

## ITERATIVE AND SEMI-ITERATIVE METHODS FOR COMPUTING STATIONARY PROBABILITY VECTORS OF MARKOV OPERATORS

IVO MAREK AND DANIEL B. SZYLD

**ABSTRACT.** Iterative and semi-iterative methods for computing stationary probability vectors of Markov-type operators are proposed and their convergence properties are analyzed. The methods studied apply to certain classes of problems in infinite-dimensional spaces as well as to classical  $n \times n$  stochastic matrices.

### 1. INTRODUCTION

The aim of this contribution is to analyze certain iterative and semi-iterative methods for computing approximate solutions to singular equations of the type

$$(1) \quad Ax = 0$$

with  $A = I - B$ , where  $I$  is the identity map,  $B$  is an operator which is non-negative in the sense that it has an invariant cone [4, 11], and is such that its spectral radius  $r(B) = 1$ . Such problems appear, for example, when modeling queueing systems; see, e.g., the references given in [1] and [3]. Another application is data fitting constrained to some eigenvalue relations [13]. The operator  $A$  in (1) is an  $M$ -operator [17], and the class of problems studied here includes in particular the solution of systems of linear equations with  $Q$ -matrices [2, 22].

From the numerical point of view, this problem leads to the calculation of a stationary probability vector (or stationary distribution) of a Markov process [14, 18]. In practical computations such problems reduce usually to constructing some or all stationary probability vectors of  $n \times n$  stochastic matrices.

Although the latter problem is very important, we analyze the methods proposed in application to general (infinite-dimensional) Markov processes [5]. In this way there is more flexibility in choosing appropriate methods. In particular, one can handle infinite-dimensional problems, say by discretizing them at each iteration step instead of discretizing a priori the original problem by reducing it to an  $n \times n$  model. Our approach offers the possibility of constructing the

---

Received by the editor February 4, 1992 and, in revised form, October 6, 1992.

1991 *Mathematics Subject Classification.* Primary 65J10, 65F10, 65F15, 47B60, 15A48.

*Key words and phrases.*  $\mathcal{N}$ -stochastic operators, Markov type operators, stationary probability vectors, iterative methods, semi-iterative methods .

This work was supported by National Science Foundation grants INT-9196077 and DMS-9201728.

required approximations without an a priori restriction to a certain size of the corresponding matrices; a reduction to finite-dimensional spaces can be done during the computational process formally working in an infinite-dimensional space. In addition, our proofs of convergence are independent of the dimension of the space. This implies that the convergence results of the methods discussed apply, e.g., to every stage of aggregation and disaggregation methods [10, 18, 26], and in general to a discretization with any number of points; see Remark 4.1.

We should add that if the discretizations of the infinite-dimensional operator are done with minimal care, the convergence rate of the iterative processes remains controlled. Let  $T$  be the iteration matrix of such an iterative process. Its spectral radius is 1, but the rate of convergence is given by the 'reduced spectral radius'  $\gamma(T) = \{\gamma \in \sigma(T) : |\gamma| < 1, |\gamma| \geq |\lambda|, \lambda \in \sigma(T), |\lambda| < 1\}$ , i.e., the eigenvalue of largest modulus inside the unit circle; see the next section for definitions and notation. Let  $h$  be the discretization parameter. The eigenvalues of the discretized problem tend to the eigenvalues of the infinite-dimensional problem as  $h \rightarrow 0$ . Thus, the reduced spectral radius of the finite-dimensional problems tend to the reduced spectral radius of the infinite-dimensional problem as the number of points increases, while the convergence results continue to apply.

## 2. DEFINITIONS AND NOTATION

Let  $\mathcal{E}$  be a real Banach space (in particular, it could be  $\mathbb{R}^n$ ),  $\mathcal{E}'$  its dual, and  $\mathcal{B}(\mathcal{E})$  the space of bounded linear operators on  $\mathcal{E}$ . It is assumed that  $\mathcal{E}'$  and  $\mathcal{B}(\mathcal{E})$  are equipped with the usual norms and are thus Banach spaces as well. Let  $\mathcal{F}$  denote the *complex extension* of  $\mathcal{E}$ , i.e.,  $\mathcal{F} = \mathcal{E} \oplus i\mathcal{E}$ , with the norm

$$\|z\|_{\mathcal{F}} = \sup\{\|x \cos \theta + y \sin \theta\|_{\mathcal{E}} : 0 \leq \theta \leq 2\pi\},$$

where  $z = x + iy$ ,  $x, y \in \mathcal{E}$ .

Let  $\mathcal{K} \subset \mathcal{E}$  be a closed *normal and generating cone* [11], i.e.,  $\mathcal{K}$  satisfies the relations (i)–(vi), where (i)  $\mathcal{K} + \mathcal{K} \subset \mathcal{K}$ , (ii)  $a\mathcal{K} \subset \mathcal{K}$ , (iii)  $\mathcal{K} \cap (-\mathcal{K}) = \{0\}$ , (iv)  $\overline{\mathcal{K}} = \mathcal{K}$  ( $\overline{\mathcal{K}}$  denotes the norm closure of  $\mathcal{K}$ ), (v)  $\mathcal{E} = \mathcal{K} - \mathcal{K}$ , i.e., for every  $y \in \mathcal{E}$ , there exist  $y_j \in \mathcal{K}$ ,  $j = 1, 2$ , such that  $y = y_1 - y_2$ , and (vi) for every pair  $x \in \mathcal{K}$ ,  $y \in \mathcal{K}$ , there exists a real  $\delta > 0$  such that  $\|x + y\|_{\mathcal{E}} \geq \delta\|x\|_{\mathcal{E}}$ . For the case of  $\mathcal{E} = \mathbb{R}^n$  one can consider  $\mathcal{K} = \mathbb{R}_+^n$ , the set of nonnegative vectors.

Let  $\mathcal{K}'$  be the *dual cone*, i.e.,

$$\mathcal{K}' = \{x' \in \mathcal{E}' : \text{for all } x \in \mathcal{K}, [x, x'] \equiv x'(x) \geq 0\}$$

and  $\mathcal{K}^d$  the *d-interior* defined as

$$\mathcal{K}^d = \{x \in \mathcal{K} : \text{for all } x' \in \mathcal{K}', x' \neq 0, [x, x'] > 0\}.$$

In the case  $\mathcal{K} = \mathbb{R}_+^n$ , the *d-interior* coincides with the topological interior consisting of vectors with all components positive.

An element  $\hat{x}' \in \mathcal{K}'$  is called *strictly positive* if  $[x, \hat{x}'] > 0$  holds for every  $x \in \mathcal{K}$ ,  $x \neq 0$ .

A partial order is introduced into  $\mathcal{E}$  by setting

$$x \leq y \text{ (or equivalently } y \geq x) \Leftrightarrow (y - x) \in \mathcal{K}.$$

Thus, in analogy to  $\mathbb{R}^n$  we denote  $x \in \mathcal{K}$  by  $x \geq 0$ .

An operator  $T \in \mathcal{B}(\mathcal{E})$  is called  $\mathcal{K}$ -nonnegative [11] if  $T\mathcal{K} \subset \mathcal{K}$ . A  $\mathcal{K}$ -nonnegative operator  $T$  is called  $\mathcal{K}$ -irreducible [25] if for every pair  $x \in \mathcal{K}$ ,  $x \neq 0$ ,  $x' \in \mathcal{K}'$ ,  $x' \neq 0$ , there is an index  $p = p(x, x') \geq 1$  such that  $[T^p x, x'] > 0$ ; see also [19].

Let  $S, T \in \mathcal{B}(\mathcal{E})$ . We let

$$T \geq S \text{ (or equivalently } S \leq T) \Leftrightarrow (T - S)\mathcal{K} \subset \mathcal{K}.$$

In particular,  $T \geq 0$  for any  $\mathcal{K}$ -nonnegative operator.

Let  $T \in \mathcal{B}(\mathcal{E})$ . By  $\mathcal{R}(T)$  we denote the range of  $T$  and by  $\ker(T)$  its null space. By  $\tilde{T}$  we denote the complex extension of  $T$ , i.e.,  $\tilde{T}z = Tx + iTy$ , where  $z = x + iy$ ,  $x, y \in \mathcal{E}$ .

Let  $I$  denote the identity operator. Let  $T \in \mathcal{B}(\mathcal{E})$  and  $\tilde{T}$  be its complex extension. The set

$$\rho(\tilde{T}) = \{\lambda \in \mathbb{C} : (\lambda I - \tilde{T})^{-1} \in \mathcal{B}(\mathcal{E})\}$$

is called the *resolvent set* of  $\tilde{T}$ . Its complement

$$\sigma(\tilde{T}) = \mathbb{C} \setminus \rho(\tilde{T})$$

is called the *spectrum* of  $\tilde{T}$ . By definition,  $\sigma(T) \equiv \sigma(\tilde{T})$ . The quantity

$$r(T) = \max\{|\lambda| : \lambda \in \sigma(T)\}$$

is called the *spectral radius* of  $T$ . We define the *peripheral spectrum* of  $T$  by setting

$$\sigma_\pi(T) = \{\lambda \in \sigma(T) : |\lambda| = r(T)\}.$$

An operator  $A \in \mathcal{B}(\mathcal{E})$  is called an *M-operator*, more precisely  $\mathcal{K}$ -*M-operator* [17], if  $A = bI - B$ , where  $b \geq r(B)$  and  $B \geq 0$ .

Let  $\tilde{T} \in \mathcal{B}(\mathcal{F})$  and let  $\mu \in \sigma(\tilde{T})$  be isolated. Then [28, pp. 305–306]

$$(\lambda I - \tilde{T})^{-1} = \sum_{k=0}^{\infty} A_k(\mu)(\lambda - \mu)^k + \sum_{k=1}^{\infty} B_k(\mu)(\lambda - \mu)^{-k},$$

where the  $A_k(\mu), B_{k+1}(\mu) \in \mathcal{B}(\mathcal{F})$  and the following relations hold:

$$[B_1(\mu)]^2 = B_1(\mu)$$

and

$$B_{k+1}(\mu) = (\tilde{T} - \mu I)B_k(\mu), \quad k = 1, 2, \dots$$

In particular, if there is an index  $q = q(\mu) < +\infty$ , such that  $B_k(\mu) = 0$  for  $k > q$ , the singularity  $\mu$  is called a *pole* of the resolvent operator  $(\lambda I - \tilde{T})^{-1}$  and  $q = q(\mu)$  is called its *multiplicity* or *order*. We write  $q(\mu) = 0$  if  $\mu \notin \sigma(\tilde{T})$ , i.e., if  $\mu \in \rho(\tilde{T})$  is a *regular point* of the resolvent operator.

An operator  $\tilde{T} \in \mathcal{B}(\mathcal{F})$  is said to have *Property P*, if the peripheral spectrum  $\sigma_\pi(\tilde{T})$  consists of poles of the resolvent operator. We say that  $T \in \mathcal{B}(\mathcal{E})$  has *Property P*, if its complex extension possesses this property. Similarly, we say

that an operator  $T \in \mathcal{B}(\mathcal{E})$  has a certain property, if its complex extension  $\tilde{T}$  possesses this property.

An operator  $\tilde{T} \in \mathcal{B}(\mathcal{F})$  is called *convergent*, if there exists an operator  $P \in \mathcal{B}(\mathcal{F})$  such that  $0 = \lim_{k \rightarrow \infty} \|\tilde{T}^k - P\|_{\mathcal{F}}$ . If  $P = 0$  (the zero operator),  $\tilde{T}$  is called *zero convergent* [24].

Let  $T$  have *Property P*, and let the value 1 be a pole of the resolvent operator  $(\lambda I - T)^{-1}$ . We denote by  $(I - T)^D$  the *Drazin generalized inverse* of  $I - T$ , i.e.,

$$(I - T)^D = \frac{1}{2\pi i} \int_{\mathcal{E}} f_0(\lambda)(\lambda I - T)^{-1} d\lambda,$$

where

$$f_0 = \begin{cases} 1/(1 - \lambda) & \text{if } |1 - \lambda| \geq \varepsilon, \\ 0 & \text{if } |1 - \lambda| < \varepsilon, \end{cases}$$

with  $\varepsilon > 0$  such that  $\{\lambda : |1 - \lambda| \leq \varepsilon\} \cap \sigma(T) = \{1\}$ . The contour of integration is

$$\mathcal{E} = \{\lambda : |1 - \lambda| = \varepsilon\} \cup \{\lambda : |\lambda| = R\},$$

with  $R > 1$  such that  $\sigma(T) \subset \{\lambda : |\lambda| \leq R\}$ .

An operator  $T \in \mathcal{B}(\mathcal{E})$ ,  $T\mathcal{H} \subset \mathcal{H}$ , is called a *Markov-type operator corresponding to  $\hat{x}'$* , if

$$(2) \quad T'\hat{x}' = \hat{x}',$$

where  $\hat{x}'$  is a strictly positive element in  $\mathcal{H}'$ . In  $\mathbb{R}^n$  the vector  $\hat{x}'$  is usually chosen as  $e' = (1, 1, \dots, 1)$ , and with this choice condition (2) means that the matrix  $T$  is column stochastic; see further §6.

A vector  $x \in \mathcal{H}$  is called a *stationary probability vector*, if

$$(3) \quad Tx = x \quad \text{and} \quad [x, \hat{x}'] = 1,$$

where  $T$  is a Markov-type operator corresponding to  $\hat{x}'$ . In  $\mathbb{R}^n$ , with the choice  $\hat{x}' = e'$ , the second condition in (3) is  $\sum_{i=1}^n x_i = 1$ .

An  $M$ -operator  $A \in B(\mathcal{E})$  is said to have *Property C*, if there is an operator  $B \in \mathcal{B}(\mathcal{E})$  such that  $b \geq r(B)$ ,  $A = bI - B$ , and  $(1/b)B = T$  is convergent. It is easy to see that each  $M$ -operator  $A = bI - B$  for which  $b > r(B)$  has *Property C*. On the other hand, it is known that even in the matrix case there are singular  $M$ -matrices not having *Property C*; see, e.g., [4, p. 152].

### 3. ITERATIVE AND SEMI-ITERATIVE METHODS

Let  $A$  be an  $M$ -operator, and let  $A = M - N$  be its splitting, i.e.,  $M^{-1} \in \mathcal{B}(\mathcal{E})$ . Given an initial guess  $x_0 \in \mathcal{E}$ , consider for the solution of (1) by iterative methods of the form

$$Mx_{k+1} = Nx_k, \quad k = 0, 1, \dots,$$

or equivalently,

$$(4) \quad x_{k+1} = Tx_k, \quad k = 0, 1, \dots,$$

where  $T = M^{-1}N$  is called the *iteration operator*.

The following result is well known; see, e.g., [4, p. 152] and also [15].

**Proposition 3.1.** *Let  $T$  have Property P. The iterative sequence  $\{x_k\}$  defined by (4) is convergent if and only if  $T = P + S$ , where  $P^2 = P$ ,  $PS = SP = 0$ , and  $r(S) < 1$ .*

These rather restrictive conditions on the iteration operator lead to the necessity of broadening the class of convergent methods. Thus, we consider semi-iterative methods corresponding to the iterative method (4), which we now define; see [6, 7, 21, 30].

Let  $p_k$  be a polynomial of degree  $k$ ,  $k \geq 0$ , i.e., let

$$(5) \quad p_k(z) = \sum_{t=0}^k \pi_{k,t} z^t, \quad \pi_{k,k} \neq 0.$$

It is assumed that

$$(6) \quad p_k(1) = 1.$$

Then

$$(7) \quad q_{k-1}(z) = \frac{1 - p_k(z)}{1 - z}$$

is a polynomial of degree  $k - 1$ , and (7) is the only polynomial of degree  $k - 1$  interpolating the function  $\frac{1}{1-z}$  on the set of the roots of the polynomial  $p_k$  [7].

We define

$$(8) \quad y_k = \sum_{t=0}^k \pi_{k,t} x_t, \quad k = 0, 1, \dots,$$

and call the sequence  $\{y_k\}$  the *semi-iterative sequence* corresponding to the basic iteration (4). It follows from (6) that  $y_0 = x_0$ .

The following result characterizes conditions for the convergence of semi-iterative methods. We study these conditions in specific examples later in §§5 and 6.

**Proposition 3.2.** *Let  $T \in \mathcal{B}(\mathcal{E})$  have Property P. Let the value 1 be a pole of the resolvent operator  $(\lambda I - T)^{-1}$  with multiplicity  $q = q(1)$ . Let the polynomials  $p_k$  and  $q_{k-1}$  be defined by (5) and (7), respectively, and let (6) hold. Then the following statements are equivalent:*

(a) *The sequence  $\{y_k\}$  is convergent for arbitrary  $y_0 \in \mathcal{R}((I - T)^q) \cup \ker(I - T)$ .*

(b)  *$\lim_{k \rightarrow \infty} p_k(T)v = 0$  for every  $v \in \mathcal{R}((I - T)^q)$ .*

(c)  *$\lim_{k \rightarrow \infty} q_{k-1}(T)v = (I - T)^D v$  for each  $v \in \mathcal{R}((I - T)^q)$ .*

*If one, and thus all, of these conditions are fulfilled, then*

$$\lim_{k \rightarrow \infty} y_k = [I - (I - T)^D(I - T)]y_0.$$

The proof of Proposition 3.2 can be given by the same arguments used in proving Theorem 1 in [6]; a generalization from the matrix case to our operator situation represents no essential difficulty.

## 4. JACOBI RELAXATION AND SUCCESSIVE RELAXATION ITERATIONS

In this section we show that some classical iterative methods are suitable for finding stationary probability vectors.

Let  $T$  be a Markov-type operator corresponding to an  $\hat{x}' \in \mathcal{X}'$ . Define the *Jacobi relaxation* operator as

$$W(\omega) = (1 - \omega)I + \omega T,$$

where  $0 < \omega < 1$ .

Similarly, cf. [30], the *successive relaxation iteration* operator is defined by

$$V(\omega) = (I - \omega L)^{-1}[(1 - \omega)I + \omega U],$$

where  $T = L + U$ ,  $L \geq 0$ , and  $U \geq 0$ .

*Remark 4.1.* These methods have been analyzed in [1] for the  $n \times n$  standard  $M$ -matrix case, and their convergence has been established for generally reducible  $M$ -matrices possessing *Property C*. The main result in [1] is based on an interesting inequality relating the  $kn$ th power of the Jacobi relaxation matrix  $W(\omega)^{kn}$  to the  $k$ th power of the successive relaxation matrix  $V(\omega)^k$ , where  $n$  is the dimension of the space, i.e., the size of the matrix  $T$ . From this relation one might conclude that the convergence result depends on the dimension of the space, and moreover, that this result may not hold in infinite dimensions. However, this is not the case. The argument used in our proof is based on a rather simple observation, telling us that the peripheral spectrum of a nonnegative operator  $T$  bounded below by a positive multiple of the identity map is a singleton; see [15].

**Proposition 4.2.** *Let  $T \in \mathcal{B}(\mathcal{E})$  be  $\mathcal{X}$ -irreducible and  $V(\omega)$  have Property P. Then  $r(V(\omega)) = 1$ .*

*Proof.* Let  $x_0 \in \mathcal{X}^d$  be an eigenvector of  $T$  corresponding to  $r(T) = 1$ . Then the relation  $V(\omega)x_0 = x_0$  and Lemma 4.1 in [17] imply that  $r(V(\omega)) = 1$ .  $\square$

**Theorem 4.3.** *Let  $T \in \mathcal{B}(\mathcal{E})$ ,  $T\mathcal{X} \subset \mathcal{X}$ , and let the spectral radius  $r(T)$  be a simple pole of the resolvent  $(\lambda I - T)^{-1}$ . Then the Jacobi relaxation method defined by  $W(\omega)$  is convergent for  $0 < \omega < 1$ . Let  $r(V(\omega)) = 1$ . Let  $T$  have the spectral decomposition  $T = P + S$ , where  $P^2 = P$ ,  $PS = SP = 0$ , and  $r(S) \leq 1$ ,  $1 \notin \sigma(S)$ . Then the successive relaxation method defined by  $V(\omega)$  is convergent for  $\omega > 0$  such that*

$$(9) \quad \omega \|PLP\| < 1.$$

*Proof.* Since  $W(\omega) \geq (1 - \omega)I$ , by Proposition 2 in [15] it follows that  $\sigma_\pi(W(\omega)) = \{1\}$ . Proposition 3.1 then implies convergence of the Jacobi relaxation.

Similarly, since  $V(\omega) \geq (1 - \omega)I$ , Proposition 2 in [15] implies that

$$\sigma_\pi(V(\omega)) = \{1\}.$$

By Proposition 3.1 it remains to show that the value 1 is a simple pole of  $(\lambda I - V(\omega))^{-1}$ . If the multiplicity of the value 1 as a pole of the resolvent

operator is not 1, then, by Theorem 6.4 in [17], there exists an element  $u(\omega)$  such that

$$u(\omega) = (I - \omega L)Pv = (I - P)w,$$

where  $v$  and  $w$  are some suitable elements. It follows that  $P(I - \omega L)Pv = 0$ , and hence,

$$Pv = \omega PLPv, \quad 0 < \omega < 1.$$

Therefore,  $\|Pv\| \leq \omega \|PLP\| \|Pv\|$ . However, this can happen only if  $Pv = 0$ . Thus,  $u(\omega) = 0$ , and the proof is complete.  $\square$

*Remark 4.4.* The validity of the condition  $r(V(\omega)) = 1$  depends upon an interplay between the cone  $\mathcal{K}$  and the operator  $T$ ; cf. Proposition 4.2. More generally, let  $\mathcal{K}$  and  $T \in \mathcal{B}(\mathcal{E})$  be such that there exists a direct decomposition  $\mathcal{K} = \sum_j \mathcal{K}_j$ , where  $\mathcal{K}_j \subset \mathcal{E}_j$  are generating normal cones in  $\mathcal{E}_j$ ,  $\mathcal{E} = \sum_j \mathcal{E}_j$ , and the reductions  $T_j$  to  $\mathcal{E}_j$  are  $\mathcal{K}_j$ -irreducible. Then, under the hypotheses of Theorem 4.3,  $r(V(\omega)) = 1$ . To see the interplay mentioned above, let  $T_j$  be the reduction of  $T$  to  $\mathcal{E}_j$ . Let  $V(\omega)_j$  be the corresponding reduction of  $V(\omega)$ . Then, by Proposition 4.2,  $r(V(\omega)_j) \leq 1$ , and thus,  $r(V(\omega)) = \sup\{r(V(\omega)_j) : j\} = 1$ .

Let us mention two examples. The first is  $\mathcal{E} = \mathbb{R}^n$ ,  $\mathcal{K} = \mathbb{R}_+^n$ , and any matrix  $T \in \mathcal{B}(\mathcal{E})$ , the second is  $\mathcal{E} = l^p$ ,  $\mathcal{K} = l_+^p$ ,  $1 \leq p \leq +\infty$ , and any compact operator  $T \in \mathcal{B}(\mathcal{E})$ .

*Remark 4.5.* The hypothesis (9) of Theorem 4.3 holds in particular if the spectral projections  $P$  are uniformly bounded, say,  $\|P\| \leq \kappa$ . In this case, the successive relaxation method defined by  $V(\omega)$  is convergent for sufficiently small  $\omega > 0$ .

*Remark 4.6.* Let  $\mathcal{E} = \mathbb{R}^n$  and  $\mathcal{K} = \mathbb{R}_+^n$ . Then the interval of  $\omega$ 's for which the successive relaxation method converges is the whole interval  $(0, 1)$ . This follows from Rothblum's Index Theorem [23] and the fact that the incidence graphs of the matrices  $V(\omega)$  are identical for all  $\omega \in (0, 1)$ .

*Remark 4.7.* An alternative proof of Theorem 4.3 for the case  $\mathcal{E} = \mathbb{R}^n$  and  $\mathcal{K} = \mathbb{R}_+^n$  can be given by using Theorem 6.5 in [17].

## 5. SOME PARTICULAR SEMI-ITERATIVE PROCESSES

From the variety of all semi-iterative methods for computing the stationary probability vectors, we present some particular examples. Our analysis shows how the methods proposed here can compete with a very effective method of conjugate gradients, the method analyzed by Tanabe in [27].

Consider first the polynomials

$$(10) \quad p_k(z) = \frac{1}{k+1} \sum_{t=0}^k z^t, \quad k = 1, 2, \dots$$

Since condition (6) is obviously satisfied, the polynomials (10) are admissible for constructing a semi-iterative method. These are the Cesaro-1 (C-1) sums; cf. [6, 24].

**Theorem 5.1.** Let  $T \in \mathcal{B}(\mathcal{E})$ ,  $T\mathcal{K} \subset \mathcal{K}$ , have Property P, let  $r(T) \leq 1$ , and let the value 1 be a simple pole of  $(\lambda I - T)^{-1}$ . Then the semi-iterative method defined by (4), (8), and (10) converges to a solution of the operator equation  $x - Tx = 0$ .

*Proof.* According to Proposition 3.2 it is enough to show that

$$\lim_{k \rightarrow \infty} p_k(T)v = 0,$$

for every  $v \in \mathcal{R}((I - T))$ . For this purpose, it is sufficient to prove that

$$(11) \quad \lim_{k \rightarrow \infty} p_k(\lambda_j) = 0,$$

for all  $\lambda_j \in \sigma_\pi(T)$ , and that

$$(12) \quad \lim_{k \rightarrow \infty} p_k^{(l)}(\lambda) = 0,$$

for all  $\lambda \in \sigma(T)$  and  $l = 0, 1, \dots, k - 1$ . However, the relations (11)–(12) are well known from elementary calculus.  $\square$

*Remark 5.2.* Let  $A = I - B$  and  $B = L + U$ , where  $B \in \mathcal{B}(\mathcal{E})$ ,  $L \geq 0$ ,  $U \geq 0$ ,  $r(L) < 1$ , and  $r(B) \leq 1$ .

Then, by choosing either  $T = B$  or  $T = (I - L)^{-1}U$ , we obtain the  $(C - 1)$ -Jacobi or the  $(C - 1)$ -Gauss-Seidel semi-iterative method, respectively. These methods have been analyzed for the  $n \times n$   $M$ -matrix case in [1].

Let  $\mathcal{E}$  be a Hilbert space with an inner product  $(\cdot, \cdot)$ , and let  $B^*$  denote the adjoint of  $B \in \mathcal{B}(\mathcal{E})$ . Let

$$(13) \quad T = I - (I - B^*)(I - B).$$

Let  $0 < \sigma_0 \leq \dots \leq \sigma_\infty$  be the nonzero singular values of  $A = I - B$ . Define

$$(14) \quad \mu_0 = \frac{4}{(\sigma_\infty + \sigma_0)^2}, \quad \mu_2 = -\frac{(\sigma_\infty - \sigma_0)^2}{(\sigma_\infty + \sigma_0)^2},$$

and

$$(15) \quad \mu_1 = 1 - \mu_0 - \mu_2.$$

Let  $y_0$  be arbitrary and consider the sequence  $\{y_k\}$  defined as follows:

$$(16) \quad y_1 = \mu_0 T y_0 + (1 - \mu_0) y_0,$$

$$(17) \quad y_k = \mu_0 T y_{k-1} + \mu_1 y_{k-1} + \mu_2 y_{k-2}, \quad k = 2, 3, \dots$$

**Theorem 5.3.** Let  $A = I - B$ , where  $B \geq 0$ ,  $r(B) = 1$ , and let  $B$  have Property P. Let  $T$  be defined by (13) and let

$$\sigma(T) \setminus \{1\} \subset [1 - \sigma_\infty^2, 1 - \sigma_0^2].$$

Then the stationary two-step iterative process (14)–(17) converges to a solution of the equation  $Ax = 0$ .

*Proof.* According to Theorem 1 in [6] and the theory of asymptotically optimal semi-iterative methods (AOSIM) developed in [8], the above two-step iteration is equivalent to a particular AOSIM related to the interval  $[1 - \sigma_\infty^2, 1 - \sigma_0^2]$ . This fact implies convergence, and completes the proof.  $\square$



*Remark 5.4.* The convergence theory is developed in [8] for the  $n \times n$  matrix case only. An appropriate generalization needed for our purposes can be obtained without any difficulty.

*Remark 5.5.* It is interesting to note that the limit vector  $y_\infty$  of the iteration (14)–(17) has the form

$$y_\infty = [I - (I - T)^+(I - T)]y_0,$$

where  $(I - T)^+$  denotes the *Moore-Penrose pseudoinverse* of  $I - T$ . This means that in the case considered, the Drazin generalized inverse and Moore-Penrose pseudoinverse coincide. For a general definition of the Moore-Penrose pseudoinverse and some of its properties the reader can consult the collection [20]. For the case of operators  $T$  having a closed graph, a brief discussion is presented in [12, p. 714].

*Remark 5.6.* The algorithm (13)–(17) should be performed in a numerically reasonable way. In particular, the vectors  $y_k$  are updated by multiplying successively by  $I - B^*$  and  $I - B$  without the explicit computation of the product  $(I - B^*)(I - B)$ ; see [6].

*Remark 5.7.* As noticed in [6], the process (13)–(17) remains convergent even if some bounds  $\hat{\sigma}_0$  and  $\hat{\sigma}_\infty$  are available in place of  $\sigma_0$  and  $\sigma_\infty$ , respectively; convergence is actually achieved even with some rather crude bounds. The rate of convergence may, however, dramatically decay in comparison with the asymptotically optimal rate of the process with the exact spectral bounds. For numerical realization some effective methods of bounding the quantities  $\sigma_0$  and  $\sigma_\infty$  are thus needed.

## 6. COMPUTATION OF ALL EXTREMAL STATIONARY PROBABILITY VECTORS OF A STOCHASTIC MATRIX

In this section we apply the results of the previous sections to the classical case of an  $n \times n$  stochastic matrix. We proceed in a way similar to that of Tanabe in [27] for the conjugate gradient method.

Let  $\mathcal{E} = \mathbb{R}^n$  with  $[x, y] = \sum_{j=1}^n x_j y_j$  and  $\mathcal{X} = \mathbb{R}_+^n$ . According to [19],  $\mathbb{R}_+^n$ -irreducibility is equivalent to the classical Frobenius irreducibility as presented, e.g., in [30, p. 18]. Let  $\hat{e}' = (1, \dots, 1) \in (\mathbb{R}^n)'$ , where  $(\mathbb{R}^n)'$  denotes the dual to  $\mathbb{R}^n$ . An  $n \times n$  (column) stochastic matrix  $(b_{j,k})$ ,  $b_{j,k} \geq 0$ ,  $\sum_{j=1}^n b_{j,k} = 1$ , defines a Markov-type operator in our terminology as follows:

$$[Bx, \hat{e}'] = [x, \hat{e}'] \quad \text{for all } x \in \mathbb{R}^n,$$

i.e.,  $\hat{x}' = \hat{e}'$ . Such a matrix  $B$  is called a *transition matrix*. It is well known that  $B$  can be expressed as  $B = Q + Z$ , where  $Q^2 = Q$ ,  $QZ = ZQ = 0$ , and  $1 \notin \sigma(Z)$ . It follows that  $x \in \mathbb{R}_+^n$  is a stationary probability vector of  $B$  if and only if  $x \in \mathcal{R}(Q)$ ,  $[x, \hat{e}'] = 1$ . Since  $\mathbb{R}^n$  is a direct sum of the ranges  $\mathcal{R}(Q)$  and  $\mathcal{R}(I - Q)$ , we deduce that

$$x \in \mathbb{R}_+^n, Qx \neq 0 \Leftrightarrow x \notin \mathcal{R}(I - Q) = \mathcal{R}(I - B).$$

A splitting  $A = M - N$  is called a *regular splitting* if  $M^{-1} \geq 0$  and  $N \geq 0$  [4, 16, 29, 30].

**Proposition 6.1.** *Let  $A = I - B = M - N$  be a regular splitting, and let  $T = M^{-1}N$ . Let  $P$  and  $R$  be such that  $T = P + R$ ,  $P^2 = P$ ,  $PR = RP = 0$ , and  $1 \notin \sigma(R)$ . If  $B$  is an irreducible transition matrix, and if the semi-iterative procedure described by (4)–(8) converges to the limit  $y_\infty$ , then*

$$(18) \quad y_\infty = \frac{[Py_0, \hat{e}']}{[\hat{x}, \hat{e}']} \hat{x},$$

where

$$(19) \quad \hat{x} = P\hat{x} \in \text{Int } \mathbb{R}_+^n,$$

i.e.,  $\hat{x} > 0$ . Moreover,  $y_\infty \neq 0$  if and only if  $y_0 \notin \mathcal{R}(I - B)$ .

*Proof.* By Proposition 3.2,

$$y_\infty = [I - (I - T)^D(I - T)]y_0.$$

It follows that [17]

$$(I - B)P = (M - N)P = M(I - T)P = 0.$$

This implies that

$$\mathcal{R}(P) \subset \ker A = \ker(I - B).$$

However,  $\ker(I - B) = \mathcal{R}(Q)$  and  $\dim \mathcal{R}(Q) = 1$ . Since  $\mathcal{R}(P) \neq \{0\}$ , we have  $\mathcal{R}(P) = \mathcal{R}(Q)$ . Therefore, since

$$P = I - (I - T)^D(I - T),$$

we have that  $I - P = (I - T)^D(I - T)$  is a projection of  $\mathbb{R}^n$  onto the range of  $I - T$ . As a direct conclusion we obtain (18) and (19).

If  $y_\infty \neq 0$ , then, since  $y_\infty = Py_\infty = Qy_\infty$ , we have

$$y_0 \notin \mathcal{R}(I - Q) = \mathcal{R}(I - B).$$

Conversely, if  $y_0 \notin \mathcal{R}(I - B) = \mathcal{R}(I - Q)$ , we see that  $y_0 \neq y_0 - Qy_0$  and thus,  $Qy_0 \neq 0$ , implying  $Py_0 \neq 0$  and consequently,  $y_\infty \neq 0$ .  $\square$

**Corollary 6.2.** *Under the same hypotheses as in Proposition 6.1, let  $y_0 \in \mathbb{R}_+^n$ . Then  $y_0 \notin \mathcal{R}(I - B)$  and the unique stationary probability vector of  $B$  is given by*

$$\hat{x} = \frac{\{I - (I - T)^D(I - T)\}y_0}{\{[I - (I - T)^D(I - T)]y_0, \hat{e}'\}} = \frac{Py_0}{[Py_0, \hat{e}']} = \frac{Qy_0}{[y_0, \hat{e}']}.$$

A computation of the stationary probability vectors of general reducible matrices is based on the following representation of  $B$  in the form [9, p. 448]

$$B = E \begin{bmatrix} F_0 & 0 & 0 & \dots & 0 \\ G_1 & F_1 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ G_r & 0 & 0 & \dots & F_r \end{bmatrix} E', \quad 1 \leq r \leq n.$$

Here,  $E$  is an  $n \times n$  permutation matrix,  $E'$  its transpose,  $F_j$ ,  $1 \leq j \leq r$ , are irreducible column stochastic matrices, and  $r(F_0) < 1$ . If  $u_j$  is the unique

stationary probability vector of  $F_j$ , then

$$(20) \quad v_j = E(0, \dots, u_j, 0, \dots, 0)', \quad j = 1, \dots, r,$$

are mutually orthogonal stationary probability vectors of  $B$ . Every  $v_j$  has  $n_j$  positive entries which correspond to the states in the same ergodic class  $\mathcal{E}_j$ , where  $n_j$  is the order of  $F_j$ . Every stationary probability vector of  $B$  is a convex combination of  $v_1, \dots, v_r$ .

For a vector  $b \in \mathbb{R}_+^n$ ,  $b = (b_1, \dots, b_n)'$ , let us define the *support* of  $b$  as

$$\text{supp}(b) = \{j : b_j > 0\}.$$

**Theorem 6.3.** *Let  $B$  be a column stochastic matrix. Let  $v_j$  be defined as in (20) and let  $\{y_k\}$  be the sequence generated by (4), (5), and (8). Then*

$$y_\infty = \lim_{k \rightarrow \infty} y_k = \sum_{j=1}^r \frac{[Py_0, \hat{e}'_j]}{[v_j, \hat{e}'_j]} v_j,$$

where  $y_0 \in \mathbb{R}_+^n$  is the initial vector of the process and

$$\hat{e}'_j = (0, \dots, 0, 1, \dots, 1, \dots, 1, 0, \dots, 0)', \quad j \geq 0,$$

so that  $\hat{e}' = \sum_{j=0}^r \hat{e}'_j$ .

If  $y_0 \notin \mathcal{R}(I - B)$ , then  $y_\infty \neq 0$ , and the vector  $\hat{y} = y_\infty/[y_\infty, \hat{e}']$  is a stationary probability vector.

Let  $e_1, \dots, e_n$  denote the elements of the standard basis,

$$e_j = (0, \dots, 1, \dots, 0)', \quad j = 1, \dots, n.$$

**Theorem 6.4.** *Let  $B$  be a column stochastic matrix, and let  $y_0 = ce_j$  with some  $c > 0$ . Then*

$$y_\infty \begin{cases} \neq 0 & \text{if } j \in \text{supp}(v_t) \text{ for some } 1 \leq t \leq r, \\ = 0 & \text{otherwise.} \end{cases}$$

If  $y_\infty \neq 0$ , there exists a unique integer  $k$  such that  $j \in \text{supp}(v_k)$ , and  $v_k$  can be expressed as

$$v_k = \frac{1}{[y_\infty, \hat{e}']} y_\infty.$$

**Corollary 6.5.** *For every  $e_j$ ,  $j = 1, \dots, n$ , one of the following conditions holds:*

1. The vector

$$\hat{x}_j = \frac{1}{[e_j, \hat{e}']} Pe_j = \frac{1}{[ \{I - (I - T)^D(I - T)\} e_j, \hat{e}' ]} \{I - (I - T)^D(I - T)\} e_j$$

coincides with one of the extremal probability vectors of  $B$  in which case the state  $j$  belongs to the ergodic class  $\text{supp}(\hat{x}_j)$ .

2. The vector  $Pe_j$  is the zero vector, in which case the state  $j$  is transient.

**Corollary 6.6.** Let  $\hat{e} = \sum_{j=1}^n e_j$ . The probability vector

$$\hat{x} = \frac{1}{[P\hat{e}, \hat{e}']} P\hat{e}$$

is a convex combination of all the extremal stationary probability vectors  $v_j$ ,  $j = 1, \dots, r$ , i.e.,  $\hat{x} = \sum_{j=1}^r c_j v_j$ , where  $c_j = [\hat{x}, E\hat{e}_j]$ .

The proofs of the theorems and corollaries stated in this section can be given in a way very similar to that of Tanabe [27] and are thus omitted.

From these results it follows that the whole approach of Tanabe can be applied to our semi-iterative methods in place of his conjugate gradient method. This concerns, in particular, both the Conceptual and Computational Algorithms. These algorithms are important tools for practical computations. This means that with our semi-iterative methods one can compute the complete basis of stationary probability vectors of  $B$  in a numerically stable way, just as it was proposed by Tanabe [27].

#### BIBLIOGRAPHY

1. G. P. Barker and S.-J. Yang, *Semi-iterative and iterative methods for singular  $M$ -matrices*, SIAM J. Matrix Anal. Appl. **9** (1988), 168–180.
2. George P. Barker and Robert J. Plemmons, *Convergent iterations for computing stationary distributions of Markov chains*, SIAM J. Algebraic Discrete Methods **7** (1986), 390–398.
3. V. A. Barker, *Numerical solution of sparse singular systems of equations arising from ergodic Markov chains*, Comm. Statist. Stochastic Models **5** (1989), 335–381.
4. Abraham Berman and Robert J. Plemmons, *Nonnegative matrices in the mathematical sciences*, Academic Press, New York, 1979.
5. Eugene B. Dynkin, *Markov processes*, Vols. I, II, Springer-Verlag, Berlin, 1965.
6. Michael Eiermann, Ivo Marek, and Wilhelm Niethammer, *On the solution of singular linear systems of algebraic equations by semiiterative methods*, Numer. Math. **53** (1988), 265–283.
7. Michael Eiermann and Wilhelm Niethammer, *On the construction of semiiterative methods*, SIAM J. Numer. Anal. **20** (1983), 1153–1160.
8. Michael Eiermann, Wilhelm Niethammer, and Richard S. Varga, *A study of semiiterative methods for nonsymmetric systems of linear equations*, Numer. Math. **47** (1985), 505–533.
9. F. R. Gantmacher, *The theory of matrices*, Goz. Izdat. Lit., Moscow, 1954; English transl., *Applications of the theory of matrices*, Interscience, New York, 1959.
10. Halim D. Kafeety, Carl D. Meyer, and William J. Stewart, *A general framework for iterative aggregation/disaggregation methods for the numerical solution of Markov chains*, Technical Report No. 06039101, Center for Research in Scientific Computation, North Carolina State University, Raleigh, North Carolina, November 1990. (See also the Proceedings of the Copper Mountain Conference on Iterative Methods (Copper Mountain, Colorado, April 9–14, 1992), Computational Mathematics Group, University of Colorado at Denver, 1992.)
11. M. G. Krein and M. A. Rutman, *Linear operators leaving a cone invariant in a Banach space*, Uspekhi Mat. Nauk **3** (1948), no. 1, 3–95; English transl., Amer. Math. Soc. Transl. **26** (1950).
12. W. Ledermann and Steven Vajda, eds., *Handbook of applicable mathematics, Vol. IV. Analysis*. Wiley Interscience, Chichester, New York, Brisbane, Toronto, and Singapore, 1982.
13. Vladimir Lelek and Ivo Marek, *Data fitting under eigenvalue constraints*, Proc. Ninth Tagung über Probleme und Methoden der Mathematischen Physik (F. Kuhnert and B. Silbermann, eds.), Karl Marx-Stadt, 1988, pp. 159–166.

14. Ivo Marek, *Aggregation methods of computing stationary distribution of Markov processes, Eigenvalue Problems and their Numerical Treatment* (W. Velte, ed.), Proc. Conf. (Mathematisches Forschungsinstitut Oberwolfach, February 26–March 2, 1990), Internat. Ser. Numer. Math. (ISNM), vol. 96, Birkhäuser, Basel, 1991, pp. 155–169.
15. ———, *On square roots of  $M$ -operators*, Linear Algebra Appl. (submitted).
16. Ivo Marek and Daniel B. Szyld, *Comparison theorems for weak splittings of bounded operators*, Numer. Math. **58** (1990), 387–397.
17. ———, *Splittings of  $M$ -operators: Irreducibility and the index of the iteration operator*, Numer. Funct. Anal. Optim. **11** (1990), 529–553.
18. ———, *Iterative aggregation with inexact correction*, Technical Report no. 91-52, Department of Mathematics, Temple University, May 1991.
19. Ivo Marek and Karel Žitný, *Equivalence of  $K$ -irreducibility concepts*, Comment. Math. Univ. Carolinae **25** (1984), 61–72.
20. M. Z. Nashed, *Generalized inverses, normal solvability and iteration for singular operator equations*, Nonlinear Functional Analysis and Applications (M. Z. Nashed, ed.), Academic Press, New York, 1971, pp. 311–359.
21. Wilhelm Niethammer and Richard S. Varga, *The analysis of  $k$ -step iterative methods for linear systems from summability theory*, Numer. Math. **41** (1983), 177–206.
22. Donald J. Rose, *Convergent regular splittings for singular  $M$ -matrices*, SIAM J. Algebraic Discrete Methods **5** (1984), 133–144.
23. Uriel G. Rothblum, *Algebraic eigenspaces of nonnegative matrices*, Linear Algebra Appl. **12** (1975), 281–292.
24. ———, *Convergence properties of powers of matrices with applications to iterative methods for solving linear systems*, Extremal Methods and Systems Analysis (Anthony V. Fiacco and Kenneth O. Kortanek, eds.), An International Symposium on the Occasion of Professor Abraham Charnes' Sixtieth Birthday (Austin, Texas, 1977), Lecture Notes in Econom. and Math. Systems, vol. 174, Springer Verlag, Berlin, Heidelberg, and New York, 1980, pp. 231–247.
25. Ikuko Sawashima, *On spectral properties of some positive operators*, Nat. Sci. Rep., Ochanomizu Univ. **15** (1964), 53–64.
26. Paul J. Schweitzer, *A survey of aggregation-disaggregation in large Markov chains*, Numerical Solution of Markov Chains (William J. Stewart, ed.), Marcel Dekker, New York, Basel, and Hong Kong, 1991, pp. 63–88.
27. Kunio Tanabe, *The conjugate gradient method for computing all extremal stationary probability vectors of a stochastic matrix*, Ann. Inst. Statist. Math. B **37** (1985), 173–187.
28. Angus E. Taylor, *Introduction to functional analysis*, Wiley, New York, 1958.
29. Richard S. Varga, *Factorization and normalized iterative methods*, Boundary Problems in Differential Equations (Rudolph E. Langer, ed.), Univ. of Wisconsin Press, Madison, WI, 1960, pp. 121–142.
30. ———, *Matrix iterative analysis*, Prentice-Hall, Englewood Cliffs, NJ, 1962.

KATEDRA NUMERICKÉ MATEMATIKY NA MATEMATICKO-FYZIKÁLNÍ FAKULTĚ, UNIVERSITY KARLOVY, MALOSTRANSKÉ NÁM. 25, 118 00 PRAHA 1, CZECH REPUBLIC  
*E-mail address:* marek@cspguk11.bitnet

DEPARTMENT OF MATHEMATICS, TEMPLE UNIVERSITY TU 038-16, PHILADELPHIA, PENNSYLVANIA 19122-2585  
*E-mail address:* szyld@euclid.math.temple.edu