Accelerating the Sinkhorn-Knopp iteration by power methods

A. Aristodemo¹, L. Gemignani⁴*

¹ Dipartimento di Matematica, Università di Pisa, Largo B. Pontecorvo, 5 - 56127 Pisa, Italy
² Dipartimento di Informatica, Università di Pisa, Largo Bruno Pontecorvo, 3 - 56127 Pisa, Italy. E-mail: luca.gemignani@unipi.it

SUMMARY

It is shown that the problem of balancing a nonnegative matrix by positive diagonal matrices can be recast as a constrained nonlinear multiparameter eigenvalue problem. Based on this equivalent formulation some adaptations of the power method and Arnoldi process are proposed for computing the dominant eigenvector which defines the structure of the diagonal transformations. Numerical results illustrate that our novel methods accelerate significantly the convergence of the customary Sinkhorn-Knopp iteration for matrix balancing in the case of clustered dominant eigenvalues. Copyright © 0000 John Wiley & Sons, Ltd.

KEY WORDS: Sinkhorn-Knopp iteration, Nonlinear Eigenvalue Problem, Power method, Arnoldi method

1. INTRODUCTION

Many important types of data, like text, sound, event logs, biological sequences, can be viewed as graphs connecting basic data elements. Networks provide a powerful tool for describing the dynamic behavior of systems in biology, computer science, information engineering. Networks and graphs are generally represented as very large nonnegative matrices describing either the network topology, quantifying certain attributes of nodes or exhibiting the correlation between certain node features. Among the challenging theoretical and computational problems with these matrices there are the balancing/scalability issues.

The Sinkhorn-Knopp (SKK) balancing problem can be stated as follows: Given a nonnegative matrix $A \in \mathbb{R}^{n \times n}$ ($A \geq 0$), find if they exist two nonnegative diagonal matrices $D_1, D_2 \in \mathbb{R}^{n \times n}$ such that $S = D_1 AD_2$ is doubly stochastic, i.e.,

$$D_2 A^T D_1 e = e, \quad D_1 A D_2 e = e, \quad e = [1, \ldots, 1]^T. \quad (1)$$

The problem was raised in three different papers [1–3] that contain the well-known iteration for matrix balancing that bears their names. Several equilibration problems exist in which row or

*Correspondence to: luca.gemignani@unipi.it

Contract/grant sponsor: The work of L. Gemignani was partially supported by the GNCS/INdAM project “Tecniche innovative per Problemi di Algebra Lineare” and by the University of Pisa (grant PRA 2017-05).
column norms are not equal but rather are specified by positive vectors. Variants of the SKK problem have attracted attention in various fields of pure and applied sciences including input-output analysis in economics [4], optimal transportation theory and its applications in machine learning [5], complex network analysis [6,7], probabilistic and statistical modeling [8], optimization of traffic and telecommunication flows [9] and matrix preconditioning [10].

By introducing the operator $\mathcal{D}: \mathbb{R}^n \to \mathbb{R}^{n \times n}$ defined by $\mathcal{D}(v) = \text{diag}[v_1, \ldots, v_n]$ the computation in (1) amounts to find two vectors $r$ and $c$ such that $D_1 = \mathcal{D}(r)$ and $D_2 = \mathcal{D}(c)$ satisfy

$$\begin{align} D(c)A^T D(r)e &= D(c)A^T e = e; \\
D(r)A D(c)e &= D(r)Ac = D(Ac)r = e 
\end{align}$$

When $A$ is symmetric we can determine $r = c = z$ to satisfy $D(Az)z = D(z)Az = e$. In [2] the authors proposed the following fixed point iteration—called Sinkhorn-Knopp (SKK) iteration— for computing the desired vectors $r$ and $c$:

$$\begin{align} c_{k+1} &= \mathcal{D}(A^T r_k)^{-1} e; \\
r_{k+1} &= \mathcal{D}(Ac_{k+1})^{-1} e 
\end{align}$$

(2)

In the symmetric case the SKK iteration reduces to

$$z_{k+1} = \mathcal{D}(Az_k)^{-1} e$$

or, equivalently, by setting $1/z_k = x_k$ with the assumption $1/0 = +\infty$,

$$x_{k+1} = A(1/x_k).$$

(4)

The SKK iterations (2),(3),(4) have been rediscovered several times in different applicative contexts. Related methods are the RAS method [4] in economics, the iterative proportional fitting procedure (IPFP) in statistics and Krithoph’s projection scheme [9] in optimization.

A common drawback of all these iterative algorithms is the slow convergence behavior exhibited even in deceivingly simple cases. To explain this performance gap we observe that the equations in (2) can be combined to get

$$c_{k+1} = \mathcal{D}(A^T D(Ac_k)^{-1} e)^{-1} e, \quad k \geq 0,$$

(5)

which can be expressed componentwise as

$$(c_{k+1})_s = \left( \sum_{m=1}^n a_{m,s} \left( \sum_{\ell=1}^n a_{m,\ell} (c_k)_\ell \right)^{-1} \right)^{-1}, \quad 1 \leq s \leq n, \quad k \geq 0.$$ 

This means that (5) is equivalent to the fixed point iteration

$$c_{k+1} = T(c_k), \quad T(x)_s = \left( \sum_{m=1}^n a_{m,s} \left( \sum_{\ell=1}^n a_{m,\ell} x_\ell \right)^{-1} \right)^{-1},$$

(6)
ACCELERATING THE SINKHORN-KNOPP ITERATION BY POWER METHODS

for solving

\[ x = T x, \quad x \geq 0, \]  

(7)

where \( T \) is the nonlinear operator introduced by Menon in [11, 12] and according to those papers we write \( T x = T(x) \).

Our first contribution consists of a novel formulation of the fixed point problem (9) as a constrained nonlinear multiparameter eigenvalue problem of the form

\[ x = J_T(x)x, \quad x \geq 0, \]  

(8)

where \( J_T(z) \) denotes the Jacobian matrix of \( T \) evaluated at the point \( z \). It follows that the local dynamics of the original SKK algorithm (5) can be described as a power method with perturbations (13) applied to the matrix \( J_T(x) \) evaluated at the fixed point. Therefore the SKK iterations (2),(3),(4) inherit the pathologies of the power process in the case of clustered dominant eigenvalues of \( J_T(x) \).

Recently, acceleration methods using nonlinear solvers applied to equation (7) have been proposed (see [14] and the references given therein). Other optimization strategies and descending techniques were considered in [15,16]. In this paper we pursue a different approach by exploiting the properties of the equivalent nonlinear multiparameter eigenvalue problem (8). Krylov methods are the algorithms of choice for the computation of a few eigenvalues of largest magnitude of matrices. They have been efficiently used in information retrieval and web search engines for accelerating PageRank computations [17]. Inspired by this, we first propose to compute an approximation of the dominant eigenvector of of \( T \) by solving a possibly infinite sequence of linear matrix eigenvalue problems. Then we design fast eigensolvers, based on the power method and the Arnoldi process that are specifically tailored to solve these problems for large-scale matrices. Numerical results show that the resulting schemes can greatly improve the performance of the SKK iterations in the case of clustered dominant eigenvalues of \( J_T(x) \).

The paper is organized as follows. In Section 2 after briefly recalling the properties of the SKK fixed point iteration (6) we exploit the eigenvalue connection by devising accelerated variants of (6) using power methods. The description and implementation of these variants together with numerical results are discussed in Section 3. Finally, in section 4 conclusion and some remarks on future work are given.

2. THEORETICAL SETUP

Let us denote by \( P, P_0 \) and \( P_\infty \) the subsets of \( \mathbb{R}^n \), \( \mathbb{R} = \mathbb{R} \cup \{\pm \infty\} \), defined by \( P = \{x \in \mathbb{R}^n : x \geq 0\} \), \( P_0 = \{x \in \mathbb{R}^n : x > 0\} \) and \( P_\infty = \{x \in \mathbb{R}^n : x \geq 0\} \), respectively. For the sake of simplicity we assume that \( A \in \mathbb{R}^{n \times n} \) is a matrix with all positive entries, that is, \( A > 0 \). Arithmetic operations are partially extended to \( \mathbb{R} \) by setting \( 1/0 = \infty, 1/\infty = 0, \infty + \infty = \infty, 0 \cdot \infty = 0, a \cdot \infty = \infty \) if \( a > 0 \), where \( \infty = +\infty \).

Under these assumptions we can introduce the nonlinear operators defined as follows:

1. \( U : P_\infty \to P_\infty, Ux = 1/x; \)

2. \( S : P_\infty \to P_\infty, Sx = UAx; \)
3. $T : \mathcal{P}_\infty \rightarrow \mathcal{P}_\infty$, $Tx = UA^TUAx$.

In this way it can be easily noticed that $T$ is the same as the operator introduced in (6) and, therefore, the Sinkhorn-Knopp problem for the matrix $A$ reduces to compute the fixed points of $T$, that is, the vectors $x \in \mathcal{P}_\infty$ such that

$$x = Tx = UA^TUAx, \quad x \in \mathcal{P}_\infty. \quad (9)$$

Summing up the results stated in [11,12] we obtain the following theorem concerning the existence and the uniqueness of the desired fixed point.

**Theorem 2.1**

Let $A \in \mathbb{R}^{n \times n}$ be a matrix with all positive entries. Then $\forall u \in \mathcal{P} \setminus \{0\}$ we have $\sup \{\lambda : Tu \geq \lambda u\} \leq 1$. Moreover, $T$ has an eigenvalue equal to 1 with a corresponding eigenvector $x \in \mathcal{P}_0$. Finally, if $u \in \mathcal{P}$ is an eigenvector of $T$ associated with the eigenvalue 1 then $u = \rho x$ with $\rho > 0$.

The basic SKK algorithm proceeds to approximate the eigenvector $x \in \mathcal{P}_0$ by means of the fixed point iteration

$$\begin{cases}
    x^{(0)} \in \mathcal{P}_0; \\
    x^{(k+1)} = Tx^{(k)}, \quad k \geq 0
\end{cases} \quad (10)$$

The iteration is shown to be globally convergent since $T$ is a contraction for the Hilbert metric associated to the cone $\mathcal{P}$. [18,19].

**Theorem 2.2**

For any $x^{(0)} \in \mathcal{P}_0$ there exists $\gamma = \gamma(x^{(0)}) \in \mathbb{R}$, $\gamma > 0$, such that

$$\lim_{k \to \infty} x^{(k)} = \gamma x.$$

The convergence is linear and the rate depends on the second singular value of the doubly stochastic matrix $P = D(Sx)AD(x)$. We have the following [6].

**Theorem 2.3**

Let $x \in \mathcal{P}_0$ denote the limit of the sequence $\{x^{(k)}\}_{k \in \mathbb{N}}$ generated according to (10). Then the matrix $P = D(Sx)AD(x)$ is doubly stochastic and, moreover, if $\sigma_2$ is the second largest singular value of $P$ it holds

$$\|x^{(k+1)} - x\|_2 \leq \sigma_2^2 \|x^{(k)} - x\|_2 + o\left(\|x^{(k)} - x\|_2\right), \quad k \geq 0.$$

The convergence can be very slow in the case of nearly decomposable matrices. The following definition is provided in [20,21].

**Definition 2.4**

For a given $\epsilon > 0$, the matrix $A \in \mathbb{R}^{n \times n}$ is $\epsilon$-nearly decomposable if there exists a permutation matrix $P$ such that $PA^T = \hat{A} + \epsilon E$ where $\hat{A}$ is block triangular with square diagonal blocks.

The relevance of nearly decomposable matrices for the study of dynamic systems in economics has been examined by Simon and Ando [22]. The role of near decomposability in queuing and computer system applications has been discussed in [23]. For a general overview of the properties of nearly decomposable graphs and networks with applications in data science and information retrieval one can see [24].
Example 2.5
Let \( A = \begin{bmatrix} 1 - \epsilon & \epsilon \\ \epsilon & 1 - \epsilon \end{bmatrix} \), \( 0 < \epsilon < 1 \) be a doubly stochastic matrix. The vector \( x = [1, 1]^T \) provides a solution of the SKK scaling problem. The singular values of the matrix \( A = D(Sx)AD(x) \) satisfy \( \sigma_1 = 1 \) and \( \sigma_2 = |1 - 2\epsilon| \). For the matrix \( A = \begin{bmatrix} 1 & \epsilon \\ \epsilon & 1 \end{bmatrix} \), \( \epsilon > 0 \), the SKK iteration (10) is convergent but for a given fixed tolerance \( (\tau = 10^{-8}) \) the number of iterations shown in Table (I) grows exponentially as epsilon becomes small.

| \( k \) | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| \( ItN \) | 16 | 46 | 132 | 391 | 1139 | 3312 | 9563 | 27360 | 77413 | 216017 |

The local dynamics of (10) depend on the properties of the Jacobian matrix evaluated at the fixed point. By using the chain rule for the composite function we obtain that

\[
J_T(z) = J_{UA}A(z) = J_T(A^TUAz) \cdot J_A(UAz) \cdot J_U(Az) \cdot J_A(z). \tag{11}
\]

Since \( J_U(z) = -\text{diag}^{-2}(z) \) we find that

\[
J_T(z) = \text{diag}^2(Tz) \cdot A^T \cdot \text{diag}^2(Sz)A.
\]

The next result gives a lower bound for the spectral radius of \( J_T(z) \) for \( z \in P_0 \).

**Theorem 2.6**
For any given fixed \( z \in P_0 \) the spectral radius of \( J_T(z) \) satisfies \( \rho(J_T(z)) \geq 1 \).

**Proof**
Let us denote \( G = \text{diag}(Tz) \cdot A^T \cdot \text{diag}(Sz) \). It holds

\[
J_T(z) = \text{diag}^2(Tz) \cdot A^T \cdot \text{diag}^2(Sz)A \\
= \text{diag}(Tz) \cdot G \cdot G^T \text{diag}^{-1}(Tz),
\]

and, hence \( J_T(z) \) and \( G \cdot G^T \) are similar. Now observe that

\[
Ge = \text{diag}(Tz) \cdot A^T \cdot \text{diag}(Sz)e = \text{diag}(Tz) \cdot A^TUAz = e.
\]

It follows that \( \| G \|_2 \geq 1 \). By using the SVD of \( G \) it is found that \( \sigma_1(G) \geq 1 \) and therefore the spectral radius of \( J_T(z) \) satisfies \( \rho(J_T(z)) \geq 1 \).

If \( x = Tz \), \( x \in P_0 \), then it is worth noting that

\[
J_T(x) = \text{diag}^2(Tz) \cdot A^T \cdot \text{diag}^2(Sz)A = \text{diag}^2(x) \cdot A^T \cdot \text{diag}^2(Sz)A,
\]

and, hence,

\[
J_T(x) = \text{diag}(x) \cdot P^T P \text{diag}^{-1}(x),
\]
where \( P \) is introduced in Theorem 2.3. This means that \( J_T(x) \) and \( F = P^T P \) are similar and therefore the eigenvalues of \( J_T(x) \) are the square of the singular values of \( P \). Since \( A > 0 \) then it is irreducible and primitive and the same holds for \( P \) and a fortiori for \( F \). By the Perron-Frobenius theorem it follows that the spectral radius of \( J_T(x) \) satisfies \( \rho(J_T(x)) = 1 \) and \( \lambda = 1 \) is a simple eigenvalue of \( J_T(x) \) with a positive corresponding eigenvector.

A characterization of such an eigenvector can be derived by the following result.

**Theorem 2.7**

For each vector \( z \in \mathcal{P}_0 \) it holds

\[
Tz = J_T(z) \cdot z.
\]

**Proof**

Let \( z \in \mathbb{R}^n, z > 0 \), then we have

\[
J_T(z) \cdot z = \text{diag}^2(Tz) \cdot A^T \cdot \text{diag}^2(Sz)Az \\
= \text{diag}^2(Tz) \cdot A^T \cdot Sz \\
= \text{diag}^{-2}(A^T Sz) \cdot A^T \cdot Sz \\
= \text{diag}^{-1}(A^T Sz)e \\
= \text{diag}(UA^T UA)z \\
= Tz.
\]

This theorem implies that

\[
x \in \mathcal{P}_0, x = T x \quad \iff \quad x \in \mathcal{P}_0, x = J_T(x)x
\]

and therefore the eigenvector of \( J_T(x) \) corresponding with the eigenvalue 1 is exactly the desired solution of the SKK problem. Furthermore, the SKK iteration (10) can equivalently be written as

\[
\begin{align*}
x^{(0)} &\in \mathcal{P}_0; \\
x^{(k+1)} &= T x^{(k)} = J_T(x^{(k)})x^{(k)}, \quad k \geq 0
\end{align*}
\]

(12)

In principle one can accelerate the convergence of this iteration without improving the efficiency of the iterative method by replacing \( T = T_1 \) with \( T_\ell = T \circ T \circ \cdots \circ T \) generated from the composition (\( \ell \) times) of \( T \) for a certain \( \ell \geq 1 \). The linearized form of the resulting iteration is

\[
\begin{align*}
x^{(0)} &\in \mathcal{P}_0; \\
x^{(k+1)} &= J_T^\ell(x)x^{(k)}, \quad k \geq 0
\end{align*}
\]

(13)

This is the power method applied to the matrix \( J_T^\ell(x) \) for the approximation of an eigenvector associated with the dominant eigenvalue \( \lambda = 1 \). A normalized variant of (13) can be more suited for
numerical computations

\[
\begin{aligned}
\begin{cases}
\begin{aligned}
x^{(0)} & \in P_0; \\
v^{(k+1)} & = J_T^\ell (x^{(k)}) x^{(k)}, \\
x^{(k+1)} & = v^{(k+1)}/(e^T v^{(k+1)})
\end{aligned}
\end{cases}
\quad k \geq 0
\end{aligned}
\]  \tag{14}

It is well known that under our assumptions (14) generates sequences such that

\[
\lim_{k \to \infty} x^{(k)} = x/(e^T x), \quad \limsup_{k \to \infty} \|x^{(k)} - x/(e^T x)\|^{1/k} \leq \lambda_2^\ell = \sigma_2^{2\ell},
\]

where \(0 \leq \lambda_2 = \sigma_2^2 < 1\) is the second largest eigenvalue of \(J_T(x)\) and \(\sigma_2\) denotes the second largest singular value of \(P\) defined as in Theorem 2.3. For practical purposes we introduce the following modified adaptation of (14) called SKK\(_\ell\) iteration:

\[
\begin{aligned}
\begin{cases}
\begin{aligned}
x^{(0)} & \in P_0; \\
v^{(k+1)} & = J_T^\ell (x^{(k)}) x^{(k)}, \\
x^{(k+1)} & = v^{(k+1)}/(e^T v^{(k+1)})
\end{aligned}
\end{cases}
\quad k \geq 0
\end{aligned}
\]

For \(\ell = 1\) SKK\(_2\) reduces to the scaled customary SKK iteration. Under suitable assumptions we can show that SKK\(_\ell\) generates a sequence converging to the desired fixed point.

**Theorem 2.8**

Let \(\{x^{(k)}\}_k\) be the sequence generated by SKK\(_\ell\) from a given initial guess \(x^{(0)} \in P_0\). Let \(x \in P_0\) be such that \(x = T x\) and \(e^T x = 1\). Assume that:

1. \(\exists \eta > 0 : J_T^\ell (x^{(k)}) = J_T^\ell (x) + E_k, \| E_k \|_2 \leq \eta \sigma_2^{2k}, \ k \geq 0;\)
2. \(\exists \gamma > 0 : \| \prod_{k=0}^m J_T^\ell (x^{(k)}) \|_2 \geq \gamma, \ m \geq 0.\)

Then we have

\[
\lim_{k \to \infty} x^{(k)} = x,
\]

and

\[
\limsup_{k \to \infty} \| x^{(k)} - x \|^{1/k} \leq \sigma_2^{2\ell}.
\]

**Proof**

Since \(\sum_{k=0}^\infty \| E_k \|_2 < \infty\) from Theorem 4.1 in [13] we obtain that the matrix sequence \(P_m = \prod_{k=0}^m J_T^\ell (x^{(k)})\) is such that

\[
\lim_{m \to \infty} P_m = x z^T, \quad z \in P.
\]

From Property 2 in view of the continuity of the norm it follows that \(z \neq 0\) and this implies the convergence of \(\{x^{(k)}\}_k\). About the rate of convergence we observe that

\[
\| x^{(k+1)} - x \|_2 = \| P_k x^{(0)} - e^T P_k x^{(0)} \|_2 - x^T z x^{(0)} x^{(0)} \|_2 \\
\leq \| P_k x^{(0)} - e^T P_k x^{(0)} \|_2 + \| P_k x^{(0)} - e^T P_k x^{(0)} \|_2 \\
\leq \| P_k - e^T P_k x^{(0)} \|_2 + \| P_k x^{(0)} \|_2 \| e^T (z x - P_k) x^{(0)} \|_2 \\
\leq \| e^T (z x - P_k) x^{(0)} \|_2 + \| P_k x^{(0)} \|_2 \| e^T P_k x^{(0)} z^T x^{(0)} \|_2.
\]
which says that $x^{(k)}$ approaches $x$ as fast as $P_k$ tends to $xz^T$. Again using Theorem 4.1 in [13] under our assumptions there follows that

$$\limsup_{k \to \infty} \| P_k - xz^T \|^{1/k}_2 \leq \sigma^2_{2} = \lambda^2_{2}.$$

which concludes the proof.

This theorem shows that in our model the speed of convergence increases as $\ell$ increases. Also, notice that the matrix $J_T(z)$, $z \in P_0$, is primitive and irreducible and therefore by the Perron-Frobenius theorem its spectral radius is a dominant eigenvalue with a corresponding positive eigenvector. From Theorem 2.6 this eigenvalue is greater than or equal to 1. These facts suggest to consider SKK$_\infty$ as an effective method for approximating the limit vector $x$. The method performs as an inner-outer procedure. In the inner phase given the current approximation $x^{(k)}$ of $x$ we apply the Power Method

\begin{equation}
\begin{aligned}
\mathbf{v}^{(0)} &= x^{(k)}; \\
\mathbf{z}^{(k+1)} &= J_T(x^{(k)})\mathbf{v}^{(k)}, \\
\mathbf{v}^{(k+1)} &= \mathbf{z}^{(k+1)}/(e^T \mathbf{z}^{(k+1)}), \quad k \geq 0
\end{aligned}
\end{equation}

until convergence to find the new approximation $x^{(k+1)}$. Numerically this latter vector solves

$$J_T(x^{(k)})x^{(k+1)} = \theta_k x^{(k+1)}, \quad \theta_k = \rho(J_T(x^{(k)})).$$

Ideally, $\theta_k$ would converge from above to 1 as well as the corresponding positive eigenvector $x^{(k+1)}$ would approach $x$; also, the convergence of the outer iteration should be superlinear. In the case of clustered eigenvalues the convergence of the inner iteration (power method) can be greatly improved by considering variants based on the Arnoldi method for approximating a few largest eigenvalues of the matrix. In the next section we report the results of numerical experiments confirming and supporting these claims.

## 3. EXPERIMENTAL SETUP

We have tested the algorithms presented above in a numerical environment using MatLab. The first method to be considered is the scaled SKK iteration implemented by Algorithm 1. The Power Method can be used to accelerate Algorithm 1. By complementing Algorithm 1 with Algorithm 2 we have derived the SKK$_\infty$ iterative method, implemented by Algorithm 3. The performance of Algorithm 3 can be improved by approximating a few largest eigenvalues simultaneously based on Arnoldi methods. In particular, it has been noted that the orthogonalization of Arnoldi process achieves effective separation of eigenvectors [17]. The following Algorithm 4 implements the Arnoldi iteration with one additional step of re-orthogonalization. Algorithm 4 can be integrated in the SKK$_\infty$ iteration by obtaining the following ArnoldiSKK iteration. The resulting Algorithm 5 is in the spirit of the Arnoldi-type accelerated power methods proposed in [17, 25].
Algorithm 1 Scaled SKK iteration

Input: \(A \in \mathbb{R}^{n \times n}, A \geq 0\) and a tolerance \(\tau > 0\)
Output: \(x\) such that \(x = Tx, x \geq 0, \sum(x) = 1\)

1: Function SKK\((A, \tau)\)
2: \(x = \text{ones}(n, 1)/n;\)
3: \(err = \infty;\)
4: while \(err > \tau\) do
5: \(z = Tx;\)
6: \(s = \sum(z);\)
7: \(z = z/s;\)
8: \(err = \text{norm}(z - x);\)
9: \(x = z;\)
10: EndFunction

Algorithm 2 Power Method

Input: \(A \in \mathbb{R}^{n \times n}, A \geq 0\), a tolerance \(\tau > 0\) and an initial guess \(x\)
Output: \((\lambda, x)\) such that \(\lambda x = Ax, \lambda = \rho(A), x \geq 0, \sum(x) = 1\)

1: Function PowerMethod\((A, \tau, x)\)
2: \(err = \infty;\)
3: while \(err > \tau\) do
4: \(z = Ax;\)
5: \(\lambda = \sum(z);\)
6: \(z = z/\lambda;\)
7: \(err = \text{norm}(z - x);\)
8: \(x = z;\)
9: EndFunction

Algorithm 3 SKK\(_\infty\) iterative method

Input: \(A \in \mathbb{R}^{n \times n}, A \geq 0\) and given tolerances \(\tau_1, \tau_2 > 0\)
Output: \(x\) such that \(x = Tx, x \geq 0, \sum(x) = 1\)

1: Function PowerSKK\((A, \tau_1, \tau_2)\)
2: \(x = \text{ones}(n, 1)/n;\)
3: \(err = \infty;\)
4: while \(err > \tau_2\) do
5: \([\lambda, z] = \text{PowerMethod}(J_T(x), \tau_1, x);\)
6: \(err = \text{norm}(Tz - x);\)
7: \(x = z;\)
8: EndFunction

The first set of test problems consists of matrices of the form

\[
A = \begin{bmatrix}
A_{1,1} & A_{1,2} \\
0 & A_{2,2}
\end{bmatrix} + kE, \quad A_{i,j} \in \mathbb{R}^{n \times n}, E \in \mathbb{R}^{2n \times 2n}.
\]

where \(k\) is a small perturbative parameter. The reduction of the adjacency matrix of a graph in a block triangular form is related with the Dulmage-Mendelsohn decomposition [26], which is a canonical decomposition of a bipartite graph based on the notion of matching. The SKK iteration applied to block triangular matrices can not converge. According to [6] we consider the perturbed matrix \(A\) for computing an approximate balancing of the initial block triangular component. In table...
Algorithm 4 Arnoldi Method

**Input:** $A \in \mathbb{R}^{n \times n}$, $b \in \mathbb{R}^n$, the number $m$ of eigenvalues to be approximated and a threshold value $tol$ for re-orthogonalization

**Output:** $H \in \mathbb{R}^{(m+1) \times m}$ Hessenberg and $Q \in \mathbb{R}^{n \times (m+1)}$ with orthonormal columns such that $H \approx Q^T AQ$

1: **Function** Arnoldi$(A, b, m, tol)$
2: $H = \text{zeros}(m + 1, m);$  
3: $q_1 = b / \|b\|_2;$  
4: for $k = 1, \ldots, m$ do
5: \hspace{1em} $v = Aq_k;$  
6: \hspace{1em} $nn = \text{norm}(v);$  
7: \hspace{1em} for $j = 1, \ldots, k$ do
8: \hspace{2em} $h_{j,k} = q_j^Tv;$  
9: \hspace{2em} $v = v - h_{j,k}q_j;$  
10: \hspace{1em} $h_{k+1,k} = \|v\|_2;$  
11: \hspace{1em} if $h_{k+1,k} \leq tol \cdot nn$ then
12: \hspace{2em} for $j = 1, \ldots, k$ do
13: \hspace{3em} $\gamma = q_j^Tv;$  
14: \hspace{3em} $v = v - \gamma q_j;$  
15: \hspace{3em} $h_{j,k} = h_{j,k} + \gamma;$  
16: \hspace{2em} $h_{k+1,k} = \|v\|_2;$  
17: \hspace{1em} $q_{k+1} = v/h_{k+1,k};$
18: **EndFunction**

Algorithm 5 SKK\(_\infty\) method with Arnoldi iteration

**Input:** $A \in \mathbb{R}^{n \times n}$, $A \geq 0$, a value of $m$ and given tolerances $\tau_1, \tau_2, \tau_3 > 0$

**Output:** $x$ such that $x = Tx$, $x \geq 0$, $\text{sum}(x) = 1$

1: **Function** ArnoldiSKK$(A, \tau_1, \tau_2, \tau_3, m)$
2: $x = z = \text{ones}(n, 1)/n;$  
3: $err = in f;$  
4: while $err > \tau_3$ do
5: \hspace{1em} $errl = in f;$  
6: \hspace{1em} while $errl > \tau_2$ do
7: \hspace{2em} $[H, Q] = \text{Arnoldi}(J_T(x), z, m, \tau_1);$  
8: \hspace{2em} $[y, d] = \text{eigs}(H(1:m, 1:m), 1);$  
9: \hspace{2em} $z = Q(:, 1:m)y;$  
10: \hspace{2em} $errl = \text{norm}(J_T(x)z - dz);$  
11: \hspace{1em} $s = \text{sum}(z);$  
12: \hspace{1em} $z = z/s;$  
13: \hspace{1em} $err = \text{norm}(Tz - z);$  
14: \hspace{1em} $x = z;$
15: **EndFunction**

II we show the overall number of iterations of Algorithm 1 and 5 for different values of $\lambda$. The blocks $A_{i,j}$ of size 1000 are generated by $A_{i,j} = \text{sprand}(1000, 1000, 0.1)$ whereas $E = \text{ones}(2000)$. The input parameters are set as follows: $\tau = \tau_3 = 1.0e - 12$, $\tau_1 = 0.1$, $\tau_2 = 1.0e - 4$, $m = 6$. The slow convergence of the SKK iteration depends on the fact that the matrix $A$ has two eigenvalues clustered around 1. The Arnoldi method makes possible a quick resolution of the dominant eigenvalue by approximating a few $m = 6$ largest eigenvalues simultaneously. In the next Figure 1 we illustrate
Table II. Number of iterations of Algorithm 1 and 5 for different values of $k$

| $k$     | 1.0e-1 | 1.0e-2 | 1.0e-3 | 1.0e-4 | 1.0e-5 | 1.0e-6 | 1.0e-7 | 1.0e-8 | 1.0e-9 |
|---------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| Alg1    | 6      | 14     | 41     | 120    | 359    | 1073   | 3197   | 9498   | 28102  |
| Alg5    | 2      | 3      | 4      | 5      | 6      | 8      | 9      | 10     | 12     |

Figure 1. Error plots generated in Algorithm 5

the convergence of the Arnoldi-based iteration for different values of $k$. The plot show the errors $err$ evaluated in the outer iterations. The slope of these plots supports the claim concerning the ultimate superlinear convergence of the resulting iterative schemes.

However, when we increase the numbers of diagonal blocks in the matrix $A$ we observe experimentally a progressive deterioration of the performance of Algorithm 5 which is not suited to deal with a large cluster of eigenvalues around 1. In that case a block variant of the Arnoldi iteration might be used. To experiment with this we consider the adjacency matrix $A \in \mathbb{R}^{198 \times 198}$ constructed from a collaboration network between Jazz musicians. Each node is a Jazz musician and an edge denotes that two musicians have played together in a band. The data was collected in 2003 [27]. The plot of the matrix is shown in Figure 2. Algorithm 1 applied to the matrix $A$ does not seem to converge. By perturbing $A = A + 1.0e - 12 \text{ones}(198)$ we find that the SKK iteration with tolerance $\tau = 1.0e - 12$ converges in 72324 iterations. The resulting balanced matrix exhibits a cluster of about eleven eigenvalues around 1. For comparison we consider Algorithm 5 where we replace the call to the Arnoldi function with one call to the function $\text{ahbeigs}$ [28] which implements a block Arnoldi method for computing a few eigenvalues of sparse matrices. By selecting $m = 16$ and $\tau_3 = 1.0e - 12$ we find that the modified Algorithm 5 converges in 14 iterations by performing an overall number of about $14 \times 100$ matrix products vs. the 72324 employed by the SKK iteration. In Figure 3 we show the error behavior as well as the singular values of the balanced matrix (compare with Theorem 2.3). The MatLab command $\text{dmperm}$ applied to the jazz matrix shown in Figure 2
computes its Dulmage-Mendelsohn decomposition. It is found that the permuted matrix illustrated in Figure 4 is block triangular with 11 diagonal blocks. It is worth stressing again the accordance of the number of blocks in the permuted matrix with the size of the cluster of singular values of the balanced permuted matrix (compare with Figure 3).

4. CONCLUSIONS AND FUTURE WORK

In this paper we have discussed some numerical techniques for accelerating the customary SKK iteration based on certain equivalent formulations of the fixed point problem as a matrix eigenvalue
problem. Variants of the power method have been proposed for the efficient solution of the matrix eigenvalue problem. There are several topics which remain to be addressed. Specifically:

1. A formal proof of the convergence for the SKK\(_\infty\) method is still missing.

2. Theoretical and computational results should be extended to nonnegative matrices under customary assumptions on their supports.

3. The efficiency of the balancing schemes depends on eigenvalue (singular value) clustering properties. Numerical experiments have revealed a close connection between the clustering of singular values of the balanced adjacency matrix and the clustering of nodes (community detection) in the corresponding graph. Relations with the block triangular form (BTF) form of adjacency matrices have also appeared. The possible use of balancing schemes for detecting the block structure of adjacency matrices is an ongoing research work.

4. Finally, we plan to study the numerical behavior of block Arnoldi based methods by performing extensive numerical experiments with large sparse and data-sparse matrices.

References

1. Sinkhorn R. A relationship between arbitrary positive matrices and doubly stochastic matrices. *Ann. Math. Statist.* 1964; 35:876–879, doi:10.1214/aoms/1177703591. URL https://doi.org/10.1214/aoms/1177703591.
14

2. Sinkhorn R, Knopp P. Concerning nonnegative matrices and doubly stochastic matrices. Pacific J. Math. 1967; 21:343–348. URL: http://projecteuclid.org/euclid.pjm/1102992505.

3. Sinkhorn R. Diagonal equivalence to matrices with prescribed row and column sums. Amer. Math. Monthly 1967; 74:402–405. doi:10.2307/2314570. URL: https://doi.org/10.2307/2314570.

4. Parikh A. Forecasts of input-output matrices using the r.a.s. method. The Review of Economics and Statistics 1979; 61(3):477–481. URL: http://www.jstor.org/stable/1926084.

5. Cuturi M. Sinkhorn distances: Lightspeed computation of optimal transport. Proceedings of the 26th International Conference on Neural Information Processing Systems - Volume 2, NIPS’13, Curran Associates Inc.: USA, 2013; 2292–2300. URL: http://dl.acm.org/citation.cfm?id=2999792.2999868.

6. Knight PA. The Sinkhorn-Knopp algorithm: convergence and applications. SIAM J. Matrix Anal. Appl. 2008; 30(1):261–275. doi:10.1137/060659624. URL: https://doi.org/10.1137/060659624.

7. Bozzo E, Franceschet M. A theory on power in networks. Commun. ACM Oct 2016; 59(11):75–83. doi:10.1145/2934665. URL: http://doi.acm.org/10.1145/2934665.

8. Rüscher L. Convergence of the iterative proportional fitting procedure. Ann. Statist. 1995; 23(4):1160–1174, doi:10.1214/aos/1176324703. URL: https://doi.org/10.1214/aos/1176324703.

9. Lamon B, Stewart NF. Bregman’s balancing method. Transportation Res. Part B 1981; 15(4):239–248, doi:10.1016/0191-2615(81)90010-2. URL: http://www.jstor.org/stable/1909285.

10. Diamond S, Boyd S. Near-decomposability, partition and aggregation, and the relevance of stability discussions. Econometrica 1961; 29(2):111–138. URL: http://www.jstor.org/stable/1909285.

11. Lemmens B, Nussbaum R. Nonlinear Perron-Frobenius theory, Cambridge Tracts in Mathematics, vol. 189. Cambridge University Press, Cambridge, 2012, doi:10.1017/CBO9781139026079. URL: http://www.jstor.org/stable/1909285.