem in s variables. If $S - I$ but not $S + I$ is singular we can exchange the roles of v and w to get essentially the same reduction with M the inverse of its definition above.

Rather than eliminating the vector x we could also split it into complementary positive and negative parts. However, especially since J can always be made nonsingular using (10) essentially doubling x would seem to introduce artificial combinatorial complexity. Since every solution of our unfolded equation $H(z) = \hat{c}$ corresponds to a solution of the LCP, the latter can be uniquely solved for any vector q if we have smooth dominance. It is well known [CPS92] that this is true if and only if M is a P-matrix. On the other hand, Rump has shown that $\rho_0^s(S) < 1$ is equivalent to M being a P-matrix, which agrees with our bijectivity result for simply switched coherently oriented systems.

8. Summary and Outlook

In this paper we have examined the properties of piecewise linear functions that are given in abs-normal form. Such a representation is always possible, but by no means unique. A key quantity is the switching depth ν, which we conjecture to be reducible to the bound $\bar{\nu}(n) = 2\,n - 1$. Of particular importance is the case of $\nu = 1$, where we call F simply switched. If such a representation exists, it is shown here that openness and bijectivity coincide provided LIKQ or the slightly weaker nondegeneracy condition of stable coherent orientation is satisfied.

The Schur complement matrix $S = L - ZJ^{-1}Y$, whose existence depends on the nonsingularity of the smooth part J, plays a central role throughout. In particular it yields the unfolded system $H(z) = [I - S\Sigma]z = \hat{c}$. This piecewise linear function $H(z)$ is simply switched and satisfies the LIKQ condition. Hence it is according to Proposition 4.1 injective if and only if it is coherently oriented, which in turn is equivalent to the the signed real spectral radius of S being less than 1. In principle this can be tested, though the evaluation of the continuous function $\rho_0^s(S)$ is generally NP hard as shown in [Rum97]. Since injectivity of $H(z)$ implies injectivity of the underlying $F(x)$ the condition $\rho_0^s(S) < 1$ is also sufficient for injectivity of $F(x)$. However, we have as yet no practical criterion for $F(x)$ to be merely open other than the theoretical possibility of exhaustively checking all Jacobians of F. Such combinatorial procedures have otherwise been avoidable throughout, thanks to the representation of F in abs-normal form. The key properties form the following chain of implications

$$\begin{array}{ccccc} \text{Absolute Contractivity} & \Longrightarrow & \text{Smooth Dominance} & \Longrightarrow & \text{Bijectvity of } H \\ \rho(|S|) < 1 & & \|DSD^{-1}\|_p < 1 & & \rho_0^s(S) < 1 \end{array}$$

So far our Linear Independence Kink Qualification (LIKQ) has only been defined in the simply switched case and it is then equivalent to the familiar linear independence constraint qualification (LICQ). However, we believe there is a generalization to general PL problems, where the kinks do not even locally consist of a set of intersecting hyperplanes, as is often envisioned. Instead there is a hierarchy of kinks with the later ones being broken into affine pieces by the earlier ones. The algorithmic handling of this structure is still not entirely clear, even in the context of minimizing a scalar valued PL function.

In order to constructively solve PL systems of equations one may apply full-step or piecewise Newton to either the original problem $F(x) = 0$ or the unfolded version $H(z) = \hat{c}$. They are guaranteed to converge if S does not deviate too much from L, which ensures at least coherent orientation. More specifically, we obtain finite convergence of generalized Newton on $H(z) = \hat{c}$ when $\|S\|_p < 1/3$ or $\rho(|S|) \leq 1/2$. The second bound is quite sharp in that divergence can occur as soon as $\rho(|S|) \geq 1/2 + 1/2^n$, as demonstrated in Proposition 7.5.

Apart from these four variants one may also apply damped versions or the fixed point iteration $z^+ = \hat{c} + S|z|$, provided one has smooth dominance, i.e., $\|S\|_p < 1$ for some $p \geq 1$, which is stronger than coherent orientation of H and thus injectivity of F, H. Piecewise smooth problems can be solved by successive piecewise linearization, yielding at least locally quadratic convergence. In this context coherent orientation of the piecewise linear model near the current outer iterate should be sufficient.

Abbreviating $\hat{\rho} = \|J^{-1}Y\|_p \|Z\|_p$ we may compile the table of solvers listed in Table 1. The effort column shows, which linear systems need to be solved, usually once per iteration. In the signed Gaussian elimination the equivalent of just one single solve is needed.

Method	Convergence condition	Rate	Effort
Generalized Newton on OPL	$2\,\hat{\rho} < (1-\|L\|_p-\hat{\rho}/2)^2$	finite	$I-S\Sigma, J$
Generalized Newton on UPL	$\|S\|_p < 1/3$	finite	$I-S\Sigma$
Signed Gauss on UPL	$\rho(\|S\|) \leq 1/2$	finite	$I-S\Sigma$ once
Fixed point Iteration on UPL	$\|S-L\|_p + \|L\|_p < 1$	linear	$I-L\Sigma, J$
Modulus Iteration on UPL	$\|S\|_p < 1$	linear	J
Piecewise Newton on OPL	coherent orient. of F	finite	$I-S\Sigma, J$
Piecewise Newton on UPL	$\rho_0^s(S) < 1$	finite	$I-S\Sigma$

Table 1: Solvers for PL systems of equations in original abs-normal or unfolded form.

Another theoretical possibility is piecewise Newton on the combined system in terms of x and z. A more promising approach would appear to be the combination of the fixed point iterations with Newton variants, which should yield finite convergence if one can get into the vicinity of a root. Without coherent orientation the fibres $\{F(x) = \lambda F(x_0) : \lambda > 0\}$ and also $\{H(z) - \hat{c} = \lambda(H(z_0) - \hat{c}) : \lambda > 0\}$ may contain turning points, which could be followed by some version of Branin's method [Bra72] originally defined by

$$\dot{x} \;=\; \pm\mathbf{adj}(F'(x))F(x) \quad \text{with} \quad \det(F'(x))I \;=\; F'(x)\,\mathbf{adj}(F'(x))$$

In the general smooth case such trajectories may converge to roots, cycle or run off to infinity. Possibly the inherent finiteness of PL functions makes it possible to avoid some of these calamities. Other globalized searches remain to be investigated. Since any Lipschitzian vector function may be approximated on compact domains by PL functions, there can be no magic solver for the general case. Numerical experiments with the various methods considered here are currently under way.

Acknowledgements

The proof of Proposition 4.2 and Lemma 4.3 was thankfully provided by our colleague Dorothee Schüth of Humboldt University. The authors are also indebted to Daniel Kressner, who pointed out the connection between the coherence condition $\det(I - \Sigma\, S) > 0$ and the sign real spectral radius $\rho_0^s(S)$ of the Schur complement S being less than 1. They are also grateful to Torsten Bosse, who contributed many insights into the piecewise linearization approach and greatly helped with the composition of this article.

References

[BC08] Luigi Brugnano and Vincenzo Casulli. Iterative solution of piecewise linear systems. *SIAM Journal on Scientific Computing*, 30(1):463–472, 2008.

[Bra72] Franklin H Branin. Widely convergent method for finding multiple solutions of simultaneous nonlinear equations. *IBM Journal of Research and Development*, 16(5):504–522, 1972.

[CPS92] Richard W. Cottle, Jong-Shi Pang, and Richard E. Stone. *The linear complementarity problem.* Computer Science and Scientific Computing. Boston, MA etc.: Academic Press, Inc.. xxiv, 762 p. , 1992.

[DER86] Iain S Duff, Albert Maurice Erisman, and John Ker Reid. *Direct methods for sparse matrices.* Clarendon Press Oxford, 1986.

[FP03] Francisco Facchinei and Jong-Shi Pang. *Finite-dimensional variational inequalities and complementarity problems*, volume 1. Springer, 2003.

[Gri13] Andreas Griewank. On stable piecewise linearization and generalized algorithmic differentiation. *Optimization Methods and Software*, 28(6):1139–1178, 2013.

[GW08] Andreas Griewank and Andrea Walther. *Evaluating derivatives: principles and techniques of algorithmic differentiation.* Siam, 2008.

[KB12] Kamil A. Khan and Paul I. Barton. Evaluating an element of the clarke generalized jacobian of a piecewise differentiable function. In *Recent Advances in Algorithmic Differentiation*, volume 87 of *Lecture Notes in Computational Science and Engineering*, pages 115–125. Springer Berlin Heidelberg, 2012.

[KB13] Kamil A. Khan and Paul I. Barton. Evaluating an element of the Clarke generalized Jacobian of a composite piecewise differentiable function. *ACM Transactions on Mathematical Software (TOMS)*, 39(4):23:1–23:28, July 2013.

[MD85] Golubitsky M. and Schaeffer D.G. *Singularities and Groups in Bifurcation Theory, volume 1.* Springer, New York., 1985.

[Nes05] Yu Nesterov. Lexicographic differentiation of nonsmooth functions. *Mathematical programming*, 104(2-3):669–700, 2005.

[PGar] Julia Piontkowski and Andreas Griewank. Solution of complementarity problems via piecewise linearization. *PAMM*, 2013 to appear.

[QS93] L. Qi and J. Sun. A nonsmooth version of Newton's method. 58(3):353–368, 1993.

[Rum97] Siegfried M Rump. Theorems of perron-frobenius type for matrices without sign restrictions. *Linear Algebra and Its Applications*, 266:1–42, 1997.

[Ruz04] M. Ruzicka. *Nichtlineare Funktionalanalysis. Eine Einfhrung.* Springer, Berlin, 2004.

[Sch12] Stefan Scholtes. *Introduction to piecewise differentiable equations.* Springer, 2012.

[Ste56] Thomas Edwin Stern. *Piecewise-linear network theory.* Tech. Rep. 315. Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, Mass., 1956.

[vB81] Wilhelmus Maria Gezinus van Bokhoven. *Piecewise-linear modelling and analysis.* Kluwer Technische Boeken, 1981.

[VBL99] Wim MG Van Bokhoven and Domine MW Leenaerts. Explicit formulas for the solutions of piecewise linear networks. *Circuits and Systems I: Fundamental Theory and Applications, IEEE Transactions on*, 46(9):1110–1117, 1999.