

Returning to evaluations involving the Taylor series representation of  $x$  (7.23) takes the form

$$f_l(x(t)) = \sum_{n=0}^{\infty} \phi_l(a)_n t^n, \quad l = 1, \dots, L \quad (7.24)$$

where

$$\phi_l(a) = \sum_{|\alpha|=0}^M c_{l,\alpha} a^\alpha = \sum_{|\alpha|=0}^M c_{l,\alpha} a_1^{\alpha_1} * \dots * a_n^{\alpha_n}. \quad (7.25)$$

Observe that this expression is just the polynomial  $f_l$  with multiplication re-interpreted as the Cauchy product  $*$ .

We need to compute derivatives and we leave it to the reader to check that

$$D_j \phi_l(a) = \frac{\partial}{\partial a_j} \phi_l(a) = \begin{cases} \sum_{|\alpha|=0}^M \alpha_j c_{\alpha,l} a_1^{\alpha_1} * \dots * a_j^{\alpha_j-1} * \dots * a_n^{\alpha_n} & \text{if } \alpha_j \geq 1 \\ 0 & \text{otherwise.} \end{cases} \quad (7.26)$$

## 7.6 Radii Polynomial Approach on Banach spaces

Consider the problem of finding a fixed point of a nonlinear map  $T: X \rightarrow X$  with  $X$  a Banach space, or of finding a zero of a function  $F: X \rightarrow Y$  with  $X$  and  $Y$  Banach spaces. In the first half of the book, when considering finite dimensional problems, we found that the contraction mapping theorem and Newton's method were the appropriate tools. We even developed a-posteriori analysis which facilitated computer assisted proofs based on these ideas. In this Section we aim to extend these tools to the setting of Banach spaces.

The following theorem facilitates a-posteriori analysis of fixed point problems.

**Theorem 7.6.1.** *Suppose that  $X$  is a Banach space, that  $T: X \rightarrow X$  is a Fréchet differentiable mapping, and that  $\bar{x} \in X$ . Let  $Y_0 \geq 0$  and  $Z: (0, \infty) \rightarrow [0, \infty)$  a non-negative function satisfying*

$$\|T(\bar{x}) - \bar{x}\|_X \leq Y_0, \quad (7.27)$$

and

$$\sup_{x \in \overline{B_r(\bar{x})}} \|DT(x)\|_{B(X)} \leq Z(r). \quad (7.28)$$

Define the radii polynomial

$$p(r) := Z(r)r - r + Y_0.$$

If there exists  $r_0 > 0$  such that  $p(r_0) < 0$ , then there exists a unique  $\tilde{x} \in \overline{B_{r_0}(\bar{x})}$  so that  $T(\tilde{x}) = \tilde{x}$ .

*Proof.* By the contraction mapping theorem, it is sufficient to show that  $T: \overline{B_{r_0}(\bar{x})} \rightarrow \overline{B_{r_0}(\bar{x})}$  is a contraction mapping.

We begin by showing that  $T$  restricted to  $\overline{B_{r_0}(\bar{x})}$  is a contraction mapping. Let  $x_1, x_2 \in \overline{B_{r_0}(\bar{x})}$ . By the Mean Value Theorem (Theorem 7.2.4),

$$\begin{aligned}\|T(x_1) - T(x_2)\|_X &\leq \sup_{x \in \overline{B_{r_0}(\bar{x})}} \|DT(x)\|_{B(X)} \|x_1 - x_2\|_X \\ &\leq Z(r_0) \|x_1 - x_2\|_X.\end{aligned}\tag{7.29}$$

Observe that

$$(Z(r_0) - 1)r_0 \leq (Z(r_0) - 1)r_0 + Y_0 = p(r_0) < 0$$

and hence  $Z(r_0) < 1$ . Therefore,  $T$  is a contraction mapping.

To complete the proof we show that  $T(\overline{B_{r_0}(\bar{x})}) \subset \overline{B_{r_0}(\bar{x})}$ . Let  $y \in \overline{B_{r_0}(\bar{x})}$ . Observe that

$$\begin{aligned}\|T(y) - \bar{x}\|_X &\leq \|T(y) - T(\bar{x})\|_X + \|T(\bar{x}) - \bar{x}\|_X \\ &\leq Z(r_0) \|y - \bar{x}\|_X + \|T(\bar{x}) - \bar{x}\|_X && \text{by (7.29)} \\ &\leq Z(r_0)r_0 + Y_0 && \text{by (7.27)} \\ &< r_0.\end{aligned}\quad \square$$

Consider a Fréchet differentiable map  $F: X \rightarrow X$ . In line with the philosophy of Chapter 2 assume that  $\bar{a} \in X$  is an approximate zero of  $F$ , i.e.,  $\|F(\bar{a})\|_X \approx 0$ , and that  $DF(\bar{a})$  is invertible. Then, fixed points of the Newton-like map

$$\hat{T}(a) = a - DF(\bar{a})^{-1}F(a)$$

are in one-to-one correspondence with zeros of  $F$ . Thus, if  $\hat{T}$  is a contraction mapping in a neighborhood of  $\bar{a}$ , then we can conclude the existence and uniqueness of a zero of  $F$ .

As is observed repeatedly in the examples of Sections 2.4 and 4.5 and Chapter 3 inverting  $DF(\bar{a})$  is in some sense “too hard.” Thus we make use of an approximate inverse, typically obtained by computing a numerical inverse. However, we are now working with infinite dimensional operators, thus finding an exact expression for  $DF(\bar{a})^{-1}$  is even more challenging and making use of a numerical inverse is no longer an option. With this in mind we adopt a slightly different approach and approximate  $DF(\bar{a})$  by an operator  $A^\dagger \in B(X)$  specifically chosen to be easier to work with. We then choose  $A \in B(X)$  to be an approximate inverse of  $A^\dagger$  and define the Newton-like operator

$$T(a) = a - AF(a).$$

Observe that if  $A$  is injective, then fixed points of  $T$  correspond to zeros of  $F$ . The following theorem provides conditions under which we can guarantee the existence of a fixed point of  $T$  and hence a zero of  $F$ . The assumptions are more general in the sense that we do not require that  $F: X \rightarrow X$  but instead that  $AF: X \rightarrow X$ .

**Theorem 7.6.2 (Radii polynomial approach on Banach spaces).** *Let  $X$  and  $Y$  be Banach spaces and  $F: X \rightarrow Y$  be a Fréchet differentiable mapping. Suppose that  $\bar{x} \in X$ ,  $A^\dagger \in B(X, Y)$ , and  $A \in B(Y, X)$ . Moreover assume that  $A$  is injective. Let  $Y_0$ ,  $Z_0$ , and  $Z_1$  be positive constants and  $Z_2: (0, \infty) \rightarrow [0, \infty)$  be a non-negative function satisfying*

$$\|AF(\bar{x})\|_X \leq Y_0, \quad (7.30)$$

$$\|I - AA^\dagger\|_{B(X)} \leq Z_0, \quad (7.31)$$

$$\|A[DF(\bar{x}) - A^\dagger]\|_{B(X)} \leq Z_1, \quad (7.32)$$

and

$$\|A[DF(c) - DF(\bar{x})]\|_{B(X)} \leq Z_2(r)r, \quad \text{for all } c \in \overline{B_r(\bar{x})} \text{ and all } r > 0. \quad (7.33)$$

Define

$$p(r) := Z_2(r)r^2 - (1 - Z_0 - Z_1)r + Y_0. \quad (7.34)$$

If there exists  $r_0 > 0$  such that  $p(r_0) < 0$ , then there exists a unique  $\tilde{x} \in B_{r_0}(\bar{x})$  satisfying  $F(\tilde{x}) = 0$ .

*Proof.* The idea of the proof is to define the Newton-like operator  $T: X \rightarrow X$  by

$$T(x) = x - AF(x)$$

and to apply Theorem 7.6.1.

Since  $F$  is Fréchet differentiable and  $A \in B(Y, X)$ ,  $T$  is Fréchet differentiable and

$$DT(x) = I - ADF(x).$$

Define  $Z(r) \stackrel{\text{def}}{=} Z_0 + Z_1 + Z_2(r)r$ , and observe that for  $x \in \overline{B_r(\bar{x})}$ ,

$$\begin{aligned} \|DT(x)\|_{B(X)} &= \|I - ADF(x)\|_{B(X)} \\ &\leq \|I - AA^\dagger\|_{B(X)} + \|A[A^\dagger - DF(\bar{x})]\|_{B(X)} + \|A[DF(\bar{x}) - DF(x)]\|_{B(X)} \\ &\leq Z_0 + Z_1 + Z_2(r)r = Z(r) \end{aligned} \quad (7.35)$$

where the last inequality follows from assumptions (7.31), (7.32) and (7.33).

Moreover, from (7.30),  $\|T(\bar{x}) - \bar{x}\|_X = \|AF(\bar{x})\|_X \leq Y_0$ . Since

$$p(r) = Z_2(r)r^2 - (1 - Z_0 - Z_1)r + Y_0 = Z(r)r - r + Y_0$$

and  $p(r_0) < 0$ , we conclude from Theorem 7.6.1 that there exists a unique  $\tilde{x} \in \overline{B_{r_0}(\bar{x})}$  so that  $T(\tilde{x}) = \tilde{x}$ . Since  $A$  is assumed to be injective,  $F(\tilde{x}) = 0$  if and only if  $T(\tilde{x}) = \tilde{x}$ .  $\square$

We conclude this section with some trivial but useful results concerning the hypothesis of Theorem 7.6.2.

**Corollary 7.6.3.** *Let  $p(r)$  be the radii polynomial (7.34). If there exists  $r_0 > 0$  such that  $p(r_0) < 0$ , then  $Z_0 + Z_1 < 1$ .*

Combining Corollary 7.6.3 with Proposition 7.6.5 gives the following result.

**Corollary 7.6.4.** *Let  $A^\dagger \in B(X, Y)$  and  $A \in B(Y, X)$  be as in Theorem 7.6.2. Let  $p(r)$  be the radii polynomial (7.34). If there exists  $r_0 > 0$  such that  $p(r_0) < 0$ , then  $AA^\dagger$  is injective.*

Observes that this suggests that if we implement the Radii Polynomial Approach and find  $r_0 > 0$  such that  $p(r_0) < 0$ , then perhaps it is not necessary to explicitly check that  $A$  is injective. The following result, which is applicable to all the problems considered in this text, provides conditions under which this is the case.

**Proposition 7.6.5.** *Consider  $A \in B(Y, X)$  and  $A^\dagger \in B(X, Y)$  where  $X$  and  $Y$  are Banach spaces such that  $\|I - AA^\dagger\|_{B(X)} < 1$ . Assume that  $X = X_0 \oplus X_1$  and  $Y = Y_0 \oplus Y_1$ , where  $X_0 \cong Y_0$  are finite dimensional. Assume that the operators  $A$  and  $A^\dagger$  can be decomposed as follows*

$$A = \iota_0 \circ A_0 \circ \pi_0 + \iota_1 \circ A_1 \circ \pi_1 \quad (7.36)$$

$$A^\dagger = \iota_0 \circ A_0^\dagger \circ \pi_0 + \iota_1 \circ A_1^\dagger \circ \pi_1 \quad (7.37)$$

where for  $i = 0, 1$ ,  $A_i: Y_i \rightarrow X_i$ ,  $A_i^\dagger: X_i \rightarrow Y_i$ , and  $\iota_i$  and  $\pi_i$  are the appropriate canonical inclusion and projection maps. Furthermore, assume that  $A_1$  is injective. Then,  $A$  is injective.

*Proof.* The assumption that  $A_1$  is injective combined with (7.36) implies that it is sufficient to prove that  $A_0$  is injective. By Corollary 7.6.4  $AA^\dagger$  is injective and therefore by (7.36) and (7.37)  $A_0B_0: X_0 \rightarrow X_0$  is injective. The assumption that  $X_0$  is finite dimensional implies that  $A_0B_0$  is an isomorphism. The assumption that  $X_0 \cong Y_0$  implies that  $A_0$  and  $B_0$  are isomorphisms and therefore  $A_0$  is injective.  $\square$

## 7.7 A first example: Taylor series solution of parameterized equilibria

Consider the simplest case of a one-parameter family of one-dimensional ODEs,

$$\dot{x} = f(x, \lambda), \quad x \in \mathbb{R}, \lambda \in \mathbb{R}.$$

Assume  $f(x_0, \lambda_0) = 0$ . If  $f_x(x_0, \lambda_0) \neq 0$ , then the implicit function theorem implies that there exist a neighborhood  $U$  of  $\lambda_0$  and a smooth function  $x: U \rightarrow \mathbb{R}$  such that  $x(\lambda_0) = x_0$  and

$$f(x(\lambda), \lambda) = 0, \quad \lambda \in U. \quad (7.38)$$

If  $f$  is analytic, then  $x$  is analytic and hence there exists  $\tau > 0$  on which  $x: (\lambda_0 - \tau, \lambda_0 + \tau) \rightarrow \mathbb{R}$  can be represented via the power series expansion

$$x(\lambda) = \sum_{n=0}^{\infty} a_n(\lambda - \lambda_0)^n \quad (7.39)$$

with  $a_0 = x_0$ .

In keeping with the philosophy of this book our goal is to provide both an explicit approximation  $\bar{x}: U \rightarrow \mathbb{R}$  of  $x$  and an explicit bound  $\|\bar{x} - x\|_\infty$  over  $U$  of the error of our approximation. The Taylor series expansion suggests how this might be done: find  $\{\bar{a}_n\}_{n=0,\dots,N}$  where  $\bar{a}_n \approx a_n$  such that

$$f\left(\sum_{n=0}^N \bar{a}_n(\lambda - \lambda_0)^n, \lambda\right) \approx 0, \quad \lambda \in U$$

and then use the Radii Polynomial approach to prove that there exists  $\tilde{a} = \{\tilde{a}_n\}_{n \in \mathbb{N}}$  such that  $\tilde{a} \approx \bar{a}$  and

$$f\left(\sum_{n=0}^{\infty} \tilde{a}_n(\lambda - \lambda_0)^n, \lambda\right) = 0, \quad \lambda \in U.$$

Before turning to the mathematical technicalities let us formally consider an explicit example,

$$f(x(\lambda), \lambda) = [x(\lambda)]^2 - \lambda = 0. \quad (7.40)$$

Substituting the power series expansion for  $x$  into Equation (7.40) leads to

$$\begin{aligned} 0 &= f(x(\lambda), \lambda) \\ &= \left( \sum_{n=0}^{\infty} a_n(\lambda - \lambda_0)^n \right)^2 - \lambda \\ &= \sum_{n=0}^{\infty} (a * a)_n(\lambda - \lambda_0)^n - \lambda \\ &= a_0^2 + 2a_0a_1(\lambda - \lambda_0) + \sum_{n=2}^{\infty} \sum_{j=0}^n a_{n-j}a_j(\lambda - \lambda_0)^n - (\lambda - \lambda_0) - \lambda_0, \end{aligned}$$

where

$$(a * a)_n \stackrel{\text{def}}{=} \sum_{j=0}^n a_{n-j}a_j$$

is the Cauchy product. Matching like powers of  $\lambda - \lambda_0$  leads to the infinite system of equations

$$\begin{aligned} a_0^2 - \lambda_0 &= 0 \\ 2a_0a_1 - 1 &= 0 \\ (a * a)_n &= 0, \quad n \geq 2, \end{aligned} \quad (7.41)$$

with infinitely many unknowns  $\{a_n\}_{n \in \mathbb{N}}$ .

Setting  $a = \{a_n\}_{n=0}^\infty$  we recast the problem of finding an explicit description of the curve of equilibria to that of solving

$$F(a) = 0 \quad (7.42)$$

where the nonlinear mapping  $F$  is defined component-wise by

$$F_n(a) \stackrel{\text{def}}{=} \begin{cases} a_0^2 - \lambda_0 & \text{if } n = 0 \\ 2a_1 a_0 - 1 & \text{if } n = 1 \\ (a * a)_n & \text{if } n \geq 2. \end{cases} \quad (7.43)$$

Observe that the map  $F$  may be more densely written as

$$F(a) = a * a - c, \quad (7.44)$$

where  $c = \{c_n\}_{n \in \mathbb{N}}$  is given by

$$c_n = \begin{cases} \lambda_0 & \text{if } n = 0 \\ 1 & \text{if } n = 1 \\ 0 & \text{if } n \geq 2. \end{cases}$$

Having derived a purely formal zero finding problem (7.42), we now turn to the mathematical technicalities that puts us into a position to apply Theorem 7.6.2 to find solutions of  $F(a) = 0$ . We treat this as a five step process. The first four steps establish the appropriate Banach spaces, an approximate solution, and the linear operators  $A$  and  $A^\dagger$ . The fifth step is to establish the  $Y_0$ ,  $Z_0$ ,  $Z_1$ , and  $Z_2$  bounds.

The first step towards applying Theorem 7.6.2 is to choose appropriate Banach spaces  $X$  and  $Y$ . Observe that since we are assuming that  $f$  is analytic the Taylor series (7.39) converges absolutely and uniformly for all  $|\lambda - \lambda_0| < \tau$ , that is

$$\sum_{n=0}^{\infty} |a_n| \nu^n < \infty \quad (7.45)$$

for all  $\nu \in [0, \tau)$ . This implies that  $a = \{a_k\}_{k \in \mathbb{N}} \in \ell_{\nu, \mathbb{N}}^1$  and therefore we choose  $X = \ell_{\nu, \mathbb{N}}^1$ . By Theorem 7.4.4,  $X$  is a commutative Banach algebra. This implies that the use of the Cauchy product in (7.44) is well defined and  $F: \ell_{\nu, \mathbb{N}}^1 \rightarrow \ell_{\nu, \mathbb{N}}^1$ . Therefore, we choose  $Y = \ell_{\nu, \mathbb{N}}^1$ . Observe that by Theorem 7.4.7  $F$  is Fréchet differentiable.

The second step towards applying Theorem 7.6.2 is to choose  $\bar{a} \in \ell_{\nu, \mathbb{N}}^1$ , an approximate zero of  $F$ . From a computational perspective we cannot work with an arbitrary infinite sequence and thus, as suggested above, we choose  $\bar{a}$  of the form

$$\bar{a}_k = \begin{cases} \bar{a}_k^{(N)} & \text{if } k = 0, \dots, N \\ 0 & \text{otherwise} \end{cases}$$

where  $\bar{a}^{(N)} = (\bar{a}_0^{(N)}, \dots, \bar{a}_N^{(N)}) \in \mathbb{R}^{N+1}$ . Observe that we can solve for  $\bar{a}^{(N)}$  by solving a truncated version of (7.41)

$$\begin{aligned} a_0^2 - \lambda_0 &= 0 \\ 2a_0 a_1 - 1 &= 0 \\ (a * a)_n &= 0, \quad 2 \leq n \leq N. \end{aligned} \tag{7.46}$$

For simplicity of presentation, for  $n = 0, \dots, N$ , we use the notation  $\bar{a}_n$  as opposed to  $\bar{a}_n^{(N)}$ .

The third step towards applying Theorem 7.6.2 is to choose an operator  $A^\dagger$  that approximates  $DF(\bar{a})$ . With this in mind consider the map  $F^{(N)}: \mathbb{R}^{N+1} \rightarrow \mathbb{R}^{N+1}$  defined by

$$F_n^{(N)}(a_0, \dots, a_N) \stackrel{\text{def}}{=} \begin{cases} a_0^2 - \lambda_0 & \text{if } n = 0 \\ 2a_1 a_0 - 1 & \text{if } n = 1 \\ (a * a)_n & \text{if } 2 \leq n \leq N. \end{cases}$$

Observe that we can explicitly compute  $DF^{(N)}(\bar{a})$ . In particular,

$$(DF(a))_{0,0} = 2\bar{a}_0,$$

for  $1 \leq i \leq n$ ,

$$(DF(a))_{n,i} = \frac{\partial F_k}{\partial a_i} = \frac{\partial}{\partial a_i} (a * a)_n = \frac{\partial}{\partial a_i} \left( \sum_{j=0}^n a_{n-j} a_j \right) = 2a_{n-i},$$

and for  $i > n$ ,

$$\frac{\partial F_n}{\partial a_i} = 0$$

since  $F_n$  only depends on the variables  $a_0, \dots, a_N$ .

Therefore we think of the operator  $DF(\bar{a})$  as a lower triangular infinite-dimensional matrix with all the diagonal terms given by  $2\bar{a}_0$ . Consider the lower diagonal terms. Without computing  $\bar{a}$  we do not have explicit information concerning  $(DF(\bar{a}))_{n,i}$  for  $0 \leq n - i \leq N$ . However,  $a \in \ell_{\nu, \mathbb{N}}^1$  implies that  $a_n \rightarrow 0$  rapidly as  $n \rightarrow \infty$ . In particular, as the difference  $n - i \geq 0$  grows – in other words as we move away from the diagonal in the “southwest” direction – the term  $(DF(a))_{n,i} = 2a_{n-i}$  decreases rapidly to zero. Our true solution  $\tilde{a} \in \ell_{\nu, \mathbb{N}}^1$ , thus we can hope that for  $0 \leq n - i \leq N$  the terms of  $(DF(\bar{a}))_{n,i} = 2\bar{a}_{n-i}$  are not consequential.

This suggests that we approximate  $DF(\bar{a})$  by the operator  $A^\dagger: \ell_\nu^1 \rightarrow \ell_\nu^1$  whose action on a vector  $h \in \ell_\nu^1$  is given by

$$(A^\dagger h)_k \stackrel{\text{def}}{=} \begin{cases} [DF^{(N)}(\bar{a})h^{(N)}]_n, & 0 \leq n \leq N \\ 2\bar{a}_0 h_n, & n \geq N + 1 \end{cases} \tag{7.47}$$

where  $h^{(N)} = (h_0, h_1, \dots, h_N) \in \mathbb{R}^{N+1}$  is the projection of  $h$  onto its first  $(N + 1)$  coordinates.

The fourth step towards applying Theorem 7.6.2 is to choose an injective linear operator  $A$  that is an approximate inverse of  $A^\dagger$ . Assuming  $a_0 \neq 0$  define  $A: \ell_\nu^1 \rightarrow \ell_\nu^1$  by

$$(Ah)_n \stackrel{\text{def}}{=} \begin{cases} [A^{(N)}h^{(N)}]_n, & 0 \leq n \leq N \\ \frac{1}{2\bar{a}_0}h_n, & n \geq N + 1, \end{cases} \quad (7.48)$$

where  $A^{(N)}$  is a numerical inverse of  $DF^{(N)}(\bar{a})$ .

Conceptually, it may be of use to think of  $A^\dagger$  and  $A$  as infinite square matrices

$$A^\dagger = \begin{bmatrix} DF^{(N)}(\bar{a}) & 0 \\ 0 & \Lambda \end{bmatrix} \quad \text{and} \quad A = \begin{bmatrix} A^{(N)} & 0 \\ 0 & \Lambda^{-1} \end{bmatrix}$$

where  $\Lambda$  is an infinite diagonal matrix with constant diagonal entries  $2\bar{a}_0$ . A direct application of Proposition 7.6.5 implies that if we identify a value  $r_0$  at which the radii polynomial  $p(r_0) < 0$ , then  $A$  is injective.

Having defined the Banach spaces  $X = Y = \ell_\nu^1$ , the nonlinear function  $F: \ell_\nu^1 \rightarrow \ell_\nu^1$ , the linear operators  $A, A^\dagger \in B(X)$ , and determining that  $A$  is invertible, we turn to fifth step towards applying Theorem 7.6.2 that of identifying the Radii Polynomial bounds  $Y_0$ ,  $Z_0$ ,  $Z_1$ , and  $Z_2$ . These are presented in the following theorem.

**Theorem 7.7.1.** *Fix  $\nu > 0$  and define the constants*

$$\begin{aligned} Y_0 &\stackrel{\text{def}}{=} \sum_{n=0}^N |[A^{(N)}F^{(N)}(\bar{a})]_n| \nu^n + \frac{1}{2|\bar{a}_0|} \sum_{n=N+1}^{2N} \sum_{j=0}^{2N-n} |\bar{a}_{N-j}| |\bar{a}_{n-N+j}| \nu^n \\ Z_0 &\stackrel{\text{def}}{=} \left\| I - A^{(N)}DF^{(N)}(\bar{a}) \right\|_{1,\nu} \\ Z_1 &\stackrel{\text{def}}{=} \frac{1}{|\bar{a}_0|} \sum_{n=1}^N |\bar{a}_n| \nu^n \\ Z_2 &\stackrel{\text{def}}{=} 2 \max \left( \|A^{(N)}\|_{1,\nu}, \frac{1}{2|\bar{a}_0|} \right) \end{aligned}$$

where, given  $B \in M_{N+1}(\mathbb{R})$ ,

$$\|B\|_{1,\nu} \stackrel{\text{def}}{=} \max_{0 \leq n \leq N} \frac{1}{\nu^n} \sum_{m=0}^N |B_{m,n}| \nu^m.$$

Then,  $Y_0$ ,  $Z_0$ ,  $Z_1$  and  $Z_2$  as defined above satisfy the hypotheses of Theorem 7.6.2.

*Proof.* First note that

$$(\bar{a} * \bar{a})_n = \begin{cases} \sum_{j=0}^n \bar{a}_{n-j} \bar{a}_j, & 0 \leq n \leq N \\ \sum_{j=0}^{2N-n} \bar{a}_{N-j} \bar{a}_{n-N+j}, & N+1 \leq n \leq 2N \\ 0, & n \geq 2N+1 \end{cases}$$

as  $\bar{a}_n = 0$  for  $n \geq N+1$ . Thus,

$$[AF(\bar{a})]_n = \begin{cases} [A^{(N)} F^{(N)}(\bar{a})]_n, & 0 \leq n \leq N \\ \frac{1}{2\bar{a}_0} \sum_{j=0}^{2N-n} \bar{a}_{N-j} \bar{a}_{n-N+j}, & N+1 \leq n \leq 2N \\ 0, & n \geq 2N+1 \end{cases}$$

and

$$\begin{aligned} \|AF(\bar{a})\|_{1,\nu} &= \sum_{n=0}^{\infty} |[AF(\bar{a})]_n| \nu^n \\ &= \sum_{n=0}^N |[A^{(N)} F^{(N)}(\bar{a})]_n| \nu^n + \sum_{n=N+1}^{2N} |[AF(\bar{a})]_n| \nu^n \\ &\leq \sum_{n=0}^N |[A^{(N)} F^{(N)}(\bar{a})]_n| \nu^n + \frac{1}{2|\bar{a}_0|} \sum_{n=N+1}^{2N} \sum_{j=0}^{2N-n} |\bar{a}_{N-j}| |\bar{a}_{n-N+j}| \nu^n = Y_0. \end{aligned}$$

Next, for  $h \in \ell_{\nu}^1$ ,

$$[(I - AA^\dagger) h]_n = \begin{cases} [(I - A^{(N)} DF^{(N)}(\bar{a})) h^{(N)}]_n, & 0 \leq n \leq N \\ 0, & n \geq N+1. \end{cases}$$

Then

$$\|I - AA^\dagger\| = \sup_{\|h\|_{1,\nu}=1} \|[I - AA^\dagger]h\|_{1,\nu} \leq \|I - A^{(N)} DF^{(N)}(\bar{a})\|_{1,\nu} = Z_0.$$

By Remark 7.4.8,

$$DF(a)h = 2a * h,$$

for  $h \in \ell_{\nu}^1$ . Moreover,

$$\begin{aligned} [(DF(\bar{a}) - A^\dagger)h]_n &= \begin{cases} [DF^{(N)}(\bar{a})h^{(N)}]_n - [DF^{(N)}(\bar{a})h^{(N)}]_n, & 0 \leq n \leq N \\ 2(\bar{a} * h)_n - 2\bar{a}_0 h_n, & n \geq N+1 \end{cases} \\ &= \begin{cases} 0, & 0 \leq n \leq N \\ 2 \sum_{j=1}^N h_{n-j} \bar{a}_j, & n \geq N+1, \end{cases} \end{aligned}$$

which exploits the fact that  $\bar{a}_n = 0$  for  $n \geq N + 1$ . Then

$$[A(DF(\bar{a}) - A^\dagger)h]_n = \begin{cases} 0, & 0 \leq n \leq N \\ \frac{1}{\bar{a}_0} \sum_{j=1}^N h_{n-j} \bar{a}_j, & n \geq N + 1. \end{cases}$$

In order to better understand this term we define  $\hat{a} \in \ell_\nu^1$  by  $\hat{a}_n = 0$  if  $n = 0$ , or  $n \geq N + 1$  and  $\hat{a}_n = \bar{a}_n^{(N)}$  for  $1 \leq n \leq N$ , i.e.  $\hat{a} = (0, \bar{a}_1, \dots, \bar{a}_N, 0, \dots)$ . Now for any  $h \in \ell_\nu^1$  with  $\|h\|_{1,\nu} = 1$ ,

$$\begin{aligned} \|A[DF(\bar{a}) - A^\dagger]h\|_{1,\nu} &= \sum_{n=N+1}^{\infty} \frac{1}{|\bar{a}_0|} \left| \sum_{j=1}^N h_{n-j} \bar{a}_j \right| \nu^n \\ &\leq \frac{1}{|\bar{a}_0|} \sum_{n=N+1}^{\infty} \left| \sum_{j=0}^n h_{n-j} \bar{a}_j \right| \nu^n \\ &\leq \frac{1}{|\bar{a}_0|} \sum_{n=0}^{\infty} \left| \sum_{j=0}^n h_{n-j} \bar{a}_j \right| \nu^n \\ &= \frac{1}{|\bar{a}_0|} \|h * \bar{a}\|_{1,\nu} \\ &\leq \frac{1}{|\bar{a}_0|} \|h\|_{1,\nu} \|\bar{a}\|_{1,\nu} \\ &\leq \frac{1}{|\bar{a}_0|} \sum_{n=1}^N |\bar{a}_n| \nu^n = Z_1. \end{aligned}$$

Hence,  $\|A[DF(\bar{a}) - A^\dagger]\| \leq Z_1$ .

Since  $DF(a)h = 2a * h$ , then

$$\|A[DF(c) - DF(\bar{x})]\| \leq 2\|A\| \|c - \bar{x}\|_{1,\nu} \leq 2\|A\|r. \quad (7.49)$$

Now, since  $A$  defined by (7.48) has the form

$$A = \begin{bmatrix} A^{(N)} & 0 \\ & \frac{1}{2\bar{a}_0} & \frac{1}{2\bar{a}_0} \\ & 0 & \ddots \end{bmatrix},$$

then by Proposition 7.3.14,  $\|A\| \leq \max \left( \|A^{(N)}\|_{1,\nu}, \frac{1}{2|\bar{a}_0|} \right)$ . From (7.49), we set

$$Z_2 = 2 \max \left( \|A^{(N)}\|_{1,\nu}, \frac{1}{2|\bar{a}_0|} \right).$$

□

Consider Equation (7.40) with  $\lambda_0 = 1/3$  truncated at  $N = 2$ . This leads to the system quadratic of equations

$$\begin{aligned} a_0^2 - 1/3 &= 0 \\ 2a_0a_1 - 1 &= 0 \\ a_2a_0 + a_1a_1 + a_0a_2 &= 0. \end{aligned}$$

An approximate solution is give by

$$\bar{a} = \begin{pmatrix} \bar{a}_0 \\ \bar{a}_1 \\ \bar{a}_2 \end{pmatrix} = \begin{pmatrix} 0.57735026918962 \\ 0.86602540378443 \\ -0.64951905283832 \end{pmatrix}.$$

Moreover the numerical matrix

$$A^{(N)} = \begin{pmatrix} 0.86602540378443 & -0.00000000000000 & 0.00000000000000 \\ -1.29903810567665 & 0.86602540378443 & -0.00000000000000 \\ 2.92283573777248 & -1.29903810567665 & 0.86602540378443 \end{pmatrix}$$

approximately inverts the matrix  $DF^{(N)}(\bar{a})$ .

We choose  $\nu = 1/4$  and check that

$$Y_0 \leq 0.016650268688483,$$

$$\|I - A^{(N)}DF^{(N)}(\bar{a})\| \leq 1.504019729180741 \times 10^{-15} =: Z_0,$$

$$Z_1 \leq 0.44531250,$$

and

$$Z_2 \leq 2.746924327628767$$

Similarly we check that if

$$0.03668029648410 = r_- \leq r \leq r_+ = 0.16525009323246,$$

the

$$p(r) \leq 0.$$

It follows that the approximate solution

$$x^{(N)}(\lambda) = \bar{a}_2(\lambda - 1/3)^2 + \bar{a}_1(\lambda - 1/3) + \bar{a}_0,$$

satisfies

$$\sup_{|\lambda| \leq 1/4} |x^{(N)}(\lambda) - \tilde{x}(\lambda)| \leq 0.03668029648410,$$

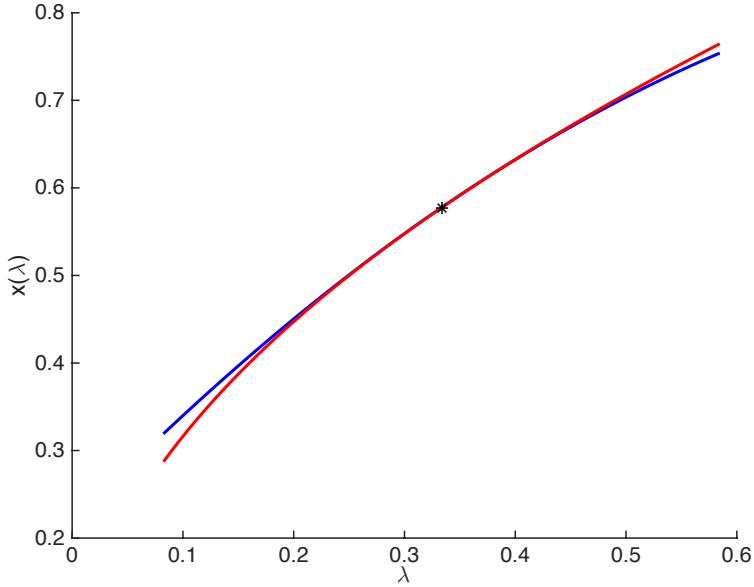


Figure 7.1: Plot of  $\bar{x}^{(N)}(\lambda)$  for  $N = 2$  shown as a blue curve versus  $\tilde{x}(\lambda) = \sqrt{\lambda}$  shown in red. The point  $(\lambda_0, x_0) = (1/3, \sqrt{1/3})$  is the black star. The approximation error is rigorously proven to be less than 0.037 in the indicated domain.

where  $\tilde{x}$  is the unique function satisfying

$$\tilde{x}(\lambda)^2 - \lambda = 0, \quad x(1/3)^2 = 1/3.$$

The example just discussed could be worked more or less by hand. Of course these results can be improved by increasing  $N$  and adjusting  $\nu$ . Results of a number of additional computations are illustrated in Table 7.2.

*Remark 7.7.2.* In this case we can compare with the true solution. Figure 8.1 illustrates the graph of  $x^{(N)}(\lambda)$  (with  $N = 2$  and  $\lambda_0 = 1/3$ ) versus the known true solution  $\tilde{x}(\lambda) = \sqrt{\lambda}$ . Indeed the explicit formula for the Taylor coefficients is given by

$$\tilde{a}_n = \frac{\sqrt{\lambda_0}}{\lambda_0^n} \frac{(-1)^n (2n)!}{(1 - 2n)(n!)^2 4^n}$$

The formula above provides valuable information about the precise decay rate of the power series coefficients. However for computational purposes it is difficult to argue that such complicated exact formulas are preferable to the validated Newton-like argument discussed above. Especially when we stop to consider that the “exact formula” for  $\tilde{a}_n$  still involves  $\sqrt{\lambda_0}$ , a quantity which must itself be computed via a rigorously validated iterative argument.

$\lambda_0$	$N$	$\nu$	$r$	$ \bar{a}_N $
1/3	10	0.25	$5.91 \times 10^{-4}$	$3.1 \times 10^2$
1/3	18	0.25	$2.9 \times 10^{-5}$	$8.5 \times 10^5$
1/3	5	0.0025	$2.22 \times 10^{-15}$	3.9
1/3	12	0.05	$8.2 \times 10^{-14}$	$2.15 \times 10^3$
0.1	10	0.01	$3.0 \times 10^{-14}$	$3 \times 10^7$
2	30	1	$2.1 \times 10^{-12}$	$2.3 \times 10^{-12}$
2	40	1	$1.8 \times 10^{-15}$	$1.5 \times 10^{-15}$
7	30	2.5	$5 \times 10^{-16}$	$3 \times 10^{-28}$

Figure 7.2: The table records the results of a number of additional computer-assisted proofs. Note that when  $\lambda_0 < 1$  the coefficients of the power series grow. This has to be balanced by taking the domain  $\nu$  smaller. It also makes high order computations numerically unstable. We achieve the best results when  $N$  is not too large. On the other hand when  $\lambda_0 > 1$  the coefficients of the series decay. This stabilizes high order computations and allows us to take both  $N$  and  $\nu$  larger.

## 7.8 Exercises

**Exercise 7.8.1.** Show that  $C^1([a, b], \mathbb{R})$  with the norm

$$\|f\|_{C^1([a,b])} = \|f\|_{C^0([a,b])} + \|f'\|_{C^0([a,b])}.$$

is a Banach space.

**Exercise 7.8.2.** Let  $X$  and  $Y$  are vector spaces, and let  $A : Y \rightarrow X$  and  $B : X \rightarrow Y$  be a linear operators. Assume that  $AB$  is invertible. Give an example where  $A$  and  $B$  are not invertible.

**Exercise 7.8.3.** Let  $X$  and  $Y$  be Banach spaces and  $T : X \rightarrow Y$  be a bijective linear map. Prove that the inverse of  $T$  is also linear.

**Exercise 7.8.4.** Let  $X$  and  $Y$  be normed linear spaces and  $T : X \rightarrow Y$  be a linear operator. Prove that the sets

$$\text{image}(T) = \{y \in Y : y = T(x) \text{ for some } x \in X\},$$

and

$$\ker(T) = \{x \in X : T(x) = 0\},$$

are normed linear (sub)spaces. Prove that  $T$  is one-to-one if and only if  $\ker(T) = 0$ .

Prove that if  $X$  and  $Y$  are Banach spaces and  $T \in B(X, Y)$ , then  $\ker(T)$  is a Banach space.

**Exercise 7.8.5.** Let  $\omega$  be a one or two sided sequence of weights (depending on the context).