Affine covariant Semi-smooth Newton in function space

Anton Schiela

March 14, 2018

These are lecture notes of my talks given for the Winter School "Modern Methods in Nonsmooth Optimization" that was held from February 26 to March 02, 2018 in Würzburg. They consist of a concise summary of the material, presented in the talks and the slides used there. Come supplementary information is also given.

1 Local Newton Methods

Consider a nonlinear, "differentiable" mapping

$$F : Z \to Y$$

for which we try to find a root: $z : \; F(z) = 0$.

One popular way is Newton's method:

$$\begin{aligned} F'(z_k)\delta z_k + F(z_k) &= 0 \\ z_{k+1} &= z_k + \delta z_k \end{aligned}$$

Newton's method is only locally convergent, so we have to globalize it:

- damping
- path-following

Affine covariance

Let $T : Y \to Z$ be a linear isomorphism. Then $F(z) = 0 \Leftrightarrow TF(z) = 0$, $(TF)'(z) = TF'(z)$ and

$$F'(z_k)\delta z_k + F(z_k) = 0 \Leftrightarrow TF'(z_k)\delta z_k + TF(z_k) = 0$$

Thus, roots and Newton corrections are not changed by a linear transformation of the image space.

Even more Let $T_x : Y \to Z_x$ be a family of linear isomorphisms onto a family of linear spaces, then the same holds true: we could transform the image space in each Newton step without changing the Newton correction.

Thus, questions of convergence of Newton's method are completely independent of quantities that live in the image space (use e.g. $T = \lambda Id$ as rescalings).

In problems that involving differential equations the norms in the domain space are readily computed, and it is easy to implement problem adjusted norms (which is often a good idea). In contrast, norms in the image space are usually dual norms, which are computationally inaccessible, and can in most cases only be approximated.

Local convergence

In order to characterize the deviation from linearity, we define an affine covariant theoretical factor

$$\Theta_{z_*}(z) := \frac{\|F'(z)^{-1}(F'(z)(z^* - z) - (F(z^*) - F(z)))\|_Z}{\|z^* - z\|_Z} \quad \text{for } z \neq z^*.$$

Obviously (for analysis only):

$$\Theta_{z_*}(z) \le \|F'(z)^{-1}\|_{Y \to Z} \frac{\|(F'(z)(z^* - z) - (F(z^*) - F(z)))\|_Y}{\|z^* - z\|_Z}$$

This looks like an affine covariant version of the usual finite differences used in the definition of Fréchet differentiability of F at z. Here, however, we will consider z^* fixed, and the limit $z \to z^*$.

Theorem 1.1. *Let R be a linear space, Z a normed space, $D \subset Z$ and $F : D \to R$. Assume there is $z_* \in D$ with $F(z_*) = 0$. For given $z \in D$ assume that there exists an invertible linear mapping $F'(z)(\cdot) : Z \to R$ such that $\Theta_{z_*}(z)$ is well defined. Assume that an inexact Newton step results in*

$$z_+ := z - F'(z)^{-1}F(z) + e,$$

where the relative error $\gamma(z) := \|e\| \,/\, \|z - z_*\|$ is bounded by

$$\gamma(z) + \Theta_{z_*}(z) \le \beta < 1. \tag{1}$$

Then

$$\|z_+ - z_*\| \le \beta \, \|z - z_*\| \, . \tag{2}$$

Proof. We compute for one inexact Newton step:

$$\begin{aligned} \|z_+ - z_*\| &= \left\|z - F'(z)^{-1}F(z) + e - z_*\right\| \\ &\le \left\|F'(z)^{-1}\left[F'(z)(z - z_*) - (F(z) - F(z_*))\right]\right\| + \|e\| \, . \end{aligned}$$

Inserting the definition of $\Theta_{z_*}(z)$ and assumption (1) we obtain

$$\|z_+ - z_*\| \le (\Theta(z) + \gamma(z)) \, \|z - z_*\| \le \beta \, \|z - z_*\| \, .$$

□

- F is called (affine covariantly) Newton differentiable or semi-smooth with respect to F', if $\Theta_{z_*}(z) \to 0$ as $z_* \to z$. For the case of exact Newton steps ($\gamma(z) = 0$) we then obtain the well known result of local superlinear convergence.
- The choice of F' is not unique. It is an algorithmic, but not an analytic quantity.

On the basis of this theorem, we denote $\Theta_{z_*}(z)$ as theoretical contraction factor.

Close relative: implicit function theorem

We know $F(z_0) = 0$, do we find a root for $F(z_0) - p = 0$?

- Difference: used to *show existence* of local solutions

- Needs *strict differentiability*:

$$\Theta^{\mathrm{IF}}_{z_0}(z,y) := \frac{\|F'(z_0)^{-1}(F'(z_0)(z-y) - (F(z) - F(y)))\|_Z}{\|z-y\|_Z},$$

  and $\lim_{z,y\to z_0} \Theta^{\mathrm{IF}}_{z_0}(z,y) = 0$.

- Then the simplified Newton method:

$$F'(z_0)\delta z_k + (F(z_k) - p) = 0.$$

  converges (for small enough p, of course) to a solution of the perturbed equation.

2 Semi-smoothness of superposition operators

Consider the following nonlinear function:

$$g : \mathbb{R}^d \times \Omega \to \mathbb{R}$$

We interpret the mapping

$$u \mapsto g(u,\omega),$$

which maps d-vectors to real numbers, as non-linear mapping $G : L_p(\Omega)^d \to L_s(\Omega)$, which maps (vector valued) *functions* to real valued *functions*, via the relation:

$$G(u)(\omega) := g(u(\omega),\omega)$$

Such a mapping is called superposition operator.

We assume that G maps measurable functions to measurable functions. A sufficient condition for that property is that g is a *Caratheodory function*, i.e., continuous in u for fixed ω and measurable in ω for fixed u. However, also pointwise limits and suprema of Caratheodory functions yield superposition operators that retain measurablility.

The following lemma is a slight refinement of a classical result of Krasnoselsky:

Theorem 2.1 (Continuity)**.** *Let Ω be a measurable subset of $\mathbb{R}^d$, and $g : \mathbb{R}^n \times \Omega \to \mathbb{R}^m$ such that the corresponding superposition operator G maps $L_p(\Omega)$ into $L_s(\Omega)$ for $1 \le p, s < \infty$.*

Let $u_* \in L_p(\Omega)$ be given. If g is continuous with respect to u at $(u_*(\omega),\omega)$ for almost all $\omega \in \Omega$, then G is continuous at u_* in the norm topology.

Proof. To show continuity of G at u_* for $s < \infty$ we consider an arbitrary sequence $\|u_n - u_*\|_{L_p} \to 0$. By picking a suitable subsequence, we may assume w.l.o.g. that $u_n(\omega) \to u_*(\omega)$ pointwise almost everywhere. Define the function

$$w(u,\omega) := |g(u,\omega) - g(u_*(\omega),\omega)|^s$$

and denote by W the corresponding superposition operator. Inserting the sequence u_n, we conclude that $W(u_n) \to 0$ pointwise a.e. because $g(u,\omega)$ is continuous in u at $(u_*(\omega),\omega)$ a.e..

Next we will show $W(u_n) \to 0$ in $L_1(\Omega)$ via the convergence theorem of Lebesgue. For this we have to construct a function $\overline{w} \in L_1(\Omega)$ that dominates the sequence $W(u_n)$. If we assume for simplicity that $|g(u,\omega)| \le M$ is bounded, and $|\Omega| < \infty$, then $\overline{w} := (2M)^s$ is such a domination function (otherwise, there is a more difficult construction possible). Hence, we obtain $W(u_n) \to 0$ in $L_1(\Omega)$ and thus $G(u_n) \to G(u_*)$ in $L_s(\Omega)$. Because u_n was arbitrary, we conclude continuity of the operator $G : L_p(\Omega) \to L_s(\Omega)$ at u_*. □

- For $p < s = \infty$, there is no continuity (except, when g is constant).
- For $p = s = \infty$ one needs *uniform* continuity of g:

$$\|u - v\|_\infty \to 0 \Rightarrow \|g(u) - g(v)\|_\infty \to 0.$$

Semi-smoothness of the max **function**

Consider the max-function:

$$m : \mathbb{R} \to \mathbb{R}$$
$$m(x) = \max(x, 0)$$

and define its "derivative" by

$$m'(x) := \begin{cases} 0 & : \quad x < 0 \\ \text{arbitrary} & : \quad x = 0 \\ 1 & : \quad x > 0. \end{cases} \tag{3}$$

Proposition 2.2. *For the* max*-function we have*

$$\lim_{x \to x^*} \Theta_{x_*}(x) = 0$$

Proof. If $x_* \neq 0$, then m is differentiable at x_* with locally constant derivative $m'(x_*)$, so the remainder term vanishes close to x_*.

If $x_* = 0$, we have two cases

$$\begin{aligned} x < 0 : \quad & m'(x)(x - x_*) - (m(x) - m(x_*)) = 0 - (0 - 0) = 0 \\ x > 0 : \quad & m'(x)(x - x_*) - (m(x) - m(x_*)) = (x - 0) - (x - 0) = 0. \end{aligned}$$

□

Semi-smoothness of superposition operators

Let $g, g' : \mathbb{R} \to \mathbb{R}$, $x_* \in L_p$ be given. Define for $x \in \mathbb{R}$

$$\gamma_{x_*}(x, \omega) := \begin{cases} \frac{|g'(x)(x - x_*(\omega)) - (g(x) - g(x_*(\omega))|}{|x - x_*(\omega)|} & : \quad x \neq x_*(\omega) \\ 0 & : \quad x = x_*(\omega) \end{cases}$$

If γ_{x_*} is continuous at x_*, $\Leftrightarrow \Theta_{x_*}(x) \to 0$, then also the superposition operator

$$\Gamma_{x_*} : L_p \to L_s$$

Then for $1/s + 1/p = 1/q$ we conclude by the Hölder-inequality:

$$\|G'(x)(x - x_*) - (G(x) - G(x_*))\|_{L_q} = \|\Gamma_{x_*}(x)|x - x_*|\|_{L_q} \leq \underbrace{\|\Gamma_{x_*}(x)\|_{L_s}}_{\to 0} \|x - x_*\|_{L_p}$$

Hence, loss of integrability, "norm-gap":

$$\frac{\|G'(x)(x - x_*) - (G(x) - G(x_*))\|_{L_q}}{\|x - x_*\|_{L_p}} \to 0 \qquad q < p.$$

To bridge this norm gap, additional structure is needed, coming, e.g., from a partial differential equation.

3 Application to problems in function space

Using function space methods for pure superposition operator problems is, of course, a bad idea. These problems can be solved pointwise. However, if there is a coupling, e.g., by a PDE, function space methods should be used.

A semi-linear equation with a semi-smooth nonlinearity

As a simple example, we consider the following problem:

$$F(x)v := \int_\Omega \nabla x \cdot \nabla v + \max(0, x)v \, d\omega \quad \forall v \in H_0^1(\Omega)$$

Then the remainder-term is a superpositon operator:

$$[F'(x)(x - x_*) - (F(x) - F(x_*)](\omega) = [M'(x)(x - x_*) - (M(x) - M(x_*))](\omega)$$

The differential operator cancels out, but inverse is smoothing:

$$(F'(x)\delta x)v = \int_\Omega \nabla \delta x \cdot \nabla v + M'(x)\delta x v \, d\omega$$

is an invertible mapping:

$$F'(x) : H_0^1(\Omega) \to H_0^1(\Omega)^*.$$

By the Sobolev embedding:

$$E : H_0^1(\Omega) \hookrightarrow L_p(\Omega), p > 2$$

we get:

$$EF'(x)^{-1}E^* : L_q(\Omega) \sim L_p(\Omega)^* \to L_p(\Omega), q < 2 < p.$$

We observe a smoothing property of $F'(x)^{-1}$ which helps to brigde the norm gap:

$$\begin{aligned}
&\|F'(x)^{-1}[F'(x)(x - x_*) - (F(x) - F(x_*)]\|_{L_p} \\
&\qquad = \|F'(x)^{-1}[M'(x)(x - x_*) - (M(x) - M(x_*))]\|_{L_p} \\
&\qquad \le \|F'(x)^{-1}\|_{L_q \to L_p} \|[M'(x)(x - x_*) - (M(x) - M(x_*))]\|_{L_q} \\
&\qquad = \|F'(x)^{-1}\|_{L_q \to L_p} o(\|x - x_*\|_{L_p}
\end{aligned}$$

Control constrained optimal control

Consider the following abstract problem:

$$\min \frac{1}{2}\|y - y_d\|_{L_2}^2 + \frac{\alpha}{2}\|u\|_{L_2}^2 \text{ s.t. } Ay - Bu = 0, \quad u \ge 0.$$

The corresponding KKT-conditions can be written as follows:

$$\begin{aligned}
0 &= J'(y) + A^* p \\
0 &= \alpha u + \lambda + B^* p \\
0 &= Ay - Bu \\
u &\ge 0, \lambda \le 0, \lambda \cdot u = 0
\end{aligned}$$

Elimination of λ yields $\alpha u = \max(B^*p, 0)$, which can be inserted into the state equation, yielding the control reduced KKT-conditions:

$$F(z) = \begin{pmatrix} J'(y) + A^*p \\ Ay - B\max(\alpha^{-1}B^*p, 0) \end{pmatrix} = 0,$$

where $F : Z \to Z^*$. $(Y \times P \to Y^* \times P^*)$. This system can be solved by semi-smooth Newton methods.

Let $w := B^*p$, $m(w) := \max(\alpha^{-1}w, 0)$, thus

$$m'(w) := \begin{cases} 0 & : \quad w \leq 0 \\ \alpha^{-1} & : \quad w > 0 \end{cases}$$

Derivative:

$$F'(z) = \begin{pmatrix} J''(y) & A^* \\ A & -B\alpha^{-1}m'(w)B^* \end{pmatrix},$$

Here $J''(y) = Id$.

As before, we will use the following strategy:

(i) Choice of appropriate framework. Here $Z = Y \times P = H^1 \times H^1$

(ii) Continuous invertibility of F' with uniform estimate

(iii) Analysis of remainder term as superposition operator

Lemma 3.1. *Assume that $m : U \to \mathbb{R}$ is bounded and non-negative, $A : H^1 \to (H^1)^*$ continuously invertible, $J''(y)$ bounded and positive semi-definite. Then the system*

$$\begin{pmatrix} J''(y) & A^* \\ A & -BmB^* \end{pmatrix} \begin{pmatrix} \delta y \\ \delta p \end{pmatrix} = \begin{pmatrix} r_1 \\ r_2 \end{pmatrix}$$

has a unique solution. Moreover for $r_1 \in L_2$, the estimate

$$\|\delta y\|_{H_1} + \|\delta p\|_{H^1} \leq C(\|r_1\|_{(H^1)^*} + \|r_2\|_{(H^1)^*})$$

holds, where C does not depend on m, but only on $\|m\|_\infty$.

Proof. Consider the optimal control problem:

$$\min \frac{1}{2}J''(y)(\delta y, \delta y) - \langle r_1, \delta y\rangle + \frac{1}{2}\|\delta u\|_U^2 \text{ s.t. } A\delta y - B\sqrt{m}\delta u - r_2 = 0,$$

where $(r_1, r_2) \in Z^*$. By standard argumentation, this problem has a solution $(\delta y, \delta u) \in H^1 \times U$, which is even unique by strict convexity (because of the δu-Term).

Further, by uniform convexity, we obtain:

$$\|u\|_U \leq c(\|r_1\|_{(H^1)^*} + \|r_2\|_{(H^1)^*}.$$

KKT-conditions read:

$$J''(y)\delta y + A^*\delta p = r_1$$
$$\delta u + \sqrt{m}B^*\delta p = 0$$
$$A\delta y + B\sqrt{m}\delta u = r_2$$

The corresponding control reduced KKT conditions (our system) also has a unique solution $(\delta y, \delta p)$.

The state equation implies:

$$\|\delta y\|_{H^1} \le c(\|\sqrt{m}\|_\infty)(\|\delta u\|_U + \|r_2\|_{(H^1)^*})$$

and by the adjoint equation we get:

$$\|\delta p\|_{H^1} \le (\|\delta y\|_{L_2} + \|r_2\|_{(H^1)^*})$$

Inserting into each other, we obtain the desired estimate. □

Remark: with regularity theory, this lemma can be refined.

Theorem 3.2. *Let $U = L_2(Q)$, (e.g. $Q = \Omega, Q = \Gamma$...) and assume that $B^* : H^1 \to L_q$ is continuous for some $q > 2$. Then Newton's method converges locally superlinearly towards the solution of our optimal control problem, where*

$$\|z\| := \|y\|_{L_2} + \|B^* p\|_{L_q}$$

Proof. Let $q > 2$ as in our assumption. We compute

$$\begin{aligned} F'(z)(z - z_*) - (F(z) - F(z_*)) &= B(m'(w)(w - w_*) - (m(w) - m(w_*))) \\ &= B \underbrace{\frac{m'(w)(w - w_*) - (m(w) - m(w_*))}{w - w_*}}_{\gamma_{w_*}(w)} (w - w_*) \end{aligned}$$

We note that $\gamma_{w_*}(w) \le \alpha^{-1}$, and thus $\Gamma_{w^*} : L_q \to L_s$ for each $s < \infty$. By the above example $\lim_{w \to w_*} \gamma_{w_*}(w) = 0$. Thus, by Lemma 2.1 we obtain $\|\Gamma_{w_*}(w)\|_{L_s} \to 0$ for $\|w - w_*\|_{L_q} \to 0$.

For $1/2 = 1/s + 1/q$ we use the Hölder inequality to get

$$\|F'(z)(z - z_*) - (F(z) - F(z_*))\|_{L_2 \times (H^1)^*} \le \|B\|_{L_2 \to (H^1)^*} \|\Gamma_{w_*}(w)\|_{L_s} \|w - w_*\|_{L_q}.$$

Now let $\delta z = F'(z)^{-1}(F'(z)(z - z_*) - (F(z) - F(z_*)))$. By Lemma 3.1

$$\begin{aligned} \|\delta z\| &= C(\|\delta y\|_{L_2} + \|B^* \delta p\|_{L_q}) \\ &\le \|\delta y\|_{H^1} + \|B^*\|_{H^1 \to L_q} \|\delta p\|_{H^1}) \\ &\le C\|B\|_{L_2 \to (H^1)^*} \|\Gamma_{w_*}(w)\|_{L_s} \|w - w_*\|_{L_q}. \qquad \le C\|\Gamma_{w_*}(w)\|_{L_s} \|B^*(p - p_*)\|_{L_q}. \end{aligned}$$

thus $\Theta_{z^*}(z) \to 0$, if $B^*(p - p_*) \to 0$ in L_q for $s < \infty$.

Hence, Newton's method converges locally superlinearly. □

Related approaches

Semi-smooth Newton methods were applied to optimal control problems in various other formulations [2, 6, 3]. Alternatively one can also eliminate the state and the adjoint state to end up with a problem in u that contains solution operators $S = A^{-1}B$ and their adjoints, together with the max-superposition operator. Here the smoothing property of S helps to bridge the norm-gap. Also, a complementarity function approach can be used to tackle the full KKT-system. Then a smoothing step has to be employed to bridge the norm gap [6], see also the pre-semi-smooth paper [8]. The approach presented here follows [5], where also convergence rates are discussed and additional details can be found. A comprehensive source on semi-smooth Newton in function space are the textbooks [7, 4].

4 Detection of local convergence

A-priori theory yields confidence in algorithms, but how to implement?

Main issues:

- detection of local convergence
- globalization

Need computational quantities, close to theoretical quantities. The ideas presented here are based on the concept of affine covariant Newton methods [1].

To assess local convergence of Newton's method we may use a one-parameter model for Θ:

$$[\Theta_z](y) := \frac{[\omega]}{2}\|z-y\|_Z,$$

where $[\omega]$ has to be estimated.

Idea: estimate $\Theta(z, z_*)$ by $\Theta(z, z_+)$:

$$\Theta_{z_+}(z) := \frac{\|F'(z)^{-1}(F'(z)(z_+ - z) - (F(z_+) - F(z)))\|_Z}{\|z_+ - z\|_Z}.$$

If $\Theta_{z_+}(z) \ll 1$ fast convergence, if $\Theta_{z_+}(z) > 1$ divergence

Computation:

$$\Theta_{z_+}(z) = \frac{\| - F'(z)^{-1}F(z_+)\|_Z}{\|\delta z\|_Z} = \frac{\|\delta \overline{z}\|_Z}{\|\delta z\|_Z},$$

where $\delta\overline{z} = F'(z)^{-1}F(z_+)$ is a *simplified* Newton correction.

This also yields an estimate for $[\omega]$:

$$[\omega] := \frac{2\Theta_{z_+}(z)}{\|z - z_+\|_Z}.$$

Under the assumption that $\Theta_{z_*}(z) \le 2\omega\|z - z_*\|_Z$ we obtain an estimate for the radius of contraction:

$$r_\Theta := \frac{2\Theta}{\omega} : \|z - z_*\|_Z \le r_\Theta \Rightarrow \Theta_{z_*}(z) \le \Theta.$$

Thus, we get a computational estimate $[r_\Theta] := \frac{2\Theta}{[\omega]}$.

Moreover, we have the error bound:

$$\|z_+ - z_*\| = \Theta_{z_*}(z)\|z - z_*\| \le \Theta_{z_*}(\|z - z_+\| + \|z_+ - z_*\|),$$

thus

$$\|z_+ - z_*\| \le \frac{\Theta_{z_*}(z)}{1 - \Theta_{z_*}(z)}\|z - z_+\|,$$

which can be estimated again via $[\Theta_{z_*}(z)] = \Theta_{z_+}(z)$

Difficulties with semi-smooth problems

Proposition 4.1. *Assume that $F'(\bar{z})^{-1}F$ is not Fréchet differentiable at $\bar{z}$. Then there there exists Θ_0, and a sequence $z_k \to \bar{z}$, such that for the radius of contraction $r^k_{\Theta_0}$ for the perturbed problems $F(z) - F(z_k) = 0$ we have*

$$\lim_{k\to\infty} r^k_{\Theta_0} = 0.$$

Proof. If F is not Fréchet differentiable at $\bar{z}$, then for each choice of $F^{-1}(\bar{z})$ there is a a constant c and a sequence $z_k \to \bar{z}$ such that

$$\left\|F^{-1}(z_0)(F'(z_0)(z_k - \bar{z}) - (F(z_k) - F(\bar{z})))\right\| \geq c\,\|z_k - \bar{z}\|\,. \tag{4}$$

This inequality, reinterpreted in terms of Newton contraction of F at z_k, directly yields the result. □

Thus, if Fréchet differentiability fails, the radius of convergence of Newton methods for semi-smooth problems may break down under perturbations. This implies difficulties in many analytic and algorithmic concepts that rely on stability of Newton's method:

- implicit function theorems
- semi-smooth Newton path-following
- detection of local convergence

Proposition 4.2. *Assume that the following* strong *semi-smoothness assumption holds:*

$$\lim_{z,y\to z_*} \frac{\left\|F'(y)^{-1}\left[F'(y)(z - z_*) - (F(z) - F(z_*))\right]\right\|}{\|z - z_*\|} = 0. \tag{5}$$

Then

$$\lim_{z_-\to z_*} \frac{[\Theta](z_-)}{\Theta(z_-)} = 1. \tag{6}$$

Proof. For a comparison of $\Theta(z_-)$ and $[\Theta](z_-)$ we introduce the auxiliary term

$$\overline{\Theta}(z_-) := \frac{\|\overline{z} - z_*\|}{\|z - z_*\|}$$

and use the inverse triangle inequality $|\,\|a\| - \|b\|\,| \leq \|a \pm b\|$ to compute

$$|\,\|z - z_*\| - \|z - \overline{z}\|\,| \leq \|z_* - \overline{z}\|\,,$$
$$|\,\|z_- - z_*\| - \|z_- - z\|\,| \leq \|z_* - z\|\,,$$

which yields

$$\|z - z_*\|\,(1 - \overline{\Theta}(z_-)) \leq \|z - \overline{z}\| \leq \|z - z_*\|\,(1 + \overline{\Theta}(z_-)),$$
$$\|z_- - z_*\|\,(1 - \Theta(z_-)) \leq \|z_- - z\| \leq \|z_- - z_*\|\,(1 + \Theta(z_-)).$$

Combination of these inequalities yields the estimate

$$\frac{1 - \overline{\Theta}(z_-)}{1 + \Theta(z_-)} \leq \frac{\|z - \overline{z}\|\,\|z_- - z_*\|}{\|z - z_-\|\,\|z - z_*\|} = \frac{[\Theta](z_-)}{\Theta(z_-)} \leq \frac{1 + \overline{\Theta}(z_-)}{1 - \Theta(z_-)}.$$

If $\lim_{z_-\to z_*} \Theta(z_-) = 0$ and $\lim_{z_-\to z_*} \overline{\Theta}(z_-) = 0$, then (6) holds. The first assumption in this statement holds, if F is semi-smooth at z_*. For the second assumption, we compute

$$\overline{\Theta}(z_-) := \frac{\|\overline{z} - z_*\|}{\|z - z_*\|} = \frac{\left\|(z - z_*) - F'(z_-)^{-1}(F(z) - F(z_*))\right\|}{\|z - z_*\|},$$

thus, $\lim_{z_-\to z_*} \overline{\Theta}(z_-) = 0$ is implied by (5). □

Such a strong semi-smoothness assumptions holds, if $y \neq \phi$ almost everywhere in Ω. This is a kind of strict complementarity assumption.

5 Globalization by damping

For bad initial guesses, Newton's method may diverge. Idea: for an iterate z_0 solve the problem: find $z(\lambda) \in Z$, such that

$$F(z(\lambda)) - (1-\lambda)F(z_0) = 0 \quad \text{for some } 0 < \lambda \leq 1,$$

This defines the *Newton path* $z(\lambda)$, with $z(0) = z_0$ and $z(1) = z_*$ (if it exists) is a zero of F. Newton step for this equation at z:

$$0 = F'(z_0)\delta z_\lambda + F(z_0) - (1-\lambda)F(z_0) = F'(z_0)\delta z_\lambda + \lambda F(z_0)$$

aka *damped Newton correction*:

$$\delta z_\lambda = \lambda \delta z = -\lambda F'(z_0)^{-1}F(z_0)$$

Choose λ such that the Newton's method for $z(\lambda)$ converges locally, i.e.

$$\Theta_{z(\lambda)}(z_0) < 1$$

The simplified Newton step for that problem reads, setting the *trial iterate* $z_t := z_0 + \lambda\delta z$

$$F'(z_0)\delta\bar{z}_\lambda + F(z_t) - (1-\lambda)F(z_0) = 0.$$

Try to control λ, such that $\Theta \approx 1/2$, but accept if $\Theta < 1$.
If $\Theta \geq 1$, then λ has to be reduced.
Estimation of Θ:

$$\begin{aligned}\Theta_{z(\lambda)}(z_0) \approx \Theta_{z_t}(z_0) &= \frac{\|F'(z_0)^{-1}(F'(z_0)\lambda\delta z - (F(z_t) - F(z_0))\|}{\|\lambda\delta z\|} \\ &= \frac{\|\delta\bar{z}_\lambda\|}{\|\lambda\delta z\|}\end{aligned}$$

Model for Θ:

$$[\Theta]_{z_0}(z) := \frac{[\omega]}{2}\|z - z_0\|.$$

If $[\Theta]$ is too large, choose λ_+ such that:

$$\frac{1}{2} = [\Theta]_{z_t}(z_0 + \lambda_+\delta z) = \frac{[\omega]\lambda_+}{2}\|\delta z\|_Z$$

where

$$[\omega] = \frac{2\Theta_{z_0}(z_t)}{\|z_t - z_0\|} = \frac{2\Theta_{z_0}(z_t)}{\|\lambda\delta z\|} = \frac{2\|\delta\bar{z}_\lambda\|}{\|\lambda\delta z\|^2}.$$

Algorithm 5.1. (Damped Newton)
Initial guesses: z, $[\omega]$ (or λ)
Parameters: λ_{fail}, TOL
do
  solve $F'(z)\delta z + F(z) = 0$
  do
    compute $\lambda = \min(1, 1/(\|\delta z\|[\omega]))$
    if $\lambda < \lambda_{fail}$
      terminate: "Newton's method failed"

solve $F'(z)\delta\bar{z} + F(z+\lambda\delta z) = 0$
compute $[\omega] = \frac{\|\delta\bar{z}-(1-\lambda)\delta z\|}{\|\lambda\delta z\|^2}$.
compute $\Theta(z, z+\lambda\delta z) = [\omega]\|\lambda\delta z\|$
while $\Theta(z, z+\lambda\delta z) \geq 1$
$z = z + \lambda\delta z$
if $\lambda = 1$ and $\Theta(z, z+\delta z) \leq 1/4$ and $\|\delta z\| \leq$ TOL
terminate: "Desired Accuracy reached", $z_{out} = z + \delta z + \delta\bar{z}$
while false

References

[1] P. Deuflhard. *Newton Methods for Nonlinear Problems. Affine Invariance and Adaptive Algorithms*, volume 35 of *Series Computational Mathematics*. Springer, 2nd edition, 2006.

[2] M. Hintermüller, K. Ito, and K. Kunisch. The primal-dual active set strategy as a semi-smooth Newton method. *SIAM J. Optim.*, 13:865–888, 2003.

[3] M. Hintermüller and M. Ulbrich. A mesh-independence result for semismooth Newton methods. *Math. Programming*, 101:151–184, 2004.

[4] K. Ito and K. Kunisch. *Lagrange Multiplier Approach to Variational Problems and Applications*. Society for Industrial and Applied Mathematics, 2008.

[5] A. Schiela. A simplified approach to semismooth newton methods in function space. *SIAM J. Optim.*, 19(3):1417–1432, 2008.

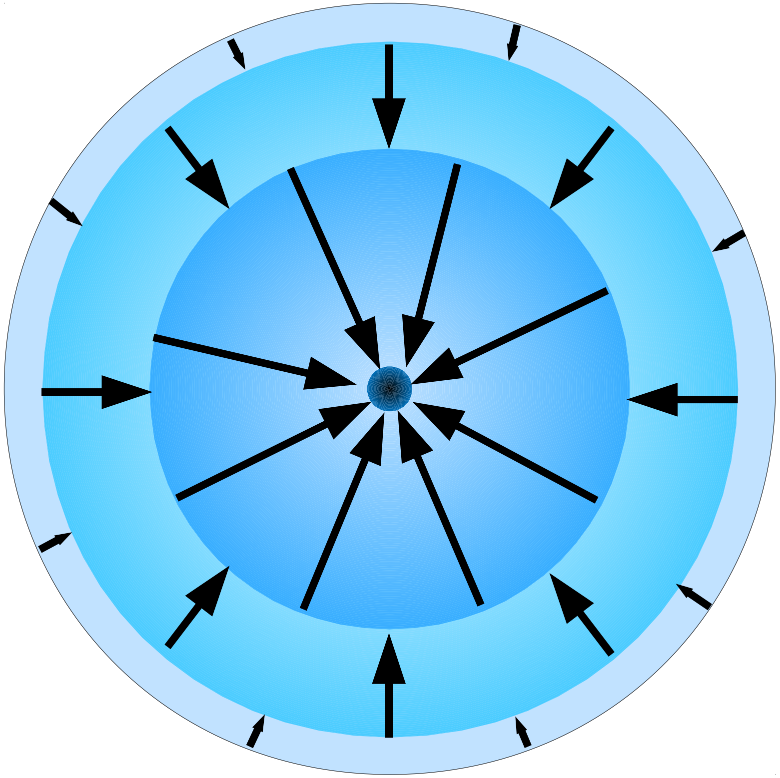
[6] M. Ulbrich. Semismooth Newton methods for operator equations in function spaces. *SIAM J. Optim.*, 13:805–842, 2003.

[7] M. Ulbrich. *Semismooth Newton Methods for Variational Inequalities and Constrained Optimization Problems in Function Spaces*. Society for Industrial and Applied Mathematics, 2011.

[8] M. Ulbrich and S. Ulbrich. Superlinear convergence of affine-scaling interior-point Newton methods for infinite-dimensional nonlinear problems with pointwise bounds. *SIAM J. Control Optim.*, 38(6):1938–1984, 2000.

Techniques for the Analysis of Semi-Smooth Newton Methods

Anton Schiela



UNIVERSITÄT BAYREUTH

Winter School Würzburg, 2. März 2018

Newton's Method in Function Space

Nonlinear Functional Analysis

Nonlinear Equations

Homotopy

Newton's Method

Optimization

Affine Invariance

[Deuflhard 2004]

Semi-Smoothness

[Kunisch, Hintermüller 2002]

[Ulbrich 2002]

Newton's Method

UNIVERSITÄT BAYREUTH

Solve the operator equation: $T(x) = 0$

$$x_* = x_* - T'(x)^{-1}T(x_*)$$

$$- \quad x_+ = x - T'(x)^{-1}T(x)$$

$$x_* - x_+ = T'(x)^{-1}(T(x) - T(x_*)) - (x - x_*)$$

Sufficient for local superlinear convergence:

$$\frac{\|x_* - x_+\|}{\|x_* - x\|} = \frac{\|T'(x)^{-1}(T(x) - T(x_*)) - (x - x_*)\|}{\|x_* - x\|} \to 0 \quad \text{for} \quad x \to x_*$$

Affine Invariant form of semi-smoothness

$$\lim_{x \to x_*} \frac{\|T'(x)^{-1}\left[T(x) - T(x_*) - T'(x)(x - x_*)\right]\|}{\|x_* - x\|} = 0$$

Remarks:

$T'(x)$ is not uniquely determined

$T'(x)$ may, but need not be a classical derivative

follows from well known semi-smoothness and regularity assumptions

computable estimate is available

$$\frac{\|T'(x)^{-1}\left[T(x) - T(x_+) - T'(x)(x - x_+)\right]\|}{\|x_+ - x\|} = \frac{\|x_+ - \overline{x}_+\|}{\|x - x_+\|}$$

$$\overline{x}_+ = x_+ - T'(x)^{-1}T(x_+)$$

The Dirichlet function is „semi-smooth“...

Example from Analysis I:

$$f(x) = \begin{cases} 1 & : \quad x \in \mathbb{R} \setminus \mathbb{Q} \\ 0 & : \quad x \in \mathbb{Q} \end{cases}$$

Definition of derivative:

$$f'(x) := \begin{cases} 1/x & : \quad x \in \mathbb{R} \setminus \mathbb{Q} \\ 0 & : \quad x \in \mathbb{Q} \end{cases}$$

Semi-smoothness at $x = 0$:

$$f(x) - f(0) - f'(x)x = f(x) - (f(x)/x)x = 0$$

Semi-smoothness is not an analytic, but an algorithmic concept

An important special case

UNIVERSITÄT BAYREUTH

$$T(x) := Lx + F(x)$$

L linear differential operator

$F : L_p(\Omega) \to L_q(\Omega)$ superposition operator

$x \mapsto F(x) : (F(x))(t) := f(t, x(t))$ a.e. e.g. $f(t,x) := \max(0,x)$

then

$$T(x) - T(x_*) - T'(x)(x - x_*) = F(x) - F(x_*) - F'(x)(x - x_*)$$

Consider

$$\lim_{x \to x_*} \frac{\|T'(x)^{-1}[F(x) - F(x_*) - F'(x)(x - x_*)]\|}{\|x_* - x\|}$$

Example: control constraints

$$\min \frac{1}{2}\|y - y_d\|_2^2 + \frac{\alpha}{2}\|u\|_2^2 \qquad Ay - Bu = 0$$

$$u \geq 0$$

$$\left.\begin{aligned} y - y_d + A^*p &= 0 \\ Ay - Bu &= 0 \\ \alpha u - \max(B^*p, 0) &= 0 \end{aligned}\right\} \Rightarrow \qquad \begin{aligned} y - y_d + A^*p = 0 \\ Ay - \alpha^{-1}B\max(B^*p, 0) = 0 \end{aligned}$$

$$Lx = \begin{pmatrix} A^*p \\ Ay \end{pmatrix} \qquad F(x) = \begin{pmatrix} y - y_d \\ -\alpha^{-1}B\max(B^*p, 0) \end{pmatrix}$$

Semi-smoothness of superposition operators

UNIVERSITÄT BAYREUTH

$$\|F(x)-F(y)-F'(x)(x-y)\|_{L_q} = \left\| \frac{F(x)-F(y)-F'(x)(x-y)}{x-y}(x-y) \right\|_{L_q}$$

$$\overset{\text{Hölder}}{\underset{\text{inequality}}{\leq}} \left\| \frac{F(x)-F(y)-F'(x)(x-y)}{x-y} \right\|_{L_r} \|x-y\|_{L_p} \qquad r^{-1}+p^{-1}=q^{-1}$$

If $G_y : L_p(\Omega) \to L_r(\Omega)$

$$G_y(x)(t) := \begin{cases} \frac{f(t,x(t))-f(t,y(t))-f'(t,x(t))(x(t)-y(t))}{x(t)-y(t)} & : \quad x(t) \neq y(t) \\ 0 & : \quad x(t) = y(t) \end{cases}$$

is continuous at y , then

$$\lim_{x \to y} \frac{\|F(x)-F(y)-F'(x)(x-y)]\|_{L_q}}{\|y-x\|_{L_p}} = 0$$

In general

UNIVERSITÄT BAYREUTH

Semi-Smoothness of outer function f

↓

Hölder Inequality $r^{-1} + p^{-1} = q^{-1}$

Local Continuity Result $r < \infty$

↓

Semi-Smoothness of superposition operator $F : L_p \to L_q$ $q < p$

Remarks:

complete analogy to Fréchet differentiability

no global Lipschitz condition necessary (growth conditions instead)

necessity of $r < \infty$ explains the well known norm-gap

$r = p = q = \infty$ not compatible with semi-smoothness

Closing the norm-gap (well known)

UNIVERSITÄT BAYREUTH

$$\lim_{x \to x_*} \frac{\|F(x) - F(x_*) - F'(x)(x - x_*)]\|_{L_q}}{\|x_* - x\|_{L_p}} = 0 \qquad \|T'(x)^{-1}\|_{L_q \to L_p} \leq M$$

$$T'(x) = Lx + F'(x)$$

$$\lim_{x \to x_*} \frac{\|T'(x)^{-1}[T(x) - T(x_*) - T'(x)(x - x_*)]\|_{L_p}}{\|x_* - x\|_{L_p}} = 0$$

Local superlinear convergence of Newton's method

Order of semi-smoothness

Same idea:

$$\|F(x) - F(y) - F'(x)(x-y)\|_{L_q} = \left\| \frac{F(x) - F(y) - F'(x)(x-y)}{(x-y)^\alpha}(x-y)^\alpha \right\|_{L_q}$$

Hölder inequalit

$$\leq \left\| \frac{F(x) - F(y) - F'(x)(x-y)}{(x-y)^\alpha} \right\|_{L_r} \|x-y\|^\alpha_{L_p} \qquad r^{-1} + \alpha p^{-1} = q^{-1}$$

If $G_y : L_p(\Omega) \to L_r(\Omega)$

$$G_y(x)(t) := \begin{cases} \frac{f(t,x(t)) - f(t,y(t)) - f'(t,x(t))(x(t)-y(t))}{(x(t)-y(t))^\alpha} & : \quad x(t) \neq y(t) \\ 0 & : \quad x(t) = y(t) \end{cases}$$

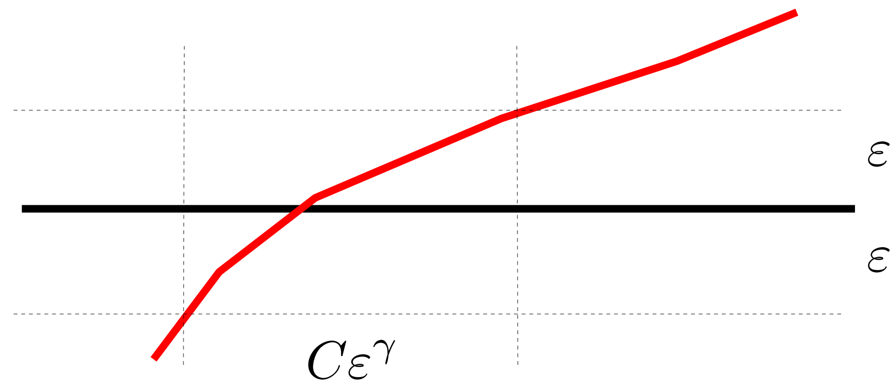
is bounded near y , then

$$\|F(x) - F(y) - F'(x)(x-y)]\|_{L_q} = O(\|y-x\|^\alpha_{L_p})$$

Strict complementarity

$$f(t,x) := \max(0,x)$$

$$\left|\{t \in \Omega : |y(t)| < \varepsilon\}\right| \leq C\varepsilon^{\gamma}$$



$$G_y(x) = \begin{cases} \frac{\max(0,x)-\max(0,y)-\iota_+(x)(x-y)}{(x-y)^{\alpha}} & : \quad x \neq y \\ 0 & : \quad x = y \end{cases}$$

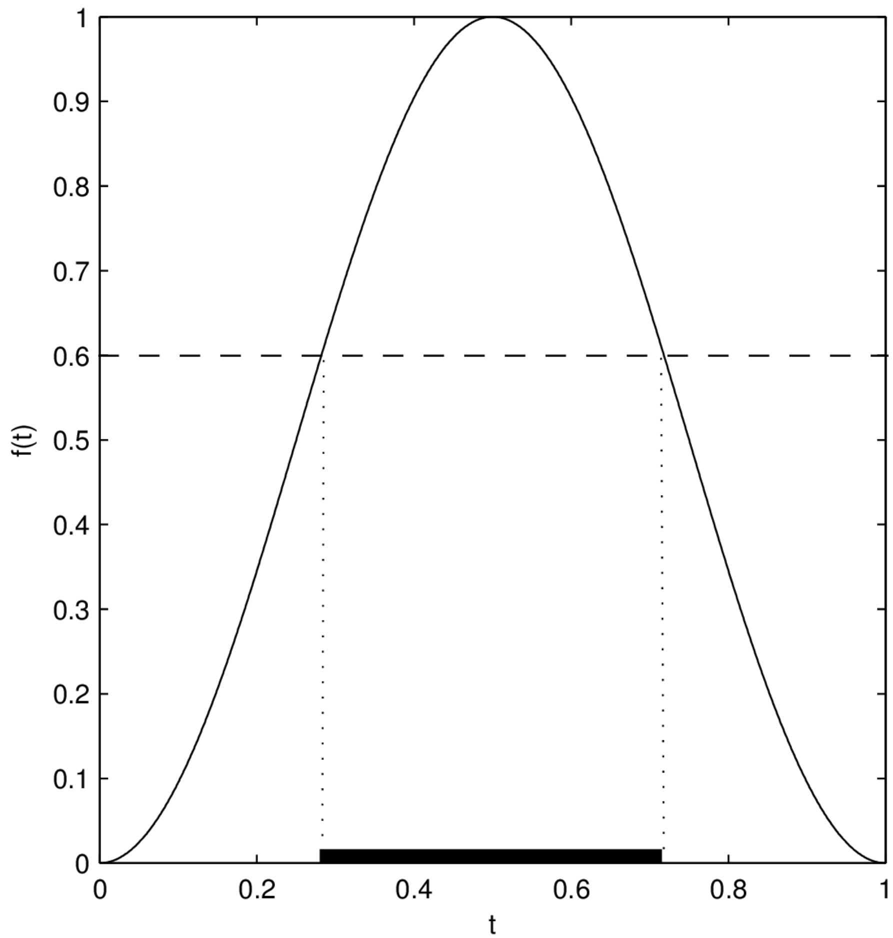
$\operatorname{sgn} x = \operatorname{sgn} y$ $\qquad$ $G_y(x) = 0$

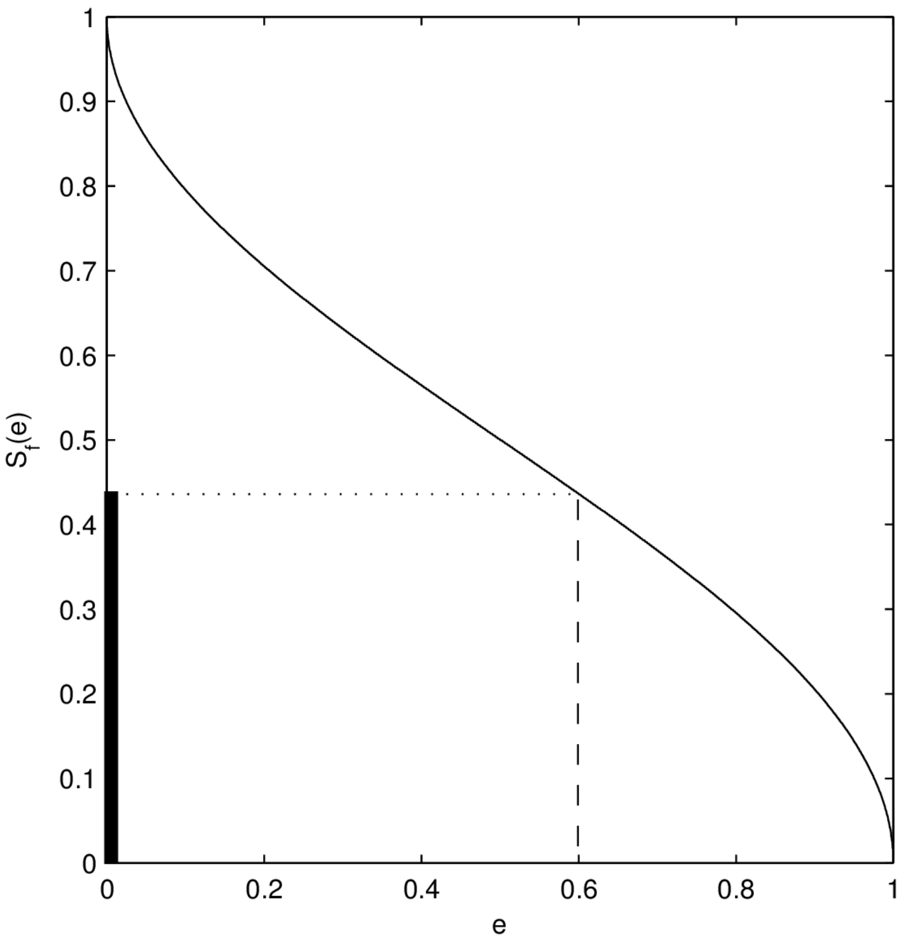
$\operatorname{sgn} x \neq \operatorname{sgn} y$ $\qquad$ $|G_y(x)| \leq \dfrac{|y|}{|x-y|^{\alpha}} = \dfrac{|y|}{(|x|+|y|)^{\alpha}} \leq |y|^{1-\alpha}$

$$\left|\{t \in \Omega : |G_y(x)(t)| > \varepsilon^{1-\alpha}\}\right| \leq C\varepsilon^{\gamma}$$

Distribution function

$$\mathcal{S}_f(e) := |\{t \in \Omega : |f(t)| > e\}|$$





Distribution function

UNIVERSITÄT BAYREUTH

$$\mathcal{S}_v(e) := |\{t \in \Omega : |v(t)| > e\}| = |\{t \in \Omega : |v(t)|^r > e^r\}| = \mathcal{S}_{v^r}(e^r)$$

$$\int_\Omega |v(t)|\, dt = \int_{[0,\infty]} \mathcal{S}_v(e)\, de$$

With strict complementarity we have the estimate:

$$\left|\{t \in \Omega : |G_y(x)(t)| > \varepsilon^{1-\alpha}\}\right| \leq C\varepsilon^\gamma \Longleftrightarrow \mathcal{S}_{|G_y(x)|}(\varepsilon^{1-\alpha}) \leq C\varepsilon^\gamma$$

$$\begin{aligned}\|G_y(x)\|_{L_r}^r &= \int_{[0,\infty]} \mathcal{S}_{|G_y(x)|^r}(e)\, de \\ &= c\int_{[0,\infty]} S_{|G_y(x)|^r}(\varepsilon^{(1-\alpha)r})\varepsilon^{(1-\alpha)r-1}\, d\varepsilon \\ &= c\int_{[0,\infty]} S_{|G_y(x)|}(\varepsilon^{1-\alpha})\varepsilon^{(1-\alpha)r-1}\, d\varepsilon \leq c\int_{[|\Omega|,\infty]} \varepsilon^{(1-\alpha)r-1+\gamma}\, d\varepsilon + C\end{aligned}$$

Orders of convergence

Lemma:

$$\left.\begin{aligned} \left|\{t \in \Omega : |G_y(x)(t)| > \varepsilon^{1-\alpha}\}\right| \leq C\varepsilon^{\gamma} \\ 1 < \alpha < 1 + \gamma r^{-1} \end{aligned}\right\} \Rightarrow \qquad G_y(x) \in L_r(\Omega) \text{ for such } \alpha$$

Conclusion:

$$\|F(x) - F(y) - F'(x)(x-y)\|_{L_q} \leq \|G_y(x)\|_{L_r} \|x-y\|_{L_p}^{\alpha} \qquad r^{-1} + \alpha p^{-1} = q^{-1}$$

Hence for the max-function:

$$\left.\begin{aligned} \|T'(x)^{-1}\|_{L_q \to L_p} \leq M \\ |\{t \in \Omega : |y(t)| < \varepsilon\}| \leq C\varepsilon^{\gamma} \\ 1 < \alpha < \frac{1+\gamma q^{-1}}{1+\gamma p^{-1}} \end{aligned}\right\} \Rightarrow$$

Local superlinear convergence of Newton's method of order α

What about globalization?

UNIVERSITÄT BAYREUTH

Assume that T is not Frechet differentiable at x_*:

$$\exists \Theta > 0, x_n \to \bar{x} : \frac{\|(T(x_n) - T(\bar{x})) - T'(\bar{x})(x_n - \bar{x})\|}{\|x_n - \bar{x}\|} > \Theta$$

What about globalization?

UNIVERSITÄT BAYREUTH

Assume that T is not Frechet differentiable at x_*:

$$\exists \Theta > 0, x_n \to \bar{x} : \frac{\|(T(x_n) - T(\bar{x})) - T'(\bar{x})(x_n - \bar{x})\|}{\|x_n - \bar{x}\|} > \Theta$$

Consider the set of good semi-smooth approximation:

$$M_\Theta(x) := \{\xi \in X : \|T(x) - T(\xi) - T'(\xi)(x - \xi)\| \leq \Theta \|x - \xi\|\}$$

Then $\bar{x} \notin M_\Theta(x_n)$

If Θ is sufficiently large, our radius of Newton contraction towards x_n can be arbitrarily small

Problem: non-differentiable points are ubiquitous in semi-smooth problems
Consequence: homotopy methods difficult to analyse

Using a merit function

Classical merit function on a Hilbert space:

$$\phi(x) := \frac{1}{2}\langle T(x), T(x)\rangle_X$$

Global convergence theory demands uniform continuity of $\phi'(x)$:

$$\phi'(x)\delta x = \langle T'(x)\delta x, T(x)\rangle$$

Problem with semi-smoothness: T' is not continuous

Remedy 1 (M. Ulbrich): use semi-smooth operator, such that

$$T(x) \neq 0 \quad \Rightarrow \quad T' \text{ is continuous at } x$$

Candidate: Fischer-Burmeister function: $f(x,y) = x + y - \sqrt{x^2 + y^2}$

Remedy 2 (Kunisch et. al.): use techniques from non-smooth optimization

Remedy 3 (Gräser): for control constrained optimal control, reformulation in adjoint state

Objective with semi-smooth derivative

UNIVERSITÄT BAYREUTH

Unconstrained minimization problem:

$$\min f(x)$$

Semi-smooth and Lipschitz continuous derivative:

$$\lim_{x \to x_*} \frac{\|f'(x) - f'(x_*) - (f')'(x)(x - x_*)\|}{\|x - x_*\|} = 0$$

Globalization with classical techniques (line-search, trust-region) easy, if you use $f(x)$ as merit-function and not $\|f'(x)\|^2$

Applications: penalized obstacle problems and state constrained problems

What about transition to fast local convergence?

Second-order semi-smoothness

Appropriate condition for transition to fast local convergence:

$$f(x+\delta x) = f(x) + f'(x)\delta x + \frac{1}{2} H_{x+\delta x}\delta x^2 + o(\|\delta x\|^2)$$

We have a quadratic approximation that may depend on $x+\delta x$

In semi-smooth Newton we have, of course: $H_{x+\delta x} = (f')'(x+\delta x)$

Theorem:

Define the local quadratic model:

$$m_x(\delta x) = f(x) + f'(x)\delta x + \frac{1}{2} H_x \delta x^2$$

Under second-order semi-smoothness we have for full Newton steps:

$$\lim_{x\to x_*} \eta(x) = \lim_{x\to x_*} \frac{f(x+\delta x) - f(x)}{m_x(\delta x) - m_x(0)} = 1$$

Thus, full Newton steps become acceptable, eventually

Second-order semi-smoothness

UNIVERSITÄT BAYREUTH

Theorem:
Define the local quadratic model:

$$m_x(\delta x) = f(x) + f'(x)\delta x + \frac{1}{2} H_x \delta x^2$$

Under second-order semi-smoothness we have for full Newton steps:

$$\lim_{x \to x_*} \eta(x) = \lim_{x \to x_*} \frac{f(x+\delta x) - f(x)}{m_x(\delta x) - m_x(0)} = 1$$

Theorem:
Under second-order semi-smoothness we have for damped Newton steps:

$$\liminf_{x \to x_*} \eta(x) \geq \frac{1}{2}$$

Thus, damped Newton steps become acceptable, eventually if $\overline{\eta} < \frac{1}{2}$

Examples:

Squared max-function:

$$f(x) = \frac{1}{2}\max(x,0)^2 \qquad H_x = \begin{cases} 1 & : \quad x \geq 0 \\ 0 & : \quad x < 0 \end{cases}$$

$$f(x+\delta x) = f(x) + \max(0,x)\delta x + \frac{1}{2}H_{x+\delta x}\delta x^2 + o(|\delta x|^2)$$

$$w \to \int_\Omega \max(0,w(\omega))^2 \, d\omega \quad \text{s.o.s.s. on } L_{2+\varepsilon}(\Omega)$$

Classical merit function:

$$\phi(x) = \frac{1}{2}\langle T(x), T(x)\rangle \qquad H_x := \langle T'(x), T'(x)\rangle$$

$$\phi'(x) = \langle T'(x)\delta x, T(x)\rangle$$

$$\begin{aligned} \phi(x_* + \delta x) \quad &\approx \langle T'(x_* + \delta x)\delta x, T'(x_* + \delta x)\delta x\rangle \\ &= \langle T(x_*) + T'(x_* + \delta x)\delta x, T(x_*) + T'(x_* + \delta x)\delta x\rangle \end{aligned}$$

Conclusions

UNIVERSITÄT BAYREUTH

Local convergence theory

Hölder Inequality & Continuity Result => Continuous Fréchet Differentiability

Hölder Inequality & Local Continuity Result => Semi-Smoothness

Issues of stability of Newton contraction

Global convergence

Difficulty with classical merit function

Optimization with semi-smooth derivatives

Transition to local convergence

Second order semi-smoothness