CHAPTER III

ROLLE'S THEOREM IN BANACH SPACES

In this chapter, we establish theorems, analogous with Rolle's theorem, for functions from a subset of $\mathbf{R}^n$ into a real Banach space $\mathbf{B}$. And as a consequence we obtained a version of mean valued theorem.

Theorem 3.1. *Let $f : D(x_o , r) \to \mathbf{B}$ be continuous on $D(x_o , r)$ and Fréchet differentiable on $B(x_o , r)$. If there is a continuous linear functional $\phi : \mathbf{B} \to \mathbf{R}$, and a point $z_o \in B(x_o , r)$ such that $\phi(f(x) - f(z_o))$ does not change sign on $S(x_o , r)$, then there exists a vector $c \in B(x_o , r)$ such that*

$$\phi(df(c)(u)) = 0 \qquad \textit{for all } u \in \mathbf{R}^n.$$

Proof. Let ϕ be a continuous linear functional on $\mathbf{B}$ and z_o a point in $B(x_o , r)$ such that $\phi (f(x) - f(z_o))$ does not change sign on $S(x_o , r)$ .Let $g : D(x_o, r) \to \mathbf{R}$ be defined by $g(x) = \phi(f(x))$ for each $x \in D(x_o, r)$. Then g is continuous on $D(x_o, r)$ and differentiable on $B(x_o , r)$. Assume that $\phi(f(x) - f(z_o)) \leq 0$ for all $x \in S(x_o , r)$. Then for any x in $S(x_o , r)$, $g(x) \leq g(z_o)$. Since g is continuous on the compact set $D(x_o, r)$, then g attains its maximum on $D(x_o , r)$. That is, there is a point $c \in B(x_o, r)$ such that $g(c) = \max \{ g(x) \mid x \in D(x_o, r) \}$. Thus $dg(c) = 0$, and hence $\phi (df(c))(u) = d(\phi \circ f)(c)(u) = dg(c)(u) = 0$ for all $u \in \mathbf{R}^n$. The proof for the case $\phi(f(x) - f(z_o)) \geq 0$ for all x in $S(x_o, r)$, is similar. ■

If the assumption in theorem 3.1 is slightly weaken by replacing the condition that there is $z_0 \in B(x_0, r)$ such that $\phi(f(x) - f(z_0))$ does not change sign on $S(x_0, r)$ with the condition that

$$\phi(f(x)) = k \quad \text{, for each } x \text{ in } S(x_0, r), \text{ where } k \text{ is a constant}$$

then the conclusion is still the same.

Theorem 3.2. *Let $f : D(x_0, r) \to$ **B** be continuous on $D(x_0, r)$ and Fréchet differentiable on $B(x_0, r)$. If $\phi :$ **B** $\to$ **R** is a continuous linear functional such that $\phi(f(x))$ is constant on $S(x_0, r)$ then there is a vector c in $B(x_0, r)$ such that*

$$\phi(df(c)(u)) = 0 \qquad \textit{for all } u \in \mathbf{R}^n.$$

Proof. Let $\phi : \mathbf{B} \to \mathbf{R}$ be a continuous linear functional such that

$$\phi(f(x)) = k \qquad \text{for all } x \in S(x_0, r), \text{ where } k \text{ is a constant}$$

Then $\phi(f(x) - f(x_0)) = \phi(f(x)) - \phi(f(x_0)) = k - \phi(f(x_0))$ for all $x \in S(x_0, r)$. Therefore the map : $x \mapsto \phi(f(x) - f(x_0))$ is a constant map on $S(x_0, r)$. By theorem 3.1, there exists a point $c \in B(x_0, r)$ such that

$$\phi(df(c)(u)) = 0 \qquad \text{for all } u \in \mathbf{R}^n.$$

■

Theorem 3.1. and theorem 3.2. are not necessarily true when the space $\mathbf{R}^n$ is replaced by an infinite dimensional Banach space, as illustrated by the following example.

Example. ([2]) Let L and R denote the continuous linear operators in l_2 given by

$$L(x) = (x_2, x_3, \ldots),$$

$$R(x) = (0, x_1, x_2, x_3, \ldots), \text{ for } x = (x_1, x_2, x_3, \ldots) \in l_2.$$

Let T be the map $T: l_2 \to l_2$ defined as

$$T(x) = ((1/2) - \| x \|^2)\, e_1 + R(x), \text{ where } e_1 = (1,0,0,\ldots), \text{ for } x \in l_2.$$

Define the function $f: l_2 \to \mathbf{R}$ by

$$f(x) = \frac{1-\|x\|^2}{\|x-T(x)\|^2}, \quad \text{for } x \in l_2.$$

Since the map T has no fixed point, $\| x - T(x) \| \neq 0$ for any $x \in l_2$ and f is well defined. Then f is continuous in l_2 and $f(x) = 0$ for every $x \in S(0,1)$. As shown in the example in chapter II, the Fréchet derivative of g where $g(x) = \|x\|^2$ at $x \in l_2$ is the continuous linear function $dg(x): u \mapsto 2\langle x,u\rangle$. So, T is Fréchet differentiable at x and, for each $u \in l_2$,

$$dT(x)(u) = -2\langle x,u\rangle e_1 + R(u).$$

Hence the Fréchet derivative of h where $h(x) = \| x - T(x) \|^2$ at $x \in l_2$ is the continuous linear function $dT(x): u \mapsto 2\langle x - T(x), u - dT(x)(u)\rangle$ for each $u \in l_2$.

Then by the corollary 2.9, f is Fréchet differentiable at every $x \in l_2$ and, for each $u \in l_2$,

$$df(x)u = \frac{1}{\|x-T(x)\|^4} \times \left[-2\| x-T(x)\|^2 \langle x,u\rangle - 2(1 - \| x \|^2)\langle x - T(x), u - dT(x)u\rangle\right]$$

Since $\langle T(x), e_1\rangle = \frac{1}{2} - \| x \|^2$ and $\langle x, Ru\rangle = \langle Lx, u\rangle$, $L(T(x)) = x$, it follow that

$$\langle x - T(x), u - dT(x)u \rangle = \langle (1 + 2x_1 + 2\|x\|^2)x - T(x) - L(x), u \rangle .$$

Therefore,

$$df(x)u = \frac{-2}{\|x - T(x)\|^4}$$

$$\times \langle \|x-T(x)\|^2 + (1-\|x\|^2)(1 + 2x_1 + 2\|x\|^2)x - (1-\|x\|^2)(L(x) + T(x), u \rangle .$$

For each x in l_2, let

$$F(x) = \frac{-2}{\|x - T(x)\|^4} \times [\ \|x-T(x)\|^2 + (1-\|x\|^2)(1 + 2x_1 + 2\|x\|^2)x - (1-\|x\|^2)(L(x) + T(x))].$$

Then $df(x)(u) = \langle F(x), u \rangle$ for all $u \in l_2$,

We note that if $df(x)(u) = 0$ for all $u \in l_2$, then $F(x) = 0$ for all $u \in l_2$.

Next, we show that there exists no x in B(0,1) such that $df(x) = 0$.

Suppose that $df(x) = 0$, for some $x \in B(0,1)$. Since $\frac{-2}{\|x - T(x)\|^4} \neq 0$, then

$$[\ \|x-T(x)\|^2 + (1-\|x\|^2)(1 + 2x_1 + 2\|x\|^2)x - (1-\|x\|^2)(L(x) + T(x))] = 0.$$

That is, $\left(\frac{\|x - T(x)\|^2}{1-\|x\|^2} + 1 + 2x_1 + 2\|x\|^2\right)x = L(x) + T(x).$

Let $s = \frac{\|x - T(x)\|^2}{1-\|x\|^2} + 1 + 2x_1 + 2\|x\|^2.$ .... (1)

Then $L(x) + T(x) = sx$ and $L^2(x) - sL(x) + x = 0$. So, for each $n \geq 1$,

$$x_{n+2} = sx_{n+1} - x_n .$$

That is, $x \in \mathrm{Ker}(L^2 - sL + I)$ is a recurrent sequence of order two in l_2. The associated characteristic equation for this type of sequence is

$$t^2 - st + 1 = 0. \qquad \text{..................................(2)}$$

This equation gives us three different alternatives according to the sign of its discriminant. That is, $|s| = 2, |s| < 2$ and $|s| > 2$.

Case 1. $|s| = 2$.

For each n $\geq$ 1,

$$x_n = A\left(\frac{s}{2}\right)^n + B(n)\left(\frac{s}{2}\right)^n, \text{ for some real numbers } A, B.$$

Since $(x_n) \in l_2$, $\lim_{n\to\infty} x_n = 0$, so we have that $A = B = 0$, i.e, $x = 0$.

Since $F(0) = 16e_1$, then $df(0)(u) = \langle F(0),u\rangle = \langle 16e_1,u\rangle = 16u_1$ for all $u \in l_2$,

and this is a contradiction.

Case 2. $|s| < 2$.

Then the characteristic equation has two complex roots given by

$$\alpha = cos\theta + i\,sin\theta,\ \beta = cos\theta - i\,sin\theta,\ sin\theta \neq 0.$$

Then, for each $n \geq 1$,

$$x_n = A(cos\theta + i\,sin\theta)^n + B(cos\theta - i\,sin\theta)^n.$$

$$= (A+B)cos\,n\theta + (A-B)i\,sin\,n\theta.$$

So $\quad x_n = C\cos n\theta + D\sin n\theta \quad$ where $C = A + B,\ D = (A - B)i$.

Since $sin\theta \neq 0$, then if C or D is nonzero, then the sequence (x_n) has no limit. But $\lim_{n\to\infty} x_n = 0$, then $C = D = 0$. That is, $x = 0$. As in the case $|s| = 2$, this is a contradiction.

Case 3. $|s| > 2$. We have two real roots

$$\alpha = \frac{s + \sqrt{s^2 - 4}}{2},\ \beta = \frac{s - \sqrt{s^2 - 4}}{2}.$$

We note that one of these roots has absolute value greater than one and the other less than one. Assume that $|\alpha| > 1$ and $|\beta| < 1$. Again, we have a solution of (2) is

$$x_n = A\alpha^n + B\beta^n, \qquad \text{for each } n \geq 1.$$

Since $\lim_{n\to\infty} x_n = 0$, we have $A = 0$.. Then $x_n = B\beta^{n-1}$, for each $n \geq 1$.

So $x = (x_1, x_1\beta, x_1\beta^2, \ldots)$, $\|x\|^2 = x_1^2\left(\frac{1}{1-\beta^2}\right)$, and

$$\|x - T(x)\|^2 = \left\| \left(x_1 - \left(\frac{1}{2} - \frac{x_1^2}{1-\beta^2}\right), x_1\beta - x_1, x_1\beta^2 - x_1\beta, \ldots \right) \right\|^2$$

$$= \left\| \left(x_1 + \frac{x_1^2}{1-\beta^2} - \frac{1}{2}, x_1(\beta - 1), x_1\beta(\beta - 1), \ldots \right) \right\|^2$$

$$= \left(x_1 + \frac{x_1^2}{1-\beta^2} - \frac{1}{2} \right)^2 + \frac{x_1^2(1-\beta)}{1+\beta}.$$

Since $\beta^2 - s\beta + 1 = 0$, then $s = (\beta + \frac{1}{\beta})$.

From $sx = T(x) + L(x)$, we have $sx_1 = \left(\frac{1}{2}\right) - \|x\|^2 + x_1\beta$.

Therefore

$$x_1^2 + \frac{1-\beta^2}{\beta}x_1 - \frac{1}{2}(1 - \beta^2) = 0. \qquad \ldots\ldots\ldots\ldots\ldots (3)$$

Since (3) implies

$$x_1^2 - \frac{1}{2}(1 - \beta^2) = -\left(\frac{1-\beta^2}{\beta}\right)x_1, \text{ then}$$

$$\|x - T(x)\|^2 = \left(\frac{x_1(1-\beta^2) + x_1^2 - \frac{1}{2}(1-\beta^2)}{1-\beta^2} \right)^2 + \frac{x_1^2(1-\beta)^2}{1-\beta^2}$$

$$= \left(\frac{x_1(1-\beta^2) + \frac{(1-\beta^2)}{\beta} x_1}{1-\beta^2} \right)^2 + \frac{x_1^2(1-\beta)^2}{1-\beta^2}$$

$$= \left(x_1 - \frac{x_1}{\beta} \right)^2 + \frac{x_1^2(1-\beta)^2}{1-\beta^2}$$

$$= x_1^2 \left[\left(1 - \frac{1}{\beta} \right)^2 + \left(\frac{1-\beta}{1+\beta} \right) \right]$$

$$= x_1^2 \left[\left(\frac{\beta - 1}{\beta} \right)^2 + \left(\frac{1-\beta}{1+\beta} \right) \right]$$

$$= x_1^2 \left[\frac{(1-\beta)^2(1+\beta) + \beta^2(1-\beta)}{\beta^2(1+\beta)} \right]$$

$$= \frac{x_1^2(1-\beta)}{\beta^2(1+\beta)} \left[(1-\beta^2) + \beta^2 \right]$$

$$= \frac{x_1^2(1-\beta)}{\beta^2(1+\beta)} .$$

From (3), we have $x_1^2 = \dfrac{(1-\beta^2)(\beta - 2x_1)}{2\beta}$ and $\dfrac{2x_1^2}{1-\beta^2} = \dfrac{\beta - 2x_1}{\beta}$.

Hence, by substituting in (1), we have

$$\beta + \frac{1}{\beta} = \frac{x_1^2(1-\beta)}{\beta^2(1+\beta)} \bullet \frac{1-\beta^2}{1-\beta^2 - x_1^2} + 1 + 2x_1 + \frac{2x_1^2}{1-\beta^2} ,$$

so

$$1 = \frac{x_1^2(1-\beta)^2}{\beta(1-\beta^2-x_1^2)} + \beta + 2x_1\beta + (\beta - 2x_1) - \beta^2$$

$$1 = \frac{x_1^2(1-\beta)^2}{\beta(1-\beta^2-x_1^2)} + (\beta - 2x_1)(1-\beta) + \beta$$

$$1 = \frac{(1-\beta^2)(\beta-2x_1)(1-\beta)^2}{2\beta^2(1-\beta^2-x_1^2)} + (\beta - 2x_1)(1-\beta) + \beta$$

$$1-\beta = \frac{(1-\beta^2)(\beta-2x_1)(1-\beta)^2}{2\beta^2(1-\beta^2-x_1^2)} + (\beta - 2x_1)(1-\beta)$$

$$1 = (\beta - 2x_1) + \frac{(1-\beta^2)(\beta-2x_1)(1-\beta)}{2\beta^2(1-\beta^2-x_1^2)}$$

$$1 = (\beta - 2x_1)\ 1 + \left[\frac{(1-\beta^2)(1-\beta)}{2\beta^2(1-\beta^2-x_1^2}\right] \qquad \text{..............................(4)}$$

From (3), we consider two subcases :

Subcase 3.1. $x_1 = \dfrac{-1+\beta^2-\sqrt{1-\beta^2}}{2\beta}$

Since $\|x\| < 1$, $1-\beta^2 > x^2$. i.e, $x_1^2 + \beta^2 < 1$.

From (4), we have that $0 < \beta - 2x_1 < 1$. Then

$$\beta - 2x_1 = \beta - 2\left(\frac{-1+\beta^2-\sqrt{1-\beta^4}}{2\beta}\right) = \frac{1+\sqrt{1-\beta^4}}{\beta},$$

so $0 < \dfrac{1+\sqrt{1-\beta^4}}{\beta} < 1$. Thus $\beta > 1 + \sqrt{1-\beta^4} > 1$.

A contradiction, since $|\beta| < 1$.

Subcase 3.2. $x_1 = \dfrac{-1+\beta^2+\sqrt{1-\beta^2}}{2\beta}$

Note that $1 - x_1^2 - \beta^2 = 1 - \left(\dfrac{-1+\beta^2+\sqrt{1-\beta^4}}{2\beta}\right)^2 - \beta^2$

$$= (1-\beta^2) - \left(\frac{1}{2\beta}\right)(1-\beta^2)\left[\frac{1-\beta^2+1+\beta^2-2\sqrt{1-\beta^4}}{2\beta}\right]$$

$$= \frac{1}{2\beta}(1-\beta^2)\left[\beta + \frac{2(-1+\beta^2+\sqrt{1-\beta^4}}{2\beta}\right]$$

$$= \frac{1}{2\beta}(1-\beta^2)(\beta - 2x_1).$$

From (4), we have that

$$1 = \frac{1}{\beta}\,(1 - \sqrt{1-\beta^4})\,\frac{2\beta^2-\beta+\sqrt{1-\beta^4}}{2\beta^2-1+\sqrt{1-\beta^4}}$$

$$\beta(2\beta^2 - 1 + \sqrt{1-\beta^4}) = (1 - \sqrt{1-\beta^4})(2\beta^2 - \beta + \sqrt{1-\beta^4})$$

$$2\beta^3 = (1 - \sqrt{1-\beta^4})(2\beta^2 - \beta + \sqrt{1-\beta^4}) + \beta(1-\sqrt{1-\beta^4})$$

$$2\beta^3 = (1 - \sqrt{1-\beta^4})(2\beta^2 + \sqrt{1-\beta^4})$$

$$2\beta^3(1+\sqrt{1-\beta^4}) = \beta^4(2\beta^2 + \sqrt{1-\beta^4})$$

$$2(1+\sqrt{1-\beta^4}) = \beta(2\beta + \sqrt{1-\beta^4})$$

$$2 + 2\sqrt{1-\beta^4} = 2\beta^3 + \beta\sqrt{1-\beta^4}$$

$$2(1-\beta^3) = (\beta - 2)\sqrt{1-\beta^4}.$$

This is also a contradiction, since $2(1-\beta^3) > 0$ but $(\beta - 2)\sqrt{1-\beta^4} < 0$

With ϕ = the identity function on **R**, and z_0 be any point in B(0,1), we have for each x in S(0,1), $\phi(f(x)) = f(x) = 0$ and $\phi(f(x) - f(z_0)) = f(x) - f(z_0) = -f(z_0)$ in S(0,1) but there is no point x in B(0,1) such that $df(x) = 0$ on l_2.

The following results are directly obtained when replace $D(x_o, r)$ by the closure of any open bounded set of $\mathbf{R}^n$ in Theorem 3.1 and Theorem 3.2.

For a subset A of $\mathbf{R}^n$, let $\overline{A}$ and ∂A denote the closure of A and the boundary of A, respectively. We have the following theorem.

Theorem 3.3. *Let A be any open and bounded subset of $\mathbf{R}^n$, $f : \overline{A} \to B$ be continuous on $\overline{A}$ and Fréchet differentiable on A. Let $\phi : \mathbf{B} \to \mathbf{R}$ be a continuous linear functional and let z_o be an element in A such that $\phi(f(x) - f(z_o))$ does not change sign for all $x \in \partial A$. Then there exists $c \in A$ such that*

$$\phi(df(c)(u)) = 0 \qquad \textit{for all } u \in \mathbf{R}^n.$$

Proof. Since $\overline{A}$ is a compact, then the proof is similar to that of theorem 3.1 with $D(x_o, r)$ is replaced by $\overline{A}$. ■

And also we have the following corollary.

Corollary 3.4. *Let* $f : \overline{A} \to \mathbf{B}$ *be continuous on* $\overline{A}$ *and Fréchet differentiable on* A. *Assume that there exists a continuous linear functional* $\phi : \mathbf{B} \to \mathbf{R}$ *such that* $\phi(f(x))$ *is constant for all* $x \in \partial A$. *Then there exists* $c \in A$ *such that*

$$\phi(df(c)(u)) = 0 \qquad \text{for all } u \in R^n.$$

Rolle's theorem can be proved by using corollary 3.4.

Corollary 3.5. (Rolle's Theorem)

Let $f : [a, b] \to \mathbf{R}$ *be continuous on* $[a, b]$ *and differentiable on* (a, b). *If* $f(a) = f(b)$ *then there is a point* $c \in (a, b)$ *such that* $df(c) = 0$.

Proof. Assume that $f(a) = f(b)$. Let $\phi : \mathbf{R} \to \mathbf{R}$ by $\phi(x) = x$. Then ϕ is a continuous linear functional such that $\phi(f(a)) = f(a) = f(b) = \phi(f(b))$. By corollary 3.4, there is $c \in (a,b)$ such that $\phi(df(c)(u)) = 0$ for all $u \in \mathbf{R}$. Then $\phi(df(c)(u)) = df(c)(u) = 0$ for all $u \in \mathbf{R}$, so $df(c) = 0$. ■

In 1996, Yamsakulna [7] proved Rolle's Theorem in Hilbert space. The corollary 3.7 is Yamsakulna's Rolle's Theorem. It can be obtained immediatly from theorem 3.2 as shown below.

Corollary 3.6. *Let* $\mathbf{H}$ *be a Hilbert space. Let* $f : D(x_o, r) \to \mathbf{H}$ *be continuous on* $D(x_o, r)$ *and differentiable on* $B(x_o, r)$. *Assume that there exists a vector* $v \in \mathbf{H}$ *such that* $\langle v, f(x)\rangle$ *is constant on* $S(x_o, r)$. *Then there exists a vector* $c \in B(x_o, r)$ *such that* $\langle v, df(c)(u)\rangle = 0$ *for all* $u \in \mathbf{R}^n$.

Proof. Let $v \in \mathbf{H}$ be such that $g(x) = \langle v, f(x)\rangle$ is constant on $S(x_o, r)$. Let $\phi : \mathbf{H} \to \mathbf{R}$ by $\phi(x) = \langle v, x\rangle$. Then ϕ is continuous linear functional and $\phi(f(x)) = \langle v, f(x)\rangle$ is constant in $S(x_o, r)$. By theorem 3.2, there exists $c \in B(x_o, r)$ such that $\phi(df(c)(u)) = 0$ for all $u \in \mathbf{R}^n$. Thus

$$\langle v, df(c)(u)\rangle = \phi(df(c)(u)) = 0 \quad \text{for all } u \in \mathbf{R}^n.$$

■

The next corollary is the Rolle's theorem of Furi and Martelli [4], mentioned in the chapter I.

Corollary 3.7. *Let* $f : D(x_o, r) \to \mathbf{R}^p$ *be continuous on* $D(x_o, r)$ *and differentiable on* $B(x_o, r)$. *Assume that there exists a vector* $v \in \mathbf{R}^p$ *such that* $v \cdot f(x)$ *is constant on* $S(x_o, r)$. *Thus there exists a vector* $c \in B(x_o, r)$ *such that*

$$v \cdot df(c)(u) = 0 \qquad \text{for every } u \in \mathbf{R}^n.$$

As a consequence of theorem 3.2 we have a theorem analogous to Sanderson's Mean Value Theorem.

Theorem 3.8. *Let* $a < b$ *and* f : [a, b] $\rightarrow$ **B** *be* k *times Fréchet differentiable. Assume that there exists a continuous linear functional* ϕ : **B** $\rightarrow$ **R** *such that* $(\phi \circ f)(a) = (\phi \circ f)(b) = 0$ *and the first* k-1 *derivatives of* f *at* a *and* $\phi \circ d^{k-1} f(a)$ *are* 0. *Then for some* $c \in (a, b)$, $\phi \circ d^k f(c) = 0$.

Proof. By theorem 3.2, there exists $c_1 \in (a, b)$ such that $\phi \circ df(c_1) = 0$. The theorem 3.2 can be applied to df in the interval $[a, c_1]$, so there exists $c_2 \in (a, c_1)$ such that $\phi \circ d(df)(c_2) = 0$. i.e. $\phi \circ d^2 f(c_2) = 0$. Again, we apply the theorem 3.2 to $d^2 f$ in the interval $[a, c_2]$, so there exists $c_3 \in (a, c_2)$ such that $\phi \circ d^3 f(c_3) = 0$. This procedure can be repeated k-1 times to obtain $c = c_k \in (a, c_{k-1}) \subseteq (a,b)$ such that $\phi \circ d^k f(c) = 0$. ■

The next theorem is the Mean Value Theorem of Sanderson [6] We show here that the theorem can be easily proved by using the theorem 3.8.

Theorem 3.9. (Sanderson's Mean Value theorem)

Suppose f : $[a, b] \rightarrow$ $\mathbf{R}^n$ *is a* k *times differentiable n-dimensional vector-valued function and* $f(a)$, $f(b)$ *and the first k-1 derivatives of* f *at* a *are orthogonal to a non-zero vector* v_0. *Then for some* c *between* a *and* b, v_0 *is orthogonal to* $d^k f(c)$.

Proof. Let $\phi : \mathbf{R}^n \to \mathbf{R}$ by $\phi(x) = x \cdot v_0$ for each $x \in \mathbf{R}^n$. Then ϕ is continuous linear functional and $\phi(f(a)) = f(a) \cdot v_0 = 0 = f(b) \cdot v_0 = \phi(f(b))$. By theorem 3.8., there exists $c \in (a, b)$ such that $\phi \circ d^k f(c) = 0$. Thus

$$d^k f(c)(u) \cdot v_0 = (\phi \circ d^k f(c))(u) = 0, \quad \text{for all } u \in \mathbf{R}^n.$$

So v_0 is orthogonal to $d^k f(c)$. ∎