The power method, Perron-Frobenius theory, Page-Rank computation

Preliminary considerations, $A \in \mathbb{C}^{n \times n}$

Given a $n \times n$ matrix $A$ with eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$, the following result holds.

Theorem (inverse power). If $\lambda_i^*$ is an approximation of the eigenvalue $\lambda_i$ of a $n \times n$ matrix $A$, i.e. $|\lambda_i - \lambda_i^*|$ is smaller than $|\lambda_j - \lambda_i^*|$, $\forall \lambda_j \neq \lambda_i$, if $m_a(\lambda_i) = m_g(\lambda_i)$, and if $v_0 \in \mathbb{C}^n$ is not orthogonal to the space spanned by the eigenvectors of $A$ corresponding to $\lambda_i$, then the sequence $\{v_k\} \text{ generated by the algorithm:}$

$$(A - \lambda_i^* I) a_k = v_{k-1}, \quad v_k = \frac{a_k}{\|a_k\|}, \quad k = 1, 2, \ldots$$

converges to an eigenvector of $A$ corresponding to the eigenvalue $\lambda_i$. If $A$ is diagonalizable, then the rate of convergence is $O((\max_{j: \lambda_j \neq \lambda_i} |\frac{\lambda_j - \lambda_i}{\lambda_j - \lambda_i^*}|)^k)$.

Remark. In the general case the rate of convergence is

$$\max_{j: \lambda_j \neq \lambda_i} \max_{s_{\lambda_j}} O((p_{s_{\lambda_j}} - 1(k))\|\frac{\lambda_j - \lambda_i}{\lambda_j - \lambda_i^*}\|^k),$$

where, given the block diagonal Jordan form of $A$, $J = X^{-1}AX$, for each $\lambda_j \neq \lambda_i$, the number $s_{\lambda_j}$ indicates the dimension of the generic Jordan block associated with $\lambda_j$, and $p_{s_{\lambda_j}} - 1(k)$ is a polynomial of degree $s_{\lambda_j}$, whose coefficients depend on $\frac{1}{\lambda_j - \lambda_i^*}$, $r = 1, \ldots, s_{\lambda_j} - 1$, and on the coefficients of $v_0$ with respect to the column vectors of $X$ corresponding to the Jordan block under consideration.

proof: See the Appendix. $\square$

Let $\lambda_1$ be such that $|\lambda_1| = \rho(A)$ and assume that all $\lambda_i$ such that $|\lambda_i| = \rho(A)$ are equal to $\lambda_1$ (in such case we say that $\lambda_1$ dominates the eigenvalues of $A$).

Assume, moreover, that the algebraic and geometric multiplicity of $\lambda_1$ are equal. Then the power method (see the Theorem below) can be used to compute $\lambda_1$ and an eigenvector corresponding to $\lambda_1$.

Theorem (power). If $\lambda_1$ dominates the eigenvalues of a $n \times n$ matrix $A$, if $m_a(\lambda_1) = m_g(\lambda_1)$, and if $v_0 \in \mathbb{C}^n$ is not orthogonal to the space spanned by the eigenvectors of $A$ corresponding to $\lambda_1$, then the sequence $\{v_k\} \text{ generated by the algorithm:}$

$$a_k = Av_{k-1}, \quad v_k = \frac{a_k}{\|a_k\|}, \quad k = 1, 2, \ldots$$

converges to an eigenvector of $A$ corresponding to the eigenvalue $\lambda_1$. Moreover,

$$\frac{u^Hv_{k+1}}{u^Hv_k} \rightarrow \lambda_1, \quad k \rightarrow +\infty$$

for any $u$ for which $u^Hv_k \neq 0$. If $A$ is diagonalizable, then the rate of convergence is $O((\max_{j: \lambda_j \neq \lambda_1} |\frac{\lambda_j - \lambda_1}{\lambda_j - \lambda_1^*}|)^k)$.

Remark. In the general case the rate of convergence is

$$\max_{j: \lambda_j \neq \lambda_1} \max_{s_{\lambda_j}} O((p_{s_{\lambda_j}} - 1(k))\|\frac{\lambda_j - \lambda_1}{\lambda_j - \lambda_1^*}\|^k),$$

where, given the block diagonal Jordan form of $A$, $J = X^{-1}AX$, for each $\lambda_j \neq \lambda_1$, the number $s_{\lambda_j}$ indicates the dimension of the generic Jordan block associated with $\lambda_j$, and $p_{s_{\lambda_j}} - 1(k)$ is...
a polynomial of degree \(s_{j} - 1\), whose coefficients depend on \(\frac{1}{\lambda_{j}}\), \(r = 1, \ldots, s_{j} - 1\), and on the coefficients of \(v_{0}\) with respect to the column vectors of \(X\) corresponding to the Jordan block under consideration.

proof: See the Appendix. □

For our purposes it is useful to recall also the following classic result on matrix deflation: how to introduce a matrix whose eigenvalues are all equal to the eigenvalues of \(A\) except one, which, instead, is zero.

**Theorem.** Let \(A\) be a \(n \times n\) matrix. Let \(\lambda_{1}\) be a nonzero eigenvalue of \(A\) and \(y_{1}\) a corresponding eigenvector, i.e. \(Ay_{1} = \lambda_{1}y_{1}\). Call \(\lambda_{2}, \lambda_{3}, \ldots, \lambda_{n}\) the remaining eigenvalues of \(A\). Then the matrix \(W = A - \frac{\lambda_{1}}{w^*y_{1}}y_{1}w^*\), \(w^*y_{1} \neq 0\), has eigenvalues \(0, \lambda_{2}, \lambda_{3}, \ldots, \lambda_{n}\).

proof. Introduce \(S = [y_{1} z_{2} \cdots z_{n}]\) non singular, and observe that \(p_{A}(\lambda) = p_{S^{-1}AS}(\lambda) = (\lambda - \lambda_{1})q(\lambda), pw(\lambda) = p_{S^{-1}WS}(\lambda) = \lambda q(\lambda)\). □

Let \(G\) be the following matrix

\[
G = \begin{bmatrix}
\frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\
\frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\
\frac{1}{16} & \frac{1}{16} & \frac{1}{16}
\end{bmatrix}.
\]

Since \(\rho(G) \leq \|G\|_{\infty} = 1\), we can say that the spectrum of \(G\) lies in the circle \(\{z \in \mathbb{C} : |z| \leq 1\}\).

Note that \(Ge = 1 \cdot e\), i.e. one eigenvalue of \(G\), \(\lambda_{1} = 1\), and its corresponding eigenvector, \(y_{1} = e = [1 \ 1 \ \cdots \ 1]^T\), are known. If \(\lambda_{2}, \lambda_{3}\) denote the remaining eigenvalues of \(G\), then we can define a matrix \(W\), in terms of \(G, \lambda_{1}, y_{1}\), whose eigenvalues are \(0, \lambda_{2}, \lambda_{3}\):

\[
W = G - \frac{\lambda_{1}}{w^*y_{1}}y_{1}w^* = G - \frac{1}{w^*e} \begin{bmatrix} w^* \\ w^* \\ w^* \end{bmatrix}, \quad \forall w, \ w^*e \neq 0.
\]

Since \((e_{i}^{T}G)e = e_{i}^{T}(1 \cdot e) = 1 \neq 0\), we choose \(w^* = e_{i}^{T}G\):

\[
W = G - \begin{bmatrix} e_{i}^{T}G \\ e_{i}^{T}G \\ e_{i}^{T}G \end{bmatrix}.
\]

For \(i = 1\) the matrix \(W\) becomes

\[
W = \begin{bmatrix}
\frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\
\frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\
\frac{1}{16} & \frac{1}{16} & \frac{1}{16}
\end{bmatrix} - \begin{bmatrix}
\frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\
\frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\
\frac{1}{16} & \frac{1}{16} & \frac{1}{16}
\end{bmatrix} = \begin{bmatrix}
0 & 0 & 0 \\
0 & -\frac{3}{8} & -\frac{3}{8} \\
\frac{1}{16} & 0 & -\frac{3}{8}
\end{bmatrix},
\]

so, the eigenvalues of \(A\) different from 1 are the eigenvalues of

\[
\tilde{W} = \begin{bmatrix}
-\frac{3}{8} & -\frac{3}{8} \\
0 & -\frac{3}{8}
\end{bmatrix},
\]

i.e. \(-\frac{3}{8}\) and \(-\frac{3}{8}\).

Thus, \(1, -\frac{3}{8}\) and \(-\frac{3}{8}\) are the eigenvalues of \(G\), and also, of course, of \(G^{T}\). In particular, 1 is eigenvalue of \(G^{T}\), but note that the eigenvector \(p\) of \(G^{T}\)
corresponding to 1 is not obvious; it must be computed. For example, it can be computed as the limit of the inverse power sequence \( \{v_k\} \) defined as follows (\( \varepsilon \) is a small positive number):

\[
v_0 \in \mathbb{R}^3, \ (G^T - (1 + \varepsilon)I) a_k = v_{k-1}, \quad v_k = \frac{a_k}{\|a_k\|}, \quad k = 0, 1, 2, \ldots
\]

(rate of convergence: \( O\left(\left|\frac{1-\beta}{1-\beta}\right|^k\right)\)). [Here a reference for the inverse power iterations]. We shall see that the vector \( p \) can be also obtained as the limit of the sequence

\[
p_0 \in \mathbb{R}^3, \ p_0 positive, \quad \|p_0\|_1 = 1, \quad p_{k+1} = G^T p_k, \quad k = 0, 1, 2, \ldots
\]

(rate of convergence: \( O\left(\left|\frac{1}{1-\beta}\right|^k\right)\)) [this result is in fact a particular case of the Perron-Frobenius theory: \( A \in \mathbb{R}^{n \times n}, \ A \geq 0 \) irreducible [Varga]].

**Lemma.** Let \( A \) be a \( n \times n \) non negative matrix, i.e. \( a_{ij} \geq 0, \forall i, j \). Assume that \( A \) is not reducible. Then \( (I + A)^{n-1} \) is a positive matrix, i.e. its entries are all positive.

**Proof.** We shall prove that the vector \( (I + A)^{n-1} x \) is positive whenever \( x \) is a non negative non null vector (prove that this is equivalent to the thesis!).

Let \( x \) be a non negative non null vector. Set \( x_0 = x, x_1 = (I + A)x_0 = x_0 + Ax_0, \ldots, x_{k+1} = (I + A)x_k = x_k + Ax_k, \quad k = 1, \ldots, n - 2. \) Note that \( x_k = (I + A)^k x, \) in particular \( x_{n-1} = (I + A)^{n-1} x. \) So, our aim is to prove that \( x_{n-1} \) is a positive vector. First observe by induction that all \( x_k \) are non negative vectors (\( x_k \) non negative and \( A \) non negative imply \( Ax_k \) non negative and \( x_{k+1} = x_k + Ax_k \) non negative). Then the thesis \( x_{n-1} \) positive is now proved by showing that \( x_{k+1} \) must have less zeros than \( x_k \) for each \( k \in \{0, \ldots, n - 2\}. \) Note that \( x_{k+1} \) cannot have more zeros than \( x_k \), since \( Ax_k \), in the definition \( x_k + Ax_k \) of \( x_{k+1} \), is non negative. Assume \( x_{k+1} \) has the same number of zero entries as \( x_k \). Of course such zeros must be in the same places; i.e. there exists a permutation matrix \( P \) such that \( Px_k = \begin{bmatrix} \alpha \\ 0 \end{bmatrix}, \quad Px_{k+1} = \begin{bmatrix} \beta \\ 0 \end{bmatrix}, \) with \( \alpha, \beta \) positive vectors of the same dimension \( m, \) \( 1 \leq m \leq n - 1 \) (why such bounds for \( m ? \)). Thus, \( Px_{k+1} = Px_k + PAX_k = PX_k + PAP^T \). Consider the following partition of the matrix \( PAP^T \)

\[
PAP^T = \begin{bmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{bmatrix}
\]

where \( M_{11} \) is \( m \times m \). Then

\[
\begin{bmatrix} \beta \\ 0 \end{bmatrix} = \begin{bmatrix} \alpha \\ 0 \end{bmatrix} + \begin{bmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{bmatrix} \begin{bmatrix} \alpha \\ 0 \end{bmatrix},
\]

3
and, in particular, \(M_{21}\alpha = 0\). The latter condition implies \(M_{21} = 0\), being \(\alpha\) a positive vector and \(M_{21}\) a non negative matrix. But this is equivalent to say that \(A\) is reducible, against the hypothesis! \(\square\)

**Example.** Set

\[
I + A = \begin{bmatrix}
1 & 0 & 0 & 1 \\
0 & 0 & b & 0 \\
0 & 0 & 0 & 1 \\
\end{bmatrix}, \quad a, b, c \text{ positive.}
\]

Note that \(A\) is a non-negative irreducible matrix, and in fact \((I + A)^3\) is a positive matrix, i.e. its entries are positive. Moreover, \(3\) is the minimum \(j\) for which \((I + A)^j\) is positive.

Let \(A\) be a \(n \times n\) non negative irreducible matrix. Let \(x\) be a non negative non null vector, and associate to \(x\) the number

\[
r_x := \min_{i: x_i > 0} \frac{\sum_j a_{ij} x_j}{x_i} = \min_{i: x_i > 0} \frac{(Ax)_i}{x_i}.
\]

**Proposition.** \(r_x\) is a non negative real number; \(r_x = r_{\alpha x}\) if \(\alpha > 0\); \(Ax \geq r_x x\);

\[
r_x = \sup\{\rho \in \mathbb{R} : Ax \geq \rho x\}.
\]

proof: easy, left to the reader.

Now associate to \(A\) the following number:

\[
r = \sup_{x \geq 0: x \neq 0} r_x = \sup_{x \geq 0: x \neq 0} \min_{i: x_i > 0} \frac{\sum_j a_{ij} x_j}{x_i}.
\]

**Proposition.** \(r\) is a positive real number; if \(w \geq 0\), \(w \neq 0\), is such that \(Aw \geq rw\), then \(Aw = rw\) and \(w > 0\).

proof. If \(e = [1 \cdots 1]^T\), then \(re = \min_{i: (e)_i > 0} \frac{\sum_j a_{ij} (e)_j}{(e)_i} = \min_i \sum_j a_{ij} \geq 0\).

Assume \(r_e = 0\). Then for some \(k\) we would have \(\sum_j a_{kj} = 0\), so the \(k\)th row of \(A\) would be null, and thus \(A\) would be reducible (exchange the \(k\) and \(n\) rows), against the hypothesis. It follows that \(r \geq r_e > 0\).

proof. Set \(\eta = Aw - rw\). We know that \(\eta \geq 0\). Assume \(\eta \neq 0\). Then, by the Lemma,

\[
0 < (I + A)^{n-1} \eta = (I + A)^{n-1} Aw - (I + A)^{n-1} rw = A(I + A)^{n-1} w - r(I + A)^{n-1} w = Ay - ry, \quad y > 0.
\]

i.e. \(r < (Ay)_i/y_i \forall i\). Thus \(r < r_y\), which is absurd. It follows that \(\eta = 0\), that is, \(Aw = rw\). Then, we also have \(w > 0\) since \(0 < (I + A)^{n-1} w = (1 + r)^{n-1} w\). \(\square\)

In the following, given \(v \in \mathbb{C}^n\) and \(M \in \mathbb{C}^{n \times n}\) we denote by \(|v|\) and \(|M|\), respectively, the column vector \((|v_k|)_{k=1}^n\) and the matrix \((|m_{ij}|)_{i,j=1}^n\).

**Theorem.** There exists a positive vector \(z\) for which \(Az = rz\); \(r = \rho(A)\); if \(B \in \mathbb{C}^{n \times n}\), \(|B| \leq A\), then \(\rho(B) \leq \rho(A)\); if \(B \in \mathbb{C}^{n \times n}\), \(|B| \leq A\), \(|B| \neq A\) then \(\rho(B) < \rho(A)\).
We now show that there exists \( z \geq 0 \) such that \( r = r_z \). Once this is proved we will have the inequality \( Az \geq r_z z = rz \) which implies, by the Proposition, \( Az = rz \) and \( z > 0 \).

Let \( A \) be an irreducible non-negative matrix. Then we know that its measure in the 1-norm is one. We also know that \( A \) is positive and is a simple eigenvalue of \( A \). Thus, \( \rho(A) \) is an eigenvalue of \( A \) such that \( \rho(A) < \lambda \). Then we can add an equality in the above arguments, \( r|y| = |\lambda||y| = |\lambda y| = |By| \leq |B||y| \leq A|y|, |y| \geq 0, |y| \neq 0 \), obtaining the inequality \( r|y| \leq A|y|, |y| \geq 0, |y| \neq 0 \). But by the above Proposition, such inequality implies \( A|y| = r|y| \) with \( |y| > 0 \). So we have

\[
r|y| = |\lambda||y| = |\lambda y| = |By| = |B||y| = A|y| = r|y|, |y| > 0,
\]

from which it follows \( |B| = A \), against the hypothesis! Thus \( \rho(B) < \rho(A) \).

We conclude this section with the following result of the Perron-Frobenius theory, stated without proof (actually we shall give a proof of such result in the case \( A \) is positive):

**Result.** If \( A \) is a non-negative irreducible \( n \times n \) matrix, then \( r = \rho(A) \) is a simple eigenvalue of \( A \), i.e. \( \lambda - r \) divides \( p_A(\lambda) \) but \( (\lambda - r)^2 \) does not divide \( p_A(\lambda) \).

**Computing the Perron-pair of \( A \) irreducible non-negative by the power method**

Let \( A \) be an irreducible non-negative \( n \times n \) matrix. Then we know that \( \rho(A) \) is positive and is a simple eigenvalue of \( A \), and the corresponding eigenvector can be chosen positive. Of course, such eigenvector is uniquely defined if we require that its measure in the 1-norm is one. We also know that \( \rho(A) = r := \sup_{x \geq 0, x \neq 0} \frac{\langle Ax, x \rangle}{\|x\|_2} \).

So, it is uniquely defined the Perron-pair \( (\rho(A), z) \), such that \( Az = \rho(A)z \), \( \rho(A) \) positive, \( z \) positive, \( \|z\|_1 = 1 \). The power method is a way to compute such Perron-pair.

**Theorem (power).** Let \( A \) be an irreducible non-negative \( n \times n \) matrix. Let \( a_0 \) be any positive vector, and set \( v_0 = a_0/\|a_0\|_1 \). Then set

\[
a_{k+1} = Av_k, \quad v_{k+1} = \frac{a_{k+1}}{\|a_{k+1}\|_1}, \quad k = 0, 1, 2, \ldots.
\]
Note that the sequences \( \{a_k\} \), \( \{v_k\} \) are well defined sequences of positive vectors, and \( \|v_k\|_1 = 1, \forall k \). Let \( X \) be the non singular matrix defining, by similarity, the Jordan block-diagonal form of \( A \), i.e.

\[
X^{-1}AX = J = \begin{bmatrix} r & 0^T \\ 0 & B \end{bmatrix}, \quad X = \begin{bmatrix} z & x_2 & \cdots & x_n \end{bmatrix},
\]

and let \( r = \rho(A), \lambda_j, j = 2, \ldots, n \), be the eigenvalues of \( A \) (\( \lambda_j \neq r, \forall j \)). If \( \alpha \) in the expression \( a_0 = \alpha z + \sum_{j=2}^{n} \alpha_j x_j \) is nonzero, then, for \( k \to +\infty \), we have

\[
v_k - z \to 0, \quad \|a_k\|_1 - \rho(A) \to 0,
\]

provided that \( |\lambda_j| \) is smaller than \( r \) for all \( j = 2, \ldots, n \). In the particular case where \( A \) is diagonalizable, the rate of convergence is

\[
\left( \max_{j=2}^{n} \frac{|\lambda_j|}{r} \right)^k
\]

[for the general case, use the Remark in Theorem(power) of the section on preliminary considerations].

**proof:** We prove the Theorem only in the case \( A \) diagonalizable, where \( Ax_j = \lambda_j x_j, j = 2, \ldots, n \). It is easy to observe that \( A^k a_0 = \alpha^k z + \sum_{j=2}^{n} \alpha_j \lambda_j^k x_j \) is a positive vector, and that

\[
v_k = \frac{\alpha^k z + \sum_{j=2}^{n} \alpha_j \lambda_j^k x_j}{\|\alpha^k z + \sum_{j=2}^{n} \alpha_j \lambda_j^k x_j\|_1} = \frac{\alpha^k z + \sum_{j=2}^{n} \alpha_j \lambda_j^k x_j}{e^T (\alpha^k z + \sum_{j=2}^{n} \alpha_j \lambda_j^k x_j)} = \frac{z + \sum_{j=2}^{n} \frac{\alpha_j}{r} \left( \frac{\lambda_j}{r} \right)^k x_j}{1 + \sum_{j=2}^{n} \frac{\alpha_j}{r} \left( \frac{\lambda_j}{r} \right)^k e^T x_j}.
\]

Moreover,

\[
a_{k+1} = A v_k = \frac{r z + \sum_{j=2}^{n} \frac{\alpha_j}{r} \left( \frac{\lambda_j}{r} \right)^k \lambda_j x_j}{1 + \sum_{j=2}^{n} \frac{\alpha_j}{r} \left( \frac{\lambda_j}{r} \right)^k e^T x_j}
\]

and, since \( a_{k+1} \) is positive,

\[
\|a_{k+1}\|_1 = e^T a_{k+1} = \frac{r + \sum_{j=2}^{n} \frac{\alpha_j}{r} \left( \frac{\lambda_j}{r} \right)^k \lambda_j e^T x_j}{1 + \sum_{j=2}^{n} \frac{\alpha_j}{r} \left( \frac{\lambda_j}{r} \right)^k e^T x_j} = \frac{r (1 + \sum_{j=2}^{n} \frac{\alpha_j}{r} \left( \frac{\lambda_j}{r} \right)^k e^T x_j)}{1 + \sum_{j=2}^{n} \frac{\alpha_j}{r} \left( \frac{\lambda_j}{r} \right)^k e^T x_j}.
\]

**Exercise.** Discuss the convergence of the power method when applied to compute the Perron-pair \((1, \begin{bmatrix} 1 \frac{1}{2} \end{bmatrix})\) of the following two \(2 \times 2\) matrices:

\[
A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad A = \begin{bmatrix} 1/4 & 3/4 \\ 3/4 & 1/4 \end{bmatrix}.
\]
In the second case, find a constant $c$ such that $\|v_k - z\| \leq c(\frac{1}{2})^k$.

**Exercise.** Prove the Theorem in case $A$ (non negative, irreducible) is $4 \times 4$, and there exists $X$ non singular for which

$$X^{-1}AX = \begin{bmatrix} r & \lambda_2 & 1 \\ \lambda_2 & 1 & \lambda_2 \\ \lambda_2 & \lambda_2 & 1 \end{bmatrix},$$

with $|\lambda_2|$ smaller than $r$. Prove that in such case the rate of convergence is

$$|p_2(k)|\left(\frac{|\lambda_2|}{r}\right)^k,$$

where $p_2$ is a degree-two polynomial.

**Exercise.** Set

$$A = \begin{bmatrix} 0 & 0 & 1 \\ \frac{1}{3} & 0 & 0 \\ -\frac{1}{4} & 1 & 0 \end{bmatrix}.$$

Prove that the eigenvalues of $A$ are $\{1, -\frac{1}{2}, -\frac{1}{2}\}$ and that $A$ is not diagonalizable. Find $X = [z \; x_2 \; x_3]$ such that

$$X^{-1}AX = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -\frac{1}{2} & 1 \\ 0 & 0 & -\frac{1}{2} \end{bmatrix}.$$

The only hypothesis $A$ irreducible, non-negative, does not assure that $r = \rho(A)$ is the unique eigenvalue of $A$ whose absolute value is equal to $r$. We only know that if $|\lambda_j| = r$, $j \in \{2, \ldots, n\}$, then $\lambda_j \neq r$. Let us see examples:

$$A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}; \text{ Eigenvalues: } -1, 1; \text{ Perron-pair: } (1, \begin{bmatrix} 1 \frac{1}{2} \end{bmatrix}).$$

$$A = \begin{bmatrix} 0 & a \\ 1 & 0 \\ 0 & 1-a \end{bmatrix}, a \in (0,1); \text{ Eig: } -1, 0, 1; \text{ Perron-pair: } (1, \begin{bmatrix} \frac{a}{1-a} \frac{1}{2} \frac{1-2a}{2} \end{bmatrix}).$$

$$A = \begin{bmatrix} 2 & 1 \\ 1 & 2 \\ 1 & 1 \end{bmatrix}; \text{ Eigenvalues: } 1, 1, 4; \text{ Perron-pair: } (4, \begin{bmatrix} 1/3 \\ 1/3 \\ 1/3 \end{bmatrix}).$$

Note that if in the second example $A$ is replaced by $A^T$, then there is no need of computation in obtaining the perron-pair, since it is clear that $A^T e = e$, $e = [1 \; 1 \; 1]^T$. Moreover, in the third example $r = \rho(A)$ (which is 4) dominates the remaining eigenvalues of $A$ (which are 1, 1). We note that this fact is true for any positive matrix $A$ (see Theorem(positive) below), even if, as the following example shows, it is not a peculiarity of positive matrices:

$$A = \begin{bmatrix} 2 & 1 \\ 1 & 2 \\ 0 & 1 \end{bmatrix}; \text{ Eig: } 2 \mp \sqrt{2}, 2; \text{ Perron-pair: } (2 + \sqrt{2}, \begin{bmatrix} 1/2 \\ \sqrt{2} \\ 1 \end{bmatrix}).$$
(here computation is required to obtain the Perron-pair).

**Theorem (positive).** If $A$ is a $n \times n$ positive matrix and $r = \rho(A)$, $\lambda_2, \ldots, \lambda_n$ are its eigenvalues, then $|\lambda_j|$ is smaller than $r$ for all $j = 2, \ldots, n$.

**proof:** Set $W = A - \text{size}^T$. The eigenvalues of $W$ are $r - s, \lambda_2, \ldots, \lambda_n$ (a sketch of the proof):

$$Y := \begin{bmatrix} z & y_2 & \cdots & y_n \end{bmatrix} \text{ non singular},$$

$$Y^{-1}AY = \begin{bmatrix} r & \cdots & 0 \\ 0 & \cdots & M \end{bmatrix}, \quad Y^{-1}WY = \begin{bmatrix} r - s & \cdots & 0 \\ 0 & \cdots & M \end{bmatrix}.$$ 

If there is a value of $s$ for which $|W| \leq A$, $|W| \neq A$, then $\rho(W)$ is smaller than $\rho(A)$, and thus $|\lambda_2|, \ldots, |\lambda_n|$ are smaller than $r = \rho(A)$. If $A$ is positive, then such $s$ exists, $s = \min_{i,j} a_{ij}$. \qed

**Irreducible non-negative stochastic-by-columns $A$ and power method**

Let $A$ be an irreducible non-negative stochastic-by-columns matrix $n \times n$ matrix. Then we know that $1 = \rho(A)$ ($A^T e = e$, $\rho(A) \leq \|A\|_1 = 1$), that $r = 1 = \rho(A)$ is a simple eigenvalue of $A$, and that the corresponding eigenvector can be chosen positive. Of course, such eigenvector is uniquely defined if we require that its measure in the 1-norm is one.

So, it is uniquely defined the Perron-pair $(1 = \rho(A), z)$, such that $Az = z$, $z$ positive, $\|z\|_1 = 1$. The computation of such Perron-pair, i.e. the computation of the vector $z$, can be performed via the power method. Actually, we now see that the power method in the particular case where $A$ is stochastic-by-columns (besides non-negative and irreducible) can be rewritten in a simpler form and converges independently from the choice of $a_0$ (provided that 1 dominates the other eigenvalues). These results follow from some remarks, reported in the following Proposition.

**Proposition.** Let $A$ be an irreducible non-negative stochastic-by-columns $n \times n$ matrix. Then

i) $$v \in \mathbb{C}^n \Rightarrow \sum_i (Av)_i = \sum_i v_i$$

ii) $$(\sum_i (Av)_i = \sum_i \sum_j a_{ij}v_j = \sum_j v_j \sum_i a_{ij} = \sum_j v_j),$$

v positive, $\|v\|_1 = 1 \Rightarrow Av$ positive, $\|Av\|_1 = 1$

(we use the irreducibility of $A$ and assertion i)),

iii) $$Av = \lambda v, \ \lambda \neq 1, \Rightarrow \sum_i v_i = 0$$

$$(\sum_i v_i = \sum_i (Av)_i = \lambda \sum_i v_i, \text{ thus } (\lambda - 1) \sum_i v_i = 0).$$

**Exercise.** Assume that $X^{-1}AX = J$ where $J$ is the Jordan block diagonal form of $A$, where $A$ is an irreducible non-negative stochastic-by-columns $n \times n$ matrix. Assume that $[J]_{11} = 1$. Prove that $\sum_i [X]_{ij} = 0, j = 2, \ldots, n$. 

**Corollary (power).** Let $A$ be an irreducible non-negative stochastic-by-columns $n \times n$ matrix. Let $a_0$ be any positive vector such that $\|a_0\|_1 = 1$. Then set

$$a_{k+1} = Aa_k, \ \ k = 0, 1, 2, \ldots .$$

8
The sequence \( \{a_k\} \) is a well defined sequence of positive vectors, and \( \|a_k\|_1 = 1 = \rho(A), \forall k \). If \( k \to +\infty \), then

\[
a_k - z \to 0,
\]

provided that \( |\lambda_j| \) is smaller than 1 for all \( j = 2, \ldots, n \). The latter event is assured by Theorem(positive) when \( A \) is positive. In the particular case where \( A \) is diagonalizable, the rate of convergence is

\[
( \max_{j=2 \ldots n} |\lambda_j| )^k
\]

[for the general case, use the Remark in Theorem(power) of the section on preliminary considerations].

proof: The sequences \( a_k \) and \( v_k \) defined in Theorem(power) coincide, because, by Proposition ii), we have \( \|A v_k\|_1 = \|v_k\|_1 \). It remains to show that \( \alpha \) in the expression \( a_0 = \alpha z + \sum_j \alpha_j x_j \) is nonzero. Note that \( e^T a_0 = \alpha e^T z + \sum_j \alpha_j e^T x_j = \alpha + \sum_j \alpha_j e^T x_j \). In the case that \( A \) is diagonalizable the thesis follows by Proposition iii) which asserts that \( e^T x_j, j = 2, \ldots, n, \) must be zero. In the generic case the proof is left to the reader . . . .

**Exercise.** Discuss existence, unicity, and computation of the Perron-pair for

\[
A = \begin{bmatrix}
0 & 1 - a & a \\
1 - a & 0 & 1 - a \\
a & a & 0
\end{bmatrix}, \ a \in [0,1], \ A = \begin{bmatrix}
0 & 0 & 1 \\
1 - a & 0 & 0 \\
a & 0 & 0
\end{bmatrix}, \ a \in [0,1],
\]

\[
A = \begin{bmatrix}
0 & b_1 & b_2 \\
1 - b_1 & 0 & \ddots \\
1 - b_2 & \ddots & b_{n-2} \\
& \ddots & 0 & 1 \\
& & 0 & b_{n-2}
\end{bmatrix}, \ b_i \in (0,1).
\]

**Page-rank computation** [Berkhin, et al]

Consider an oriented graph with set of verteces \( V = \{1, 2, \ldots, n\} \), and set of edges \( E, i, j \in E \) if there is a link from \( i \) to \( j \).

Associate to such graph the adiacency matrix:

\[
L = \begin{cases}
1 & ij \in E \\
0 & ij \notin E
\end{cases}
\]

Call \( \deg (i) \) the number of edges starting from \( i \). Of course, \( \deg (i) = 0 \) (no edge starts from \( i \)) iff \( L_{ij} = 0, \forall j \). Moreover, \( \deg (i) = \sum j L_{ij} \).

Associate to the graph the transition matrix:

\[
P = \begin{cases}
\frac{1}{\deg (i)} & ij \in E \\
0 & ij \notin E
\end{cases}
\]
Note that \( P_{ij} = L_{ij} / \deg(i) \) if \( i \) is such that \( \deg(i) > 0 \), and \( P_{ij} = L_{ij} = 0 \) otherwise. Moreover, \( \sum_j P_{ij} = 1 \) if \( \deg(i) > 0 \) and \( \sum_j P_{ij} = 0 \) otherwise. So, the matrix \( P \) is a non negative matrix quasi-stochastic by rows.

**Remark.** Row \( i \) of \( P \) is null iff no edge starts from \( i \); column \( j \) of \( P \) is null iff no edge points to \( j \).

Let \( \mathbf{p} \in \mathbb{R}^n \) be the vector whose entry \( p_j \), is the importance (authority) of the vertex \( j \). Then

\[
p_j = \sum_{i: i \to j} \frac{p_i}{\deg(i)} = \sum_{i=1}^{n} P_{ij}p_i = \sum_{i=1}^{n} P^T_i p_i = (P^T \mathbf{p})_j, \quad \mathbf{p} = P^T \mathbf{p}
\]

(note that such fixed point \( \mathbf{p} \) may not exist, or, if exists, may be not unique or with zero entries; see below).

Let \( p_j^{(k+1)} \) be the probability that at step \( k+1 \) of my visit of the graph (navigation on the web) I am on the vertex (page) \( j \). Then

\[
p_j^{(k+1)} = \sum_{i: i \to j} \frac{p_i^{(k)}}{\deg(i)} = \sum_{i=1}^{n} P_{ij}p_i^{(k)} = \sum_{i=1}^{n} P^T_{i} p_i^{(k)} = (P^T \mathbf{p}^{(k)})_j, \quad \mathbf{p}^{(k+1)} = P^T \mathbf{p}^{(k)}
\]

(note that such sequence of vector probabilities exists and is uniquely defined, once \( \mathbf{p}^{(0)} \) is given, but the \( \mathbf{p}^{(k)} \) may lose the possible ddp property of \( \mathbf{p}^{(0)} \); see below). [A vector \( \mathbf{w} \) is said ddp (discrete distribution of probability) if \( \mathbf{w} \) is positive and \( \|\mathbf{w}\|_1 = 1 \).

Note that it is natural to require that: \( p_i^{(k)} > 0 \) (at step \( k \) there is a probability that I am on vertex \( i \)), \( \sum_j p_j^{(k)} = 1 \) (at step \( k \) I am on some vertex): \( p_i > 0 \) (any vertex has a portion of importance ...), \( \sum_i p_i = 1 \) (... of the total 1).

So, the following facts must be true:

a) \( \mathbf{p} \) such that \( \mathbf{p} = P^T \mathbf{p}, \mathbf{p} > 0, \|\mathbf{p}\|_1 = 1 \), exists and is uniquely defined,

b) the method \( \mathbf{p}^{(0)} = \text{ddp}, \mathbf{p}^{(k+1)} = P^T \mathbf{p}^{(k)} \) converges to \( \mathbf{p} \).

We now show that in order to have a) and b), the matrix \( P^T \) must be both stochastic by columns and irreducible.

**Theorem (stoc).** If the above facts a) and b) are true, then \( P^T \) must be stochastic by columns.

**proof:** We know that \( \exists ! \mathbf{p} \) such that \( \mathbf{p} = P^T \mathbf{p}, \mathbf{p} > 0, \|\mathbf{p}\|_1 = 1 \). Assume that \( P^T \) is quasi-stochastic but not stochastic by columns. Then

\[
\|P^T \mathbf{p}\|_1 = \sum_i (P^T \mathbf{p})_i = \sum_j \sum_j (P^T)_{ij} p_j = \sum_j \sum_i P_{ij} p_j = \sum_{j: \deg(j)>0} 1 \cdot p_j + \sum_{j: \deg(j)=0} 0 \cdot p_j < \sum_j p_j = \|\mathbf{p}\|_1,
\]

analogously,

\[
\|\mathbf{p}^{(k+1)}\|_1 = \|P^T \mathbf{p}^{(k)}\|_1 \leq \|\mathbf{p}^{(k)}\|_1 \leq \|\mathbf{p}^{(1)}\|_1 < \|\mathbf{p}^{(0)}\|_1 = 1.
\]

Thus \( \mathbf{p} \) cannot be equal to \( P^T \mathbf{p} \), and \( \mathbf{p}^{(k)} \) cannot converge to a ddp. □

**Theorem (irred).** If the above facts a) and b) are true, then \( P^T \) must be irreducible.
proof: assume $P^T$ reducible. Then there exists a permutation matrix $Q$ such that

$$Q^T P^T Q = \begin{bmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{bmatrix}, \quad (*)$$

with $A_{11}$ and $A_{22}$ at least $1 \times 1$ square matrices. Note that $Q^T P^T Q$ is stochastic by columns, like $P^T$. We know that $\exists! \mathbf{p}$ such that $\mathbf{p} = P^T \mathbf{p}$, $\mathbf{p} > 0$, $\|\mathbf{p}\|_1 = 1$, but this is equivalent to say that $\exists! Q^T \mathbf{p}$ such that

$$Q^T \mathbf{p} = \begin{bmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{bmatrix} Q^T \mathbf{p}, \quad Q^T \mathbf{p} > 0, \quad \|Q^T \mathbf{p}\|_1 = 1. \quad (***)$$

Case 1: assume $A_{12} = 0$. Then $A_{11}$ and $A_{22}$ are non negative stochastic by columns matrices. We can assume they are also irreducible (why?). Then, by the Perron-Frobenius theory,

$$\exists! \mathbf{y}_1, \mathbf{y}_2 > 0, \quad \|\mathbf{y}_1\|_1 = \|\mathbf{y}_2\|_1 = 1, \quad \mathbf{y}_1 = A_{11} \mathbf{y}_1, \quad \mathbf{y}_2 = A_{22} \mathbf{y}_2.$$

and, as a consequence, for all $\alpha \in (0, 1)$ we have

$$\begin{bmatrix} \alpha \mathbf{y}_1 \\ (1-\alpha) \mathbf{y}_2 \end{bmatrix} = \begin{bmatrix} A_{11} & 0 \\ 0 & A_{22} \end{bmatrix} \begin{bmatrix} \alpha \mathbf{y}_1 \\ (1-\alpha) \mathbf{y}_2 \end{bmatrix} > 0, \quad \|\begin{bmatrix} \alpha \mathbf{y}_1 \\ (1-\alpha) \mathbf{y}_2 \end{bmatrix}\|_1 = 1.$$

So, all vectors $\mathbf{p}$ such that $Q^T \mathbf{p} = \begin{bmatrix} \alpha \mathbf{y}_1 \\ (1-\alpha) \mathbf{y}_2 \end{bmatrix}$ satisfy the properties $\mathbf{p} > 0$, $\|\mathbf{p}\|_1 = 1, \mathbf{p} = P^T \mathbf{p}$, which is against the hypothesis of unicity.

Case 2: assume $A_{12} \neq 0$. Then $A_{22}$ in $(*)$ is not stochastic by columns, thus $A_{22}$ may have no eigenvalue equal to 1, i.e. the equations in $(***)$ involving $A_{22}$ may be verified only if part of $\mathbf{p}$ is null, against the hypothesis of positiveness of $\mathbf{p}$.

Viceversa, we know that if the non negative matrix $P^T$ is irreducible and stochastic by columns, then $1 = \rho(P^T)$ is a simple eigenvalue of $P^T$ and there exists a unique vector $\mathbf{p}$ such that $\mathbf{p} = P^T \mathbf{p}$, $\mathbf{p} > 0$, $\|\mathbf{p}\|_1 = 1$. We also know that such hypotheses are not sufficient to assure the convergence (to $\mathbf{p}$) of the sequence $\mathbf{p}^{(k+1)} = P^T \mathbf{p}^{(k)}$, $\mathbf{p}^{(0)} > 0$, $\|\mathbf{p}_0\|_1 = 1$, or, equivalently, to assure that the remaining eigenvalues of $P^T$ have absolute value smaller than 1. We can only say that the sequence $\{\mathbf{p}^{(k)}\}$ is a well defined sequence of ddp.

We have to modify $P$ (the graph) so to make well posed (exists!) the mathematical problem and to make convergent the algorithm for solving it.

Make $P^T$ stochastic:

$$P' = P + d \mathbf{v}^T, \quad d = \begin{bmatrix} \delta_{\text{deg}(1),0} \\ \vdots \\ \delta_{\text{deg}(n),0} \end{bmatrix}, \quad \mathbf{v} = \begin{bmatrix} \frac{1}{n} \\ \vdots \\ \frac{1}{n} \end{bmatrix}$$

i.e, where $P$ has null rows $P'$ has the row vector $\mathbf{v}^T$. The vertex $i$ with $\text{deg}(i) = 0$ now links to all vertices of the graph. [We discuss the uniform case, but what follows can be repeated for the more general case $\mathbf{v} = d \mathbf{d}$.]

11
Observe that \((P')^T\) is stochastic by columns, \((P')^T \geq 0\), thus 1 is eigenvalue of \((P')^T\) and the other eigenvalues of \((P')^T\), \(\lambda_2', \ldots, \lambda_n'\), are such that \(|\lambda_j'| \leq 1\).

Make \((P')^T\) irreducible:

\[
P'' = cP' + (1-c)e e^T, \quad e = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}, \quad c \in (0, 1).
\]

Since

\[
e_i^T P'' = \begin{cases} cv^T + (1-c)v^T = v^T & \text{deg}(i) = 0 \\ c[\cdots 0 \frac{1}{\text{deg}(i)} 0 \cdots] + (1-c)v^T & \text{deg}(i) > 0 \end{cases}
\]

we are assuming that a visitor of the graph can go from the vertex \(i\) to one of its neighborhoods with probability \(c/\text{deg}(i) + (1-c)/n\), and with probability \((1-c)/n\) to an arbitrary vertex of the graph. Of course the parameter \(c\) must be chosen near 1, in order to maintain our model faithful to the way the graph (web) is visited.

Observe that \((P'')^T\) is stochastic by columns and positive, therefore, in particular, it is non negative and irreducible. So, we have all we need.

**Theorem (Page-Rank).** \(1 = \rho((P'')^T)\) is a simple eigenvalue of \((P'')^T\), there exists a unique vector \(p\) such that \(p = (P'')^T p, p > 0, \|p\|_1 = 1\) (i.e. we have fact (a)), and the other eigenvalues of \((P'')^T\), \(\lambda_2', \ldots, \lambda_n'\), are such that \(|\lambda_j'| < 1\) (by Theorem(positive)). Thus, \(p^{(k+1)} = (P'')^T p^{(k)}, p^{(0)} > 0, \|p^{(0)}\|_1 = 1\), is a sequence of ddp convergent to \(p\) and, in case \(P''\) is diagonalizable,

\[
\|p^{(k)} - p\| = O((\max_{j=2, \ldots, n} |\lambda_j'|)^k)
\]

[for the general case see the Remark in Theorem(power) of Section 1] (i.e. we have fact (b)). Moreover, for the particular choice of \((P'')^T\), the cost of each step of the power method is \(O(n)\) and is dominated by the cost of the matrix-vector multiplication \(P^T z\), and if \(A\) is diagonalizable, then the rate of convergence is

\[
\|p^{(k)} - p\| = O(c^k)
\]

[for the general case see the Remark in Theorem(power) of Section 1] (Google-search engine sets \(c = 0.85\) [Berkhin]).

**proof:** We have to prove only the final assertions.

\(O(n)\) arithmetic operations are sufficient to perform each step of the power method. We have

\[
(P'')^T = c(P^T + vd^T) + (1-c)ve^T,
\]

and, if \(x \geq 0\), then

\[
(P'')^T x = cP^T x + \gamma v, \\
\gamma = c d^T x + (1-c)e^T x = e^T x - c[e^T x - d^T x] = \|x\|_1 - c\|P^T x\|_1
\]

(why the latter equality holds?). Thus, in order to compute \(p^{(k+1)}\) from \(p^{(k)}\) one can use the following function

\[
y = cP^T x, \\
\gamma = \|x\|_1 - \|y\|_1, \\
(P'')^T x = y + \gamma v
\]

12
where the dominant operation is the matrix-vector multiplication $P^T x$. Note that each row $j$ of $P^T$ has (on the average) a very small number of nonzero entries, i.e. exactly the number of vertices pointing to $j$, so $P^T x$ can be computed with $O(n)$ arithmetic operations. Note also that in order to implement the above function one needs only $2n$ memory allocations.

**Rate of convergence** $\|p^{(k)} - p\| = O(c^k)$. We first prove that if $e^T v = 1$ then

$$p_{P'}(\lambda) = (\lambda - 1)p_{n-1}(\lambda) \Rightarrow p_{P'} + \frac{1}{c} e v^T(\lambda) = (\lambda - \frac{1}{c})p_{n-1}(\lambda). \quad (***)$$

In fact,

$$S = \begin{bmatrix} e & y_2 & \cdots & y_n \end{bmatrix}, \quad \det(S) \neq 0,$$

$$S^{-1}P' S = \begin{bmatrix} 1 & u^T \\ 0 & M \end{bmatrix}, \quad p_{P'}(\lambda) = (\lambda - 1)p_{M}(\lambda),$$

$$S^{-1}(P' + \frac{1-c}{c} e v^T) S = \begin{bmatrix} 1 & u^T \\ 0 & M \end{bmatrix} + \frac{1-c}{c} S^{-1} e v^T S$$

$$= \begin{bmatrix} 1 & u^T \\ 0 & M \end{bmatrix} + \frac{1-c}{c} \begin{bmatrix} 1 & \cdots & * \\ 0 & \cdots & 0 \end{bmatrix} = \begin{bmatrix} 1 & \frac{1}{c} u^T \\ 0 & M \end{bmatrix}.$$

As a consequence of (***) if $1, \lambda'_2, \ldots, \lambda'_n$ are the eigenvalues of $P'$ (recall that $|\lambda'_j| \leq 1, j = 2, \ldots, n$), then the eigenvalues of $P' + \frac{1-c}{c} e v^T$ are $\frac{1}{c}, \lambda'_2, \ldots, \lambda'_n$, and thus the eigenvalues of $cP' + (1-c)e v^T$ are $1, c\lambda'_2, \ldots, c\lambda'_n$.

It follows that if $A$ is diagonalizable, then $\|p^{(k)} - p\| = O((\max_{j=2,\ldots,n} |c\lambda'_j|^k)) = O(c^k). \quad \Box$

Why to compute $p$? [Berkhin].

**QUERY**: Berkhin survey.

Go in the inverted terms document file, which is a table containing a row for each term of a collection’s dictionary. In such file, for each term there is a list of all documents that contain such term

<table>
<thead>
<tr>
<th>term</th>
<th>LISTA$_{\text{term}}$ = {i$_1$, i$_2$, \ldots, i$_k$} $\subset$ {1, 2, \ldots, n}</th>
</tr>
</thead>
<tbody>
<tr>
<td>Berkhin</td>
<td>LISTA$_{\text{Berkhin}}$ = {1, 4, 6}</td>
</tr>
<tr>
<td>survey</td>
<td>LISTA$_{\text{survey}}$ = {1, 3}</td>
</tr>
</tbody>
</table>

Define the set of relevance of the query

$$\bigcup_{\text{term} \in \text{QUERY}} \text{LISTA}_{\text{term}} = \{1, 3, 4, 6\}$$

Reading $p$ ($p$ is updated once each month) considers and orders the corresponding set of authorities $\{p_1, p_3, p_4, p_6\}$, for example $p_4 \geq p_6 \geq p_3 \geq p_1$.

Finally, show the titles of the documents 1, 3, 4, 6 in the order 4, 6, 3, 1, from the one with greatest authority to the one with smallest authority.
Main criticism: This procedure, being independent from the query allows a fast answer, but does not make distinction between pages with authority from pages with authority on a specific subject.

**Exercise.** Draw the graph whose transition matrix is

\[
P = \begin{bmatrix}
0 & 1/2 & 1/2 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1/3 & 0 \\
0 & 0 & 0 & 0 & 1/2 & 1/2 \\
0 & 0 & 0 & 1/2 & 0 & 1/2 \\
0 & 0 & 0 & 1 & 0 & 0
\end{bmatrix}.
\]

Note that \(P\) is non negative, reducible and quasi-stochastic, but not stochastic, by rows. Prove that there is no positive vector \(p\) such that \(p = PTp\). Starting from \(P\) and proceeding as indicated in the theory, introduce a non negative matrix \(P^0\) stochastic by rows. Note that \(P^0\) is reducible, like \(P\). Prove that there is no positive vector \(p\) such that \(p = (P^0)^Tp\). Starting from \(P^0\) and proceeding as indicated in the theory, introduce a non negative matrix \(P^00\) irreducible and stochastic by rows. Prove that there exists a unique positive vector \(p\) such that \(p = (P^00)^Tp\), \(\|p\|_1 = 1\), and describe an algorithm for the computation of \(p\).

Here below is an approximation of such vector \(p\):

\[p^T = [0.03721 \, 0.05396 \, 0.04151 \, 0.3751 \, 0.206 \, 0.2862].\]

Assume, for example, that the set of relevance of a query is \(\{1, 3, 4, 6\}\). Then the documents \(1, 3, 4, 6\) are listed in the order \(4, 6, 3, 1\), being \(p_4 \geq p_6 \geq p_3 \geq p_1\).

**APPENDIX**

*Proof (Theorem (inverse power)):*

Assume \(A\) diagonalizable, \(A\)x\(_j\) = \(\lambda_j\)x\(_j\), \(\{x_j\}\)\(_j=1^n\) linearly independent. Then

\[Ax_j - \lambda_j^*x_j = (\lambda_j - \lambda_j^*)x_j \Rightarrow (A - \lambda_j^*I)^{-m}x_j = \frac{1}{(\lambda_j - \lambda_j^*)^m}x_j.
\]

Call \(\alpha_j\) the numbers for which \(v_0 = \sum_{j: \lambda_j = \lambda_i} \alpha_j x_j + \sum_{j: \lambda_j \neq \lambda_i} \alpha_j x_j\). Note that the first sum is a non null vector (by the assumption on \(v_0\)). Then the following equality holds

\[(\lambda_i - \lambda_j^*)m(A - \lambda_j^*I)^{-m}v_0 = \sum_{j: \lambda_j = \lambda_i} \alpha_j x_j + \sum_{j: \lambda_j \neq \lambda_i} \alpha_j \left(\frac{\lambda_j - \lambda_j^*}{\lambda_i - \lambda_j^*}\right)^m x_j\]

which, for \(m \to +\infty\), yields the thesis.

Assume that \(A\) is not diagonalizable. In this case call \(x_j, j = 1, \ldots, n\), the columns of the non singular matrix \(X\) for which \(X^{-1}AX = J\) where \(J\) is the block diagonal Jordan form of \(A\). Assume that the upper-left diagonal block of \(J\) is diagonal and its diagonal contains all \(\lambda_j\) such that \(\lambda_j = \lambda := \lambda_i\). Assume that there are \(t\) of such eigenvalues. Then consider any other diagonal block of \(J\); of course it corresponds to an eigenvalue \(\lambda_j\) different from \(\lambda\), call such
eigenvalue $\mu$. The block under consideration has an order, say $s = s_\mu$.

\[
J = \begin{bmatrix}
\lambda & \lambda & \cdots & \lambda \\
\mu & 1 & \cdots & \mu \\
\ddots & \ddots & \ddots & \ddots \\
\mu & \mu & \cdots & 1
\end{bmatrix}.
\]

Restrict the matrix equation $X^{-1}AX = J$ to this block, thus, for some $r \geq t$ we have

\[
A[x_{r+1} \ x_{r+2} \ x_{r+3} \ \cdots \ x_{r+s}] = [x_{r+1} \ x_{r+2} \ x_{r+3} \ \cdots \ x_{r+s}]
\begin{bmatrix}
\mu & 1 & \cdots & \mu \\
\ddots & \ddots & \ddots & \ddots \\
\mu & \mu & \cdots & 1
\end{bmatrix}.
\]

It follows that

\[
A_{x_{r+1}} = \mu x_{r+1}, \quad (A - \lambda^* I)x_{r+1} = (\mu - \lambda^*)x_{r+1},
\]

\[
(A - \lambda^*)^{-1}x_{r+1} = \frac{1}{\mu - \lambda^*}x_{r+1}, \quad (A - \lambda^*)^{-m}x_{r+1} = \frac{1}{(\mu - \lambda^*)^m}x_{r+1}.
\]

\[
A_{x_{r+2}} = \mu x_{r+2} + x_{r+1}, \quad (A - \lambda^*)^{-1}x_{r+2} = (\mu - \lambda^*)x_{r+2} + x_{r+1},
\]

\[
(A - \lambda^*)^{-m}x_{r+2} = \frac{1}{(\mu - \lambda^*)^m}x_{r+2} + \frac{1}{(\mu - \lambda^*)^m}x_{r+1}.
\]

\[
A_{x_{r+3}} = \mu x_{r+3} + x_{r+2}, \quad (A - \lambda^*)^{-1}x_{r+3} = (\mu - \lambda^*)x_{r+3} + x_{r+2},
\]

\[
(A - \lambda^*)^{-m}x_{r+3} = \frac{1}{(\mu - \lambda^*)^m}x_{r+3} - \frac{1}{(\mu - \lambda^*)^m}x_{r+2} + \frac{1}{(\mu - \lambda^*)^m}x_{r+1}.
\]

\[
A_{x_{r+4}} = \mu x_{r+4} + x_{r+3}, \quad (A - \lambda^*)^{-1}x_{r+4} = (\mu - \lambda^*)x_{r+4} + x_{r+3},
\]

\[
(A - \lambda^*)^{-m}x_{r+4} = \frac{1}{(\mu - \lambda^*)^m}x_{r+4} - \frac{1}{(\mu - \lambda^*)^m}x_{r+3} + \frac{1}{(\mu - \lambda^*)^m}x_{r+2} - \frac{1}{(\mu - \lambda^*)^m}x_{r+1},
\]

\[
A_{x_{r+s}} = \mu x_{r+s} + \cdots + (-1)^{s-1} \frac{q_{s-1}(m)}{(\mu - \lambda^*)^{m+s-1}}x_{r+1},
\]

\[
(A - \lambda^*)^{-m}[x_{r+1} \ x_{r+2} \ x_{r+3} \ \cdots] = [x_{r+1} \ x_{r+2} \ x_{r+3} \ \cdots]
\begin{bmatrix}
\frac{1}{(\mu - \lambda^*)^m} & -\frac{m}{(\mu - \lambda^*)^m} & \frac{1}{(\mu - \lambda^*)^m} & \cdots \\
0 & \frac{1}{(\mu - \lambda^*)^m} & \frac{1}{(\mu - \lambda^*)^m} & \cdots \\
\ddots & \ddots & \ddots & \ddots \\
0 & 0 & \frac{1}{(\mu - \lambda^*)^m} & \frac{1}{(\mu - \lambda^*)^m} & \cdots
\end{bmatrix}.
\]
Set also $\lambda^* := \lambda^*_i$ ($\lambda^*_i$ is the given approximation of $\lambda = \lambda_i$). Choose $v_0$, and call $\alpha_j$ the numbers for which $v_0 = \sum_{j: \lambda_j = \lambda_i} \alpha_j x_j + \{ \ldots + \sum_{j=r+s} \alpha_j x_j + \ldots \}$.

Note that $\sum_{j: \lambda_j = \lambda} \alpha_j x_j = \sum_{j=1}^t \alpha_j x_j$. Then

$\begin{align*}
(A - \lambda^* I)^{-m} v_0 &= \sum_{j: \lambda_j = \lambda} \alpha_j (A - \lambda^* I)^{-m} x_j + \{ \ldots + \sum_{j=r+s} \alpha_j (A - \lambda^* I)^{-m} x_j + \ldots \} \\
&= \sum_{j: \lambda_j = \lambda} \alpha_j x_j + \{ \ldots + \sum_{j=r+s} \alpha_j \frac{(m^2 + m)}{\mu - \lambda^*} x_j + \ldots \} \\
&+ \alpha_{r+1}^{+} \frac{1}{(n - \lambda^*)^{m+1}} x_{r+1} + \ldots + \alpha_{s}^{+} \frac{1}{(n - \lambda^*)^{m+s}} x_{s+1} \\
&+ \ldots + \alpha_{r+s}^{+} \frac{1}{(n - \lambda^*)^{m+s+r+1}} x_{s+r+1} + \ldots \\
&= \sum_{j: \lambda_j = \lambda} \alpha_j \frac{1}{\mu - \lambda^*}^m x_j + \{ \ldots + \sum_{j=r+s} \alpha_j \frac{1}{(n - \lambda^*)^{m+r+s}} x_j + \ldots \}.
\end{align*}$

But this implies

$\begin{align*}
(\lambda - \lambda^*)^m (A - \lambda^* I)^{-m} v_0 &= \sum_{j: \lambda_j = \lambda} \alpha_j x_j + \{ \ldots + \frac{1}{(n - \lambda^*)^{m+r+s}} x_{r+s} + \ldots \} \\
&= \sum_{j: \lambda_j = \lambda} \alpha_j x_j + \{ \ldots + \sum_{j=r+s} \alpha_j \frac{1}{(n - \lambda^*)^{m+r+s}} x_{r+s} + \ldots \}.
\end{align*}$

from which it is clear that the sequence $(\lambda - \lambda^*)^m (A - \lambda^* I)^{-m} v_0$ converges to an eigenvector of $A$ associated with $\lambda$. The assertions about the rate of convergence follow by setting

$\begin{align*}
p_{n-1}(\lambda) &= \alpha_{r+1}^{+} - \alpha_{r+2}^{+} \frac{m}{\mu - \lambda^*} + \alpha_{r+3}^{+} \frac{1}{\mu - \lambda^*} + \ldots + (-1)^{s-1} \alpha_{r+s}^{+} \frac{q_{r-s}(m)}{(\mu - \lambda^*)^{m+r+s}} x_{r+s} \\
&- \alpha_{r+4}^{+} \frac{(m^2 + m)}{\mu - \lambda^*} + \alpha_{r+5}^{+} \frac{1}{\mu - \lambda^*} + \ldots + (-1)^{s-1} \alpha_{r+s}^{+} \frac{q_{r-s}(m)}{(\mu - \lambda^*)^{m+r+s}} x_{r+s} + \ldots.
\end{align*}$

Proof (Theorem (Power)):

Assume that $A$ is diagonalizable, so $A x_j = \lambda_j x_j, \{x_j\}_{j=1}^n$ linearly independent. Choose $v_0$, and call $\alpha_j$ so that $v_0 = \sum_{j: \lambda_j = \lambda_1} \alpha_j x_j + \sum_{j: \lambda_j \neq \lambda_1} \alpha_j x_j$. Note that the first sum is non null by assumption. Then we have the equalities:

$\begin{align*}
A^m v_0 &= \sum_{j: \lambda_j = \lambda_1} \alpha_j \lambda_1^m x_j + \sum_{j: \lambda_j \neq \lambda_1} \alpha_j \lambda_j^m x_j \\
&= \sum_{j: \lambda_j = \lambda_1} \alpha_j \lambda_1^m x_j + \sum_{j: \lambda_j \neq \lambda_1} \alpha_j \lambda_j^m x_j.
\end{align*}$

The thesis follows letting $m$ go to infinite.

Assume that $A$ is not diagonalizable. Set $\mu := \lambda_j$, where $\lambda_j \neq \lambda_1$, and restrict, as above, the Jordan matrix equation $X^{-1} A X = J$ to a diagonal Jordan block of order $s$ associated with $\mu$. Then

$\begin{align*}
A x_{r+1} &= \mu x_{r+1}, \; A^m x_{r+1} = \mu^m x_{r+1}, \\
A x_{r+2} &= \mu x_{r+2} + x_{r+1}, \; A^m x_{r+2} = \mu^m x_{r+2} + \mu x_{r+1}, \\
A x_{r+3} &= \mu x_{r+3} + x_{r+2}, \; A^m x_{r+3} = \mu^m x_{r+3} + \mu x_{r+2} + \frac{1}{2}(m^2 - m) \mu^{m-2} x_{r+1},
\end{align*}$
\[ A_{x_{r+4}} = \mu x_{r+4} + x_{r+3}, \]
\[ A^m x_{r+4} = \mu^m x_{r+4} + m \mu^{m-1} x_{r+3} + \frac{1}{2} (m^2 - m) \mu^{m-2} x_{r+2} + \left( \frac{1}{6} m^3 - \frac{1}{2} m^2 + \frac{1}{3} m \right) \mu^{m-3} x_{r+1}, \]

and so on. Set \( \lambda := \lambda_1 \) and \( t = m_a(\lambda_1) = m_q(\lambda_1) \), so we can assume that the upper-left \( t \times t \) submatrix of \( J \) is diagonal with diagonal entries all equal to \( \lambda \). Then

\[
A^m v_0 = A^m \left( \sum_{j=1}^t \alpha_j x_j + \cdots + \sum_{j=r+1}^{r+s} \alpha_j x_j + \cdots \right) \\
= \sum_{j=1}^t \alpha_j \lambda^m x_j + \cdots + \sum_{j=r+1}^{r+s} \alpha_j A^m x_j + \cdots \\
= \lambda^m \sum_{j=1}^t \alpha_j x_j + \cdots + \sum_{j=r+1}^{r+s} \alpha_j A^m x_j + \cdots \\
= \lambda^m \sum_{j=1}^t \alpha_j x_j \\
+ \cdots + \left[ \alpha_{r+1} \mu^m x_{r+1} + \alpha_{r+2} (\mu^m x_{r+2} + m \mu^{m-1} x_{r+1}) + \alpha_{r+3} \left( \mu^m x_{r+3} + m \mu^{m-1} x_{r+2} + \frac{1}{2} (m^2 - m) \mu^{m-2} x_{r+1} + \frac{1}{6} m^3 - \frac{1}{2} m^2 + \frac{1}{3} m \right) \mu^{m-3} x_{r+1} \right] + \cdots, \\
\frac{1}{\lambda^m} A^m v_0 = \sum_{j=1}^t \alpha_j x_j + \cdots + \left( \frac{\mu}{\lambda} \right)^m \sum_{j=r+1}^{r+s} \alpha_j \mu^m x_j + \cdots + \left( \frac{\mu}{\lambda} \right)^m \sum_{j=r+1}^{r+s} \alpha_j \mu^m x_j + \cdots \\
= \sum_{j=1}^t \alpha_j x_j + \cdots + \left( \frac{\mu}{\lambda} \right)^m \left[ \alpha_{r+1} + \alpha_{r+2} \frac{m}{\mu} + \alpha_{r+3} \frac{1}{2} \frac{(m^2 - m)}{\mu^2} + \cdots + \alpha_{r+s} \frac{g_{s-1}(m)}{\mu^{s-1}} \right] x_{r+1} + \cdots \\
+ \left( \frac{\mu}{\lambda} \right)^m \left[ \alpha_{r+2} + \alpha_{r+3} \frac{m}{\mu} + \alpha_{r+4} \frac{1}{2} \frac{(m^2 - m)}{\mu^2} + \cdots + \alpha_{r+s} \frac{g_{s-1}(m)}{\mu^{s-1}} \right] x_{r+2} + \cdots + \left( \alpha_{r+s} \right) x_{r+s} + \cdots. \\
\]

It is clear that, as \( m \) goes to infinity, the sequence \( \frac{1}{\lambda^m} A^m v_0 \) converges to an eigenvector of \( A \) associated with \( \lambda \), the dominant eigenvalue of \( A \). Finally, the assertions on the rate of convergence follow by setting

\[
p_{s-1}(m) = \alpha_{r+1} + \alpha_{r+2} \frac{m}{\mu} + \alpha_{r+3} \frac{1}{2} \frac{(m^2 - m)}{\mu^2} + \cdots + \alpha_{r+s} \frac{g_{s-1}(m)}{\mu^{s-1}}.
\]