Iterative $k$-Step Methods for Computing Possibly Repulsive Fixed Points in Banach Spaces

MARTIN H. GUTKNECHT AND ANDRÉ KAISER

Seminar für Angewandte Mathematik, Eidgenössische Technische Hochschule Zürich, ETH-Zentrum, CH-8092 Zürich, Switzerland

Submitted by K. Fan

Received August 29, 1985

We discuss iterative methods of the form

$$x_n := \mu_0 \Phi(x_{n-1}) + \mu_1 x_{n-1} + \cdots + \mu_k x_{n-k} \quad (n = k, k + 1, \ldots)$$

for computing a fixed point $x^*$ of a Fréchet-differentiable self-mapping $\Phi$ of a subset in a Banach space. By suitably choosing the coefficients $\mu_0, \mu_1, \ldots, \mu_k$ the local convergence is established under certain assumptions on the spectrum of $\Phi_*$. This spectrum need not be contained in the unit disk however; if it is, convergence can often be speeded up considerably compared to Picard iteration. The methods are generalizations of the methods of V. N. Kublanovskaya and W. Niethammer for linear systems of equations.

A more general type of iteration with nonstationary coefficients is considered also. For the proof of their local convergence a generalization of the convergence theorems of Perron, Ostrowski and Kitchen is presented. The connection with other methods, in particular Mann’s iterative processes, is also discussed.

1. INTRODUCTION

Let $(X, \| \cdot \|)$ be a Banach space over the field of complex numbers, and let $\Phi$ be a continuous self-mapping of some subset $D_\Phi$ of $X$. Both in theory and practice the contraction principle is the standard tool for computing a fixed point $x^*$ of $\Phi$. Unfortunately, if $\Phi$ is not a contraction, the Jacobi iteration (or Picard iteration) $x_{n+1} := \Phi(x_n)$ $(n = 0, 1, \ldots)$ is most likely to diverge. In 1953 Mann [16] proposed to use instead an averaging iteration of the form

$$v_n := \sum_{i=0}^n a_n x_i, \quad x_{n+1} := \Phi(v_n), \quad n = 0, 1, \ldots, \quad (1.1)$$

where $A = (a_n)_{n, i \geq 0}$ is an infinite lower triangular real matrix with non-negative elements summing to 1 in each row. (If $D_\Phi$ is convex, these
assumptions guarantee that $v_n, x_n \in \mathcal{D}_\Phi$ for all $n$.) Mann proved in particular that if $\mathcal{D}_\Phi$ is a compact real interval, if $\Phi$ has a unique fixed point there, and if $a_{nk} = (n + 1)^{-1} (k = 0, \ldots, n)$, then both $\{x_n\}$ and $\{v_n\}$ converge to $x^*$.

On the other hand, Krasnoselskii [15] proved in 1955 that if $\Phi$ is a non-expansive mapping (i.e., $\|\Phi(x) - \Phi(y)\| \leq \|x - y\|$) defined on a closed convex subset $\mathcal{D}_\Phi$ of a uniformly convex Banach space and if $\Phi(\mathcal{D}_\Phi)$ is compact, the iteration $v_n := \frac{1}{2}\Phi(v_{n-1}) + \frac{1}{2}v_{n-1}$ converges always to a fixed point of $\Phi$. Shaefer [26] extended this result to $v_n := \omega\Phi(v_{n-1}) + (1 - \omega)v_{n-1}$ with fixed $\omega \in (0, 1)$ and Reinermann [23] proved it for nonstationary iterations of the form

$$
v_n := \omega^{(n)}\Phi(v_{n-1}) + (1 - \omega^{(n)})v_{n-1}, \quad n = 1, 2, \ldots, \quad (1.2)
$$

with $\omega^{(0)} = 1$, $0 < \omega^{(n)} < 1$, $\omega^{(n)} \downarrow$, $\sum \omega^{(n)} = \infty$. Moreover, if $\omega^{(n)} \to 0$, this iteration was also seen to converge in the above-mentioned case of a function on a compact interval [23]. Many further results on (1.1) and (1.2) and on relations between these two iterations have since been obtained by various authors, starting with Outlaw and Groetsch [21], Dotson [3], and Ishekawa [12]; see the surveys by Mann [17], Bruck [2], Guzzardi et al. [11] and Kirk [14]. In particular, it was shown that if the condition

$$
a_{nl} = (1 - a_{nn})a_{n-1,l}, \quad l = 0, \ldots, n - 1, \quad (1.3)
$$

holds, then the sequence $\{v_n\}$ of iteration (1.1) can as well be generated by (1.2) with $\omega^{(n)} := a_{nn}$.

Iterations of the form (1.2) are particularly easy to investigate if $\Phi$ is affine, i.e., $\Phi(v) := Tv + c$ with a bounded linear operator $T$. Numerical analysts then call it (first-order) Richardson iteration. If $\omega^{(n)} = \omega$ is fixed, one has the stationary Richardson iteration as it was proposed by L. F. Richardson in 1911; this method is sometimes, in particular if $\omega \in (1, 2)$, also called the Jacobi overrelaxation (JOR) method. See [24, 27] for surveys of the early work in this area of numerical analysis.

Here we discuss first (nonlinear) stationary iterative k-step methods of the form

$$
v_n := \mu_0\Phi(v_{n-1}) + \mu_1v_{n-1} + \cdots + \mu_{n-1}v_1 + (1 - \mu_0 - \mu_1 - \cdots - \mu_{n-1})v_0, \quad n = 1, \ldots, k - 1, \quad (1.4a)
$$

$$
v_n := \mu_0\Phi(v_{n-1}) + \mu_1v_{n-1} + \cdots + \mu_kv_{n-k}, \quad n = k, k + 1, \ldots, \quad (1.4b)
$$

with in general complex coefficients $\mu_j$ satisfying

$$
\mu_0 + \mu_1 + \cdots + \mu_k = 1, \quad \mu_0 \neq 0, \quad \mu_k \neq 0. \quad (1.4c)
$$
Under certain assumptions on $\Phi$, in particular its Fréchet differentiability at the possibly repulsive fixed point $x^*$, we find parameters $\mu_0, \ldots, \mu_k$ such that the iteration \eqref{1.4} converges locally at $x^*$, i.e., if $v_0$ is close enough to $x^*$. In practice, this method is often also very effective for the computation of attractive fixed points since the rate of convergence may improve drastically compared to the one of Jacobi iteration. Our basic result is

**Theorem 1.** Assume $\Phi: D_\Phi \subset X \to X$ has a fixed point $x^*$ in the interior of $D_\Phi$, and is Fréchet differentiable at $x^*$, with the spectrum $\sigma(\Phi'_x)$ of the F-derivative $\Phi'_x$ lying in a compact set $S_\eta \subset \mathbb{C}$, which has the following properties: $1 \notin S_\eta$, and the complement $\mathbb{C} \setminus S_\eta$ is the image of $E_\eta := \{z \in \mathbb{C}; |z| > 1/\eta\}$ (with $\eta > 1$) under a map $g$ which is of the form

$$g(z) := \frac{1}{\mu_0} \left[ z - \mu_1 - \mu_2 z^{-1} - \cdots - \mu_k z^{-k+1} \right], \quad (1.5)$$

where the coefficients satisfy \eqref{1.4c}. Then, if $v_0$ is sufficiently close to $x^*$, the iteration \eqref{1.4} converges linearly to $x^*$, and its root-convergence factor $\kappa$ satisfies $\kappa \leq 1/\eta$.

(The root-convergence factor $\kappa$ of the method is defined by

$$\kappa := \sup_{n \to \infty} (\limsup_n \|v_n - x^*\|^{1/n}),$$

where the supremum has to be taken over all initial values $v_0$ sufficiently close to $x^*$.)

In the case $X = \mathbb{C}^m$ Theorem 1 has already been given by Gekeler [8] and Gutknecht, Niethammer and Varga [10]. Under the additional restrictive assumption that $\Phi$ is affine ($\Phi(x) =Tx + c$ with a $m \times m$ matrix $T$) the idea for using methods of the form \eqref{1.4} or the more general form

$$v_n := \mu_0 \Phi(v_{n-1}) + \mu_1 v_{n-1} + \cdots + \mu_{n-1} v_1$$

$$+ (1 - \mu_0 - \mu_1 - \cdots - \mu_{n-1}) v_0, \quad n = 1, 2, \ldots, \quad (1.6)$$

where the coefficients $\mu_0, \mu_1, \ldots$ are determined by a conformal (or at least meromorphic) map, is due to V. N. Kublanovskaya (see Sections 95–97 of [6] for a presentation in English of her work on this). If $g$ is conformal (i.e., meromorphic and bijective) and $v_0 := c$, these linear iterations can be understood as Euler summation processes and have been studied exhaustively by Niethammer and Varga [18, 20]. Therefore these methods are also called Euler methods. This approach has been extended from $\mathbb{C}^n$ to Hilbert and Banach spaces by Niethammer and Schempp [19, 25]. Our approach here, which does not rely on Euler summation, treats iteration \eqref{1.4} in a simpler way, but does not extend to \eqref{1.6}. (However, iteration
(1.6) is worthless in practice since it requires storing all previously computed iterates. The same approach was used in [8] and [10]. Since it does not require that \( g \) be bijective, some of our results are even in the affine case extensions of their analogs in [18–20, 25].

Using the forward difference operator \( \Delta \) defined by \( \Delta v_n = v_{n+1} - v_n \) we may write (1.6) (and thus also (1.4a), (1.4b) if we let \( \mu_n := 0 \) for \( n > k \)) as

\[
\begin{align*}
\Delta v_n &= \mu_0 \Delta \Phi(v_{n-1}) + \mu_1 \Delta v_{n-1} + \cdots + \mu_n \Delta v_0, \\
v_{n+1} &= v_n + \Delta v_n, \quad n = 0, 1, \ldots.
\end{align*}
\]

In this form the iteration is likely to be less affected by rounding errors. If \( \Phi \) is affine, \( \Delta \Phi(v_{n-1}) = T \Delta v_{n-1} \), and there are further ways to rewrite the iteration formulas, see [4, 6, Sect. 95, 18, 20] and also our remarks at the end of Section 7.

A discussion of the maps \( g \) involved and examples for such maps, i.e., for methods of the form (1.5), are given in Section 2. We also cite an optimality result. In Section 3 we present the proof of Theorem 1. It consists of putting together three pieces: Bittner’s results on the spectral mapping properties of \( k \)-step methods of a more general type [1], the special mapping properties of the function \( g \), and Kitchen’s Banach space generalization of Ostrowski’s theorem [13].

In Section 4 we then turn to asymptotically stationary iterative \( k \)-step methods

\[
\begin{align*}
v_n &= \mu_0^{(n)} \Phi(v_{n-1}) + \mu_1^{(n)} v_{n-1} + \cdots + \mu_n^{(n)} v_0, \quad n = 1, \ldots, k - 1, \quad (1.7a) \\
v_n &= \mu_0^{(k)} \Phi(v_{n-1}) + \mu_1^{(k)} v_{n-1} + \cdots + \mu_k^{(k)} v_{n-k}, \quad n = k, k + 1, \ldots, \quad (1.7b)
\end{align*}
\]

with

\[
\begin{align*}
\sum_{j=0}^{\min\{k, n\}} \mu_j^{(n)} &= 1, & \mu_0^{(n)} \neq 0, & n = 1, 2, \ldots, \quad (1.7c) \\
\mu_j^{(n)} &\to \mu_j \quad \text{as} \quad n \to \infty & (j = 0, \ldots, k). \quad (1.7d)
\end{align*}
\]

References to examples of such methods are given at the end of Section 2.

In the case \( X = \mathbb{C}^n \) the local convergence of iteration (1.7) has been proved in [10] by applying a theorem of Perron [22, Theorem 5]. In order to prove the analog in Banach spaces we present here first a generalization of Perron’s result to nonlinear iteration in Banach spaces. This new result is also an extension of Kitchen’s theorem.

If \( \Phi \) is a contraction, the Jacobi iteration converges globally. In Section 5 we point out that the same is true for iterations (1.4) and (1.7) if, additionally, \( \mathcal{A}_g \) is convex and \( \mu_j \geq 0 \) \( (j = 0, \ldots, k) \), \( \mu_j^{(n)} \geq 0 \) \( (j = 0, \ldots, k; \ n = 1, 2, \ldots) \), respectively.
In Section 6 we then investigate the connection between iterations of the form (1.4) or (1.7) and Mann's processes (1.1). It turns out that our iterations can always be written in the form (1.1) though the coefficients $a_{nl}$ may be complex (or, if the $\mu_j$ are real, may be negative). On the other hand, an iteration of the form (1.1) with $a_{n0} + a_{n1} + \cdots + a_{nm} = 1, a_{nn} \neq 0$ ($n = 1, 2, \ldots$) can always be written in the form of (1.7a), valid for all $n$. Finally, some remarks concerning the analysis of all these methods when applied to an affine function $\Phi$, i.e., to a linear system, are made in Section 7.

In applications it often occurs that the function $\Phi$ whose fixed point is sought is not differentiable with respect to the whole space $X$, but only along an invariant dense subspace. Our results can also be applied in this situation: Let $Y$ be a dense subspace of the Banach space $X$, and let $\Phi$ be a continuous self-mapping of $\mathcal{D}_\phi \subseteq Y$ with a fixed point $x^*$ which is an interior point of $\mathcal{D}_\phi$ with respect to the induced topology in $Y$. Assume $\Phi$ is Fréchet differentiable at $x^*$ along $Y$, i.e., there exists $L \in \mathcal{B}(Y)$ such that

$$ \|\Phi(x^* + h) - \Phi(x^*) - Lh\| = o(\|h\|) \quad \text{as } h \to 0, x^* + h \in \mathcal{D}_\phi \subseteq Y. \quad (1.8) $$

Then, $L$ can be uniquely extended to $\bar{L} \in \mathcal{B}(X)$, and the spectra of $L$ and $\bar{L}$ satisfy $\sigma(\bar{L}) \subseteq \sigma(L)$. Consequently, given the information that $\sigma(L) \subseteq S_\eta$, one knows that $\sigma(\bar{L}) \subseteq S_\eta$ also, and, hence, by Theorem 1, that the iterative method (1.4) converges locally with a root-convergence factor $\kappa \leq 1/\bar{\eta}$. By the same arguments our further results on the methods (1.4) and (1.7) are generalized to this situation.

2. Discussion and Examples

The function $g$ in (1.5) is a rational function with a pole of order 1 at $\infty$ and another fixed point at 1. Thus, $g$ maps a neighbourhood of $\infty$ conformally onto another neighbourhood of $\infty$. In fact, there is a maximum value $\hat{\eta}$ of $\eta$ such that $g$ is conformal (i.e., meromorphic and bijective) in $E_\eta$. If $\eta \leq \hat{\eta}$, $S_\eta = \overline{\mathbb{C}} \setminus g(E_\eta)$ is the complement of the simply connected region $g(E_\eta)$. However, in Theorem 1, where we need $\eta > 1$, $g$ is not required to be conformal, and thus $\eta$ may be larger than $\hat{\eta}$, unless $S_\eta$ is empty; there is a smallest value $\check{\eta}$ such that $S_\eta$ is empty for $\eta > \check{\eta}$.

On the other hand, if $S \subseteq \mathbb{C}$ is compact and consists of more than one point, if $1 \notin S$, and if $\overline{\mathbb{C}} \setminus S$ is simply connected in $\overline{\mathbb{C}}$, there is a unique conformal map $\tilde{g}$ of a set $E_\eta$ (with appropriate $\eta > 1$) onto $\overline{\mathbb{C}} \setminus S$ such that $\tilde{g}(\infty) = \infty$ and $\tilde{g}(1) = 1$. Thus $S = S_\eta$ for this $\eta$. At $\infty$ the function $\tilde{g}$ has an expansion

$$ \tilde{g}(z) = \frac{1}{\mu_0} \left[ z - \sum_{j=1}^{\infty} \mu_j z^{-j+1} \right], \quad (2.1) $$

which yields the coefficients of a method of type (1.6). In practice, such a method is only useful if the series has but a finite number of terms, so that \( \tilde{g} \) is of the form (1.5). If this is not the case, one should replace \( S \) by another set \( S_{\eta} \subseteq S \) which actually corresponds to a function \( g \) of the form (1.5) and for which \( \eta \) is as large as possible (given a maximum value for \( k \)).

As already noted by Kulanovskaya [6, Sect. 951 the conformal map just mentioned is optimal for \( S = S_{\eta} \) in the sense that the quantity \( 1/\eta \) determining the rate of convergence is minimal. \( 1/\eta \) determines \( \kappa \) in the following sense: on one hand, \( \kappa \leq 1/\eta \) according to Theorem 1; on the other hand, given \( g \) and \( \eta \leq \hat{\eta} \), there exists \( \Phi \) satisfying the assumptions of Theorem 1 such that \( \kappa = 1/\eta \); cf. [10].) This optimality persists with respect to nonbijective functions of the form (1.5) with arbitrary large \( k \). In fact, the following comparison theorem, which is a slight generalization of its analogs in [8, 20], holds:

**Theorem 2.** Assume the meromorphic functions \( g_j \) \( (j = 1, 2) \) map \( E_{\eta_j} \) onto the complements \( \mathbb{C} \setminus S_{\eta_j} \) of nonempty sets \( S_{\eta_j} \), with \( 1 \notin S_{\eta_j} \), satisfying

\[
S_{\eta_1} \subseteq S_{\eta_2},
\]

and assume that \( g_1 \) and \( g_2 \) have the two fixed points 1 and \( \infty \), and that \( g_1 \) is bijective. Then

\[
1/\eta_1 \leq 1/\eta_2.
\]

Moreover, if equality does not hold in (2.2), then strict inequality holds in (2.3).

**Proof** (cf. [20]). The map \( h: z \mapsto [\eta_1 g_1^{-1}(g_2(1/[\eta_2 z]))]^{-1} \) (where \( g_1^{-1} \) denotes the inverse map of \( g_1 \)) maps the open unit disk into itself, is holomorphic, and satisfies \( h(0) = 0, \ h(1/\eta_2) = 1/\eta_1 \). By Schwarz’ lemma, \( |h(z)| \leq |z| \); setting \( z = 1/\eta_2 \) we obtain (2.3). If equality does not hold in (2.2), strict equality holds in Schwarz’ lemma, and thus also in (2.3).

The standard examples for the function \( g \) of (1.5) and the iteration (1.4) concern the cases \( k = 1 \) and \( k = 2 \).

**Case** \( k = 1 \) (stationary first-order Richardson iteration). \( g(z) = [z - (1 - \mu_0)]/\mu_0 \) maps the sets \( E_{\eta} \) \( (1 < \eta < \infty) \) conformally onto the complements of the concentric disks \( S_{\eta} \) with center \( 1 - 1/\mu_0 \) and radii \([\eta/\mu_0]^{-1} \). Obviously, every closed disk that does not contain 1 is obtained this way by suitably choosing \( \mu_0 \) and \( \eta \).
Case \( k = 2 \) (stationary second-order Richardson iteration). This method was first proposed in 1950 by S. Frankel for positive definite linear systems, i.e., \( S_\eta = [\alpha, \beta], \alpha < \beta < 1; \) see [24, p. 31]. The function

\[
g(z) = \left[ z - \mu_1 - \mu_2/z \right]/\mu_0
\]

(2.4)

defines a mapping of Joukowski type discussed in detail in [20, Sect. 7]. If \( |\mu_2| < 1 \), the sets \( E_\eta, 1 < \eta < \eta := 1/|\mu_2| \) are mapped conformally onto the complements of the elliptical domains \( S_\eta \) with the common foci \( \alpha, \beta := (-\mu_1 \pm 2\sqrt{-\mu_2})/\mu_0 \). For \( \eta = \eta \) the elliptical domain degenerates to the complex interval \([\alpha, \beta]\). Therefore, \( g(\overline{E_\eta}) = \emptyset \) and \( S_\eta = \emptyset \) for \( \eta > \eta := \eta \). On the other hand, any closed elliptical domain (and any closed interval) not containing 1 can be obtained this way, see [20, Sect. 7] for details. Setting \( \gamma := \frac{1}{2}(\beta - \alpha), \delta := \frac{1}{2}(\beta + \alpha) \) and denoting by \( \zeta \) the absolutely larger root (satisfying \( |\zeta| > 1 \)) of

\[
\zeta^2 - \frac{2(1-\delta)}{\gamma} \zeta + 1 = 0
\]

one gets the parameters

\[
\mu_0 := \frac{2}{\gamma \zeta}, \quad \mu_1 := -\frac{2\delta}{\gamma \zeta}, \quad \mu_2 := -\frac{1}{\zeta^2}.
\]

(2.6)

The two most useful cases are

(i) \( S_\eta = [\alpha, \beta], \alpha < \beta < 1, \) a real interval;

(ii) \( S_\eta = [-i\rho, i\rho] \) an interval symmetric about 0 on the imaginary axis.

For further examples of functions \( g \) and corresponding sets \( S_\eta \) see [6, 20].

An asymptotically stationary 2-step method associated to the stationary second-order Richardson iteration just discussed is the Chebyshev semi-iterative method proposed and proved to be optimal (even in a non-asymptotic sense) for positive definite linear systems of equations by D. A. Flanders and G. A. Shortly in 1950 also; see, e.g., [24, 27]. Its extension to systems with complex eigenvalues was discussed by Wrigley [29] and others; recent related results are due to Freund and Ruscheweyh [7]. Besides Euler methods many other asymptotically optimal methods (i.e., methods with optimal root-convergence factor) for linear systems have been proposed; see, e.g., [4, 9, 24, 27].
3. Proof of Theorem 1

The nonlinear $k$th-order difference equation (1.4b) is equivalent to the first-order difference equation

$$y_n := \mathcal{P}(y_{n-1}), \quad n = k, k + 1, \ldots, \quad (3.1)$$

with $y_n := [v_n, v_{n-1}, \ldots, v_{n-k+1}]^T \in \mathcal{X}^k$ and

$$\mathcal{P}(y_{n-1}) := \begin{bmatrix} u'(y_{n-1}) + \mu_1 v_{n-1} + \cdots + \mu_k v_{n-k} \\ v_{n-1} \\ \vdots \\ v_{n-k+1} \end{bmatrix}. \quad (3.2)$$

The initial value $v_0$ and the values computed from (1.4a) are now contained in $y_{k-1} = [v_{k-1}, \ldots, v_0]^T$. In $\mathcal{X}^k$ we use, e.g., the norm defined by

$$\|v_{k-1}, \ldots, v_0\| = \max_{0 \leq j \leq k-1} \|v_j\|. \quad (3.3)$$

$\mathcal{P}$ has the fixed point $y^* := [x^*, x^*, \ldots, x^*]^T$, and it is Fréchet differentiable there, the derivative being

$$\mathcal{P}'(y^*) = \begin{bmatrix} \mu_0 \Phi(x^*) + \mu_1 I & \mu_2 I & \cdots & \mu_{k-1} I & \mu_k I \\ I & 0 & \cdots & 0 & 0 \\ 0 & I & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & I & 0 \end{bmatrix}. \quad (3.4)$$

Moreover, the relation between the spectra $\sigma(\Phi(x^*))$ and $\sigma(\mathcal{P}'(y^*))$ is well known [1, Lemma 2]:

$$\sigma(\mathcal{P}'(y^*)) = \{ \lambda ; q(\tau, \lambda) = 0 \text{ for some } \tau \in \sigma(\Phi(x^*)) \}, \quad (3.5)$$

where

$$q(\tau, \lambda) := \mu_k + \mu_{k-1} \lambda + \cdots + \mu_1 \lambda^{k-1} + \mu_0 \lambda^{k-1} \tau - \lambda^k. \quad (3.6)$$

Since, with the function $g$ of (1.5), there holds

$$q(\tau, \lambda) = \mu_0 \lambda^{k-1}[\tau - g(\lambda)], \quad (3.7)$$

it follows (under our assumptions on $g$) that $q(\tau, \lambda) \neq 0$ if $\lambda \in \mathcal{E}_n$, $\tau \in \mathcal{S}_n$. Consequently, $\sigma(\mathcal{P}'(y^*)) \subset \mathbb{C}\setminus \mathcal{E}_n$, i.e., the spectral radius $r_\sigma$ satisfies

$$r_\sigma(\mathcal{P}'(y^*)) \leq 1/\eta < 1. \quad (3.8)$$
Kitchen's generalization of Ostrowski's theorem [13] implies that $y_n \to y^*$ if $y_{k-1}$ is sufficiently close to $y^*$, which, in view of the continuity of $\Phi$, is guaranteed if $v_0$ is sufficiently close to $x^*$. Thus $v_n \to x^*$ if $v_0$ is sufficiently close to $x^*$.

From Kitchen's proof it is not difficult to conclude that the root-convergence factor of $\{y_n\}$ is at most $r_n(\Psi'_{y^*})$. (The argument is given in the proof of our Theorem 3 below.) Finally, as explicitly shown in [28], $\{v_n\}$ has the same root-convergence factor.

Remark. The iteration (1.4) is well-defined (i.e., $v_n \in \mathcal{D}_\Phi$ for all $n$) if $v_0$ is sufficiently close to $x^*$. This follows from the above proof; the argument is made explicit in Kitchen's proof [13].

4. A Generalization of the Local Convergence Theorems of Perron, Ostrowski and Kitchen, and Its Application to Asymptotically Stationary Iterative $k$-Step Methods

We turn now to asymptotically stationary iterations (1.7). In order to prove for them the analog of Theorem 1 we need a generalization of both Kitchen's theorem [13] and Perron's theorem [22, Theorem 5]. The latter has been applied in [10] to prove this result in the case $X = \mathbb{C}^n$.

**Theorem 3.** Let $(Y, \| \cdot \|)$ be a Banach space, and let $\Psi, \chi_n (n = 0, 1, 2, \ldots)$ be continuous self-mappings of $\mathcal{D} \subseteq Y$. Let $y^*$ be a fixed point of $\Psi$ in the interior of $\mathcal{D}$ and assume that

(i) $\Psi$ is Fréchet differentiable at $y^*$, and the spectral radius of its derivative $\Psi'_{y^*}$ at $y^*$ is less than 1;

(ii) $\|\chi_n(y)\| = O(\|y - y^*\|)$ as $y \to y^*$ $(n = 0, 1, \ldots)$, and

$$\lim_{\substack{n \to \infty \\ \ n \to y^*}} \frac{\|\chi_n(y)\|}{\|y - y^*\|} = 0.$$

Then there is a neighbourhood $\mathcal{N} \subseteq \mathcal{D}$ of $y^*$ such that the sequence $\{y_n\}$ generated by

$$y_{n+1} := \Psi(y_n) + \chi_n(y_n), \quad n = 0, 1, 2, \ldots, \quad (4.1)$$

with arbitrary initial value $y_0 \in \mathcal{N}$, converges to $y^*$. The root-convergence factor $\kappa$ of $\{y_n\}$ is less than or equal to the spectral radius of $\Psi'_{y^*}$.
**Proof.** First the problem can be reduced to the case where \( \mathcal{Y} \) is a bounded linear operator (with the trivial fixed point 0). In fact,

\[
y_{n+1} - y^* = \mathcal{Y}(y_n) + \chi_n(y_n) - y^* \]

\[
= \mathcal{Y}_n^*(y_n - y^*) + \tilde{\chi}_n(y_n), \tag{4.2a}
\]

where

\[
\tilde{\chi}_n(y):= \chi_n(y) + \psi(y), \tag{4.2b}
\]

\[
\psi(y):= \mathcal{Y}(y) - \mathcal{Y}(y^*) - \mathcal{Y}_n^*(y - y^*). \tag{4.2c}
\]

Since \( \psi(y) = o(\|y - y^*\|) \) holds by the Fréchet differentiability of \( \mathcal{Y} \), \( \tilde{\chi}_n \) again satisfies assumption (ii). Hence, we may assume that \( \mathcal{Y} \) is linear, i.e., \( \mathcal{Y} = \mathcal{Y}^*_n \), and \( y^* = 0 \).

By assumption (i) and the spectral radius formula,

\[
\lim_{n \to \infty} \|\mathcal{Y}^n\|^{1/n} = r_n(\mathcal{Y}) < 1.
\]

Thus, for any \( r \) satisfying \( r_n(\mathcal{Y}) < r < 1 \) there is \( m \in \mathbb{N} \) such that \( \|\mathcal{Y}^m\| < r^m \). We consider now for every \( i \in \{0, ..., m-1\} \) the sequence \( \{z_i^{(n)}\} \) defined by

\[
z_0^{(i)}:= y_i, \quad z_{n+1}^{(i)}:= \pi_{m, m_0}(z_n^{(i)}), \quad n = 0, 1, ..., \tag{4.3a}
\]

where

\[
\pi_{l, l}(z):= (\mathcal{Y} + \chi_{l+i-1})(\mathcal{Y} + \chi_{l+i-2}) \cdots (\mathcal{Y} + \chi_l)(z). \tag{4.3b}
\]

There is a neighbourhood \( \mathcal{V}_0 \) of \( y^* = 0 \) where all functions \( \pi_{l, l} \) \((i = 1, ..., m; \ l \in \mathbb{N})\) are defined. Clearly, \( z_n^{(i)} = y_{m+i}^{(n)} \) \((n = 0, 1, ...; \ i = 0, ..., m-1)\). Hence, in order to prove \( y_n \to y^* \) it suffices to prove \( z_n^{(i)} \to y^* \) for each \( i \).

We claim that in analogy to assumption (ii) on \( \{\chi_n\} \) the functions

\[
\omega_{i, l}:= \mathcal{Y}_i^* - \pi_{i, l} \quad (i = 1, ..., m; \ l \in \mathbb{N}), \tag{4.4}
\]

(defined at least in \( \mathcal{V}_0 \)) satisfy

\[
\|\omega_{i, l}(z)\| = O(\|z\|) \quad \text{as } z \to 0 \tag{4.5a}
\]

and

\[
\lim_{l \to \infty} \frac{\|\omega_{i, l}(z)\|}{\|z\|} = 0 \quad (i = 1, ..., m). \tag{4.5b}
\]

The proof is by induction on \( i \): If \( i = 1 \), \( \omega_{1, l} = \chi_l \), so that (4.5a) and (4.5b)
reduce to assumption (ii). For the induction step we use first (4.3b), (4.4), and the linearity of \( \Psi \) to obtain

\[
\omega_{i,i} = \Psi^i - (\Psi + \chi_{i+i-1}) \pi_{i-1,i} \\
= \Psi(\Psi^{i-1} - \pi_{i-1,i}) + \chi_{i+i-1} \pi_{i-1,i} \\
= \Psi(\omega_{i-1,i} + \chi_{i+i-1} \pi_{i-1,i}).
\]

By assumption, \( \omega_{i-1,i} \) satisfies (4.5). Therefore, \( \|\pi_{i-1,i}(z)\| = \|\Psi^{i-1}z - \omega_{i-1,i}(z)\| = O(\|z\|) \) uniformly in \( i \), and

\[
\|\omega_{i,i}(z)\| = \|\omega_{i-1,i}(z)\| + \|\chi_{i+i-1}(\pi_{i-1,i}(z))\| = O(\|z\|),
\]

\[
\|\omega_{i,i}(z)\| = \|\omega_{i-1,i}(z)\| + \|\chi_{i+i-1}(\pi_{i-1,i}(z))\| \|\pi_{i-1,i}(z)\| \|\pi_{i-1,i}(z)\| \\
\rightarrow 0 \quad \text{as } l \rightarrow \infty \quad \text{and } z \rightarrow 0 \text{ (independently)}.
\]

This establishes our claim (4.5).

There remains to study (for \( i = 0, \ldots, m - 1 \)) the iterations

\[
z_n^{(i+1)} = \pi_{m,m}(z_n^{(i)}) = \Psi^m z_n^{(i)} - \omega_{m,m}(z_n^{(i)}), \quad n = 0, 1, \ldots, \tag{4.6}
\]

where \( \Psi^m \) is a linear operator whose norm is bounded by \( r^m < 1 \) and \( \{\omega_{m,m}\}_{n \in \mathbb{N}} \) is a sequence of functions (defined in \( \mathcal{N}_0 \)) satisfying (4.5). Let \( \mathcal{B}_\varepsilon = \{z \in X; \|z\| < \varepsilon\} \). Given \( \varepsilon > 0 \) such that \( \mathcal{B}_{\varepsilon/2} \subseteq \mathcal{N} \) and \( \varepsilon < 1 - r^m \), then as a consequence of (4.5b) there is \( M \in \mathbb{N} \) such that for \( l \geq M \)

\[
\|\omega_{m,l}(z)\| \leq \varepsilon \|z\| \quad \text{if } z \in \mathcal{B}_{\varepsilon/2}. \tag{4.7}
\]

Moreover, by the continuity of \( \Psi \) and \( \chi_l \) (\( l \in \mathbb{N} \)), the functions \( \pi_{i,l} \) (\( i = 1, \ldots, m; l \in \mathbb{N} \)) are continuous, and by (4.4), (4.5a), they have \( y^* = 0 \) as a fixed point. Therefore, there is a neighborhood \( \mathcal{N} \subseteq \mathcal{N}_0 \) of 0 such that \( y_0 \in \mathcal{N} \) implies \( z^{(i)}_M \in \mathcal{B}_{\varepsilon/2} \) (\( i = 1, \ldots, m \)). Let us assume for induction that \( z^{(i)}_{M+j-1} \in \mathcal{B}_{\varepsilon/2} \). Then, by (4.6) and (4.7),

\[
\|z^{(i)}_{M+j+1}\| = \|\Psi^m z^{(i)}_{M+j} - \omega_{m,m}(z^{(i)}_{M+j})\| \\
\leq (r^m + \varepsilon) \|z^{(i)}_{M+j}\| \\
\leq (r^m + \varepsilon) \frac{\varepsilon}{2} \leq \frac{\varepsilon}{2}.
\]

Hence \( z^{(i)}_{M+j} \in \mathcal{B}_{\varepsilon/2} \) for all \( l \in \mathbb{N} \), and

\[
\|z^{(i)}_{M+j}\| \leq (r^m + \varepsilon)^l \varepsilon/2 \rightarrow 0 \quad \text{as } l \rightarrow \infty.
\]
Moreover, since \( y_{mn+i} = x_n^{(i)} \),

\[
\limsup_{n \to \infty} \|y_{mn+i}\|^{1/(mn+i)} \leq \limsup_{n \to \infty} \left( (r^m + \varepsilon)^n - M \frac{\varepsilon}{2} \right)^{1/(mn+i)} - (r^m + \varepsilon)^{1/m}.
\]

This holds for all \( y_0 \in \mathcal{N} \), for \( i = 1, \ldots, m \), and for all \( \varepsilon > 0 \). Consequently \( \kappa \leq r \).

Now the proof given in [10] for the local convergence of asymptotically stable \( k \)-step methods can be repeated word by word. It yields

**Theorem 4.** Theorem 1 holds also for the asymptotically stationary iterative \( k \)-step method (1.7) related to the stationary method (1.4) by (1.7d).

### 5. Global Convergence of Iterative \( k \)-Step Methods for Contractions

If the parameters are chosen appropriately, the iterative \( k \)-step methods (1.4) and (1.7) have in general a better root-convergence factor than Jacobi iteration even if \( \Phi \) is a contraction. However, in order to guarantee global convergence we must assume nonnegative parameters \( p^{(n)} \). On the other hand, the Fréchet differentiability of \( \Phi \) at \( x^* \) can be dropped. The following result can be proved as the corresponding one, Theorem 4, in [10]:

**Theorem 5.** Let \( \Phi \) be a continuous self-mapping of the convex set \( \mathcal{D}_\Phi \subseteq X \), and let \( \Phi \) satisfy a Lipschitz condition with constant \( L \in (0, 1) \). Assume the parameters \( p^{(n)} \) of the iterative method (1.7) are nonnegative real.

Then this method converges for arbitrary initial value \( v_0 \in \mathcal{D}_\Phi \).

The crucial point of the proof becomes apparent in our argument for

**Theorem 6.** Let \( \Phi \) satisfy the same assumptions as in Theorem 5, and let \( \mu_0 > 0, \mu_1 \geq 0, \ldots, \mu_k \geq 0, \mu_0 + \cdots + \mu_k = 1 \). Then there is a norm in \( X^k \) such that the function \( \Psi \) defined in (3.2) is a contraction.

**Proof.** In contrast to our notation used so far let \( y \in X^k \) and \( \psi(y) \in X^k \) have the components \( y_1, \ldots, y_k \) and \( \psi_1(y), \ldots, \psi_k(y) \). Then, for \( y, \tilde{y} \in X^k \), the inequality

\[
\begin{bmatrix}
\|\psi_1(y) - \psi_1(\tilde{y})
\| \\
\vdots \\
\|\psi_k(y) - \psi_k(\tilde{y})
\| \\
\end{bmatrix}
\leq L
\begin{bmatrix}
\|y_1 - \tilde{y}_1\| \\
\vdots \\
\|y_k - \tilde{y}_k\|
\end{bmatrix}
\] (5.1)
holds componentwise. In view of (3.5) the spectrum of the companion matrix \( L \) consists of the \( n \) zeros of the function \( \lambda \mapsto q(L, \lambda) \) defined by (3.6). Now, for \( |\lambda| \geq 1 \), by (3.7) and since \(|\mu_0 + \mu_1 \lambda^{-1} + \cdots + \mu_k \lambda^{-k}| \leq 1\),

\[
|q(L, \lambda)| \geq \mu_0 |\lambda|^{k-1} \left| \frac{\lambda}{\mu_0} (1 - \mu_1 \lambda^{-1} - \cdots - \mu_k \lambda^{-k}) - L \right|
\]

\[
\geq \mu_0 |\lambda|^k \left| \frac{1}{\mu_0} (1 + \mu_0 - 1) - \frac{L}{|\lambda|} \right|
\]

\[
= \mu_0 |\lambda|^k \left| 1 - \frac{L}{|\lambda|} \right| > 0.
\]

Consequently, all eigenvalues of \( L \) lie inside the unit disk, and by a well-known result on \( k \times k \) matrices there is a norm of \( \mathbb{R}^k \) such that the associated norm of \( L \) is less than 1. If this norm of \( \mathbb{R}^k \) is used to define the norm in \( X^k \) from the norm in \( X \), the claimed result follows from (5.2).

If \( x_0 = X \), it is easy to see that the assertions of Theorems 5 and 6 hold likewise under the assumption that

\[
|\mu_0^{(n)}| L + |\mu_1^{(n)}| + \cdots + |\mu_k^{(n)}| < L' < 1
\]

for the now possibly complex parameters \( \mu_j^{(n)} \) of the iterative method (1.7).

6. THE RELATION TO MANN’S AVERAGING PROCESS

If we let \( v := (v_0, v_1, \ldots)^T \) and \( x := (x_0, x_1, \ldots)^T \), Mann’s averaging process (1.1) may be written as

\[
v = Ax,
\]

\[
x_{n+1} := \Phi(v_n), \quad n = 0, 1, \ldots
\]

If we assume that \( A \) has nonvanishing diagonal elements, its inverse \( B := (b_{mn})_{n,m \geq 0} := A^{-1} \) exists and is also an infinite lower triangular matrix. Hence, \( x = Bv \), i.e.,
\[ \Phi(v_{n-1}) = x_n = \sum_{i=0}^{n} b_{ni}v_i. \quad (6.2) \]

Solving for \( v_n \) we get
\[ v_n = \frac{1}{b_{nn}} \left[ \Phi(v_{n-1}) - \sum_{i=0}^{n-1} b_{ni}v_i \right], \quad n = 1, 2, \ldots \quad (6.3) \]

The iterative processes (1.4), (1.6) and (1.7) are also of this form. Actually one can replace in (1.7c) the index bound \( \min\{k, n\} \) by \( n \) in order to obtain in analogy to (1.6) more general nonstationary methods of the type
\[ v_n = \mu_0^{(n)} \Phi(v_{n-1}) + \mu_1^{(n)} v_{n-1} + \cdots + \mu_n^{(n)} v_0. \quad (6.4a) \]

with
\[ \mu_0^{(n)} + \mu_1^{(n)} + \cdots + \mu_n^{(n)} = 1, \quad \mu_0^{(n)} \neq 0 \quad (6.4b) \]

\( n = 1, 2, \ldots \), which are identical to the iterations of the form (6.3) if
\[ \mu_0^{(n)} = \frac{1}{b_{nn}}, \quad \mu_i^{(n)} = -\frac{b_{ni}}{b_{nn}} (i = 1, \ldots, n). \quad (6.5) \]

In fact, the conditions (6.4b) translate into
\[ b_{n0} + b_{n1} + \cdots + b_{nn} = 1, \quad b_{nn} \neq 0. \quad (6.6) \]

which in view of the lower triangularity of \( B \) and \( A = B^{-1} \) are equivalent to Mann's assumption
\[ a_{n0} + a_{n1} + \cdots + a_{nn} = 1, \quad a_{nn} \neq 0. \quad (6.7) \]

The invariance of the row sums in (6.6) and (6.7) follows because, e.g., the one in (6.7) is equivalent to the implication \( x = (c, c, \ldots)^T \Rightarrow v = Ax = (c, c, \ldots)^T \), which in turn is equivalent to \( v = (c, c, \ldots)^T \Rightarrow x = Bv = (c, c, \ldots)^T \), which is then equivalent to (6.6). Formulas (6.4b)–(6.7) holds for \( n \geq 1 \) and persist for \( n = 0 \) if we set \( a_{00} = -b_{00} = -\mu_0^{(0)} = -1 \).

Hence, we have proved:

**Theorem 7.** Each method of the type (6.4) is equivalent to a method of the form (1.1) satisfying (6.7), and vice versa. (More precisely, the two methods, if started with the same \( v_0 \), produce the same iterates \( v_n \).) The coefficients \( \mu_i^{(n)} \in \mathbb{C} \) and \( a_{nn} \in \mathbb{C} \) are related by (6.5), where the lower triangular infinite matrix \( B = (b_{ni})_{n,i \geq 0} \) is the inverse of \( A = (a_{ni})_{n,i \geq 0} \).

An example of such an equivalence is the one mentioned in the Introduction between iteration (1.2) and iteration (1.1) satisfying (1.3).
As mentioned, the iteration methods (1.4) and (1.7), which are the subject of this paper, are special cases of (6.4). The first one is obtained by requiring

\[ \mu_l^{(n)} := \mu_l, \quad l = 0, \ldots, \min\{k, n\} - 1, \quad (6.8a) \]

\[ \mu_{\min\{k, n\}}^{(n)} := 1 - \mu_0 - \mu_1 - \cdots - \mu_{\min\{k, n\} - 1}, \quad (6.8b) \]

\[ \mu_{l}^{(n)} := 0, \quad l = k + 1, \ldots, n. \quad (6.8c) \]

(Then also \( \mu_k^{(n)} = \mu_k \) if \( n \geq k \).) The second one is characterized by (6.8c) and (1.7d).

We want to investigate the effect of these conditions (6.8) on the matrix \( A \) of the associated equivalent method (1.1). First, they mean that \( B \) is of the form

\[
B = \frac{1}{\mu_0} \begin{bmatrix}
\mu_0 & -\mu_1^{(1)} & & & \\
-\mu_1^{(1)} & 1 & & & \\
& -\mu_2^{(2)} & -\mu_1 & 1 & \\
& & \vdots & \ddots & \\
& & & -\mu_{k-1}^{(k-1)} & -\mu_{k-2} \cdots -\mu_1 \end{bmatrix} \begin{bmatrix} 0 \\ \vdots \end{bmatrix}
\]

\[ (6.9a) \]

with

\[ \mu_j^{(n)} = 1 - \mu_0 - \mu_1 - \cdots - \mu_{n-1}, \quad n = 1, \ldots, k - 1. \quad (6.9b) \]

\( B \) is a rank 1 modification of the banded Toeplitz matrix

\[ \tilde{B} := (\tilde{b}_{n})_{n, l \geq 0}, \quad \tilde{b}_{n, n-j} := \begin{cases} 1/\mu_0 & \text{if } j = 0, \\ -\mu_j/\mu_0 & \text{if } 1 < j \leq k, \\ 0 & \text{if } j < 0 \text{ or } j > k, \end{cases} \quad (6.10) \]

since

\[ B = \tilde{B} - be^T, \quad (6.11a) \]

with

\[ b^T := \frac{1}{\mu_0} (1 - \mu_0, 1 - \mu_0 - \mu_1, \ldots, 1 - \mu_0 - \cdots - \mu_{k-1}, 0, 0, \ldots), \quad (6.11b) \]

\[ e^T := (1, 0, 0, 0, \ldots). \quad (6.11c) \]
By the Sherman–Morrison formula, if \( \tilde{A} = \tilde{B}^{-1} \),

\[
A = (\tilde{B} - b e^T)^{-1} = \tilde{A} + \frac{\tilde{A} b e^T \tilde{A}}{1 - e^T \tilde{A} b} \tag{6.12}
\]

Owing to the simplicity of \( e \) and the triangularity of \( \tilde{A} \) we can simplify this to

\[
A = \tilde{A} + a e^T \quad \text{with} \quad a := \tilde{A} b. \tag{6.13}
\]

Hence, \( A \) is also a rank 1 modification of the inverse \( \tilde{A} \) of \( \tilde{B} \), the modification again being restricted to the first column. Now, as is well known, the inverse of a regular lower triangular infinite Toeplitz matrix is again a matrix of the same kind: Associated to \( \tilde{B} \) is the Taylor series at \( x \)

\[
\frac{1}{\mu_0} \left( 1 - \mu_1 z^{-1} - \mu_2 z^{-2} - \cdots - \mu_k z^{-k} \right) = \frac{g(z)}{z}, \tag{6.14}
\]

and associated to \( \tilde{A} \) is then the Taylor series at \( \infty \) of \( z/g(z) \),

\[
\alpha_0 + \alpha_1 z^{-1} + \alpha_2 z^{-2} + \cdots = \frac{z}{g(z)} = \frac{\mu_0}{(1 - \mu_1 z^{-1} - \cdots - \mu_k z^{-k})}. \tag{6.15}
\]

Multiplying through with the denominator and comparing coefficients yields the recursion

\[
\alpha_0 := \mu_0, \quad \alpha_n := \sum_{l=1}^{\min\{k,n\}} \mu_l \alpha_{n-l}, \quad n = 1, 2, \ldots \tag{6.16}
\]

Now, \( A \) differs from \( \tilde{A} \) only in its first column, which could be computed from (6.13). However, since the row sums in \( A \) are 1, it is clear that

\[
a_{nl} = \begin{cases} 
0 & \quad \text{if} \quad l > n, \\
\alpha_{n-l} & \quad \text{if} \quad 0 < l \leq n, \\
1 - \alpha_0 - \cdots - \alpha_{n-1} & \quad \text{if} \quad l = 0.
\end{cases} \tag{6.17}
\]

Summarizing we get

\[\text{THEOREM 8.} \quad \text{The method of the form (1.1) (satisfying (6.7)) which is equivalent to the method (1.4) has the coefficients (6.17), where the } \alpha_i \text{ are obtained from the } \mu_i \text{ according to recursion (6.16).}\]

So far we have not paid any attention to the nonnegativity condition \( a_{nl} \geq 0 \), which together with \( a_{n1} + \cdots + a_{nn} = 1 \) has the important effect that
the iterates (1.1) remain always in the domain $\mathcal{D}_g$ if the latter is convex. In general, a method (1.1) equivalent to a method (1.4) does not satisfy this nonnegativity condition. (In fact, the coefficients $\mu_i$ and $a_{nl}$ may be complex.) However, if $\mu_i \geq 0$, $l=0, 1, \ldots, k$, as in Theorems 5 and 6, then (6.16) clearly implies $a_j \geq 0$, $j=0, 1, \ldots$. Moreover, if $g$ is conformal in $E_g$, $\hat{n} > 1$, then (6.15) implies $a_0 + a_1 + \cdots = g(1) = 1$, so that $a_{nl} \geq 0$ ($\forall l, \forall n$) by (6.17). We get

\textbf{Theorem 9.} If $g$ is conformal in $|z| > 1/\hat{n}$, $\hat{n} > 1$, and if $\mu_i \geq 0$ ($l=0, \ldots, k$), the method of the form (1.1) which is equivalent to the method (1.4) satisfies $a_{nl} \geq 0$ ($n, l=0, 1, \ldots$).

Theorems 8 and 9 extend to more general Euler methods defined for all $n$ ($\geq 1$) by the recursion (1.4a) and generated by a function

$$g(z) = \frac{1}{\mu_0} (z - \mu_1 - \mu_2 z^{-1} - \mu_3 z^{-2} - \cdots),$$

(6.18)

which is conformal (i.e., meromorphic and one-to-one) in a set $1/\hat{n} < |z| \leq \infty$ for some $\hat{n} > 1$ and satisfies $g(1) = 1$, $g(\infty) = \infty$ (see [18–20, 25] for a treatment of these methods in the linear case). The equivalent iteration of type (1.1) satisfying (6.7) has the coefficients $a_{nl}$ of (6.17), where the $\alpha_j$ result from relation (6.16) modified by replacing the summation bound $\min\{k, n\}$ by $n$. These extensions of Theorems 8 and 9 are immediate since in our derivation we made use neither of the band structure of $B$ nor of the fact that $g$ is rational.

7. \textbf{Comparing Methods By Their Spectral Damping Properties For Linear Systems}

As we have seen the effectiveness of the $k$-step methods (1.4) and (1.7) depends asymptotically only on the spectral mapping properties of the function $g$ of (1.5). If $\Phi$ is affine, $\Phi(x) = Tx + c$, and $T$ is either a diagonalizable $m \times m$ matrix or a normal bounded linear operator $T \in \mathcal{L}(X)$ in a Hilbert space $X$, the effect of these spectral mapping properties can easily be understood in a nonasymptotic sense. At least in the finite dimensional case this is often used in the discussion of iterative methods, see, e.g., [4, 6, 7, 9, 20, 24, 27].

In fact, introducing the errors $e_n := v_n - x^*$ we get from (6.4), which includes (1.4), (1.6), and (1.7) as special cases,

$$e_n = \mu_0^{(n)}Te_{n-1} + \mu_1^{(n)}e_{n-1} + \mu_2^{(n)}e_{n-2} + \cdots + \mu_n^{(n)}e_0, \quad n = 1, 2, \ldots$$

(7.1)
It follows by induction that \( e_n = p_n(T) e_0 \), where \( \{ p_n \} \) is the sequence of polynomials defined by the recurrence

\[
p_0(t) \equiv 1, \\
p_n(t) := (\mu_0^{(n)} t + \mu_1^{(n)} p_{n-1}(t) + \mu_2^{(n)} p_{n-2}(t) + \cdots + \mu_n^{(n)} p_0(t)).
\]

Due to (6.4b) these polynomials satisfy \( p_n(1) = 1 \). Under the stated special assumptions on \( T \)

\[
\| p_n(T) \| = \| p_n \|_{\infty, \sigma(T)} := \sup_{\lambda \in \sigma(T)} | p_n(\lambda) | 
\]

for a suitable matrix norm or the operator norm on \( \mathcal{L}(X) \), respectively. In view of \( \| e_n \| \leq \| p_n \|_{\infty, \sigma(T)} \| e_0 \| \) it is clear that the goal would be to minimize (7.3) under the restriction \( p_n(1) = 1 \). However, since \( \sigma(T) \) cannot be expected to be known exactly, the aim is to minimize \( \| p_n \|_{\infty, S} \) for a set \( S \) known to contain \( \sigma(T) \).

The Chebyshev semi-iterative method mentioned in Section 2 is optimal in this sense for the real interval \( S = [\alpha, \beta], \ 1 \notin [\alpha, \beta] \). Its associated polynomials \( p_n \) are up to the normalization the Chebyshev polynomials transformed to \([\alpha, \beta]\). The polynomials \( p_n \) of Frankel’s stationary second-order Richardson iteration are, e.g., given in [24, p. 31].

If \( p_0 \) can be expanded in terms of eigenfunctions \( u_j \) of \( T \) (as, in particular, if \( T \in \mathcal{L}(X) \) is a normal operator with countable spectrum or, if \( T \) is a diagonalizable \( m \times m \) matrix), the effect of an iteration of the form (6.4) can be understood in an explicit way as an error component suppression [6, 24]: Let

\[
e_0 = \sum_j e_j^{(0)} u_j, \quad Tu_j = \lambda_j u_j,
\]

then

\[
e_n - p_n(T) e_0 = \sum_j p_n(\lambda_j) e_j^{(0)} u_j,
\]

i.e., the component \( e_j^{(0)} u_j \) of \( e_0 \) is damped by a factor \( p_n(\lambda_j) \).

Note that these remarks apply for any linear iterative method for which \( e_n = p_n(T) e_0 \) \( (n = 1, 2, \ldots) \) with a sequence of polynomials \( p_n \) of degree at most \( n \). Such a method is determined uniquely by the sequence \( \{ p_n \} \) and therefore—as is seen from (7.2) and (7.1)—it can be written in the form (6.4) if, for all \( n \), \( p_n \) has exact degree \( n \) and \( p_n(1) = 1 \). Hence, these methods (6.4) are equivalent to Varga’s semi-iterative methods [27, 4] and Rutishauser’s gradient methods [24, 5] (which include the conjugate gradient method of M. R. Hestenes and E. Stiefel as a special case). However, as mentioned in the Introduction, (6.4) may not be the optimal formula for the computation of the iterates subject to roundoff.
ACKNOWLEDGMENTS

The authors have profited from the first author's work with Professors W. Niethammer and R. S. Varga on a related paper on k-step methods in finite dimensional space. Additionally, they are indebted to Professor W. Niethammer for pointing out the work of Mann and others on averaging iteration processes.

REFERENCES

29. H. E. Wrigley, Accelerating the Jacobi method for solving simultaneous equations by Chebyshev extrapolation when the eigenvalues of the iteration matrix are complex, Comput. J. 6 (1963), 169–176.