Treating Non-Smooth Functions as Smooth Functions in Global Optimization and Nonlinear Systems Solvers

R. Baker Kearfott

0 Introduction

Techniques of interval extensions and interval Newton methods have been developed for verified solution of nonlinear systems of equations and for global optimization. In most of the literature to date, such interval extensions and interval Newton methods are applicable when the functions are given by smooth expressions, without conditional branches. In fact, however, many practical problems, in particular those containing expressions such as $|E(X)|$ and $\max\{E(X), F(X)\}$, $E, F : \mathbb{R}^n \rightarrow \mathbb{R}$, or expressions defined by `IF-THEN-ELSE` branches, result in functions whose derivatives have jump discontinuities. However, in [4], continuous, order-1 interval extensions were proposed for such continuous but non-smooth functions.

Such interval extensions of non-smooth functions can be used in the same contexts as other interval extensions. However, the width of a non-smooth derivative extension $\boldsymbol{F}'(\boldsymbol{X})$ or slope extension of a function $F : \mathbb{R}^n \rightarrow \mathbb{R}^n$ over an interval vector $\boldsymbol{X}$ does not, in general, tend to zero as the width vector $\mathrm{w}(\boldsymbol{X})$ tends to zero. Because of this, interval Newton iterations are only convergent in certain cases. However, ϵ-inflation algorithms can always be devised to prove existence or uniqueness for $\mathrm{w}(\boldsymbol{X})$ sufficiently small, provided merely that $\boldsymbol{F}'(\boldsymbol{X})$ is extended appropriately.

Special algorithms can be developed to handle non-smooth problems such as l_1 optimization and l_∞ optimization. However, simplicity is a major advantage of treating such non-smooth problems with the same techniques as smooth problems. This simpler, unified treatment is possible within the context described in this paper.

First, required properties of interval extensions for non-smooth functions are discussed. Then, selected formulas from [4] are presented. The formulas for slopes presented here represent improvements (sharper bounds) over those in [4]. A simple, illustrative example is then given. Fourth, a convergence and existence / uniqueness verification theory for interval Newton methods using non-smooth extensions is developed. Finally, formulas and examples are given for which interval Newton methods can find critical points, even if the *gradient itself is discontinuous*.

1 Extensions of Non-Smooth Functions

1.1 Required Properties

- An interval extension $\boldsymbol{F}(\boldsymbol{X})$ of F must *contain the range*, i.e.

$$\{F(X) \mid X \in \boldsymbol{X}\} \subseteq \boldsymbol{F}(\boldsymbol{X}).$$

- Interval extensions $\boldsymbol{F}'(\boldsymbol{X})$ of the Jacobi matrix must be *Lipschitz sets* in the sense that

$$F(X_2) - F(X_1) = A(X_2) - A(X_1)$$

  for some $A \in \boldsymbol{F}'(\boldsymbol{X})$, for every $X_1, X_2 \in \boldsymbol{X}$ [6].

- Slope extensions $\mathbf{S}(F, \boldsymbol{X}, \check{\boldsymbol{X}})$ must be *slope sets* in the sense that

$$F(X) - F(\check{X}) = A(X - \check{X})$$

  for some $A \in \mathbf{S}(F, \boldsymbol{X}, \check{\boldsymbol{X}})$, for every $X \in \boldsymbol{X}$ and $\check{X} \in \check{\boldsymbol{X}}$ [6].

Such interval extensions should be as *sharp as possible* (i.e. should overestimate actual ranges as little as possible), subject to the above conditions, and subject to the condition that they can be computed automatically. The formulas in the next section give such extensions.

1.2 Some Example Formulas

Selected formulas satisfying the above conditions appear here. More complete listings of formulas appear in [4], and will appear with additional explanation in [5].

1.2.1 `IF-THEN-ELSE`

Branches can be handled in operator overloading and automatic differentiation with the function

$$\chi(x_s, x_q, x_r) = \begin{cases} x_q & \text{if } x_s < 0, \\ x_r & \text{otherwise.} \end{cases}$$

For example, if

$$f(x) = \begin{cases} x^2 & \text{if } x < 1 \\ 2x - 1 & \text{if } x \geq 1, \end{cases} \tag{1}$$

then an internal representation for f may be generated within our `INTLIB_90` system [3], with the following program.

```
PROGRAM CHI_EXAMPLE
  USE OVERLOAD
      TYPE(CDLVAR), DIMENSION(1):: X
      TYPE(CDLLHS), DIMENSION(1):: F
      OUTPUT_FILE_NAME='CHI_EXAMPLE.CDL'
      CALL INITIALIZE_CODELIST(X)
  F(1) = CHI(X(1)-1,X(1)**2,2*X(1)-1)
      CALL FINISH_CODELIST
END PROGRAM CHI_EXAMPLE
```

Once the internal representation is generated, the automatic differentiation process introduced in [3] and explained in [5] can make use of the following formulas.

- Interval extension:

$$\chi(\boldsymbol{x}_s, \boldsymbol{x}_q, \boldsymbol{x}_r) = \begin{cases} \boldsymbol{x}_q & \text{if } \boldsymbol{x}_s < 0; \\ \boldsymbol{x}_r & \text{if } \boldsymbol{x}_s > 0; \\ \boldsymbol{x}_q \underline{\cup} \boldsymbol{x}_r & \text{otherwise,} \end{cases} \tag{2}$$

  where $\boldsymbol{x} \underline{\cup} \boldsymbol{y}$ is the interval hull of the intervals $\boldsymbol{x}$ and $\boldsymbol{y}$.

- Derivative extension when $\chi(0^-, x_q, x_r) = \chi(0^+, x_q, x_r)$, i.e. when $f(X) = \chi(E(X), F(X), G(X))$ is continuous:

$$\begin{aligned} \frac{\partial \chi(\boldsymbol{x}_s,\boldsymbol{x}_q,\boldsymbol{x}_r)}{\partial x_q} &= \begin{cases} 1 & \text{if } \boldsymbol{x}_s < 0; \\ 0 & \text{if } \boldsymbol{x}_s > 0; \\ [0,1] & \text{otherwise.} \end{cases} \\ \frac{\partial \chi(\boldsymbol{x}_s,\boldsymbol{x}_q,\boldsymbol{x}_r)}{\partial x_r} &= \begin{cases} 0 & \text{if } \boldsymbol{x}_s < 0; \\ 1 & \text{if } \boldsymbol{x}_s > 0; \\ [0,1] & \text{otherwise.} \end{cases} \\ \frac{\partial \chi(\boldsymbol{x}_s,\boldsymbol{x}_q,\boldsymbol{x}_r)}{\partial x_s} &= 0 \end{aligned} \tag{3}$$

Within the automatic differentiation process, these formulas lead to enclosures of the range and to Lipschitz sets.

1.2.2 Absolute Value $x_p = |x_q|$

Although $|x| = \chi(x, -x, x)$, $\chi(\boldsymbol{x}, -\boldsymbol{x}, \boldsymbol{x})$ overestimates the range of $|\circ|$ over the interval $\boldsymbol{x}$. For example, the range of $|\circ|$ over $[-1, 2]$ is $[0, 2]$, whereas $\chi([-1, 2], [-2, 1], [-1, 2]) = [-2, 2]$. The following formulas are sharper.

- Interval extension:

$$|\boldsymbol{x}| = \begin{cases} [0, \max\{|\underline{x}|, |\overline{x}|\}] & \text{if } 0 \in \boldsymbol{x}; \\ [\min\{|\underline{x}|, |\overline{x}|\}, \max\{|\underline{x}|, |\overline{x}|\}] & \text{otherwise.} \end{cases} \tag{4}$$

- Derivative enclosure:

$$\frac{d|\boldsymbol{x}_q|}{dx_q} = \begin{cases} -1 & \text{if } \boldsymbol{x}_q < 0; \\ 1 & \text{if } \boldsymbol{x}_q > 0; \\ [-1,1] & \text{otherwise.} \end{cases} \tag{5}$$

- Slope enclosure:

$$\mathbf{S}(|\boldsymbol{x}_q|), \boldsymbol{X}, \check{\boldsymbol{X}}) = \begin{cases} -\mathbf{S}(x_q, \boldsymbol{X}, \check{\boldsymbol{X}}) & \text{if } \boldsymbol{x}_q \underline{\cup} \check{\boldsymbol{x}}_q < 0; \\ \mathbf{S}(x_q, \boldsymbol{X}, \check{\boldsymbol{X}}) & \text{if } \boldsymbol{x}_q \underline{\cup} \check{\boldsymbol{x}}_q > 0; \\ \mathbf{S}^{(d)}(|x_q|, \boldsymbol{x}_q, \check{\boldsymbol{x}}_q)\mathbf{S}(x_q, \boldsymbol{X}, \check{\boldsymbol{X}}) & \text{otherwise,} \end{cases} \tag{6}$$

where

$$\mathbf{S}^{(d)}(|x|, \boldsymbol{x}, \check{\boldsymbol{x}}) = h(\underline{x}) \underline{\cup} h(\overline{x}) \quad \text{with} \quad h(x) = \begin{cases} \frac{|x| - |\check{\boldsymbol{x}}|}{x - \check{\boldsymbol{x}}} & \text{for } x \notin \check{\boldsymbol{x}}; \\ [-1,1] & \text{otherwise.} \end{cases} \tag{7}$$

Formula (7) is derived from formulas explained by Rump in [8, 9]. In particular, it is valid to compute slopes by taking differences at end points, whenever the function, such as $|\cdot|$, is convex.

1.2.3 $x_p = \max\{x_q, x_r\}$

- Interval extension:

$$\max\{\boldsymbol{x}_q, \boldsymbol{x}_r\} = [\max\{\underline{x}_q, \underline{x}_r\}, \max\{\overline{x}_q, \overline{x}_r\}] \tag{8}$$

- Derivative enclosure:

$$\begin{aligned} \frac{\partial \max\{\boldsymbol{x}_q, \boldsymbol{x}_r\}}{\partial x_q} &= \begin{cases} 1 & \text{if } \boldsymbol{x}_q > \boldsymbol{x}_r; \\ 0 & \text{if } \boldsymbol{x}_q < \boldsymbol{x}_r; \\ [0,1] & \text{otherwise.} \end{cases} \\ \frac{\partial \max\{\boldsymbol{x}_q, \boldsymbol{x}_r\}}{\partial x_r} &= \begin{cases} 0 & \text{if } \boldsymbol{x}_q > \boldsymbol{x}_r; \\ 1 & \text{if } \boldsymbol{x}_q < \boldsymbol{x}_r; \\ [0,1] & \text{otherwise.} \end{cases} \end{aligned} \tag{9}$$

For additional formulas, see [4, 5].

2 Example – Univariate Interval Newton Method

Suppose roots of the function

$$f(x) = |x^2 - x| - 2x + 2 = 0$$

are sought. The function f has a root and a cusp at 1: $f(1) = 0$, but

$$\lim_{x\to 1^-} f'(x) = -3, \qquad \lim_{x\to 1^+} f'(x) = -1.$$

For illustration purposes, a slope enclosure when $1 \in \boldsymbol{x}$ may be given by $\mathbf{S}(f, \boldsymbol{x}, x) = [-1, 1](\boldsymbol{x} + x - 1) - 2$. (This enclosure is cruder, but simpler, than that given above by $\mathbf{S}^{(d)}$.) Apply the interval Newton method

$$\tilde{\boldsymbol{x}} \leftarrow \check{x}^{(k)} - f(\check{x}^{(k)})/\mathbf{S}(f, \boldsymbol{x}^{(k)}, \check{x}^{(k)}), \qquad \boldsymbol{x}^{(k+1)} \leftarrow \boldsymbol{x}^{(k)} \cap \tilde{\boldsymbol{x}},$$

with $\check{x}^{(k)}$ equal to the midpoint of $\boldsymbol{x}^{(k)}$, and $\boldsymbol{x}^{(0)} = [0.7, 1.1]$. The initial slope enclosure is $\mathbf{S}(f, [0.7, 1.1], 0.9) = [-3, -1]$, $\tilde{\boldsymbol{x}} = .9 - .29/[-3, -1] = [.99\overline{6}, 1.19]$, and $\boldsymbol{x}^{(1)} = [0.99\overline{6}, 1.1]$. Continuing iteration leads to the table:

k	$\boldsymbol{x}^{(k)}$	$\mathrm{w}(\boldsymbol{x}^{(k)})$	$\frac{\mathrm{w}(\boldsymbol{x}^{(k)})}{\mathrm{w}(\boldsymbol{x}^{(k-1)})}$
0	[0.699999, 1.1000]	4.0E-1	—
1	[0.996666, 1.1000]	1.0E-1	0.25
2	[0.996666, 1.0337]	3.7E-2	0.37
3	[0.999461, 1.0102]	1.1E-2	0.28
4	[0.999949, 1.0032]	3.3E-3	0.30
5	[0.999994, 1.0010]	1.1E-3	0.33
6	[0.999999, 1.0003]	3.6E-4	0.33

(Here, $\mathrm{w}(\boldsymbol{x})$ represents the width of an interval $\boldsymbol{x}$.) On iteration 3, existence of a root within $\boldsymbol{x}^{(3)}$ is proven, since $\boldsymbol{x}^{(3)}$ is in the interior of $\boldsymbol{x}^{(2)}$. The observed convergence is linear, with a convergence factor of about 0.33. (Somewhat faster convergence would be observed if $\mathbf{S}^{(d)}$ were used instead.) In any case, convergence of the interval Newton method, requiring derivative information, occurs, even though the derivative is discontinuous. Also, programming of the optimization problem requires, besides variable declarations and the program template, only the statement:

```
F(1) = ABS(X(1) ∗ ∗2 − X(1)) − 2 ∗ X(1) + 2
```

Preliminary experiments in n-space are reported in [4].

3 Convergence Theory

The consequences of using the above formulas in nonlinear equations and global optimization algorithms are briefly discussed here.

3.1 Objective Function Approximation

Interval extensions involving χ, $|\circ|$, and max, computed with formulas as in §1.2 and in [4, 5], are first-order extensions in the sense of [7, p. 38]. Thus, these interval extensions may be used to prove that no roots exist within a box $\boldsymbol{X}$. Similarly, since the formulas for the derivative extensions, if they lead to finite intervals, lead to Lipschitz matrices or slope matrices, an image of $\boldsymbol{X}$ under any interval Newton method that employs these extensions must contain all roots within $\boldsymbol{X}$; hence, the interval Newton method can also be used to prove that no roots exist within $\boldsymbol{X}$.

However, the mean value extension based on the above does not, in general, lead to a second-order interval extension. For an example, see [5].

3.2 Verification and Interval Convergence

Convergence of interval Newton methods occurs, and existence or uniqueness verification is possible in many cases with the non-smooth extensions described above. Conditions under which convergence occurs or under which verification is possible are examined in this section.

3.2.1 Interval Newton Methods – General Convergence

We consider the general multivariate interval Newton method:

$$\tilde{\boldsymbol{X}} = \boldsymbol{N}(F; \boldsymbol{X}, \check{X}) = \check{X} + \boldsymbol{V},$$

where $\boldsymbol{X}$, $\boldsymbol{V}$, and $\tilde{\boldsymbol{X}}$ are boxes (i.e. interval vectors) in $\mathbb{R}^n$, where

$$\Sigma\big(\boldsymbol{A}, -F(\check{X})\big) \subset \boldsymbol{V},$$

where $\boldsymbol{A}$ is either a Lipschitz matrix or a slope enclosure $\mathbf{S}(F, \boldsymbol{X}, \check{\boldsymbol{X}})$ for F over $\boldsymbol{X}$ and centered at $\check{\boldsymbol{X}}$, and where $\Sigma\big(\boldsymbol{A}, -F(\check{X})\big)$ is the exact solution set of the interval linear system $\boldsymbol{A}X = -F(\check{X})$.

Convergence analysis of

$$\boldsymbol{X} \leftarrow \boldsymbol{N}(F; \boldsymbol{X}, \check{X}),$$

as well as analysis of existence/uniqueness verification with

$$\boldsymbol{N}(F; \boldsymbol{X}, \check{X}) \subset \mathrm{int}(\boldsymbol{X}),$$

may proceed by bounding the width norm:

$$\left\| \mathrm{w}\big(\boldsymbol{N}(F; \boldsymbol{X}, \check{X})\big) \right\|.$$

Define

$$\Sigma_0 = \Sigma\big(\boldsymbol{A}, -F(\check{X})\big) \qquad \text{and} \qquad \square\Sigma_0 = \square\Sigma\big(\boldsymbol{A}, -F(\check{X})\big), \tag{10}$$

where $\square\Sigma$ is the smallest interval vector containing Σ. First note that the width norms obey

$$\|\mathrm{w}\big(\Sigma_0\big)\| = \|\mathrm{w}\big(\square\Sigma_0\big)\|, \tag{11}$$

from the definition of $\square\Sigma_0$ and since $\|\cdot\| = \|\cdot\|_\infty$. Also write

$$\boldsymbol{V} = \square\Sigma_0 + \boldsymbol{E}, \quad \text{so} \quad \mathrm{w}\big(\boldsymbol{N}(F;\boldsymbol{X},\check{X})\big) = \mathrm{w}(\boldsymbol{V}) = \mathrm{w}\big(\square\Sigma_0\big) + \mathrm{w}(\boldsymbol{E}). \tag{12}$$

The following assumption is non-restrictive.

Assumption 1 *There exists a K, depending only on the particular interval Newton method, such that*

$$\|\mathrm{w}(\boldsymbol{E})\| = \|\mathrm{w}(\boldsymbol{V}) - \mathrm{w}\big(\Sigma\big)\| \le K\|\mathrm{w}\big(\Sigma\big)\|. \tag{13}$$

In fact, for most methods of bounding Σ, Assumption 1 follows from (12) and considerations as in [6, §4.2, p. 124 and §4.3.5, p. 138]. Thus, $\|\mathrm{w}\big(\boldsymbol{N}(F;\boldsymbol{X},\check{X})\big)\|$ can be bounded by bounding $\|\mathrm{w}(\Sigma_0)\|$. To bound $\|\mathrm{w}(\Sigma_0)\|$, a sensitivity bound for real linear systems will be used, namely

Theorem 1 (A modification of Theorem 2.7.2 in [2]) *Assume*

$$\begin{aligned} AX &= B, & A \in \mathbb{R}^{n\times n}, B \in \mathbb{R}^n, \text{ and} \\ (A+\Delta A)(X+\Delta X) &= B, & \Delta A \in \mathbb{R}^{n\times n}, \end{aligned}$$

assume $\|\Delta A\| \le \delta\|A\|$, and assume $\delta\kappa(A) = r < 1$, where $\kappa(A)$ is the condition number $\|A\|\cdot\|A^{-1}\|$. Then

$$\begin{aligned} \|\Delta X\| &\le r\frac{1}{1-r}\|X\| \\ &\le r\frac{1}{1-r}\|A^{-1}\|\cdot\|B\|. \end{aligned}$$

To use Theorem 1, identify:

$$B = -F(\check{X}) \text{ and } A = \check{A} = \mathrm{m}(\boldsymbol{A}), \text{ so } \Delta A = \frac{\mathrm{w}(\boldsymbol{A})}{2},$$

(where $\mathrm{m}(\boldsymbol{A})$ is the midpoint matrix of an interval matrix $\boldsymbol{A}$) so that

$$\delta = \frac{1}{2}\frac{\|\mathrm{w}(\boldsymbol{A})\|}{\|\check{A}\|} \text{ and } r = \frac{1}{2}\|\mathrm{w}(\boldsymbol{A})\|\cdot\|\check{A}^{-1}\|. \tag{14}$$

Assume $X_1 \in \Sigma_0$ and $X_2 \in \Sigma_0$. Then (14) and Theorem 1 give

$$\|X_1 - X_2\| \le \|\mathrm{w}(\boldsymbol{A})\|\cdot\|-F(\check{X})\|\left\{\frac{\|\check{A}^{-1}\|}{2}\cdot\frac{1}{1-r}\right\}. \tag{15}$$

Combining (15) and (13) gives

Lemma 2 *Make Assumption 1. Then*

$$\left\|\mathrm{w}\big(\boldsymbol{N}(F;\boldsymbol{X},\check{X})\big)\right\| \le C\|\mathrm{w}(\boldsymbol{A})\|\cdot\|F(\check{X})\|,$$

where

$$C = K\left\{\frac{1}{2}\cdot\frac{\|\check{A}^{-1}\|}{1-\frac{1}{2}\|\mathrm{w}(\boldsymbol{A})\|\cdot\|\check{A}^{-1}\|}\right\}. \tag{16}$$

The next assumption is necessary for proof that verification is possible.

Assumption 2 *Assume there exists an $X^* \in \boldsymbol{X}$ with $F(X^*) = 0$. Also assume that $\check{X} = \mathrm{m}(\boldsymbol{X})$.*

Now, since $\boldsymbol{A}$ is either a Lipschitz matrix or a slope enclosure centered at $\check{X}$,

$$F(X^*) = 0 = F(\check{X}) + A(X^* - \check{X}) \quad \text{for some } A \in \boldsymbol{A}.$$

Thus,

$$\|F(\check{X})\| \le \|\boldsymbol{A}\|\frac{\|\mathrm{w}(\boldsymbol{X})\|}{2}. \tag{17}$$

Combining (17) and Lemma 2 thus gives

Theorem 3 *Make assumptions 1 and 2, and let C be as in (16) of Lemma 2. Then*

$$\left\|\mathrm{w}\big(\boldsymbol{N}(F;\boldsymbol{X},\check{X})\big)\right\| \le C\|\mathrm{w}(\boldsymbol{A})\|\cdot\|\boldsymbol{A}\|\frac{\|\mathrm{w}(\boldsymbol{X})\|}{2}.$$

Corollary 4 *Make assumptions 1 and 2, and let C be as in (16) of Lemma 2. Then*

$$\left\|\mathrm{w}\big(\boldsymbol{N}(F;\boldsymbol{X},\check{X})\big)\right\| \le \|\mathrm{w}(\boldsymbol{X})\|$$

provided

$$C\|\mathrm{w}(\boldsymbol{A})\|\cdot\|\boldsymbol{A}\| < 2 \quad \text{(for convergence)}. \tag{18}$$

Furthermore, if we assume $\check{X} \in \boldsymbol{N}(F;\boldsymbol{X},\check{X})$, such as when $F(\check{X}) \approx 0$ and ϵ-inflation is used, then $\boldsymbol{N}(F;\boldsymbol{X},\check{X}) \subset \boldsymbol{X}$, provided

$$C\|\mathrm{w}(\boldsymbol{A})\|\cdot\|\boldsymbol{A}\|\frac{\|\mathrm{w}(\boldsymbol{X})\|}{\min_{1\le i\le n}\mathrm{w}(\boldsymbol{x}_i)} < 1 \quad \text{(for inclusion)}. \tag{19}$$

For smooth problems and Lipschitz or slope matrices formed with the usual computations,

$$\|\mathrm{w}(\boldsymbol{A})\| \le K_{F'}\|\mathrm{w}(\boldsymbol{X})\|, \tag{20}$$

so Condition (18) (and the more restrictive Condition 19 for the inclusion $\boldsymbol{N}(F;\boldsymbol{X},\check{X}) \subset \boldsymbol{X}$) hold for sufficiently small $\boldsymbol{X}$ centered about an approximate solution $\check{X} \approx X^*$. Thus, neglecting roundout error, it should always be possible to verify existence or uniqueness with ϵ-inflation. In fact, the following is a direct consequence of Theorem 3.

Theorem 5 *Make assumptions 1 and 2, and let* C *be as in Lemma 2. Also assume* $\|\mathrm{w}(\boldsymbol{A})\| \le K_{F'}\|\mathrm{w}(\boldsymbol{X})\|$. *Then*

$$\big\|\mathrm{w}\big(\boldsymbol{N}(F;\boldsymbol{X},\check{X})\big)\big\| \le \tilde{C}\|\mathrm{w}(\boldsymbol{X})\|^2,$$

where $\tilde{C} = CK_{F'}\|\boldsymbol{A}\|/2$.

3.2.2 Convergence for Non-Smooth Problems

For non-smooth problems, $\mathrm{w}(\boldsymbol{A}) \not\to 0$ as $\mathrm{w}(\boldsymbol{X}) \to 0$ (i.e. Condition (20) does not hold), Condition 18 is not in general satisfied, and the interval Newton method does not converge. However, when the non-smoothness does not occur in dominant terms, Condition 18 is sometimes still satisfied for small $\mathrm{w}(\boldsymbol{X})$, and the interval Newton method converges linearly:

Theorem 6 *Make assumptions 1 and 2, and let* C *be as in Lemma 2. Also assume that (for small enough* $\mathrm{w}(\boldsymbol{X})$*)*

$$\|\mathrm{w}(\boldsymbol{A})\| \le \frac{2}{CM'_F}.$$

Then the width of $\boldsymbol{X}$ *under iteration of*

$$\boldsymbol{X} \leftarrow \boldsymbol{X} \cap \boldsymbol{N}(F;\boldsymbol{X},\check{X})$$

tends to zero linearly, with convergence factor $c = C\|\mathrm{w}(\boldsymbol{A})\| \cdot \|\boldsymbol{A}\|/2$. *That is,*

$$\big\|\mathrm{w}\big(\boldsymbol{N}(F;\boldsymbol{X},\check{X})\big)\big\| \le c\|\mathrm{w}(\boldsymbol{X})\|.$$

In some cases, Theorem 6 does not give reasonably sharp bounds, since the convergence factor c incorporates a worst-case bound $\|X^* - \check{X}\| \le \mathrm{w}(\boldsymbol{X})/2$. Actually, in ϵ-inflation algorithms, $\boldsymbol{X}$ can be *constructed* so

$$\big\|X^* - \check{X}\big\| \le \nu\|\mathrm{w}(\boldsymbol{X})\|^2 \tag{21}$$

For example, as explained in [5], X^* can be computed with an approximate solver with a stopping tolerance that is proportional to the square of the width of the tolerance of the box constructed around it. In such cases, (17) can be replaced by

$$\|F(\check{X})\| \le \|\boldsymbol{A}\|\nu\|\mathrm{w}(\boldsymbol{X})\|^2, \tag{22}$$

and Theorem 6 may be replaced by

Theorem 7 *Make assumptions 1 and 2, and let* C *be as in Lemma 2. Also assume (21) for some* $\nu > 0$. *Then, for small enough initial* $\mathrm{w}(\boldsymbol{X})$, *the width of* $\boldsymbol{X}$ *under iteration of*

$$\boldsymbol{X} \leftarrow \boldsymbol{X} \cap \boldsymbol{N}(F;\boldsymbol{X},\check{X})$$

tends to zero quadratically. In particular,

$$\left\|\mathrm{w}\big(\boldsymbol{N}(F;\boldsymbol{X},\check{X})\big)\right\| \le C\|\mathrm{w}(\boldsymbol{A})\|\cdot\|\boldsymbol{A}\|\nu\|\mathrm{w}(\boldsymbol{X})\|^2,$$

where C is as in (16).

Theorem 7 follows directly from Lemma 2.

Remark 1 *Theorem 7 implies that, even for non-smooth functions, existence and uniqueness can* always *be verified with ϵ-inflation, provided a sufficiently accurate approximate solution can be obtained.*

Now we will use Theorem 6 to analyze the convergence for the example of §2. The limiting value of the slope bound set as $\boldsymbol{x}\to[1,1]$, for $\check{\boldsymbol{x}}\subseteq\boldsymbol{x}$, is $\boldsymbol{A}=[-3,-1]$, and the limiting value of its center is $\check{A}=-2$. From this, we compute limiting values for C as defined in (16) and Assumption 1: Assumption 1 implies that, for an $\epsilon>0$ of our choosing and $\mathrm{w}\big(\square\Sigma_0\big)$ sufficiently small (i.e. $\mathrm{w}(\boldsymbol{X})$ sufficiently small), K can be taken to be $(1+\epsilon)$. Furthermore, $\mathrm{w}(\boldsymbol{A})\approx -1-(-3)=2$, and $\|\check{A}^{-1}\|\approx|1/(-2)|=\frac{1}{2}$, so

$$C\approx(1+\epsilon)\cdot\frac{1}{2}\cdot\frac{\frac{1}{2}}{1-\frac{1}{2}(2)\frac{1}{2}}\approx\frac{1}{2}.$$

Also, $\|\boldsymbol{A}\|\approx\max\{|-3|,|-1|\}=3$, and

$$c\approx\frac{1}{2}\cdot(2)\cdot\frac{3}{2}=\frac{3}{2}$$

in Theorem 6. Thus, Theorem 6 is not sharp enough in this case to predict inclusion. Also, we observe that $|\check{x}-x^*|$ is tending to zero at about the same rate as $\mathrm{w}(\boldsymbol{x})$, so Theorem 7 would not give an appropriate conclusion either. However, after iteration no. 2, we observe $|f(x^{(2)})|\le 0.015$ and $\mathrm{w}\big(x^{(2)}\big)\approx 0.04$. Plugging these values directly into Lemma 2 gives

$$\mathrm{w}\Big(\boldsymbol{x}^{(3)}\Big)\le\frac{1}{2}\cdot 2\cdot 0.015=0.015<0.04/2=.02,$$

thus proving that verification must occur on this step. (In fact, the new width $\mathrm{w}\big(\boldsymbol{x}^{(3)}\big)\approx 0.01$. The left endpoints of $\boldsymbol{x}^{(k)}$ are converging faster than the right endpoints.)

4 Slopes for Discontinuous Functions

In computations such as l_1 or l_∞ optimization, the objective function $\phi:\mathbb{R}^n\to\mathbb{R}$ typically is continuous but non-smooth, and consists of compositions of smooth operations and $|\cdot|$, max, or, more generally, χ. For example, if

$$\phi(x)=\max\{2-x^2,x^2\},\tag{23}$$

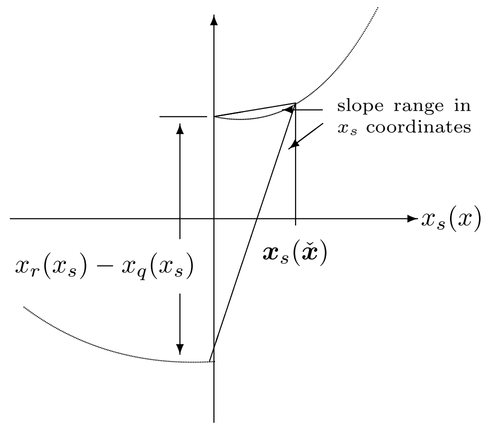

Figure 1: Computation of slope bounds of a discontinuous function

then ϕ has a minimum at $x = 1$. This minimum is at a cusp of ϕ, where the gradient

$$\nabla\phi(x) = \chi\big(x^2 - (2 - x^2), -2x, 2x\big) \tag{24}$$

has a jump discontinuity. Thus, the interval extension of the second derivative of ϕ (corresponding to the Hessian matrix) is $\mathbb{R}$, and an interval Newton method would not be a useful computation. On the other hand, the actual gradient (24), rather than, say, a slope that would not have the discontinuity, is necessary to detect the critical point. Fortunately, interval *slopes*[1] of functions such as (24) contain meaningful information.

Definition of the interval slopes of discontinuous functions is illustrated in Figure 1. There, the slope $\mathbf{S}(\chi(x_s(x), x_q(x), x_r(x)), \boldsymbol{x}, \check{\boldsymbol{x}})$ is shown in terms of x_s coordinates, that is, assuming $x_s = x$. It is seen that, if $0 \notin \check{\boldsymbol{x}}$, the slope bound is finite. Furthermore, the slope bounds will include the jump (a finite change over an infinitely small interval), and interval Newton methods will converge on critical points, be they zeros or breaks in the gradient.

Based on Figure 1, the following formula can be derived.

$$\begin{aligned}
&\mathbf{S}(\chi(x_s(x), x_q(x), x_r(x)), \boldsymbol{x}, \check{\boldsymbol{x}}) = \\
&\begin{cases}
\mathbf{S}(x_r(x), \boldsymbol{x}, \check{\boldsymbol{x}}) \underline{\cup} \left\{ \frac{1}{\boldsymbol{x}_s(\check{\boldsymbol{x}})} \big(\boldsymbol{x}_r(\boldsymbol{x}) - \boldsymbol{x}_q(\boldsymbol{x})\big) \mathbf{S}(x_s(x), \boldsymbol{x}, \check{\boldsymbol{x}}) \right\} & \text{if } \boldsymbol{x}_s(\check{\boldsymbol{x}}) > 0 \\
\mathbf{S}(x_q(x), \boldsymbol{x}, \check{\boldsymbol{x}}) \underline{\cup} \left\{ \frac{-1}{\boldsymbol{x}_s(\check{\boldsymbol{x}})} \big(\boldsymbol{x}_r(\boldsymbol{x}) - \boldsymbol{x}_q(\boldsymbol{x})\big) \mathbf{S}(x_s(x), \boldsymbol{x}, \check{\boldsymbol{x}}) \right\} & \text{if } \boldsymbol{x}_s(\check{\boldsymbol{x}}) < 0 \\
\left\{ \left[\frac{1}{\overline{x}_s(\check{\boldsymbol{x}})}, \infty \right) \cup \left[\frac{-1}{\underline{x}_s(\check{\boldsymbol{x}})}, \infty \right) \right\} \big(\boldsymbol{x}_r(\boldsymbol{x}) - \boldsymbol{x}_q(\boldsymbol{x})\big) \mathbf{S}(x_s, \boldsymbol{x}, \check{\boldsymbol{x}}) & \\
\qquad\qquad \underline{\cup} \mathbf{S}(x_q(x), \boldsymbol{x}, \check{\boldsymbol{x}}) \underline{\cup} \mathbf{S}(x_r(x), \boldsymbol{x}, \check{\boldsymbol{x}}) & \text{if } 0 \in \boldsymbol{x}_s(\check{\boldsymbol{x}}).
\end{cases}
\end{aligned} \tag{25}$$

The first branch of Formula (25), for $\boldsymbol{x}_s(\check{\boldsymbol{x}}) > 0$, corresponds to Figure 1: The factor $\frac{1}{\boldsymbol{x}_s(\check{\boldsymbol{x}})}$ represents the distance of $\boldsymbol{x}_s(\check{\boldsymbol{x}})$ to the break point, while $\big(\boldsymbol{x}_r(\boldsymbol{x}) - \boldsymbol{x}_q(\boldsymbol{x})\big)$

[1]explained in [1] and later works

represents the jump at the break point, so the product $\frac{1}{\boldsymbol{x}_s(\check{\boldsymbol{x}})}\big(\boldsymbol{x}_r(\boldsymbol{x}) - \boldsymbol{x}_q(\boldsymbol{x})\big)$ represents the slope of the line from the point above $\boldsymbol{x}_s(\check{\boldsymbol{x}})$ to the break point. The factor $\mathbf{S}(x_s(x), \boldsymbol{x}, \check{\boldsymbol{x}})$ comes from the analogue of the chain rule for slopes. The part $\mathbf{S}(x_r(x), \boldsymbol{x}, \check{\boldsymbol{x}})$ takes account of the portion of the curve to the right of the break point. The other two branches of Formula (25) are analogous.

Formula (25) can be applied to the example of (24). Suppose $\boldsymbol{x} = [1, 5]$ and $\check{\boldsymbol{x}} = 3$. Then $\boldsymbol{x}_s(\check{\boldsymbol{x}}) = 2\check{\boldsymbol{x}}^2 - 2 = 16 > 0$, so the first branch of (25) applies. Proceeding, $1/\boldsymbol{x}_s(\check{\boldsymbol{x}}) = 1/16$, $\big(\boldsymbol{x}_r(\boldsymbol{x}) - \boldsymbol{x}_q(\boldsymbol{x})\big) = 2\boldsymbol{x} + 2\boldsymbol{x} = [4, 20]$, and $\mathbf{S}(x_s, \boldsymbol{x}, \check{\boldsymbol{x}}) = \mathbf{S}(2x^2 - 2, [1, 5], 3) = [8, 16]$ (computed using convexity). Thus, the slope bound is

$$\begin{aligned} \mathbf{S}(\nabla\phi, \boldsymbol{x}, \check{\boldsymbol{x}}) &= \mathbf{S}\left(\chi\big(x^2 - (2 - x^2), -2x, 2x\big), \boldsymbol{x}, \check{\boldsymbol{x}}\right) \\ &= \{2\} \underline{\cup} \frac{1}{16}[4, 20][8, 16] = [2, 20]. \end{aligned} \tag{26}$$

Applying the bound (25) in an interval Newton method, we obtain

$$\tilde{\boldsymbol{x}} \leftarrow 3 - 6/[2, 20] = 3 - \left[\frac{3}{10}, 3\right] = [0, 2.7] \subset \boldsymbol{x}. \tag{27}$$

This computation proves that there is a critical point of ϕ in $\tilde{\boldsymbol{x}}$. Further analysis, such as that in [8], could possibly prove that this critical point is unique. Also, iteration of the interval Newton method may lead to convergence to the critical point.

References

[1] G. Alefeld. Bounding the slope of polynomial operators and some applications. *Computing*, 26:227–237, 1980.

[2] G. H. Golub and C. F. Van Loan. *Matrix Computations (Second Edition)*. Johns Hopkins University Press, Baltimore, MD, 1989.

[3] R. B. Kearfott. A Fortran 90 environment for research and prototyping of enclosure algorithms for nonlinear equations and global optimization. *ACM Trans. Math. Software*, 21(1):63–78, March 1995.

[4] R. B. Kearfott. Interval extensions of non-smooth functions for global optimization and nonlinear systems solvers, preprint, 1995.

[5] R. B. Kearfott. *Verified Global Optimization – Interval Methods and Software.* Kluwer, Dordrecht, Netherlands, to be published.

[6] A. Neumaier. *Interval Methods for Systems of Equations*. Cambridge University Press, Cambridge, England, 1990.

[7] H. Ratschek and J. Rokne. *New Computer Methods for Global Optimization*. Wiley, New York, 1988.

[8] S. M. Rump. Verification methods for dense and sparse systems of equations. In J. Herzberger, editor, *Topics in Validated Computations*, pages 63–135, Amsterdam, 1994. Elsevier Science Publishers.

[9] S. M. Rump. New results on validation algorithms for large systems of equations, presented at SCAN'95.

Addresses:

R. B. Kearfott, University of Southwestern Louisiana, Department of Mathematics, U.S.L. Box 4-1010, Lafayette, Louisiana 70504-1010, USA.