Revisiting Superlinear Convergence of Proximal Semismooth Newton Methods to Degenerate Solutions

Academia Sinica *LEE Ching-pei
University of Wisconsin-Madion Stephen J. Wright

1. Introduction

Consider the generalized equation problem:

$$\text{Find } x \in \mathcal{H} \quad \text{such that} \quad 0 \in (A+B)(x), \tag{1}$$

where $\mathcal{H}$ is a real Hilbert space with an inner product $\langle \cdot, \cdot \rangle$ and induced norm $\|\cdot\|$; A, B are set-valued operators; A is single-valued and continuous; B is maximal monotone; the solution set Ω is nonempty; and A is locally L-Lipschitz continuous in a neighborhood of Ω.

Clearly, a point x is a solution to (1) if and only if the forward-backward step

$$R(x) \coloneqq x - (\mathrm{Id} + B)(\mathrm{Id} - A)(x)$$

is zero, where Id is the identity operator. We define the norm of the forward-backward step as $r(x) \coloneqq \|R(x)\|$. We further assume that the order-q Hölderian error bound condition holds locally to Ω:

$$\begin{aligned} \operatorname{dist}(x, \Omega) &= \operatorname{dist}\big(x, (A+B)^{-1}(0)\big) \\ &\leq \kappa r(x)^q, \quad \forall x \in \{x \mid r(x) \leq \epsilon\}, \end{aligned} \tag{2}$$

for some positive values of κ, ϵ, and q, where $\operatorname{dist}(x, \Omega)$ is the distance between the point x and the set Ω. In our algorithm design, knowledge of the values of κ, ϵ, or q is not assumed, but in our analysis, fast local rates are attained only for a certain range of q (but has no restriction on κ and ϵ). In the presence of (2), it is clear that $x \in \Omega$ if and only $r(x) = 0$.

In our analysis, we will also assume that A is semismooth [1] of order $p \in (0,1]$ in a neighborhood of Ω. A function A is semismooth of order p at x if it is locally Lipschitz continuous in a neighborhood of x, directionally differentiable at x, and for any $A'(x + \Delta x) \in \partial(A(x+\Delta x))$ with $\Delta x \to 0$, we have

$$A(x+\Delta x) - A(x) - A'(x+\Delta x)\Delta x = O\left(\|\Delta x\|^{1+p}\right).$$

In this work, we analyze a damped variant of inexact forward-backward-semismooth-Newton method for (1) with the degenerate conditions such that A is differentiable only almost everywhere, and the generalized Jacobian of A at any $x^* \in \Omega$ could be singular. We first analyze local convergence rates for (1) under such degeneracy, but with the Hölderian error bound (2) and the semismoothness condition hold in a neighborhood of Ω, and show that superlinear convergence to 0 for both $r(x)$ and $\operatorname{dist}(x, \Omega)$ as well as full convergence of the iterates can be obtained for a range of p and q. We will then discuss the special case of regularized optimization

$$\min_{x \in \mathcal{H}} \quad F(x) \coloneqq f(x) + \Psi(x), \tag{3}$$

where $\Psi : \mathcal{H} \to [-\infty, \infty]$ is convex, proper, and closed; and f is differentiable in the domain of Ψ, with its gradient Lipschitz continuous locally in a neighborhood of Ω. Finally, a globalization strategy for (3) is discussed, which ensures global convergence even without the error bound or semismoothness condition, and can guarantee fast local superlinear convergence for not only $r(x)$ and $\operatorname{dist}(x, \Omega)$ but also $F(x)$.

2. Algorithm

Let us first define $\partial A(x)$ as the Clarke generalized Jacobian of A at x. At iteration t with iterate x_t, we then select an element $A'(x_t) \in \partial A(x_t)$, and update the iterate by

$$\begin{aligned} x_{t+1} &\approx (H_t + B)^{-1}(H_t - A)(x_t) \\ H_t &\coloneqq (\mu_t \mathrm{Id} + J_t), \quad \text{with} \end{aligned} \tag{4}$$

$$\mu_t \coloneqq c r(x_t)^\rho, \quad r(x_t) \coloneqq \|R(x_t)\|, \tag{5}$$

where $c > 0$ and $\rho \geq 0$ are parameters, and J_t is a positive semidefinite but possibly non-Hermitian linear operator satisfying

$$\left\|J_t - A'(x_t)\right\| = O\big(r(x_t)^\theta\big) \tag{6}$$

for some $\theta \geq \rho$. For the approximate solution in (4), let

$$\hat{x}_{t+1} \coloneqq (H_t + B)^{-1} (H_t - A)(x_t)$$

be the exact solution, we define $r_t(x) \coloneqq \|R_t(x)\|$,

$$\begin{aligned} R_t(x) &\coloneqq x- \\ &(\mathrm{Id} + B)^{-1} \left((H_t - A)(x_t) - (H_t - \mathrm{Id})(x) \right), \end{aligned} \tag{7}$$

and consider the following criterion with some $\nu \geq 0$ for the precision of x_{t+1} using (7):

$$r_t(x_{t+1}) \leq \nu r(x_t)^{1+\rho}. \tag{8}$$

3. Local Convergence

We define $d_t \coloneqq \mathrm{dist}(x_t, \Omega), r_t \coloneqq r(x_t), p_t \coloneqq x_{t+1} - x_t$, and have the following result.

Theorem 1 *Fix an iterate x_t and consider the update scheme (4)–(6) for (1) with A single-valued and continuous, B maximal monotone, and $\Omega \neq \emptyset$. Assume that x_{t+1} satisfies (8) for some given $\nu, \rho \geq 0$, (2) holds for some $q > 0$ in a neighborhood V of Ω, and A is locally Lipschitz continuous and semismooth of order p for some $p \in (0, 1]$ within the same neighborhood. If in addition to $\rho \geq 0$, the following are satisfied:*

$$\begin{cases} (1+\rho)q & > 1, \\ (1+p)q & > 1, \\ \left(1 + p - \frac{\rho}{q}\right)(1+p)q & > 1, \end{cases} \tag{9}$$

then we obtain Q-superlinear convergence

$$r_{t+1} = O\left(r_t^{1+s}\right), \quad d_{t+1} = O\left(d_t^{1+s}\right)$$

within V, where

$$s \coloneqq \min\left\{(1+\rho)q, (1+p)q, \left(1+p-\frac{\rho}{q}\right)(1+p)q\right\} - 1.$$

If further r_0 is small enough such that $\{r_t\}$ is a decreasing sequence, we have that $\{x_t\}$ converges strongly to some $x^* \in \Omega$.

4. Regularized Optimization

Our globalization strategy for (3) is in Algorithm 1.

Lemma 1 *Given $\beta, \gamma \in (0, 1)$, assume that f is L-Lipschitz-continuously differentiable for some $L > 0$, Ψ is convex, proper, and closed, and F is lower bounded by some $F^* > -\infty$. Then we have for Algorithm 1 that*

$$\lim_{t \to \infty} r_t = 0.$$

Algorithm 1: A Proximal-semismooth-Newton Method with Strict Decrease

input : $x_0 \in \mathcal{H}$, $\beta, \gamma \in (0,1)$, $\nu \in [0,1)$, $\rho \in (0,1]$, $c > 0$, $\delta \geq 0$

Compute an upper bound $\hat{L}$ for the Lipschitz constant L of ∇f

for $t = 0, 1, \ldots$ **do**

  Select H_t satisfying (5) and (6) and find an approximate solution $\hat{x}_{t+1}$ of (4) satisfying (8)

  $\alpha_t \leftarrow 1$

  while *True* **do**

    $y_{t+1} \leftarrow x_t + \alpha_t (\hat{x}_{t+1} - x_t)$

    Compute

$$\bar{x}_{t+1} \leftarrow \mathrm{prox}_{\frac{\Psi}{\hat{L}}}\left(y_{t+1} - \frac{\nabla f(y_{t+1})}{\hat{L}}\right)$$

    if $F(\bar{x}_{t+1}) \leq F(x_t) - \gamma \alpha_t^2 \|p_t\|^{2+\delta}$ **then**

      $x_{t+1} \leftarrow \bar{x}_{t+1}$ and break

    else $\alpha_t \leftarrow \beta \alpha_t$

Theorem 2 *Consider the setting of Theorem 1 for (3) and assume in addition that f is convex and the optimal objective value of (3) is $F^* > -\infty$. Consider Algorithm 1. If (9) holds and δ is large enough such that $\|p_t\|^{2+\delta} = o(d_t^{(q+1)/q})$, then there is $t_0 \geq 0$ such that for all $t \geq t_0$, $\alpha_t = 1$ is accepted and*

$$\begin{cases} d_{t+1} & = O\left(d_t^{1+s}\right), \\ r_{t+1} & = O\left(r_t^{1+s}\right), \\ F(x_{t+1}) - F^* & = O\big((F(x_{t+1}) - F^*)^{1+s} \big). \end{cases}$$

References

[1] Robert Mifflin. Semismooth and semiconvex functions in constrained optimization. 15(6):959–972, 1977.