Electronic Transactions on Numerical Analysis.
Volume 17, pp. 102-111, 2004.

ISSN 1068-9613.

ETNA
Kent State University
etna@mcs.kent.edu

ON THE EXISTENCE THEOREMS OF KANTOROVICH, MIRANDA AND BORSUK*

GÖTZ ALEFELD†, ANDREAS FROMMER‡, GERHARD HEINDL‡, AND JAN MAYER†

Abstract. The theorems of Kantorovich, Miranda and Borsuk all give conditions on the existence of a zero of a nonlinear mapping. In this paper we are concerned with relations between these theorems in terms of generality in the case that the mapping is finite-dimensional. To this purpose we formulate a generalization of Miranda's theorem, holding for arbitrary norms instead of just the l_∞-norm. As our main results we then prove that the Kantorovich theorem reduces to a special case of this generalized Miranda theorem as well as to a special case of Borsuk's theorem. Moreover, it turns out that, essentially, the Miranda theorems are themselves special cases of Borsuk's theorem.

Key words. nonlinear equations, existence theorems, fixed points, Newton-Kantorovich theorem, Miranda theorem, Borsuk theorem.

AMS subject classifications. 47H10, 47J05, 65H10.

1. Introduction. In this paper we are concerned with two well-known classical theorems, both of which guarantee the existence of a zero of a nonlinear mapping f from a norm ball in $\mathbb{R}^n$ to $\mathbb{R}^n$. These theorems are Kantorovich's theorem and Borsuk's theorem. Miranda's theorem, which we also consider, is essentially a special version of Borsuk's theorem in the case that the norm ball is a box, i.e. the norm is the maximum norm. Kantorovich's theorem and Borsuk's theorem apparently are very different in nature: Kantorovich's theorem is motivated by the analysis of Newton's iteration to approximate a zero of f and it gives an a priori criterion for the convergence of this iteration, in this manner proving that there is a zero of f within a certain ball centered at the initial guess for Newton's method. The major ingredients in its hypotheses are the Lipschitz-continuity of the derivative of f in a sufficiently large neighbourhood of the starting point and the assumption that the function value at the starting point is sufficiently small. On the other hand, Borsuk's theorem only requires the mapping f to be continuous on the ball and to fulfill a non-colinearity condition for the function values at all pairs of antipodal points on the boundary of the ball.

The purpose of this paper is to prove the remarkable fact that the Kantorovich theorem is (essentially) a special case of Borsuk's theorem in the sense that the hypotheses of the Kantorovich theorem imply those of Borsuk. For the case that the norm is the l_∞- norm this result was essentially already obtained in [1], where it was shown that the hypotheses of Kantorovich's theorem imply those of Miranda's theorem. In the present paper we formulate a version of Miranda's theorem holding for arbitrary norms, which is then proven to be 'in between' Kantorovich and Borsuk, i.e. more general than Kantorovich's but (essentially) more special than Borsuk's theorem.

Let us remark here that our results heavily rely on the finite dimension of the underlying vector space. While the Kantorovich theorem immediately extends to general Banach spaces, Borsuk's theorem does so only if one substantially restricts the class of mappings to be considered, for example to compact modifications of the identity or generalizations thereof.

The rest of this paper is organized as follows: In the next section we give precise formulations of Kantorovich's theorem, of Miranda's theorem and of Borsuk's theorem. In section 3 we

* Received April 1, 2003. Accepted for publication December 19, 2003. Recommended by Bill Gragg.
†Institut für Angewandte Mathematik Universität Karlsruhe, D-76128 Karlsruhe, Germany. E-mail: {goetz.alefeld,jan.mayer}@math.uni-karlsruhe.de.
‡Fachbereich Mathematik und Naturwissenschaften, Universität Wuppertal, D-42097 Wuppertal, Germany. E-mail: {frommer,heindl}@math.uni-wuppertal.de.

ETNA
Kent State University
etna@mcs.kent.edu

formulate and prove our generalization of Miranda's theorem for arbitrary norms. In section 4 we then come to the major result of this paper establishing the hierarchy of the theorems with respect to generality. Some conclusions are formulated in section 5.

2. The theorems of Kantorovich, Miranda and Borsuk. We start with Kantorovich's theorem. It can be stated in its 'standard' form and in an 'affine invariant' form. Although the latter is the more general one, the standard form is the one that can usually be found in textbooks. We therefore give both versions.
Here, as in the sequel, $\|\cdot\|$ denotes some arbitrary norm in $\mathbb{R}^n$ and its corresponding operator norm. The closed ball with radius $\rho \geq 0$, centered at x^0, is

$$\overline{B}(x^0, \rho) = \{x \in \mathbb{R}^n : \|x - x^0\| \leq \rho\}.$$

$B(x^0, \rho)$ and $\partial B(x^0, \rho)$ denote the topological interior and boundary of $\overline{B}(x^0, \rho)$, respectively.
THEOREM 2.1. *(Kantorovich, standard form [8]) Let $f : D \subseteq \mathbb{R}^n \to \mathbb{R}^n$ be differentiable in the open convex set D. Assume that for some point $x^0 \in D$ the Jacobian $f'(x^0)$ is invertible with*

$$\|f'(x^0)^{-1}\| \leq \beta, \quad \|f'(x^0)^{-1} f(x^0)\| \leq \eta.$$

Let there be a Lipschitz constant $\kappa > 0$ for f' such that

$$\|f'(u) - f'(v)\| \leq \kappa \cdot \|u - v\| \text{ for all } u, v \in D.$$

If $h = \eta\beta\kappa \leq \frac{1}{2}$ and $\overline{B}(x^0, \rho_-) \subseteq D$, where

$$\rho_- = \frac{1 - \sqrt{1 - 2h}}{\beta\kappa}$$

then f has a zero x^* in $\overline{B}(x^0, \rho_-)$. Moreover, this zero is the unique zero of f in $(\overline{B}(x^0, \rho_-) \cup B(x^0, \rho_+)) \cap D$ where $\rho_+ = \frac{1+\sqrt{1-2h}}{\beta\kappa}$ and the Newton iterates x^k with

$$x^{k+1} = x^k - f'(x^k)^{-1} f(x^k)$$

are well-defined, remain in $\overline{B}(x^0, \rho_-)$ and converge to x^*.
The following affine invariant form of the Kantorovich theorem is a generalization of the standard form as can be seen immediately by setting $\omega = \beta\kappa$.
THEOREM 2.2. *(Kantorovich, affine invariant form [4, 5]) Let $f : D \subseteq \mathbb{R}^n \to \mathbb{R}^n$ be differentiable in the open convex set D. Assume that for some point $x^0 \in D$ the Jacobian $f'(x^0)$ is invertible with*

$$\|f'(x^0)^{-1} f(x^0)\| \leq \eta.$$

Let there be a Lipschitz constant $\omega > 0$ for $f'(x^0)^{-1} f'$ such that

$$\|f'(x^0)^{-1}(f'(u) - f'(v))\| \leq \omega \cdot \|u - v\| \text{ for all } u, v \in D.$$

If $h = \eta\omega \leq \frac{1}{2}$ and $\overline{B}(x^0, \rho_-) \subseteq D$, where

$$\rho_- = \frac{1 - \sqrt{1 - 2h}}{\omega}$$

ETNA
Kent State University
etna@mcs.kent.edu

then f has a zero x^* in $\overline{B}(x^0,\rho_-)$. Moreover, this zero is the unique zero of f in $(\overline{B}(x^0,\rho_-)\cup B(x^0,\rho_+))\cap D$ where $\rho_+ = \frac{1+\sqrt{1-2h}}{\omega}$ and the Newton iterates x^k with

$$x^{k+1} = x^k - f'(x^k)^{-1}f(x^k)$$

are well-defined, remain in $\overline{B}(x^0,\rho_-)$ and converge to x^*.

Note that in this theorem one may leave the values of η and ω unchanged after transformations $f \to A\cdot f$ for any non-singular matrix $A \in \mathbb{R}^{n\times n}$. Therefore, the theorem holds irrespective of linear transformations whence the name 'affine invariant form'.

Note also that ω will often be much smaller than $\beta\kappa$. It is therefore not difficult to construct examples where for a given differentiable mapping f and a given point x^0, the main assumption $\eta\beta\kappa \le \frac{1}{2}$ of the standard theorem is not fulfilled, whereas the assumption $\eta\omega \le \frac{1}{2}$ in the affine invariant theorem is met. In this sense, Theorem 2.2 is more general than Theorem 2.1.

We now turn to formulate Borsuk's theorem. Let us say that a set $B \subseteq \mathbb{R}^n$ is symmetric with respect to $x^0 \in \mathbb{R}^n$, if for all $y \in \mathbb{R}^n$ we have $x^0 + y \in B \Rightarrow x^0 - y \in B$.

Then Borsuk's theorem can be stated as follows.

THEOREM 2.3. *(Borsuk [2, 3]) Let $B \subseteq \mathbb{R}^n$ be open, bounded, convex and symmetric with respect to $x^0 \in B$. Let $f : \overline{B} \to \mathbb{R}^n$ be a continuous mapping, assume that $f(x) \neq 0$ on ∂B and that*

$$f(x^0+y) \neq \lambda f(x^0-y) \text{ for all } \lambda > 0 \text{ and all } x^0+y \in \partial B. \tag{2.1}$$

Then f has a zero in B.

Often, this theorem is stated in terms of the mapping h defined by $h(y) = f(x^0+y)$ on $B' = B - x^0 = \{y \in \mathbb{R}^n : y = x - x^0,\ x \in B\}$. Condition (2.1) then reads

$$h(y) \neq \lambda h(-y) \text{ for all } \lambda > 0 \text{ and all } y \in \partial B',$$

where B' is open, bounded, convex and symmetric with respect to the origin $0 \in B'$.

Very interestingly, Borsuk's theorem is 'naturally' affine invariant: If f satisfies (2.1), then $A\cdot f$ satisfies (2.1), too, for any non-singular matrix $A \in \mathbb{R}^{n\times n}$.

In our comparisons to the Kantorovich theorem, it will be useful to consider Borsuk's theorem on balls $B = B(x^0,\rho)$, $\rho > 0$, with respect to a given norm $\|\cdot\|$. This is just an apparent restriction, since in our finite-dimensional setting, any set B satisfying the assumptions of Borsuk's theorem, Theorem 2.3, is in fact a norm ball with respect to the Minkowski functional corresponding to B (and its center x^0).

If, in addition, we do not want to exclude a zero on the boundary of B, we arrive at the following immediate corollary to Theorem 2.3.

COROLLARY 2.4. *Let $f : \overline{B}(x^0,\rho) \to \mathbb{R}^n$ be a continuous mapping. Assume that*

$$f(x^0+y) \neq \lambda f(x^0-y) \text{ for all } \lambda > 0 \text{ and all } x^0+y \in \partial B(x^0,\rho). \tag{2.2}$$

Then f has a zero in $\overline{B}(x^0,\rho)$.

We finally formulate Miranda's theorem. This theorem works with the l_∞-norm and looks at components of f on the faces of an l_∞-ball which is a hypercube. We write $B_\infty(x^0,\rho)$ to denote such a ball centered at x^0 with its faces given as

$$B_\infty^{i,+}(x^0,\rho) = \{x \in \mathbb{R}^n : \|x-x^0\|_\infty = \rho, x_i - x_i^0 = \rho\}$$
$$B_\infty^{i,-}(x^0,\rho) = \{x \in \mathbb{R}^n : \|x-x^0\|_\infty = \rho, x_i - x_i^0 = -\rho\}$$

Then Miranda's theorem can be stated as follows.

THEOREM 2.5. *(Miranda [7]) Let $f : \overline{B}_\infty(x^0,\rho) \subseteq \mathbb{R}^n \to \mathbb{R}^n$ be a continuous mapping. Assume that*

$$f_i(x) \begin{cases} \geq 0 & \text{for all } x \in B_\infty^{i,+}(x^0,\rho) \\ \leq 0 & \text{for all } x \in B_\infty^{i,-}(x^0,\rho) \end{cases} \quad \text{for } i = 1,\ldots,n. \tag{2.3}$$

Then f has at least one zero x^* in $\overline{B}_\infty(x^0,\rho)$.

In [12] it is shown that Miranda's theorem is equivalent to Brouwer's fixed point theorem (for l_∞-balls).

3. Generalization of Miranda's Theorem. Miranda's original theorem has been generalized in several different directions before, see e.g. [11], [14] and [6]. For any of these generalizations, however, it has not been shown that it contains Kantorovich's theorem as a special case, i.e. that whenever Kantorovich's theorem guarantees the existence of a zero then the respective generalization would guarantee the existence of such a zero, too. Indeed, there are examples which show that this is not always the case.

In Theorem 3.2 we present a new generalization of Miranda's theorem which does contain Kantorovich's theorem: As we will show in Theorem 4.1, whenever Kantorovich's theorem guarantees the existence of a zero, our generalization of Miranda's theorem does so, too.

Observe that (2.3) can be interpreted as saying that at each point x on the boundary of $B_\infty(x^0,\rho)$ the image $f(x)$ points in an 'outside' direction. This interpretation is the basis for our generalization of Miranda's theorem to balls with respect to an arbitrary norm formulated as Theorem 3.2 below.

This generalization uses the concept of normal vectors. Let $\langle\cdot,\cdot\rangle$ denote the usual inner product on $\mathbb{R}^n$. We say that the vector $a \in \mathbb{R}^n$ is normal to the open convex set $C \subseteq \mathbb{R}^n$ at $x \in \partial C$ iff $\langle a, x-y\rangle > 0$ for all $y \in C$, i.e. if a is a nonzero vector normal to $\overline{C}$ at x in the sense of [10]. By the Hahn-Banach-Theorem, there exists at least one vector normal to C at each $x \in \partial C$. If C is a ball $B(x^0,\rho)$, $\rho > 0$, with respect to some norm $\|\cdot\|$ and $\|\cdot\|_d$ its dual norm

$$\|y\|_d = \max_{\|x\|=1} \langle x, y\rangle = \max_{\|x\|=1} |\langle x, y\rangle|, \tag{3.1}$$

then the vectors normal to C at x can be characterized as follows:

LEMMA 3.1. *For any $\rho > 0$, the vector $a \in \mathbb{R}^n$ is normal to $B(x^0,\rho)$ at $x \in \partial B(x^0,\rho)$ iff a is a positive multiple of some $a' \in \mathbb{R}^n$ for which*

$$\|a'\|_d = 1 \quad \text{and} \quad \langle a' \,,\, x - x^0\rangle = \rho.$$

Proof. a is normal to $B(x^0,\rho)$ at $x \in \partial B(x^0,\rho)$ iff there is a $\lambda > 0$ such that $a \in \lambda \cdot \partial\varphi(x)$, where $\partial\varphi(x)$ denotes the subdifferential of the convex function

$$\varphi : \mathbb{R}^n \to \mathbb{R}, \ y \mapsto \frac{1}{\rho}\|y - x^0\|,$$

see e.g. [10, Cor. 23.7.1]. We therefore show

$$\partial\varphi(x) = \frac{1}{\rho}\left\{a' \in \mathbb{R}^n : \|a'\|_d = 1 \text{ and } \langle a' \,,\, x - x^0\rangle = \rho\right\}.$$

To this purpose we first observe that if $a \in \partial\varphi(x)$, i.e. $\varphi(y) \geq \varphi(x) + \langle a \,,\, y - x\rangle$ for all $y \in \mathbb{R}^n$, then

$$\langle a \,,\, h\rangle \leq \frac{1}{\rho}\|x - x^0 + h\| - 1 \leq \frac{1}{\rho}\|h\|$$

ETNA
Kent State University
etna@mcs.kent.edu

holds for all $h \in \mathbb{R}^n$. Hence $\|a\|_d \le (1/\rho)$ and $\langle a\,,\, x^0 - x\rangle \le -1$. But this implies $\|a\|_d = (1/\rho)$ and $\langle a\,,\, x - x^0\rangle = 1$ since $1 \le \langle a\,,\, x - x^0\rangle \le \|a\|_d \|x - x^0\| \le 1$. Consequently, for $a' = \rho a$ we have $\|a'\|_d = 1$ and $\langle a'\,,\, x - x^0\rangle = \rho$.
Conversely, if $\|a'\|_d = 1$, $\langle a'\,,\, x - x^0\rangle = \rho$ and $a = (1/\rho)a'$, then for any $y \in \mathbb{R}^n$ we have

$$\varphi(x) + \langle a\,,\, y - x\rangle = 1 + \frac{1}{\rho}\langle a'\,,\, y - x^0 + x^0 - x\rangle$$
$$= \frac{1}{\rho}\langle a'\,,\, y - x^0\rangle \ \le \frac{1}{\rho}\|a'\|_d\|y - x^0\| \ = \ \varphi(y),$$

showing $a \in \partial\varphi(x)$. □

THEOREM 3.2. *Let $B(x^0,\rho)$, $\rho > 0$, be an open ball with respect to an arbitrary norm $\|\cdot\|$. Let $f : \overline{B}(x^0,\rho) \to \mathbb{R}^n$ be continuous and assume that for all $x \in \partial B(x^0,\rho)$ there exists a vector a normal to $B(x^0,\rho)$ at x such that*

$$\langle f(x), a\rangle \ge 0\,. \tag{3.2}$$

Then

a) the relation (3.2) actually holds for all *vectors a normal to $B(x^0,\rho)$ at x,*

b) f has a zero in $\overline{B}(x^0,\rho)$.

Proof. To prove part a), we first note that [10, Cor. 23.7.1], which we already used in the proof of Lemma 3.1, implies that it is sufficient to show that for the function $\varphi : y \mapsto (1/\rho)\|y - x^0\|$ we have

$$\langle f(x)\,,\, a\rangle \ge 0 \text{ for all } x \in \partial B(x^0,\rho) \text{ and all } a \in \partial\varphi(x).$$

Secondly, we remark that it is known (see e.g. [10, Th. 25.6]) that

$$\partial\varphi(x) = \overline{\text{conv}\, S(x)}$$

where $a \in S(x)$ iff there is a sequence x^i in $\mathbb{R}^n$ such that for all i the convex function φ is differentiable in x^i, i.e. $\partial\varphi(x^i) = \{\varphi'(x^i)\}$, and such that $\lim_{i\to\infty} x^i = x$ and $\lim_{i\to\infty}\varphi'(x^i) = a$.
For $x \in \partial B(x^0,\rho)$ let x^i denote such a sequence. W.l.o.g. we can assume $x^i \ne x^0$ for all i. Due to the homogeneity of $\|\cdot\|$, the function φ is differentiable not only in x^i but also in $\tilde{x}^i = x^0 + (1/\varphi(x^i))(x^i - x^0) \in \partial B(x^0,\rho)$ with $\varphi'(\tilde{x}^i) = \varphi'(x^i)$. Hence $\lim_{i\to\infty}\varphi'(\tilde{x}^i) = a$ and obviously also $\lim_{i\to\infty}\tilde{x}^i = x$. At $\tilde{x}^i$ every vector normal to $B(x^0,\rho)$ is a positive multiple of $\varphi'(\tilde{x}^i)$, so by assumption we have $\langle f(\tilde{x}^i)\,,\, \varphi'(\tilde{x}^i)\rangle \ge 0$ and since f is continuous in x we can conclude $\langle f(x)\,,\, a\rangle \ge 0$. We therefore have $\langle f(x)\,,\, a\rangle \ge 0$ for all $a \in S(x)$, and since $\langle\cdot,\cdot\rangle$ is linear and continuous in its second argument we even have $\langle f(x)\,,\, a\rangle \ge 0$ for all $a \in \overline{\text{conv}\, S(x)}$.
Part b) is now easily proved via Borsuk's theorem. Let $\varepsilon > 0$ and take

$$f_\varepsilon(x) = f(x) + \varepsilon(x - x^0).$$

By a) we know that $\langle f(x), a\rangle \ge 0$ for all vectors a normal to $B(x^0,\rho)$ at x, and thus $\langle f_\varepsilon(x), a\rangle > 0$ for all such a. By Corollary 2.4 and Lemma 3.3 below the mapping f_ε has a zero in $\overline{B}(x^0,\rho)$. So any limit point of the sequence of these zeros for $\varepsilon = \frac{1}{n}$, $n \in \mathbb{N}$ is a zero of f. □

The proof above makes use of the following auxiliary result which we will need again in section 4.

LEMMA 3.3. *Let $B(x^0,\rho)$, $\rho > 0$, be an open ball with respect to an arbitrary norm $\|\cdot\|$. Let $f : \overline{B}(x^0,\rho) \to \mathbb{R}^n$ be continuous and assume that for all $x \in \partial B(x^0,\rho)$ and for all*

ETNA
Kent State University
etna@mcs.kent.edu

vectors a normal to $B(x^0,\rho)$ at x we have

$$\langle\, f(x), a\rangle\, > 0\ .$$

Then for all $x = x^0 + y \in \partial B(x^0,\rho)$ and for all $\lambda > 0$ we have

$$f(x^0 + y) \neq \lambda f(x^0 - y)\ .$$

Proof. If a is normal to $B(x^0,\rho)$ at $x = x^0 + y$, its negative $-a$ is normal to $B(x^0,\rho)$ at $x^0 - y$. So, if we had $f(x^0 + y) = \lambda f(x^0 - y)$ for some $\lambda > 0$, we would arrive at

$$0 < \langle\, f(x^0 + y), a\,\rangle = \langle\, \lambda f(x^0 - y), a\,\rangle = -\lambda\langle\, f(x^0 - y), -a\,\rangle\ ,$$

which is impossible since $\langle\, f(x^0 - y), -a\,\rangle > 0$. □

Just in passing, let us note that Theorem 3.2 remains valid if we replace $B(x^0,\rho)$ by an arbitrary non-empty open bounded convex set. Part a) can then be proved by replacing $\varphi(x)$ by $p(x - x^0)$ where x^0 is an arbitrary point of C and p is the Minkowski functional of $C - x^0$. A proof for b) can easily be derived from the following proposition, which is just the application of the Leray-Schauder-Theorem [8, 6.3.3] to the mapping g with $g(x) = x - f(x^0 + x)$, $x \in \overline{C - x^0}$.

PROPOSITION 3.4. *Let C be a non-empty open bounded subset of $\mathbb{R}^n$ and $f : \overline{C} \to \mathbb{R}^n$ a continuous mapping. Assume that there is an $x^0 \in C$ such that for all $x \in \partial C$*

$$f(x) \notin \{\lambda(x - x^0) : \lambda < 0\}.$$

Then f has a zero in $\overline{C}$.

4. The hierarchy with respect to generality. In this section we prove our central results. We show that the Kantorovich theorem for an arbitrary norm is a special case of the generalized Miranda theorem, Theorem 3.2. In addition we show that the Kantorovich theorem for an arbitrary norm is also a special case of Borsuk's theorem. This hierarchy is illustrated in Figure 1.

We will use the numbers ρ_-, ρ_+ which have been defined in Theorem 2.2.

THEOREM 4.1. *Let f satisfy all assumptions of the affine invariant form of the Kantorovich theorem (Theorem 2.2). Put*

$$g : D \to \mathbb{R}^n,\ x \mapsto f'(x^0)^{-1} f(x)$$

and consider any positive $\rho \in [\rho_-, \rho_+]$ such that $\overline{B}(x^0,\rho) \subseteq D$. Then for all $x \in \partial B(x^0,\rho)$ and for all normals a to $B(x^0,\rho)$ at x we have

$$\langle g(x)\ ,\ a\rangle \geq 0, \tag{4.1}$$

which is the hypothesis of the generalized Miranda theorem (Theorem 3.2) for the mapping g and the ball $B(x^0,\rho)$.

Proof. Let $x \in \partial B(x^0,\rho)$ and a a normal to $B(x^0,\rho)$ at x. Because of Lemma 3.1, we can assume $||a||_d = 1$ and $\langle a\ ,\ x - x^0\rangle = \rho$. We first prove

$$\left|\big\langle g(x) - g(x^0)\ ,\ a\big\rangle - \rho\right| \leq \frac{\omega}{2}\rho^2. \tag{4.2}$$

In order to do so, we define the continuously differentiable function $\varphi : [0,1] \to \mathbb{R}$ as

$$\varphi(t) = \big\langle\, g(x^0 + t(x - x^0)) - tg'(x^0)(x - x^0)\ ,\ a\,\big\rangle\ .$$

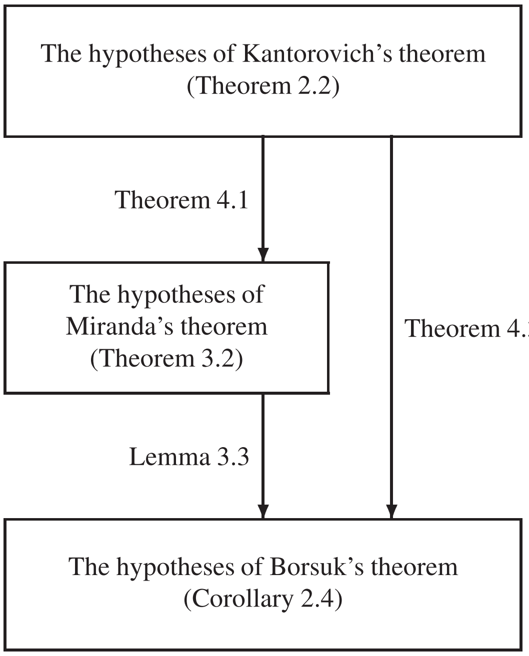

FIG. 4.1. *The hierarchy with respect to generality*

Then

$$\varphi'(t) = \left\langle \, \big(g'(x^0 + t(x - x^0)) - g'(x^0)\big)\,(x - x^0)\,,\, a \,\right\rangle$$

and

$$\begin{aligned}
|\varphi(1) - \varphi(0)| &= \left| \int_0^1 \varphi'(t)dt \right| \\
&\le \int_0^1 |\varphi'(t)|\, dt \\
&= \int_0^1 \left| \left\langle \, \big(g'(x^0 + t(x - x^0)) - g'(x^0)\big)\,(x - x^0)\,,\, a \,\right\rangle \right| \, dt \\
&\le \int_0^1 \left\| \big(g'(x^0 + t(x - x^0)) - g'(x^0)\big)\,(x - x^0) \right\| \cdot \|a\|_d \, dt \\
&\le \int_0^1 \|g'(x^0 + t(x - x^0)) - g'(x^0)\| \cdot \|x - x^0\| \, dt \\
&\le \int_0^1 \omega \cdot t \cdot \|x - x^0\|^2 dt \\
&= \frac{\omega}{2} \cdot \|x - x^0\|^2 \\
&= \frac{\omega}{2}\rho^2 \, .
\end{aligned}$$

On the other hand

$$\varphi(1) - \varphi(0) = \big\langle g(x) - g'(x^0)(x - x^0)\,,\, a \big\rangle - \big\langle g(x^0)\,,\, a \big\rangle$$

and $g'(x^0) = \mathrm{id}$ as well as $\langle x - x^0\,,\, a\rangle = \rho$, which shows (4.2).
¿From (4.2) we conclude

$$\begin{aligned}\langle\, g(x)\,,\, a\,\rangle &\geq -\frac{\omega}{2}\rho^2 + \rho + \langle\, g(x^0)\,,\, a\,\rangle \\ &\geq -\frac{\omega}{2}\rho^2 + \rho - \|g(x^0)\| \cdot \|a\|_d \\ &= -\frac{\omega}{2}\rho^2 + \rho - \|g(x^0)\| \\ &\geq -\frac{\omega}{2}\rho^2 + \rho - \eta \left\{ \begin{array}{ll} \geq 0 & \text{if } \rho \in [\rho_-, \rho_+] \\ > 0 & \text{if } \rho \in (\rho_-, \rho_+) \end{array} \right. .\end{aligned}$$

□

In the proof of Theorem 4.1 we have actually shown that (4.1) holds with strict inequality as soon as $\rho \in (\rho_-, \rho_+)$. Using Lemma 3.3 we thus arrive at the following result.

THEOREM 4.2. *Let f satisfy all assumptions of the affine invariant form of the Kantorovich theorem. Then f satisfies (2.2) for all balls $\overline{B}(x^0, \rho) \subseteq D$ with $\rho \in (\rho_-, \rho_+)$.*

Proof. The discussion above already showed that $g = f'(x^0)^{-1} f$ satisfies (2.2). But (2.2) remains invariant under affine transformations, so f satisfies (2.2), too. □

Of course, it would be more satisfying if in the above theorem one could take the closed interval $[\rho_-, \rho_+]$ for ρ instead of just the open interval. As we will show now, this is indeed so, as soon as $\eta > 0$, i.e. if $f(x^0) \neq 0$. This result is proved in our final theorem, where we have to go a direct way without using the generalized Miranda Theorem.

THEOREM 4.3. *Let f satisfy all assumptions of the affine invariant form of the Kantorovich theorem (Theorem 2.2) and exclude the case $\eta = 0$. Then f satisfies (2.2) for all $\rho \in [\rho_-, \rho_+]$ such that $\overline{B}(x_0, \rho) \subseteq D$.*

Proof. We will show that (2.2) holds for $g = f'(x^0)^{-1} f$. It then also holds for f. To start, let $x^0 + y \in \partial B(x^0, \rho)$ and assume that (2.2) does not hold, i.e. we have $\lambda \in (0, \infty)$ such that

$$g(x^0 + y) = \lambda g(x^0 - y). \tag{4.3}$$

Replacing, if necessary, y by $-y$, we can even assume $\lambda \geq 1$. Let a be an arbitrary vector to be specified later, and set

$$\varphi(t) = \langle\, g(x^0 + ty) - tg'(x^0)y\,,\, a\,\rangle\,.$$

Then

$$\varphi'(t) = \langle\, \left(g'(x^0 + ty) - g'(x^0)\right) y\,,\, a\,\rangle$$

and, since $g'(x^0) = \mathrm{id}$, we have

$$\begin{aligned}&\int_0^1 \langle\, \left(g'(x^0 + ty) - g'(x^0)\right) y\,,\, a\,\rangle\; dt \\ &= \varphi(1) - \varphi(0) \\ &= \langle\, g(x^0 + y) - g'(x^0)y\,,\, a\,\rangle - \langle\, g(x^0)\,,\, a\,\rangle \\ &= \langle\, g(x^0 + y)\,,\, a\,\rangle - \langle\, y\,,\, a\,\rangle - \langle\, g(x^0)\,,\, a\,\rangle,\end{aligned}$$

which gives

$$\langle\, g(x^0 + y)\,,\, a\,\rangle = \langle\, y\,,\, a\,\rangle + \langle\, g(x^0)\,,\, a\,\rangle + \int_0^1 \langle\, \left(g'(x^0 + ty) - g'(x^0)\right) y\,,\, a\,\rangle\; dt, \tag{4.4}$$

ETNA
Kent State University
etna@mcs.kent.edu

and, similarly for $-y$ instead of y

$$\langle\, g(x^0-y)\,,\,a\,\rangle = -\langle\, y\,,\,a\,\rangle + \langle\, g(x^0)\,,\,a\,\rangle - \int_0^1 \langle\, \left(g'(x^0-ty)-g'(x^0)\right)y\,,\,a\,\rangle\; dt. \tag{4.5}$$

By (4.3) we have

$$\langle\, g(x^0+y)\,,\,a\,\rangle = \lambda\langle\, g(x^0-y)\,,\,a\,\rangle,$$

which, using (4.4) and (4.5) and after rearranging terms gives

$$\begin{aligned}(1+\lambda)\langle\, y\,,\,a\,\rangle = (\lambda-1)\langle\, g(x^0)\,,\,a\,\rangle &- \int_0^1 \langle\, \left(g'(x^0+ty)-g'(x^0)\right)y\,,\,a\,\rangle\; dt\\ &- \lambda\int_0^1 \langle\, \left(g'(x^0-ty)-g'(x^0)\right)y\,,\,a\,\rangle\; dt.\end{aligned} \tag{4.6}$$

Now take a such that $||a||_d = 1$ and $\langle\, y\,,\,a\,\rangle = ||y||$. Such a exists, since $(\mathbb{R}^n, ||\cdot||)$ is identical to its bidual, i.e.

$$||y||_{dd} = ||y|| = \max_{||a||_d=1}\langle\, a\,,\,y\,\rangle.$$

In exactly the same manner as in the proof of Theorem 4.1, bounding the terms on the right hand side of (4.6), we get

$$(1+\lambda)||y|| \le (\lambda-1)||g(x^0)|| + \frac{\omega}{2}\cdot||y||^2 + \lambda\frac{\omega}{2}||y||^2$$

where $||y|| = \rho$ and $||g(x^0|| \le \eta$. We therefore have

$$(1+\lambda)\rho \le (\lambda-1)\eta + (1+\lambda)\frac{\omega}{2}\rho^2$$

or

$$(1+\lambda)\left(\frac{\omega}{2}\rho^2-\rho+\eta\right) - 2\eta \ge 0.$$

But this is impossible, because for $\rho \in [\rho_-, \rho_+]$ the first summand is non-positive and $\eta > 0$. Therefore, (4.3) does not hold, and since x^0+y was chosen to be an arbitrary point from $\partial B(x^0,\rho)$ we have shown that g satisfies (2.2). □

As was suggested by an anonymous referee, there is another well-known theorem on the existence of zeros which adds an additional stage to the hierarchy presented so far. This theorem is Smale's theorem [13], which in the finite-dimensional case may be stated as follows:

THEOREM 4.4. *Let $f:\mathbb{R}^n \to \mathbb{R}^n$ be analytic in $\mathbb{R}^n$ and for $x^0 \in \mathbb{R}^n$ let*

$$\beta(x^0,f) = \left|f'(x^0)^{-1}f(x^0)\right|,\quad \gamma(x^0,f) = \max_{k>1}\left|\frac{1}{k!}f'(x^0)^{-1}f^{(k)}(x^0)\right|^{1/(k-1)}.$$

Then, if $\beta(x^0,f)\gamma(x^0,f) < \alpha_0$, where α_0 is an invariant, approximately equal to 0.130707, the iterates of Newton's method starting with x^0 converge to a zero x^* of f.

This result is interesting because it proves convergence of Newton's method from information at just one point x^0.

As was already mentioned in [13], this result is in fact a special case of the affine invariant version of the Newton-Kantorovich theorem (Theorem 2.2), meaning that the hypotheses of Theorem 4.4 imply those of Theorem 2.2. As was shown by Rheinboldt [9], Theorem 4.4 may be generalized to a local version where f is assumed to be analytic only on some open subset of $\mathbb{R}^n$. The proof of that result shows that, again, we are in the presence of a special case of the Newton-Kantorovich theorem.

ETNA
Kent State University
etna@mcs.kent.edu

5. Conclusions. In this paper we have established a hierarchy with respect to generality between Kantorovich's theorem (which contains Smale's theorem), Miranda's theorem and its generalization and Borsuk's theorem. This hierarchy is meant only with respect to the existence of a zero. While the Kantorovich theorem also guarantees the *uniqueness* of the zero (and the convergence of Newton's method), the other theorems only partly address this aspect: they actually guarantee that the topological degree of the mapping is odd, see [3].

We have proven two major results. The first (Theorem 4.1) shows that if Kantorovich's theorem (Theorem 2.2) guarantees the existence of a zero of f in a ball $B(x^0, \rho)$, $\rho > 0$, $\rho_- \leq \rho \leq \rho_+$ with $\overline{B}(x^0, \rho) \subseteq D$, then the generalized Miranda theorem (Theorem 3.2), applied to $f'(x^0)^{-1} f$ also guarantees the existence of a zero in the same ball. In this sense, the generalized Miranda theorem is the more general theorem. Our second major result, Theorem 4.3, establishes a similar relationship between Kantorovich's theorem and Borsuk's theorem. Here, we can even use the same mapping f in both theorems, i.e. the transition to $f'(x^0)^{-1} f$ is not necessary. On the other hand, we have to restrict ourselves to the case $\eta > 0$, i.e. $f(x^0) \neq 0$. If $f(x^0) = 0$, Theorem 4.2 shows that f still satisfies the hypothesis of Borsuk's theorem for all $\rho \in (0, \rho_+)$ (note that $\rho_- = 0$ for $\eta = 0$). Hence, the only situation where we did not establish the connection between Kantorovich and Borsuk is when we simultaneously have $\eta = 0$ and $\rho = \rho_+$. The following example shows that then there indeed need not be such a connection.

EXAMPLE 1. Let $f : \mathbb{R} \to \mathbb{R}$, $f(x) = x(1 - |x|)$. Then $f'(x) = 1 - 2|x|$. Take $x^0 = 0$ so that $f(x^0) = 0$, $f'(x^0) = 1$, and take $\eta = 0, \omega = 2$ in Kantorovich's theorem which thus gives $\rho_- = 0$, $\rho_+ = 1$. Therefore, the Kantorovich theorem guarantees that x^0 is the unique zero of f in $(-1, 1)$. But Borsuk's theorem cannot be applied to the ball $[-1, 1]$, since $f(-1) = f(1) = 0$, i.e. we have $f(x) = \lambda f(-x)$ for all $\lambda > 0$ and all $x \in \partial(-1, 1) = \{-1, 1\}$.

REFERENCES

[1] G. Alefeld, F.A. Potra, and Z. Shen, *On the existence theorems of Kantorovich, Moore and Miranda*, Comput. Suppl., 15 (2001), pp. 21–28.
[2] Karol Borsuk, *Drei Sätze über die n-dimensionale Sphäre*, Fund. Math., 20 (1933), pp. 177–190.
[3] Klaus Deimling, *Nonlinear Functional Analysis*, Springer, Berlin, Heidelberg, New York, 1985.
[4] Peter Deuflhard and Gerhard Heindl, *Affine invariant convergence theorems for Newton's method and extensions to related methods*, SIAM J. Numer. Anal., 16 (1980), pp. 1–10.
[5] L. Kantorovich, *On Newton's method for functional equations (Russian)*, Dokl. Akad. Nauk SSSR, 59 (1948), pp. 1237–1240.
[6] J. Mayer, *A generalized theorem of Miranda and the theorem of Newton-Kantorovich*, Numer. Funct. Anal. Optim., 23 (2002), nos. 3&4, pp. 333–357.
[7] C. Miranda, *Un'osservatione su un teorema di Brouwer*, Boll. Un. Mat. Ital., 3 (1940), no. 2, pp. 5–7.
[8] James Ortega and Werner Rheinboldt, *Iterative Solution of Nonlinear Equations in Several Variables*, Academic Press, New York, 1970.
[9] Werner C. Rheinboldt, *On a theorem of S. Smale about Newton's method for analytic mappings*, Appl. Math. Lett., 1 (1988), pp. 69–72.
[10] T. Rockafellar, *Convex Analysis*, Princeton University Press, Princeton, New Jersey, 1970.
[11] J. Rohn, *An existence theorem for systems of nonlinear equations*, Z. Angew. Math. Mech., 60 (1980), no. 8, pp. 345.
[12] G. Sansone and R. Conti, *Nonlinear Differential Equations*, Pergamon Press, Oxford, 1964.
[13] Steve Smale, *Algorithms for solving equations*, in Proceedings of the International Congress of Mathematicians, Vol. 1, 2 (Berkeley, 1986), A. Gleason, ed., Amer. Math. Soc., Providence, RI, 1987, pp. 172–195.
[14] M. Vrahatis, *A short proof and a generalization of Miranda's existence theorem*, Proc. Amer. Math. Soc., 107 (1989), no. 3, pp. 701–703.