PLISKA
STUDIA MATHEMATICA
BULGARICA

ПЛИСКА
БЪЛГАРСКИ
МАТЕМАТИЧЕСКИ
СТУДИИ

Pliska Stud. Math. Bulgar. **12** (1998), 31-38

PLISKA
STUDIA MATHEMATICA
BULGARICA

THE GRAVES THEOREM REVISITED II: ROBUST CONVERGENCE OF THE NEWTON METHOD*

Asen L. Dontchev

Based on the original proof of the Graves theorem [9] we study the convergence of the Newton method for the solution of the equation $f(x) = y$, uniform with respect to the starting point and the parameter y. We show that the surjectivity of the Jacobian implies the Aubin continuity, relative to the supremum norm, of the map taking the starting point and the parameter y to the set of all Newton sequences. These results complement our previous paper [4].

Keywords: Newton's method, Aubin property, robust convergence.

AMS subject classification: 65J15, 47H04, 90C30.

1 The Graves theorem

As in our previous paper [4] devoted to the Graves theorem, we start with the formulation and the proof of the result given in the original paper of Graves [9]. It uses the following version of the Banach open mapping theorem: Let X and Y be Banach spaces and $A : X \to Y$ be a linear and continuous. Then A is onto if and only if there is a constant $M > 0$ such that for every operator $y \in Y$ there exists $x \in X$ such that $y = A(x)$ and

$$\| x \| \leq M \| y \| . \tag{1}$$

In the sequel X, Y are Banach spaces and $B_a(x)$ is the closed ball centered at x with radius a.

Theorem 1.1 *(Graves [9]). Let f be a continuous function from X to Y defined in $B_\varepsilon(0)$ for some $\varepsilon > 0$ with $f(0) = 0$. Let A be a continuous and linear operator from*

*This work was supported by The National Science Foundation. The revised version of this paper was prepared during author's visit at the University of Zürich, Switzerland.

X onto Y and let M be the corresponding constant in (1). Suppose that there exists a constant $\delta < M^{-1}$ such that

$$\| f(x_1) - f(x_2) - A(x_1 - x_2) \| \leq \delta \| x_1 - x_2 \| \tag{2}$$

whenever $x_1, x_2 \in B_\varepsilon(0)$. Then the equation $y = f(x)$ has a solution $x \in B_\varepsilon(0)$ whenever $\| y \| \leq c\varepsilon$, where $c = M^{-1} - \delta$.

Proof. Let $y \in Y$, $\| y \| \leq c\varepsilon$, and let $x_0 = 0$. Since A is surjective, by (1) there exists $x_1 \in X$ such that

$$A(x_1) = y \text{ and } \| x_1 \| \leq M \| y \| \leq \varepsilon.$$

Suppose that for $n1$ we are given $x_i, i = 1, \cdots, n-1$, satisfying

$$A(x_i) = y - f(x_{i-1}) + A(x_{i-1}) \text{ and } \| x_i - x_{i-1} \| \leq M(M\delta)^{i-1} \| y \| .$$

Then

$$\begin{aligned} \| x_i \| &\leq \sum_{j=1}^{i} \| x_j - x_{j-1} \| \\ &\leq M \| y \| \sum_{j=1}^{i} (M\delta)^{j-1} \leq M \| y \| /(1 - M\delta) = \|y\|/c \leq \varepsilon. \end{aligned}$$

By the surjectivity of A and (1), there exists an x_n such that

$$A(x_n) = y - f(x_{n-1}) + A(x_{n-1}) \tag{3}$$

and

$$\| x_n - x_{n-1} \| \leq M \| y - f(x_{n-1}) \| .$$

Since $y = A(x_{n-1}) - A(x_{n-2}) + f(x_{n-2})$, from (2) we have

$$\| x_n - x_{n-1} \| \leq M\delta \| x_{n-1} - x_{n-2} \| .$$

Hence,

$$\| x_n - x_{n-1} \| \leq M(M\delta)^{n-1} \| y \| .$$

Thus x_n is a Cauchy sequence, hence it is convergent to some $x \in B_\varepsilon(0)$. Passing to the limit in (3) with $n \to \infty$ we obtain $y = f(x)$. $\square$

Let us assume that the function f is continuously differentiable (for short, $f \in C^1$) around 0. Then the Graves theorem can be stated as follows: If $\nabla f(0)$ is onto, then there exist a neighborhood U of 0 and a constant $c > 0$ such that for every $x \in U$ and $\tau > 0$ with $B_\tau(x) \subset U$,

$$B_{c\tau}(f(x)) \subset f(B_\tau(x)). \tag{4}$$

The property (4) is called *linear openness, openness with linear rate around a point* or *covering in a neighborhood* and can be extended to set-valued mappings acting in metric spaces. In very general circumstances the linear openness of a mapping is equivalent to both the Aubin property and the metric regularity of its inverse. Recall that a set-valued map F from X to the subsets of Y is *Aubin continuous at* $(y^*, x^*) \in \text{graph}F$ if there exist constants a, b, and M such that for every $y', y'' \in B_b(y^*)$ and for every $x' \in F(y') \cap B_a(x^*)$ there exists $x'' \in F(y'')$ with

$$\|x' - x''\| \leq M\|y' - y''\|.$$

Remark 1.1 *By the definition, if a map F is Aubin continuous at (y^*, x^*) with constants a, b, and M, then for every $0 < a' < a$ and every $0 < b' < min\{b, a'/M\}$ the map F is Aubin continuous at (y^*, x^*) with constants a', b', and M, and moreover $F(y) \cap B_{a'}(x^*) \neq \emptyset$ for all $y \in B_{b'}(y^*)$.*

We showed in [4] that the Graves theorem is a special case of the following general observation: the Aubin property is stable under perturbations of the inverse of order $o(x)$. In particular, for smooth functions we have the following characterization:

Theorem 1.2 *Let $f \in C^1$ around 0. Then the following are equivalent:*
(i) $\nabla f(0)$ is onto;
(ii) the map f^{-1} is Aubin continuous at $(0, 0)$.

The Graves theorem is sometimes viewed as a version of a theorem by Lyusternik [11] which is stated as follows: if a function f from Banach space X into a Banach space Y is Fréchet differentiable near x_0, its derivative ∇f is continuous at x_0, and $\nabla f(x_0)$ is onto, then the tangent manifold to $f^{-1}(0)$ at x_0 is exactly $x_0 + \text{Ker}\nabla f(x_0)$. Actually, the Lyusternik theorem can be deduced from the Graves theorem, see [4]. On the other hand, it is important to note that in his proof Lyusternik used the iterative process (3), with $A = \nabla f(0)$. The Lyusternik theorem and the Graves theorem and their various extensions have numerous applications in variational analysis and optimization. Here we shall not go into this further, for samples of results in this direction see [1, 2, 6, 7, 8, 10, 12, 13, 14, 15].

2 The Newton method

Let us go back to the proof of the Graves theorem. Suppose that $f \in C^1$ around 0. Then for every $\delta > 0$ there exists $\varepsilon > 0$ such that for every $x_1, x_2 \in B_\varepsilon(0)$,

$$\| f(x_1) - f(x_2) - \nabla f(0)(x_1 - x_2) \| \leq \delta \| x_1 - x_2 \| .$$

If $\nabla f(0)$ is surjective, with a constant M in (1), then the operator A in the proof of Graves can be replaced by $\nabla f(0)$ provided that $\delta < M^{-1}$. The points x_n, $n = 1, 2, \cdots$, obtained by the iterative procedure (3) satisfy

$$y = f(x_{n-1}) + \nabla f(0)(x_n - x_{n-1}).$$

This procedure is a version of the Newton method, sometimes called the *modified Newton method*, for solving the equation $y = f(x)$. That is, to prove the linear openness of the map f, Graves applied a Newton-type method proving simultaneously its geometric convergence. Note that this convergence is independent of the choice of the starting point x_0 and the parameter y. The result is not of the conventional type where the Jacobian $\nabla f(0)$ is assumed invertible. In his proof Graves assumes surjectivity of the Jacobian and proves the existence of a convergent Newton-type sequence. Recall that the standard version of the Newton method is:

$$y = f(x_{n-1}) + \nabla f(x_{n-1})(x_n - x_{n-1}). \tag{5}$$

In the following theorem we show that, by modifying the proof of Graves, one obtains the existence of a sequence x_n satisfying (5) which converges Q-superlinearly to a solution, uniformly in the starting point and the parameter.

Theorem 2.1 *Suppose that $f(0) = 0$, $f \in C^1$ around 0 and $\nabla f(0)$ is onto. Then there exist positive constants ρ and ε such that for every $y \in B_\rho(0)$ and for every initial point $x_0 \in B_\varepsilon(0)$ there exists a Newton sequence $x_n \in B_\varepsilon(0)$ which converges Q-superlinearly, uniformly in $x_0 \in B_\varepsilon(0)$ and $y \in B_\rho(0)$, to a solution $x(y)$ of the equation*

$$y = f(x). \tag{6}$$

Proof. Let $M' > M$ and let $\delta > 0$ be such that $\delta(M'+1) < 1$. Choose $\varepsilon > 0$ such that for every $x, x_1, x_2 \in B_\varepsilon(0)$,

$$\| f(x_1) - f(x_2) - \nabla f(x)(x_1 - x_2) \| \leq \delta \| x_1 - x_2 \| \tag{7}$$

and, moreover, for every $x \in B_\varepsilon(0)$ the following holds: for every $y \in Y$ there exists $z \in X$ such that

$$y = \nabla f(x) z \quad \text{and} \quad \|z\| \leq M'\|y\|. \tag{8}$$

Let $\rho > 0$ be such that

$$\frac{M'\rho + \delta\varepsilon}{1 - M'\delta} \leq \varepsilon.$$

Let $x_0 \in B_\varepsilon(0)$ and $y \in B_\rho(0)$. From (8) there exists a Newton step x_1,

$$x_1 \in \nabla f(x_0)^{-1}(y - f(x_0) + \nabla f(x_0)x_0), \tag{9}$$

such that

$$\begin{aligned} \|x_1\| &\leq M'\|y - f(x_0) + \nabla f(x_0)x_0\| \\ &\leq M'(\|y\| + \|f(x_0) - f(0) - \nabla f(x_0)(x_0 - 0)\|) \\ &\leq M'(\|y\| + \delta\|x_0\|) \leq M'\rho + \delta\varepsilon \leq \varepsilon. \end{aligned}$$

Then, again from (8), we can find a Newton step x_2, that is,

$$x_2 - x_1 \in \nabla f(x_1)^{-1}(y - f(x_1)),$$

with

$$\|x_2 - x_1\| \le M'\|y - f(x_1)\| = M'\|f(x_0) + \nabla f(x_0)(x_1 - x_0) - f(x_1)\| \le M'\delta\|x_1 - x_0\|.$$

By induction, suppose that the sequence $x_1, x_2, \cdots, x_{n-1}$ satisfies

$$\|x_i - x_{i-1}\| \le (M'\delta)^{i-1}\|x_1 - x_0\|$$

and $\|x_i\| \le \varepsilon$ for all $i = 1, 2, \cdots, n-1$. Then there exists

$$x_n \in x_{n-1} + \nabla f(x_{n-1})^{-1}(y - f(x_{n-1}))$$

such that

$$\|x_n - x_{n-1}\| \le M'\|y - f(x_{n-1})\| \le M'\delta\|x_{n-1} - x_{n-2}\|.$$

Hence

$$\|x_n - x_{n-1}\| \le (M'\delta)^{n-1}\|x_1 - x_0\|.$$

Further, from

$$x_n \in \nabla f(x_{n-1})^{-1}(y - f(x_{n-1}) + \nabla f(x_{n-1})x_{n-1})$$

we have

$$\begin{aligned} \|x_n\| &\le M'\|y - f(x_{n-1}) + \nabla f(x_{n-1})x_{n-1}\| \\ &\le M'(\|y\| + \delta\|x_{n-1}\|) \le M'\rho + \delta\varepsilon \le \varepsilon. \end{aligned}$$

Thus for every n, x_n is in $B_\varepsilon(0)$ and the sequence $\{x_n\}$ is convergent. Passing to the limit with $n \to \infty$ in (5) we obtain that $\{x_n\}$ is convergent to a solution $x(y)$ of (6). Furthermore, observe that for every n,

$$\begin{aligned} \|x_{n+1} - x_n\| &\le M'\|y - f(x_n)\| \\ &= M'\|f(x_{n-1}) + \nabla f(x_{n-1})(x_n - x_{n-1}) - f(x_n)\| \\ &\le M'\|\int_0^1 (\nabla f(x_{n-1} + t(x_n - x_{n-1})) - \nabla f(x_{n-1}))dt\|\|x_n - x_{n-1}\|, \end{aligned}$$

that is,

$$\frac{\|x_{n+1} - x_n\|}{\|x_n - x_{n-1}\|} \le M'\|\int_0^1 (\nabla f(x_{n-1} + t(x_n - x_{n-1})) - \nabla f(x_{n-1}))dt\|.$$

Since ∇f is continuous in $B_\varepsilon(0)$ and $x_n \to x(y)$ uniformly in x_0 and y, the convergence is Q-superlinear and uniform. □

Of course, the origin 0 of X does not play any special role in the above proof; it can be replaced by any fixed point x^* and then the origin of Y will be replaced by $f(x^*)$. Further, if we assume that the Jacobian ∇f is Lipschitz continuous around x^*, then we obtain the existence of a Q-quadratic convergent Newton sequence, see [3].

We note that Theorem 2.1 is an existence theorem; it does not show how to construct a Newton sequences. A natural choice is to take the iterate with minimal norm; this means at each iteration to solve a least squares problem.

Theorem 2.1 is a special case of Theorem 1 in [5], where we considered a generalized equation of the form $y \in f(x) + F(x)$, with f a smooth function and F a set-valued map acting in Banach spaces. In [5], however, no connection to the Graves theorem had been made and the proof was much more involved. In further lines we will not discuss more general models (e.g. variational inequalities) but rather focus on an issue of different nature: how a Newton sequence depends on changes of the parameter y.

3 Well-posedness of Newton sequences

Let $\mathcal{X}$ be the space of infinite sequences $x_1, x_2, \cdots, x_n, \cdots$, denoted $\{x_n\}$, of elements of X equipped with the norm

$$\|\{x_n\}\| = \sup_{n \geq 1} \|x_n\|.$$

We denote by $N(x_0, y)$ the set of all Newton sequences in $\mathcal{X}$ starting at the initial point x_0 and associated with the value y of the parameter. Obviously, the sequence $\{0\}$ with all components zero is an element of $N(0,0)$. We have the following result.

Theorem 3.1 *Suppose that $f(0) = 0$, $f \in C^1$ around 0, the derivative ∇f is Lipschitz continuous around 0. Let $\nabla f(0)$ be onto. Then the map N is Aubin continuous at $((0,0),\{0\})$;*

In the proof we use the following lemma.

Lemma 3.1 *Suppose that the assumptions of Theorem 3.1 hold and let M be the constant of surjectivity of $\nabla f(0)$, as in (1). Then the map*

$$\mathcal{P} : (x, y) \mapsto (f(x) - \nabla f(x)(\cdot - x))^{-1}(y)$$

has the following property: for every $M' > M$ and for every $\gamma > 0$ there exists $a > 0$ such that for every $y', y'' \in Y$, $x', x'' \in B_a(0)$ and for every $z' \in \mathcal{P}(x', y') \cap B_a(0)$ there exists $z'' \in \mathcal{P}(x'', y'')$ with

$$\|z' - z''\| \leq M'\|y' - y''\| + \gamma\|x' - x''\|.$$

Proof. Let $M' > M$ and $\gamma > 0$, and let δ and ε be chosen as at the beginning of the proof of Theorem 2.1 so that (7) and (8) hold. Taking ε smaller if necessary, let L be the Lipschitz constant of ∇f in $B_\varepsilon(0)$. Take δ smaller if necessary and choose a such that

$$0 < a \leq \varepsilon \quad \text{and} \quad \frac{5}{2}M'La \leq \gamma. \tag{10}$$

Let $x', x'' \in B_a(0)$, $y', y'' \in Y$ and let $z' \in \mathcal{P}(x', y') \cap B_a(0)$. Then

$$\begin{aligned} y' &= f(x') + \nabla f(x')(z' - x') \\ &= f(x') + \nabla f(x'')(z' - x') + (\nabla f(x') - \nabla f(x''))(z' - x'). \end{aligned}$$

From (8) there exists z'' such that

$$y'' = f(x'') + \nabla f(x'')(z'' - x'')$$

and moreover

$$\begin{aligned}
\|z' - z''\| \;\le\; & M'\|y' - y'' - f(x') + \nabla f(x'')x' \\
& -(\nabla f(x') - \nabla f(x''))(z' - x') + f(x'') - \nabla f(x'')x''\| \\
\le\; & M'\|y' - y''\| + M'\|f(x') - f(x'') - \nabla f(x'')(x' - x'')\| \\
& +M'\|(\nabla f(x'') - \nabla f(x'))(x' - z')\| \\
\le\; & M'\|y' - y''\| + M'\|\int_0^1 (\nabla f(x'' + t(x' - x'')) - \nabla f(x''))dt\|\|x' - x''\| \\
& +M'\|(\nabla f(x'') - \nabla f(x'))(x' - z')\| \\
\le\; & M'\|y' - y''\| + \frac{1}{2}M'La\|x' - x''\| + M'L\|x' - x''\|(\|x'\| + \|z'\|) \\
\le\; & M'\|y' - y''\| + \frac{5}{2}M'La\|x' - x''\|.
\end{aligned}$$

Taking into account (10) the proof is complete. □

Proof of Theorem 3.1. Let $\nabla f(0)$ be onto and let M be the constant in (1). Choose $M' > M$ and $0 < \gamma < 1$ and let a be a constant satisfying (10) with ε and L as in the beginning of the proof of Lemma 3.1. Let $0 < \alpha \le a$ and $0 < \beta \le \min\{\rho, \varepsilon\}$, where ρ is as in the statement of Theorem 2.1. By Theorem 2.1, for every $x_0 \in B_\beta(0)$ and $y \in B_\beta(0)$, the set $N(x_0, y) \cap B_\alpha(0)$ is nonempty. Let $y', y'' \in B_\beta(0)$, $x'_0, x''_0 \in B_\beta(0)$, and let $\{x'_n\} \in N(x'_0, y') \cap B_\alpha(\{0\})$. By Lemma 3.1, for every n there exists a Newton step x''_n from x''_{n-1} associated with y'' and such that

$$\|x''_n - x'_n\| \le M'\|y' - y''\| + \gamma\|x''_{n-1} - x'_{n-1}\|.$$

Hence, for every n

$$\|x''_n - x'_n\| \le \frac{M'}{1-\gamma}\|y' - y''\| + \gamma^n\|x'_0 - x''_0\|.$$

Thus the map N is Aubin continuous at $((0,0), \{0\})$ with constants α, β, and $\lambda = \max\{M'/(1-\gamma), \gamma\}$. □

If the Jacobian $\nabla f(0)$ is nonsingular, then, by a parallel analysis one can show that the map N can be locally identified with a Lipschitz continuous function. In other words, in this case the Newton sequence is not only unique and Q-quadratically convergent, but also it depends in a Lipschitz way on parameters.

The observations presented in this paper can be extended for more elaborate versions of the Newton method for solving variational inequalities and optimization problems.

Acknowledgment. The author wishes to thank anonymous referee for inspiring comments.

REFERENCES

[1] J.-P. Aubin, H. Frankowska. Set-Valued Analysis. Birkhäuser, Boston 1990.

[2] A. V. Dmitruk, A. A. Milyutin, N. P. Osmolovskiĭ. The Lyusternik theorem and the theory of extremum. *Uspekhi Math. Nauk* **35** (1980) 11-46.

[3] A. L. Dontchev. Local convergence of the Newton method for generalized equations. *C. R. Acad. Sci. Paris* **322** Série I (1996) 327-331.

[4] A. L. Dontchev. The Graves theorem revisited. *J. of Convex Analysis* **3** (1996) 45-53.

[5] A. L. Dontchev. Uniform Convergence of the Newton Method for Aubin Continuous Maps. *Serdica Math. Journal* **22** (1996) 385-398.

[6] A. L. Dontchev, W. W. Hager. An inverse function theorem for set-valued maps. *Proc. Amer. Math. Soc.* **121** (1994) 481-489.

[7] A.. L. Dontchev, R. T. Rockafellar. Characterizations of strong regularity for variational inequalities over polyhedral convex sets. *SIAM J. Optim.* **6** (1996) 1087-1105.

[8] H. Frankowska. An open mapping principle for set-valued maps. *J. Math. Anal. Appl.* **127** (1987) 172-180.

[9] L. M. Graves. Some mapping theorems. *Duke Math. J.* **17** (1950) 111-114.

[10] A. D. Ioffe. On the local surjection property. *Nonlinear Anal., Theory, Methods and Appl.* **11** (1987) 565-592.

[11] L. A. Lyusternik. On the conditional extrema of functionals. *Math. Sbornik* **41** (1934) 390-401.

[12] B. Mordukhovich. Stability theory for parametric generalized equations and variational inequalities via nonsmooth analysis. *Trans. Amer. Math. Soc.* **343** (1994) 609-657.

[13] J.-P. Penot. Metric regularity, openness and Lipschitz multifunctions. *Nonlinear Anal., Theory, Methods and Appl.* **13** (1989) 629-643.

[14] S. M. Robinson. Stability theory for systems of inequalities. Part II: Differentiable nonlinear systems. *SIAM J. Num. Anal.* **13** (1976) 497-513.

[15] R. T. Rockafellar, J.-B, Wets. Variational Analysis, Springer, Berlin, 1998.

Mathematical Reviews (AMS)
Ann Arbor
MI 48107
USA