A Generalized Univariate Newton Method Motivated by Proximal Regularization

Regina S. Burachik[*] C. Yalçın Kaya[†] Shoham Sabach[‡]

April 14, 2012

Abstract

We devise a new generalized univariate Newton method for solving nonlinear equations, motivated by Bregman distances and proximal regularization of optimization problems. We prove quadratic convergence of the new method, a special instance of which is the classical Newton's method. We illustrate the possible benefits of the new method over classical Newton's method by means of test problems involving the Lambert W function, Kullback-Leibler distance and a polynomial. These test problems provide insight as to which instance of the generalized method could be chosen for a given nonlinear equation. Finally, we derive a closed-form expression for the asymptotic error constant of the generalized method and make further comparisons involving this constant.

[*]School of Mathematics and Statistics, University of South Australia, Mawson Lakes, Australia 5095. Email: regina.burachik@unisa.edu.au

[†]School of Mathematics and Statistics, University of South Australia, Mawson Lakes, Australia 5095. Email: yalcin.kaya@unisa.edu.au

[‡]Department of Mathematics, The Technion – Israel Institute of Technology, 32000 Haifa, Israel. Email: ssabach@tx.technion.ac.il.

Key words. Newton-Raphson method, generalized Newton methods, numerical analysis, nonlinear equations, Bregman distances, anti-resolvent.

AMS subject classifications. 65H05, 49M15, 90C30, 65H04

1 Introduction

Newton's method and its variants are essential in solving nonlinear equations arising from problems in many disciplines such as mathematical programming, engineering, physics, and economics, just to name a few. As a result, new properties of Newton's method are studied and novel extensions constantly developed and tested, see, e.g. [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13] and the references therein. In the present paper, we consider the problem of finding a solution of a nonlinear equation in a single variable:

$$f(x) = 0\,, \tag{1}$$

where $f : X \subset \mathbb{R} \to \mathbb{R}$ is at least once continuously differentiable. In order to find a solution to (1) for a given initial approximation x_0, Newton's method generates a sequence $\{x_n\}_{n\in\mathbb{N}}$ by the rule

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}\,, \tag{2}$$

which is referred to as the *Newton iteration*.

The problem of finding a zero of a function of a single variable is widely encountered in optimization methods, e.g. as part of a subproblem in line-search algorithms, where powerful techniques are needed. It is well-known that, under standard assumptions, the sequence generated by the Newton iteration (2) is well-defined and it converges to a solution quadratically, provided x_0 is chosen close enough to the solution. In practice, modifications of Newton's method and its variants are needed to ensure global convergence. In this paper, we focus on local convergence properties.

Our generalized version of the Newton iteration formula (2) is given by

$$x_{n+1} := s^{-1}\left(s(x_n) - s'(x_n)\,\frac{f(x_n)}{f'(x_n)}\right), \tag{3}$$

where the function s is an invertible function with Lipschitz derivative in an interval containing the solution. The generalized method defined by (3) shares all the virtues of (2): it is well-defined and quadratically convergent near a solution (see Theorem 4.1). Note also that the choice $s(x) = x$ reduces to (2), hence (3) provides extra flexibility simply because Newton's method is one of its specific instances. On the other hand, (3) is a particular instance of (2), in which f is replaced by $f \circ s^{-1}$. More precisely, set $\tilde{f} := f \circ s^{-1}$, $y_n := s(x_n)$, and apply (2) to $\tilde{f}$, to get

$$y_{n+1} := y_n - \frac{\tilde{f}(y_n)}{(\tilde{f})'(y_n)}. \tag{4}$$

Because s is invertible in a neighbourhood of the solution, classical results on Newton's method imply that iteration (4) can be used to solve (1). In summary, both (3) and (4) can be used to solve (1), and both reduce to (2) when $s(x) = x$. This is why we call iteration (3) a "generalized Newton method". In our analysis, we will focus on iteration (3).

In the literature, "generalized Newton method" is also a description used for an analogue of Newton's method for solving nonlinear equations, but with a nondifferentiable f. In these non-differentiable approaches, the derivative f' in (2) is usually replaced by a suitable analogue; see, for example, [11] for semi-smooth Newton methods and [5] for generalized Newton methods based on the so-called graphical derivatives. A recent literature survey on nonsmooth Newton-like methods can also be found in [5]. Our generalized Newton's method is for solving equations with a differentiable f as in (1).

Our study is motivated by proximal regularization of convex optimization problems via Bregman distances, which lead to proximal point methods (see, e.g., [14, 15, 16] or [17, Sections 6.2-6.4]). It is well-known that these methods result in fixed point iterations whose convergence rate can be at most linear. In the present paper, an anti-resolvent formulation (see [15, Equation 3.1]) and the replacement of a parameter in the proximal regularization turn out to be the key in obtaining quadratically convergent fixed-point iterations. Numerical experiments suggest that some choices of s in (3) result in a better (i.e. faster and more robust) behaviour than that exhibited by (2), which is the case when $s(x) = x$ in (3).

At this point, we should say a few words regarding the practicality of the generalized Newton method we propose. At first sight, it may appear impractical having to choose a function s based on f for an instance of the generalized method. However, it should be kept in mind that Newton's method, where simply $s(x) = x$, does not provide the same flexibility. As long as $s(x)$ and its inverse $s^{-1}(x)$ are easy, i.e. cheap, to compute, our generalized method offers a flexibility which in turn can result in a faster and more robust algorithm, as we illustrate in the numerical examples.

The paper is organized as follows. In Section 2, standard convergence results for fixed-point methods are recalled. In Section 3, proximal regularization for convex problems using Bregman distances is described, and an anti-resolvent formulation of the fixed point iteration is introduced as a motivation for designing the generalized Newton's method. In Section 4, well-definedness and convergence of the new generalized method are proved. In Section 5, we provide a list of particular instances of our generalized Newton method, and illustrate some of these instances using three test problems. We also derive a closed-form expression for the asymptotic error constant of the generalized method and make further comparisons. Finally, in Section 6, we discuss possible further questions and extensions of our analysis.

2 Fixed Point Methods

Let $f : X \subset \mathbb{R} \to \mathbb{R}$ be a differentiable function. It is easy to see that Newton's method (2) is a fixed point method: Define

$$g(x) = x - \frac{f(x)}{f'(x)}. \tag{5}$$

Then a *fixed point* of g, that is, a point x that satisfies $g(x) = x$, is a zero of f.

We recall the following standard definitions about the rate of convergence of iterative methods, which we will use in the sequel.

Suppose that $\{x_n\}_{n\in\mathbb{N}}$ is a sequence that converges to x^*, with $x_n \neq x^*$ for all $n \in \mathbb{N}$. If there exists two positive constants α and λ such that

$$\lim_{n\to\infty} \frac{|x_{n+1} - x^*|}{|x_n - x^*|^\alpha} = \lambda, \tag{6}$$

then $\{x_n\}_{n\in\mathbb{N}}$ *converges to* x^* *of order* α *with asymptotic error constant* λ. Two cases are given special attention:

(i) If $\alpha = 1$ and $\lambda < 1$ then the sequence is *linearly convergent*;

(ii) if $\alpha = 2$ then the sequence is *quadratically convergent*.

Theorems 2.1 and 2.2 below furnish conditions under which fixed point iteration methods converge with a linear or a quadratic rate, respectively (see [18]).

Theorem 2.1 (Linear Rate of Convergence)**.** *Let* $g \in C^1[a, b]$ *be such that* $g(x) \in [a, b]$ *for any* $x \in [a, b]$*. Suppose, in addition, that there exists a positive constant* $L < 1$ *such that*

$$|g'(x)| \leq L, \quad \forall x \in (a, b).$$

If $g'(x^*) \neq 0$*, then for any number* $x_0 \neq x^*$ *in* $[a, b]$*, the sequence* $\{x_n\}_{n\in\mathbb{N}}$ *which is generated by* $x_{n+1} = g(x_n)$ *converges only linearly to the unique fixed point* x^* *in* $[a, b]$*.*

Theorem 2.2 (Quadratic Rate of Convergence)**.** *Let $g \in C^2[a,b]$ be such that $g(x^*) = x^*$, $g'(x^*) = 0$ and $|g''(x)| \le M$ for all $x \in I$, where $I \subset [a,b]$ is an open interval containing x^*. Then there exists $\delta > 0$ such that for $x_0 \in [x^* - \delta, x^* + \delta]$, the sequence defined by $x_{n+1} = g(x_n)$, for any $n \ge 0$, converges at least quadratically to x^*. Moreover, for sufficiently large values of n, the following inequality holds:*

$$|x_{n+1} - x^*| < \frac{M}{2} |x_n - x^*|^2 .$$

Theorems 2.1 and 2.2 together imply that for a fixed point method to converge quadratically one needs to have $g'(x^*) = 0$ with $g(x^*) = x^*$. In this situation, if $f(x^*) = 0$ and $f'(x^*) \neq 0$, then for starting values sufficiently close to x^*, Newton's method (2) will converge at least quadratically.

Theorem 2.2 requires the iteration function g to be twice differentiable. Quadratic convergence, however, can be established under weaker hypotheses, as shown in [21, Theorem 2.4.3]. We quote this result below.

Theorem 2.3 (Quadratic Rate of Convergence under Lipschitzian assumptions)**.** *Let $a, b \in \mathbb{R}$ such that $a < b$ and consider $F : (a,b) \to \mathbb{R}$ such that F' is Lipschitz in (a,b) with constant γ. Assume that for some $\rho > 0$, we have $|F'(x)| \ge \rho$ for all $x \in (a,b)$. If $F(x^*) = 0$ has a solution $x^* \in (a,b)$, then there exists some $\eta > 0$ such that: If $|x_0 - x^*| < \eta$ then the sequence $\{x_k\}$ generated by*

$$x_{k+1} := x_k - \frac{F(x_k)}{F'(x_k)}, \quad k = 0, 1, \ldots ,$$

exists and converges to x^*. Furthermore, for $k = 0, 1, \ldots,$

$$|x_{k+1} - x^*| \le \frac{\gamma}{2\rho} |x_k - x^*|^2. \tag{7}$$

3 Motivation by Proximal Regularization

Consider the problem of minimizing a convex function $\phi : \mathbb{R} \to \mathbb{R} \cup \{+\infty\} =: \overline{\mathbb{R}}$, namely,

$$\min_{x \in X} \phi(x), \tag{8}$$

where X is a closed interval in $\mathbb{R}$. Let $h : X \to \overline{\mathbb{R}}$, with $X \subset \mathbb{R}$, be a strictly convex function which is twice continuously differentiable. The *Bregman distance* (see, e.g., [14, 16, 17, 19] and the references therein) induced by a function h is the function $D_h : X \times \operatorname{int} X \to [0, +\infty]$ given by

$$D_h(y, x) = h(y) - h(x) - h'(x)(y - x). \tag{9}$$

If $h(t) = t^2$ then $X = \mathbb{R}$ and $D_h(y, x) = (x - y)^2$ for every $x, y \in \mathbb{R}$. It is common practice to "regularize" the minimization problem in (8) by adding a positive multiple of the Bregman distance to the objective function ϕ. The regularized problem will certainly have a solution because the Bregman distance has bounded level sets. Indeed, the function $D_h(\cdot, x)$ is convex and its level set $\{y : D_h(y, x) \leq 0\} = \{x\}$ is bounded. Hence all its level sets are bounded (see [20, Corollary 8.7.1]). The resulting numerical methods, when applied to the "regularized" problem, are expected to exhibit a better behaviour. Starting at a given $x_0 \in X$, the Bregman-proximal sequence $\{x_n\}$ is defined as follows. Given $x_n \in X$, let $x_{n+1} \in X$ be the solution of the problem:

$$\min_{x \in X} \phi(x) + \alpha_n D_h(x, x_n), \tag{10}$$

where $\alpha_n > 0$ and bounded above. The resulting sequence $\{x_n\}$ is well defined and contained in X under standard assumptions. Moreover, this sequence converges to a solution from any starting point x_0 if and only if solutions exist (see [16]). Under our differentiability assumptions, the convergence rate with positive α_n can at most be linear, as the following simple proposition states.

Proposition 3.1. *Let $h : X \to \mathbb{R}$ be a strictly convex function which is twice continuously differentiable, with $X \subset \mathbb{R}$ a closed convex and nonempty set. Let $\phi : \mathbb{R} \to \overline{\mathbb{R}}$ be convex and twice* *continuously differentiable. Consider the sequence $\{x_n\}$ defined by (10), where $0 < \alpha \leq \alpha_n \leq \beta$ for all n. If $\{x_n\}$ converges to a solution x^*, it does so at most linearly.*

Proof. The necessary and sufficient condition of optimality for problem (10) is

$$\frac{1}{\alpha_n}\,\phi'(x_{n+1}) + h'(x_{n+1}) - h'(x_n) = 0\,, \tag{11}$$

which can be solved implicitly for x_{n+1}, for $n \geq 1$. Indeed, since $\alpha_n > 0$, the iteration function $\eta_n(x) := (\phi'/\alpha_n + h')^{-1} \circ h'(x)$ associated with the proximal iteration (11) is well defined (because $(\phi'/\alpha_n + h')$ is invertible) and invertible for every $x \in \operatorname{int} X$. In other words, we can write

$$x_{n+1} = \eta_n(x_n) = \left(\frac{\phi'}{\alpha_n} + h'\right)^{-1} \circ h'(x_n)\,,$$

for all n. Denote by x^* the limit of $\{x_n\}$. Without loss of generality, we can assume that $\{\alpha_n\}$ converges to some $\alpha^* > 0$. Taking limits in the expression above yields

$$x^* = \left(\frac{\phi'}{\alpha^*} + h'\right)^{-1} \circ h'(x^*) =: \bar{\eta}(x^*)\,.$$

Note that x^* is a fixed point of both $\bar{\eta}$ and its inverse (which are well defined because $\alpha^* > 0$). Differentiating the equality $x = (\bar{\eta} \circ (\bar{\eta})^{-1})(x)$ and evaluating at x^* we obtain

$$1 \;=\; \bar{\eta}'(\bar{\eta}^{-1}(x^*))(\bar{\eta}^{-1})'(x^*) = \bar{\eta}'(x^*)(\bar{\eta}^{-1})'(x^*),$$

where we have used the fact that $\bar{\eta}^{-1}(x^*) = x^*$. Therefore, we cannot have $\bar{\eta}'(x^*) = 0$. Using Theorem 2.1 we conclude that convergence of (11) can at most be linear. □

As a result of the proposition above, we must use a negative parameter α in the iteration (11) if we wish to obtain quadratic convergence. This leads us to consider the iteration function

$$\widetilde{g}(x) := (h')^{-1}\left(h'(x) + \frac{1}{\alpha}\,\phi'(x)\right)\,. \tag{12}$$

Note that this function is well defined for any nonzero α, since h is strictly convex, and that it can be seen as the inverse of the proximal iteration function $\eta := (\phi'/\alpha + h')^{-1} \circ h'$, which may not exist for $\alpha < 0$. For $\alpha = -1$, the expression obtained above coincides with the *anti-resolvent* of ϕ with respect to h as defined by Butnariu and Kassay (see [15, Equation 3.1]).

Finding a solution of the minimization problem we stated in (8) is then equivalent to finding a fixed point of $\widetilde{g}$. By Theorem 2.2, quadratic convergence is achieved only in the case when $\widetilde{g}'(x^*) = 0$. Assume now that ϕ, h are strongly convex. Then we can differentiate both sides of (12) and using $\widetilde{g}(x^*) = x^*$, one obtains

$$\widetilde{g}'(x^*) = \frac{1}{h''(x^*)}\left(h''(x^*) + \frac{1}{\alpha}\,\phi''(x^*)\right).$$

From the above equality, we see that $\widetilde{g}'(x^*) = 0$ if and only if

$$\alpha = -\frac{\phi''(x^*)}{h''(x^*)}\,.$$

Hence the above choice of α guarantees quadratic convergence. Of course, in practice, this choice of α cannot be used in (12), because it requires the knowledge of x^*, which is not available. However, if the same functional form is substituted for α in (12), we obtain what we call here the *generalized Newton function*:

$$g(x) := (h')^{-1}\left(h'(x) - h''(x)\,\frac{\phi'(x)}{\phi''(x)}\right). \tag{13}$$

The function g depends only on the first and second derivatives of ϕ and h. So, for notational convenience, let

$$f := \phi' \quad \text{and} \quad s := h'\,,$$

and re-write the generalized Newton function g as

$$g(x) = s^{-1}\left(s(x) - s'(x)\,\frac{f(x)}{f'(x)}\right). \tag{14}$$

4 The Generalized Newton Method

The convexity assumptions on the functions ϕ and h have been posed in Section 3 mainly to relate the generalized method we propose with the Bregman distances and the associated proximal regularization of the problem of minimizing $\phi(x)$ given in (8). In fact, the strong convexity assumption we used for deriving (13) can be replaced by the weaker conditions $\phi''(x^*) \neq 0$ and $h''(x^*) \neq 0$, in other words, by $f'(x^*) \neq 0$ and $s'(x^*) \neq 0$.

The following trivial lemma formalizes some claims we stated in Introduction regarding iteration (4).

Lemma 4.1. *Let $y_n = s(x_n)$. The iteration formula (3) can be re-written in terms of the new variable iterate y_n as*

$$y_{n+1} = y_n - \frac{(f \circ s^{-1})(y_n)}{(f \circ s^{-1})'(y_n)}\,. \tag{15}$$

Proof. The iteration formula (3) can be re-written as

$$s(x_{n+1}) := s(x_n) - s'(x_n)\,\frac{f(x_n)}{f'(x_n)}\,, \tag{16}$$

Let $y_n = s(x_n)$. Then the iteration formula (16) becomes

$$y_{n+1} := y_n - \frac{f(s^{-1}(y_n))}{f'(s^{-1}(y_n))/s'(s^{-1}(y_n))}\,,$$

which simply yields (15). □

The iteration formula in Lemma 4.1 is nothing but the classical Newton iteration as applied to

$$(f \circ s^{-1})(y) = 0\,.$$

So our generalized method can be interpreted as the classical Newton's method applied for finding the transformed solution y^*, such that $(f \circ s^{-1})(y^*) = 0$, and that a solution of the original problem is obtained simply by $x^* = s^{-1}(y^*)$.

We will establish quadratic convergence by invoking Theorem 2.3, which requires Lipschitzianity of the first derivative of the iteration function. The following simple lemma establishes conditions under which $F := (f \circ s^{-1})'$ is Lipschitz.

Lemma 4.2. *Let $f, s \in C^1[a,b]$. Assume that for some $\theta > 0$, we have $0 < \theta < |s'(x)|$ for every $x \in (a,b)$. If f', s' are Lipschitz in (a,b), then $(f \circ s^{-1})'$ is well defined and Lipschitz in $I := s([a,b])$. Moreover, if L_f and L_s are the Lipschitz constants of f' and s', respectively, then the Lipschitz constant for $(f \circ s^{-1})'$ over I is*

$$M_{fs} := \frac{M_f\, L_s + M_s\, L_f}{\theta^3}\,, \tag{17}$$

where $M_f \geq \max\{|f'(x)| \,:\, x \in [a,b]\}$ and $M_s \geq \max\{|s'(x)| \,:\, x \in [a,b]\}$.

Proof. Note first that M_f can be chosen finite because by assumption f' is continuous on $[a,b]$. The assumptions on s' ensure that s is a bijection from $[a,b]$ to I. Therefore, $(f \circ s^{-1})$ is well defined on I. To establish the Lipschitzianity, take $y_1, y_2 \in I$, and denote by $J := f([a,b])$, the range of f restricted to the interval $[a,b]$. Then $(f \circ s^{-1}) : I \to J$, and there exist unique $z_1, z_2 \in [a,b]$ such that $s(z_1) = y_1$ and $s(z_2) = y_2$. We can write

$$\begin{aligned}
\left|(f \circ s^{-1})'(y_1) - (f \circ s^{-1})'(y_2)\right| &= \left|\frac{(f'(s^{-1}(y_1))}{s'(s^{-1}(y_1)} - \frac{(f'(s^{-1}(y_2))}{s'(s^{-1}(y_2)}\right| \\
&= \left|\frac{f'(z_1)s'(z_2) - f'(z_2)s'(z_1)}{s'(z_1)\, s'(z_2)}\right| \\
&= \left|\frac{f'(z_1)s'(z_2) - f'(z_1)s'(z_1) - f'(z_1)s'(z_1) - f'(z_2)s'(z_1)}{s'(z_1)\, s'(z_2)}\right| \\
&\leq \frac{1}{\theta^2}\left[\,|f'(z_1)|\; |s'(z_2) - s'(z_1)| + |s'(z_1)|\; |f'(z_1) - f'(z_2)|\,\right] \\
&\leq \frac{M_f\, L_s + M_s\, L_f}{\theta^2}\, |z_1 - z_2|
\end{aligned} \tag{18}$$

Use the Mean Value Theorem to write

$$|y_1 - y_2| = |s(z_1) - s(z_2)| = |s'(\psi)|\, |z_1 - z_2| \geq \theta |z_1 - z_2|,$$

for some $\psi \in (a,b)$, which yields $|z_1 - z_2| \le |y_1 - y_2|/\theta$. Combine the latter inequality with (18) to obtain

$$|(f \circ s^{-1})'(y_1) - (f \circ s^{-1})'(y_2)| \le \frac{M_f\, L_s + M_s L_f}{\theta^3} |y_1 - y_2| = M_{fs}|y_1 - y_2|,$$

as claimed. □

The following theorem states that the fixed-point iteration $x_{n+1} = g(x_n)$ generates a sequence $\{x_n\}_{n\in\mathbb{N}}$ which is convergent to a zero of f at least quadratically.

Theorem 4.1. *Let $f, s \in C^1[a,b]$ satisfy the assumptions of Lemma 4.2 and assume that there exists $x^* \in (a,b)$ such that $f(x^*) = 0$. Assume that for some $\rho > 0$, $|f'(x)| \ge \rho$ for every $x \in (a,b)$. Then there exists $\eta > 0$ such that if $x_0 \in (x^* - \eta, x^* + \eta)$ then the sequence $\{x_n\}$ defined recursively as*

$$x_{n+1} = g(x_n),\ k = 0, 1, 2, \ldots\,, \tag{19}$$

where g is as given in (14), is well-defined. In this situation, the sequence $\{x_n\}$ converges quadratically to x^*; in particular,

$$|x_{n+1} - x^*| \le \frac{M_{fs}\,(M_s)^3}{2\,\rho\,\theta}\,|x_n - x^*|^2\,, \tag{20}$$

where θ, M_{fs} and M_s are as in Lemma 4.2.

Proof. Since $s'(x) \ne 0$ for every $x \in (a,b)$,

$$s : I_1 := (a,b) \longrightarrow I_2\,, \tag{21}$$

is a C^1-diffeomorphism, where I_2 is either $(s(a), s(b))$ or $(s(b), s(a))$. In other words, $s : I_1 \to s(I_1) = I_2$ is C^1 with inverse $s^{-1} : I_2 \to I_1$ also C^1. Moreover, since $x^* \in (a,b)$, $y^* = s(x^*) \in I_2$, by virtue of the diffeomorphism of s. By Lemma 4.2, $(f \circ s^{-1})'$ is well defined and Lipschitz in (a,b) with constant M_{fs}. Since $|f'(x)| > \rho$ and $0 \ne |s'(x)| \le M_s$ for every $x \in (a,b)$, one has that, with

$y = s(x)$,

$$|(f \circ s^{-1})'(y)| = |f'(s^{-1}(y))/s'(s^{-1}(y))| = |f'(x)/s'(x)| > \rho/M_s,$$

for every $y \in I_2$. Consider the iteration formula given in (15) for y_n. Therefore, by applying Theorem 2.3 to $F = (f \circ s^{-1})$, there exists $\gamma > 0$ such that if $y_0 \in (y^* - \gamma, y^* + \gamma)$ then the sequence $\{y_n\}$ defined recursively as in (15) is well-defined. In this situation, the sequence $\{y_n\}$ converges quadratically to y^*; in particular,

$$|y_{n+1} - y^*| \leq \frac{M_{fs}\, M_s}{2\,\rho}\, |y_n - y^*|^2\,. \tag{22}$$

Clearly, y_n is in the domain of s^{-1}. Furthermore, by Lemma 4.1, the sequence $\{x_n\} = \{s^{-1}(y_n)\}$ is generated by (19). Since $\{y_n\}$ is well-defined, the sequence $\{x_n\}$ is also well-defined. Re-write (22) with $y_n = s(x_n)$:

$$|s(x_{n+1}) - s(x^*)| \leq \frac{M_{fs}\, M_s}{2\,\rho}\, |s(x_n) - s(x^*)|^2\,. \tag{23}$$

Using the Mean Value Theorem as in Lemma 4.2 we can write

$$|s(x_{n+1}) - s(x^*)| = |s'(\psi_1)|\, |x_{n+1} - x^*| \geq \theta |x_{n+1} - x^*|\,,$$

and we can also write

$$|s(x_n) - s(x^*)| = |s'(\psi_2)|\, |x_n - x^*| \leq M_s\, |x_n - x^*|\,,$$

for some $\psi_1, \psi_2 \in (a, b)$. Using the last two inequalities in (23), one gets

$$|x_{n+1} - x^*| \leq \frac{M_{fs}\,(M_s)^3}{2\,\rho\,\theta}\, |x_n - x^*|^2, \tag{24}$$

as desired. □

It should be pointed that iteration (19), with g defined as in (14), is in itself an iterative procedure unless s^{-1} is known analytically. In other words, for general s, evaluation of $g(x_n)$ using (14) requires carrying out an additional iterative scheme. As discussed in the forthcoming sections, our method is practical, because we choose functions s for which s^{-1} can be written down easily and computed cheaply.

4.1 Instances of the generalized Newton function

Several choices for the function h and the resulting generalized Newton function g, using (14), are listed in Table 1. Note that the choice of the Bregman function $h(x) = x^2/2$ in (13), which is the case when $p = 2$ in the first row of Table 1, yields the classical Newton function (5).

In Table 1, we have chosen h in such a way that the inverse of its derivative could be written in terms of elementary functions. Although our motivation was furnished by convex functions, the following proposition illustrates that h does not have to be convex.

Proposition 4.1. *Both functions h and $-h$ yield the same generalized Newton function g.*

Proof. Suppose that g is obtained by using $-h$. Then $s(x) = -h'(x)$, $s^{-1}(x) = (h')^{-1}(-x)$ and $s'(x) = -h''(x)$. Substituting these into (14) one gets

$$g(x) := (h')^{-1}\left(-\left(-h'(x) + h''(x)\,\frac{f(x)}{f'(x)}\right)\right) = (h')^{-1}\left(h'(x) - h''(x)\,\frac{\phi'(x)}{\phi''(x)}\right),$$

which is the same as the expression in (13), namely, the expression for $g(x)$ obtained using $h(x)$. □

In Table 1, the domain on which h is defined is also listed. We have taken $\mathbb{R}_+$ as the domain of $h(x) = x\log x - x$, since $\lim_{x\to 0^+} h(x) = 0$. It should be noted that some g in the list have a domain different from that of h. For example, while the domain of $h(x) = e^x$ is $\mathbb{R}$, the domain of the resulting g depends on the function f: g is defined whenever $f(x) < f'(x)$ (see the second row in Table 1). This might be restrictive for some problems; in other words, it may make g not applicable to some problems.

$h(x)$	$\operatorname{dom} h$	$s(x)$	$s^{-1}(x)$	$s'(x)$	$g(x)$
x^p/p, $p \in \mathbb{Z}\backslash\{0,1\}$	$\mathbb{R}$, for $p \geq 2$; $\mathbb{R}_{--}$ or $\mathbb{R}_{++}$, for $p \leq -1$	x^{p-1}	$x^{1/(p-1)}$	$(p-1)\,x^{p-2}$	$x^{(p-2)/(p-1)}\left(x-(p-1)\,\dfrac{f(x)}{f'(x)}\right)^{1/(p-1)}$
e^x	$\mathbb{R}$	e^x	$\log x$	e^x	$x+\log\left(1-\dfrac{f(x)}{f'(x)}\right)$
$x\,\log x - x$	$\mathbb{R}_+$	$\log x$	e^x	$1/x$	$x\,e^{-f(x)/(xf'(x))}$
$\log x$	$\mathbb{R}_{++}$	$1/x$	$1/x$	$-1/x^2$	$x\left(1+\dfrac{f(x)}{x\,f'(x)}\right)^{-1}$
$\cosh x$	$\mathbb{R}$	$\sinh x$	$\operatorname{arcsinh} x$	$\cosh x$	$\operatorname{arcsinh}\left(\sinh x-\cosh x\,\dfrac{f(x)}{f'(x)}\right)$
$\log(\cos x)$	$\left(-\dfrac{\pi}{2},\dfrac{\pi}{2}\right)$	$\tan x$	$\arctan x$	$\sec^2 x$	$\arctan\left(\tan x-\sec^2 x\,\dfrac{f(x)}{f'(x)}\right)$
$x\arctan x-(\log(x^2+1))/2$	$\mathbb{R}$	$\arctan x$	$\tan x$	$1/(1+x^2)$	$\tan\left(\arctan x-\dfrac{f(x)}{(1+x^2)f'(x)}\right)$

Table 1: Generalized Newton function g for several choices of function h.

4.2 Asymptotic error constant

We observe in the forthcoming sections 5.1-5.3 that a small asymptotic error constant λ is desirable for the generalized Newton method. With twice continuous differentiability of f and s, we can provide a closed-form expression for λ as follows.

Proposition 4.2. *Let $f, s \in C^2[a,b]$. Then*

$$\lambda = \frac{1}{2}\left|\frac{f''(x^*)}{f'(x^*)} - \frac{s''(x^*)}{s'(x^*)}\right| . \tag{25}$$

Proof. From the proof of Theorem 2.9 in [18], one has

$$\lambda = \frac{1}{2}\,|g''(x^*)| . \tag{26}$$

Now re-write (14) as

$$s(g(x)) = s(x) - s'(x)\,\frac{f(x)}{f'(x)} ,$$

and differentiate both sides to get

$$s'(g(x))\,g'(x) = s'(x) - s''(x)\,\frac{f(x)}{f'(x)} - s'(x)\left[\frac{(f'(x))^2 - f(x)\,f''(x)}{(f'(x))^2}\right] . \tag{27}$$

Because $f(x^*) = 0$, $f'(x^*) \neq 0$, $g(x^*) = x^*$ and $s'(x^*) \neq 0$, the expression in (27), when all functions in it are evaluated at $x = x^*$, simply yields $g'(x^*) = 0$.

Next, differentiate both sides of (27). It is not difficult to show that, with $f(x^*) = 0$, $g'(x^*) = 0$, $f'(x^*) \neq 0$ and $s'(x^*) \neq 0$, one gets, after manipulations and rearranging,

$$g''(x^*) = \frac{f''(x^*)}{f'(x^*)} - \frac{s''(x^*)}{s'(x^*)} , \tag{28}$$

which, combined with (26), completes the proof. □

It is interesting to note that the asymptotic error constant for the iteration formula given in (15) is given by

$$\widetilde{\lambda} = \frac{1}{2}\left|\frac{1}{s'(x^*)}\,\frac{f''(x^*)}{f'(x^*)} - \frac{s''(x^*)}{s'(x^*)}\right| , \tag{29}$$

which is different from λ given in (25).

5 Numerical Experiments

In this section, we discuss the behaviour of the generalized Newton iteration formula in (14) by means of three examples and various choices of function s. Our aim here is two-fold.

(i) Illustrate that the generalized Newton method works; and

(ii) gain insight as to which function s needs to be chosen for a particular problem.

In Tables 3-5, we tabulate the intervals $[a, b]$ in which the generalized Newton method converges in at most 5 and 10 iterations, respectively, under different choices of function s. All cases of the generalized Newton method, associated with different choices of function s, converge quadratically to a solution. We also tabulate the asymptotic error constant, λ, obtained through iterations numerically, for each function s.

In the intervals $[a, b]$ mentioned above, a and b are given as numbers with two decimal places. Convergence in at most 5 and 10 iterations are shown to be achieved for every initial approximation (or initial guess) x_0 chosen in the following way:

$$x_0 \in \{a + 0.01\, i \;:\; i = 0, 1, 2, \ldots, 100\,(b - a)\}. \tag{30}$$

For each of the examples in Sections 5.1-5.3, one gets

$$\lambda = \begin{cases} \dfrac{1}{2}\left|\dfrac{x^*+2}{x^*+1} - \dfrac{s''(x^*)}{s'(x^*)}\right|, & \text{for Example in Section 5.1;} \\[2ex] \dfrac{1}{2}\left|\dfrac{8}{x^*\,(x^*-8)} - \dfrac{s''(x^*)}{s'(x^*)}\right|, & \text{for Example in Section 5.2;} \\[2ex] \dfrac{1}{2}\left|\dfrac{6x^*}{3(x^*)^2+1} - \dfrac{s''(x^*)}{s'(x^*)}\right|, & \text{for Example in Section 5.3;} \end{cases}$$

Table 2 calculates the term $(s''(x)/s'(x))$ for those functions s that we used in the three examples in Sections 5.1-5.3.

$s(x)$	$s'(x)$	$s''(x)$	$s''(x)/s'(x)$
x	1	0	0
e^x	e^x	e^x	1
$\log x$	$1/x$	$-1/x^2$	$-1/x$
$1/x$	$-1/x^2$	$2/x^3$	$-2/x$
$\sinh x$	$\cosh x$	$\sinh x$	$\tanh x$
x^3	$3x^2$	$6x$	$2/x$
x^5	$5x^4$	$20x^3$	$4/x$
$1/x^3$	$-3/x^4$	$12/x^5$	$-4/x$
$\tan x$	$\sec^2 x$	$2\tan x \sec^2 x$	$2\tan x$
$\arctan x$	$1/(1+x^2)$	$-2x/(1+x^2)^2$	$-2x/(1+x^2)$

Table 2: Expressions with several choices of function s.

In Tables 3-5, we tabulate the exact values of λ, denoted by λ^*, correct to four decimal places. Note that the tabulated values of λ^* agree with those obtained numerically in Sections 5.1-5.3.

5.1 An example with the Lambert W function

Consider the function $\phi(x) = (x-1)\,e^x - x$, which is depicted in Figure 1. Then $\phi'(x) = x\,e^x - 1 = f(x)$ and $\phi''(x) = (x+1)\,e^x = f'(x)$. Note that $\phi(x)$ gives a nonconvex problem; in fact, $x = -1$ is a point of inflection, where $\phi''(x) = 0$ and concavity of the function changes: for $x > -1$, ϕ is convex, and, for $x < -1$, ϕ is concave. If x is optimal then $x\,e^x - 1 = 0$ which has a unique solution, $x^* = 0.567143290409784$ (correct to 15 decimal places), which is the value of the so-called Lambert w function [22] at 1. Note that x^* is the global solution. The Lambert w function arises

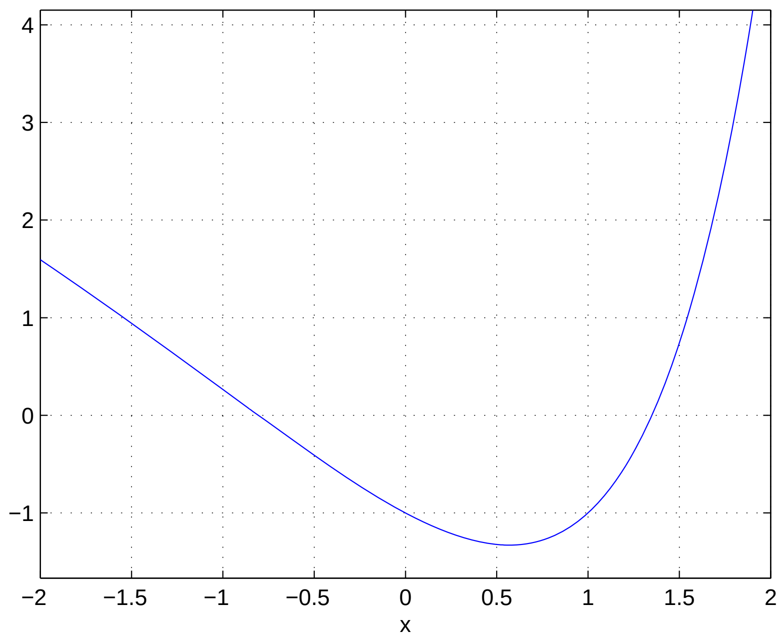

Figure 1: The graph of $\phi(x) = (x-1)\,e^x - x$.

in many applications involving exponential growth, e.g. rumour processes [23], where an efficient computation of its values is important.

Table 3 contains the results of numerical experiments for the function $\phi(x) = (x-1)\,e^x - x$ and several choices of function s.

For any x_0 chosen as in (30) in an interval in the third column of Table 3, the methods all take at most five iterations to find the solution correct to 15 decimal places. For example, when $s(x) = x$ is chosen, i.e. when the classical Newton's method is taken (see the first entry of Table 1 for $p = 2$), with any initial approximation x_0 in the interval $[0.44, 0.72]$ the classical Newton's method converges to the solution in at most five iterations. On the other hand, with any initial approximation x_0 in the interval $[-0.42, 3.78]$, the classical Newton's method converges to the solution in at most 10 iterations.

When the lengths of intervals in which a method converges in at most a given number of iterations

$s(x)$	λ	Interval for $N \le 5$	Int. length for $N \le 5$	Interval for $N \le 10$	Int. length for $N \le 10$	λ^*
$\tan x$	0.18	$[0.05, 0.92]$	0.87	$[-0.31, 1.17]$	1.48	0.1821
e^x	0.32	$[0.25, 0.92]$	0.67	$[-0.99, 9.40]$	10.39	0.3191
$\sinh x$	0.56	$[0.39, 0.78]$	0.39	$[-0.97, 8.40]$	9.37	0.5624
x	0.82	$[0.44, 0.72]$	0.28	$[-0.42, 3.78]$	4.20	0.8191
x^3	0.94	$[0.45, 0.69]$	0.24	$[0.01, 1.33]$	1.32	0.9442
$\arctan x$	1.2	$[0.48, 0.66]$	0.18	$[-0.06, 2.82]$	2.88	1.2482
$\log x$	1.7	$[0.51, 0.64]$	0.13	$[0.22, 2.84]$	2.62	1.7007
$1/x$	2.6	$[0.53, 0.61]$	0.08	$[0.35, 2.27]$	1.92	2.5823
x^5	2.7	$[0.52, 0.61]$	0.09	$[0.03, 0.73]$	0.70	2.7074
$1/x^3$	4.3	$[0.55, 0.59]$	0.04	$[0.45, 1.63]$	1.18	4.3455

Table 3: Numerical experiments for $f(x) = x\,e^x - 1 = \phi'(x) = 0$.

is taken into account (as some measure of robustness), the generalized Newton method with $s(x) = e^x$ and $s(x) = \sinh x$ looks certainly superior over the classical Newton's method. It is interesting to note that these two instances also have smaller asymptotic error constants, λ. Although the generalized Newton method with $s(x) = \tan x$ has the smallest λ of all choices in Table 3, it doesn't appear as robust as the methods with $s(x) = e^x$ and $s(x) = \sinh x$.

The interval in which the classical Newton's method converges is rather large for this problem. On the other hand, for $x_0 = 10$, while Newton's method takes 18 iterations to converge, the generalized method with $s(x) = e^x$ and $s(x) = \sinh x$ takes 11 iterations in each instance. The method with $s(x) = \tan x$ does not converge at all with $x_0 = 10$. For $x_0 = 100$, while Newton's method converges in 110 iterations, the generalized method with $s(x) = e^x$ and $s(x) = \sinh x$ converges in 34 and 35 iterations, respectively.

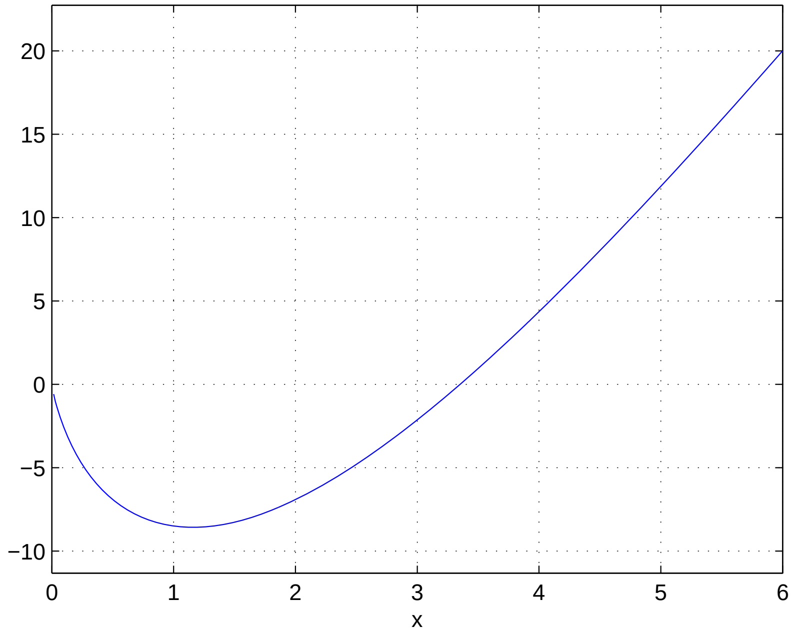

Figure 2: The graph of $\phi(x) = 8\,(x\,\log x - x) - x^2/2$.

In $\phi(x)$, e^x certainly looks like the "dominating" function, and so "prompts" one to consider either e^x or $\sinh x$ as a candidate for s, as an alternative to the classical Newton method's $s(x) = x$.

5.2 An example with Kullback-Leibler distance

Consider the function $\phi(x) = 8\,(x\,\log x - x) - x^2/2$, which is depicted in Figure 2. Then $\phi'(x) = 8\,\log x - x = f(x)$ and $\phi''(x) = 8/x - 1 = f'(x)$. The term $(x\,\log x - x)$ is referred to as the Kullback-Leibler distance. The problem has two stationary points: a minimum at $x^* = 1.155370825100078$ and a maximum at $x^* = 26.093485476611910$, both correct to 15 decimal places. We will focus on the behaviour of the generalized Newton method with various choices of $s(x)$ around the minimum point.

Table 4 provides insight as to which $s(x)$ should be employed in the generalized Newton's method. Clearly the method with $s(x) = \arctan x$, results in the "best" behaviour in terms of both small λ and robustness. The generalized Newton method with $s(x) = \log x$, which appears as the "dominating" term in f, seems to be equally preferable, over the classical Newton's method.

$s(x)$	λ	Interval for $N \leq 5$	Int. length for $N \leq 5$	Interval for $N \leq 10$	Int. length for $N \leq 10$	λ^*
$\arctan x$	0.01	$[0.42, 5.95]$	5.53	$(0.00, 7.18]$	7.18	0.0110
$\log x$	0.07	$[0.05, 2.57]$	2.52	$(0.00, 7.98]$	7.98	0.0730
$1/x$	0.36	$[0.92, 1.54]$	0.62	$[0.44, 7.46]$	7.02	0.3597
x	0.51	$[0.95, 1.37]$	0.42	$[0.01, 2.71]$	2.70	0.5058
$\sinh x$	0.92	$[1.03, 1.27]$	0.24	$[0.01, 1.86]$	1.85	0.9156
e^x	1.0	$[1.04, 1.26]$	0.22	$[0.01, 1.78]$	1.77	1.0058
$1/x^3$	1.2	$[1.08, 1.26]$	0.18	$[0.83, 7.97]$	7.14	1.2252
x^3	1.4	$[1.07, 1.23]$	0.16	$[0.11, 1.59]$	1.48	1.3713
x^5	2.2	$[1.10, 1.20]$	0.10	$[0.28, 1.40]$	1.12	2.2369
$\tan x$	2.8	$[1.11, 1.19]$	0.08	$[0.03, 1.34]$	1.31	2.7729

Table 4: Numerical experiments for $f(x) = 8\,\log x - x = \phi'(x) = 0$.

5.3 A polynomial example

Consider the problem of minimizing the polynomial function, $\phi(x) = x^4/4 + x^2/2 - 3x + 1$. Then $\phi'(x) = x^3 + x - 3 = f(x)$ and $\phi''(x) = 3\,x^2 + 1 = f'(x)$. The problem has the unique solution, $x^* = 1.213411662762230$ (correct to 15 decimal places).

Table 5 provides insight as to which generalized Newton methods can be favored in this case. Clearly the method with $s = x^3$, which is the highest power term in the polynomial in f, yields the "best" behaviour in terms of small λ and robustness. The generalized Newton method with $s(x) = e^x$ and $s(x) = \sinh x$ looks favourable over the classical Newton's method, because of a

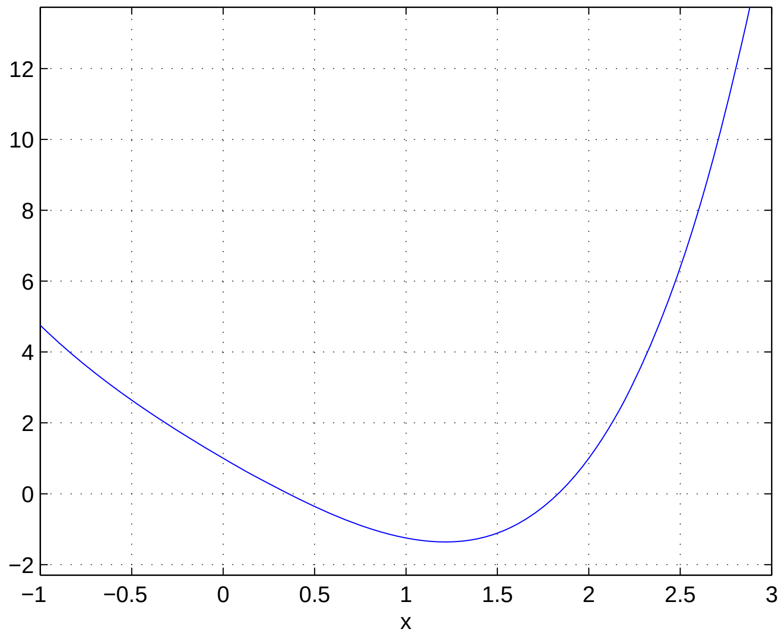

Figure 3: The graph of $\phi(x) = x^4/4 + x^2/2 - 3x + 1$.

smaller λ, for finding the solution in a small neighbourhood of the solution.

Starting far from the solution, say at $x_0 = 10$, while Newton's method takes 11 iterations to converge, the generalized Newton method with $s(x) = x^3$ takes 7 iterations; on the other hand, the method with $s(x) = e^x$ and $s(x) = \sinh x$ does not yield a solution. Further away with $x_0 = 100$, while Newton's method takes 16 iterations, the method with $s(x) = x^3$ still takes 7 iterations. Robustness of the method with $s(x) = x^3$ is evident from the length of interval, 10^8, for finding the solution in at most 10 iterations.

$s(x)$	λ	Interval for $N \leq 5$	Int. length for $N \leq 5$	Interval for $N \leq 10$	Int. length for $N \leq 10$	λ^*
x^3	0.15	$[0.71, 2.05]$	1.34	$[0.01, 10^8]$	10^8	0.1521
e^x	0.17	$[0.72, 2.32]$	2.60	$[-3.46, 4.67]$	8.13	0.1720
$\sinh x$	0.25	$[0.87, 2.38]$	1.51	$[-4.95, 5.14]$	10.09	0.2531
x	0.67	$[1.06, 1.39]$	0.33	$[-4.50, 9.18]$	13.68	0.6720
x^5	0.98	$[1.09, 1.32]$	0.23	$[0.02, 1.86]$	1.84	0.9763
$\log x$	1.1	$[1.12, 1.33]$	0.21	$[0.58, 5.82]$	6.40	1.0841
$\arctan x$	1.2	$[1.12, 1.32]$	0.20	$[0.61, 4.87]$	4.26	1.1628
$1/x$	1.5	$[1.14, 1.29]$	0.15	$[0.79, 4.40]$	5.19	1.4961
$\tan x$	2.0	$[1.15, 1.26]$	0.11	$[-1.23, 1.43]$	2.66	2.0060
$1/x^3$	2.3	$[1.17, 1.27]$	0.10	$[0.97, 3.17]$	4.14	2.3202

Table 5: Numerical experiments for $f(x) = x^3 + x - 3 = \phi'(x) = 0$.

6 Discussion and Conclusion

We have introduced a new generalized Newton method with quadratic convergence for solving univariate equations. The method is given by a single iteration formula with a general function s to choose in the formula. The only theoretical requirement on s is that it has to be continuously differentiable once, Lipschitz, and also that $s'(x^*) \neq 0$, which are the same requirements as those on f. We practically require $s^{-1}(y)$ to be computed easily, or cheaply, for a given y. In this sense, the range of choices for s is vast. In the numerical experiments, it has been observed that an s chosen to be "similar" to the "dominating" functional term in f seems to result in a faster (i.e. a smaller asymptotic constant, λ) and a more robust behaviour of the generalized method.

Future work should include a more comprehensive investigation as to which choices of s would yield a better behaviour of the generalized method. For the case when f and s are both twice continuously differentiable, the expression (25) derived in this paper for λ can perhaps be used (over an interval containing a solution) to find out *a priori* which choice of s would work better for the generalized method.

As a line of further research, we aim to study an extension of Kantorovich's sufficient condition for the convergence of the classical Newton's method [24] to our setting of the generalized Newton method. Our analysis also opens the way to developing an extension of our method for nonlinear systems of equations in n variables.

7 Acknowledgments

Authors would like to thank Gábor Kassay for motivating discussions on anti-resolvents of proximal-projection methods during his visit to the School of Mathematics and Statistics at the University of South Australia. The iteration formula (15) in Lemma 4.1 is an interpretation of Roberto Cominetti after seeing our generalized Newton iteration formula in (3). The authors are indebted to him for his observation, which consequently made the proof of Theorem 4.1 neater. They also thank the referees whose comments and suggestions improved the paper.

References

[1] Aragón-Artacho, F. J., Dontchev, A. L., Gaydu, M., Geoffroy, M. H., Veliov, V. M.: Metric regularity of Newton's iteration. SIAM J. Control Optim. 49, 339–362 (2011)

[2] Diniz-Ehrhardt, M. A., Gomes-Ruggiero, M. A., Lopes, V. L. R., Martínez, J. M.: Discrete Newton's method with local variations for solving large-scale nonlinear systems. Optimization 52, 417–440 (2003)

[3] Dontchev, A. L., Hager, W. W., Veliov, V. M.: Uniform convergence and mesh independence of Newton's method for discretized variational problems. SIAM J. Control Optim. 39, 961–980 (2000)

[4] Ford, W. F., Pennline, J. A.: Accelerated convergence in Newton's method. SIAM Review 38, 658–659 (1996)

[5] Hoheisel, T., Kanzow, C., Mordukhovich, B. S., Phan, H.: Generalized Newton's method based on graphical derivatives. Nonlinear Analysis, Theory, Methods and Applications 75, 1324–1340 (2012)

[6] Kelley, C. T.: Solving Nonlinear Equations with Newton's Method. SIAM, Philadelphia (2003)

[7] Martínez, J. M.: Fixed-point quasi-Newton methods. SIAM J. Num. Anal. 29, 1413–1434 (1992)

[8] Martínez, J. M.: Practical quasi-Newton methods for solving nonlinear systems. Journal of Computational and Applied Mathematics 124, 97–122 (2000)

[9] Ortega, J. M., Rheinboldt, W. C.: Iterative Solution of Nonlinear Equations in Several Variables. Academic Press, New York (1970)

[10] Polyak, B. T.: Newton's method and its use in optimization. European J. Operational Research 181, 1086–1096 (2007)

[11] Qi, L., Sun, J.: A nonsmooth version of Newton's method. Math. Prog. 58, 353–367 (1993)

[12] Thorlund-Petersen, L.: Global convergence of Newton's method on an interval. Mathematical Methods of Operations Research 59, 91–110 (2004)

[13] Tseng, C.-L.: A Newton-type univariate optimization algorithm for locating the nearest extremum. European J. of Operational Research 105, 236–246 (1998)

[14] Burachik, R. S., Iusem, A. N., A generalized proximal point algorithm for the variational inequality problem in a Hilbert space. SIAM J. Optim. 8, 197–216 (1998)

[15] Butnariu, D., Kassay, G.: A proximal-projection method for finding zeroes of set-valued operators. SIAM J. Control Optim. 47, 2096–2136 (2008)

[16] Chen, G., Teboulle, M.: Convergence analysis of a proximal-like minimization algorithm using Bregman functions. SIAM J. Optim. 3, 583–543 (1993)

[17] Burachik, R. S., Iusem, A. N.: Set-Valued Mappings and Enlargements of Monotone Operators. Springer, Series: Optimization and Its Applications 8, New York, USA (2008)

[18] Burden, R. L., Faires, J. D.: Numerical Analysis, 9th edition. Thompson Brooks/Cole, Belmont, CA, USA (2011)

[19] Bregman, L.: The relaxation method of finding the common points of convex sets and its application to the solution of problems in convex programming. U.S.S.R. Comput. Math. and Math. Phys. 7, 200–217 (1967)

[20] Rockafellar, R.T.: Convex Analysis. Princeton Univ. Press (1970)

[21] Dennis Jr., J. E., Schnabel, R. B.: Numerical Methods for Unconstrained Optimization and Nonlinear Equations. Prentice Hall, Englewood Cliffs, NJ (1983)

[22] Corless, R. M., Hare, D. E. G., Jeffrey, D. J., Knuth, D. E.: On the Lambert W function. Advances in Computational Mathematics 5, 329–359 (1996)

[23] Belen, S., Kaya, C. Y., Pearce, C. E. M.: Impulsive control of rumours with two broadcasts. ANZIAM J. 46, 379–391 (2005)

[24] Ferreira O. P., Svaiter, B. F.: Kantorovich's majorants principle for Newton's method. Comput. Optim. Appl. 42, 213–229 (2009)