Matrix-Vector Splitting for Eigenvector Solution: An Extended Abstract

Douglas V. de Jager*

Abstract

In this paper, we present a theorem concerning the eigenvector equation $\overline{\mathbf{x}} = \mathbf{M}\overline{\mathbf{x}}$, where $\mathbf{M} \in \mathbb{R}^{n\times n}$ is irreducible and $\rho(|\mathbf{M}|) = 1$. The theorem is of significance for a number of reasons. It permits novel iterative solution techniques—in particular, asynchronous iterative solution. It permits accelerated solution using existing techniques—by way of improved sparsity patterns and/or reduced spectral radius for the principal matrix. It informs us that if $\mathbf{M}$ comprises negative entries and there is some suitable eigenvector $\overline{\mathbf{x}}$, then this eigenvector $\overline{\mathbf{x}}$ is necessarily unique (up to scalar multiples). Also, the theorem provides an alternative explanation of why the $\mathbf{D}$ matrix in the PageRank definition (used to deal with dangling nodes) changes only the scale of the PageRank vector—the $\mathbf{D}$ matrix has no effect on PageRank order.

In this paper, we present a theorem concerning a particular type of eigenvector equation. Given a matrix $\mathbf{M} \in \mathbb{R}^{n\times n}$, let $|\mathbf{M}| \in \mathbb{R}^{n\times n}$ denote the matrix whose entries are the absolute values of the corresponding entries in $\mathbf{M}$; also, given such a square matrix $\mathbf{M}$, let $\rho(\mathbf{M})$ denote the spectral radius of the matrix (i.e. $\rho(\mathbf{M}) := max\{|\lambda| : \lambda \text{ is an eigenvalue of } \mathbf{M}\}$). The theorem we present here concerns eigenvectors $\overline{\mathbf{x}} \in \mathbb{R}^n$ satisfying:

$$\overline{\mathbf{x}} = \mathbf{M}\overline{\mathbf{x}}, \text{ where } \mathbf{M} \in \mathbb{R}^{n\times n} \text{ is irreducible and } \rho(|\mathbf{M}|) = 1 \tag{1}$$

When $\mathbf{M}$ is non-negative (i.e. $\forall i, j(M_{ij} \geq 0)$), we term such an $\overline{\mathbf{x}}$ a Perron vector. When $\mathbf{M}$ is column-stochastic (i.e. $\forall i, j(M_{ij} \geq 0)$ and $\forall j(\sum_i M_{ij} = 1)$) and $\overline{\mathbf{x}}$ is normalised (i.e. $\sum_i |x_i| = 1$)), we term such an $\overline{\mathbf{x}}$ a DTMC steady-state distribution.

We present the theorem below, and address some of the implications thereafter.

Theorem 1 *Let the following hold.*

(P1) $\mathbf{M} \in \mathbb{R}^{n\times n}$ *is irreducible.*

(P2) $\rho(|\mathbf{M}|) = 1$.

(P3) $\overline{\mathbf{v}} \in \mathbb{R}^n$.

(P4) $\mathbf{V}^{(K)} \in \mathbb{R}^{n\times n}$ $(K \subseteq \{1, 2, \ldots, n\})$ *is defined by* $V_{ij}^{(K)} := \begin{cases} v_i & \text{if } j \in K \\ 0 & \text{if } j \notin K \end{cases}$.

(P5) $\overline{\mathbf{v}}, K$ *are such that* $\begin{cases} \forall i, j \ 0 \leq |V_{ij}^{(K)}| \leq |M_{ij}|; \\ \nexists i, j \text{ such that } V_{ij}^{(K)}, M_{ij} \text{ are of different sign;} \\ \exists i, j \text{ such that } 0 < |V_{ij}^{(K)}| \leq |M_{ij}|. \end{cases}$

*Department of Computing, Imperial College London, United Kingdom. Email: dvd03@doc.ic.ac.uk

Then, we know the following to be true.

(C1) $\rho(|\mathbf{M}-\mathbf{V}^{(K)}|)<1$*;*

(C2) *If 1 is an eigenvalue of* $\mathbf{M}$*, then it is a simple eigenvalue of* $\mathbf{M}$ *and there is a corresponding right eigenvector* $\overline{\mathbf{x}}$ *of* $\mathbf{M}$*, which comprises no zero entries, and is the unique fixed-point of*

$$\overline{\mathbf{x}}=(\mathbf{M}-\mathbf{V}^{(K)})\overline{\mathbf{x}}+\overline{\mathbf{v}}. \tag{2}$$

Proof:

(C1) For $\alpha\in[0,1)$, let matrix $\mathbf{P}^{(\alpha)}\in\mathbb{R}^{n\times n}$ be defined by:

$$\forall i,j(P_{ij}^{(\alpha)}:=|M_{ij}-\alpha\mathbf{V}_{ij}^{(K)}|). \tag{3}$$

When $\alpha=0$ we have that

$$\rho(\mathbf{P}^{(\alpha)})=\rho(|\mathbf{M}|)=1. \tag{4}$$

As $\mathbf{M}$ is irreducible, so too is $\mathbf{P}^{(\alpha)}$, for all α.
By the Perron–Frobenius theorem [11, 7] we know that as α as increased, $\rho(\mathbf{P}^{(\alpha)})$ is necessarily reduced.
So, as α tends to 1 we may bound $\rho(\mathbf{P}^{(\alpha)})$ strictly away from $\rho(|\mathbf{M}|)=1$.
By continuity of spectral radius with respect to matrix entries [10] (in particular, continuity when entries become zero—and irreducibility is potentially compromised), we have

$$\rho(|\mathbf{M}-\mathbf{V}^{(K)}|)<\rho(|\mathbf{M}|)=1. \quad \square \tag{5}$$

(C2) Suppose $\overline{\mathbf{y}}=\mathbf{M}\overline{\mathbf{y}}$.
Then,

$$\begin{array}{rcl}\overline{\mathbf{y}} & = & \mathbf{M}\overline{\mathbf{y}}\\ & = & (\mathbf{M}-\mathbf{V}^{(K)})\overline{\mathbf{y}}+\mathbf{V}^{(k)}\overline{\mathbf{y}}\\ & = & (\mathbf{M}-\mathbf{V}^{(K)})\overline{\mathbf{y}}+(\sum_{k\in K}y_k)\overline{\mathbf{v}}.\end{array} \tag{6}$$

The last equality holds because

$$\mathbf{V}^{(K)}\overline{\mathbf{y}}=\begin{pmatrix} v_1\sum_{k\in K}y_k\\ \vdots\\ v_n\sum_{k\in K}y_k\end{pmatrix}=\sum_{k\in K}y_k\begin{pmatrix}v_1\\ \vdots\\ v_n\end{pmatrix}=(\sum_{k\in K}y_k)\overline{\mathbf{v}}. \tag{7}$$

Now, by the preceding, we know that $\rho(|\mathbf{M}-\mathbf{V}^{(K)}|)<1$.
And, we know that $\forall\mathbf{N}\in\mathbb{R}^{n\times n}(\rho(|\mathbf{N}|)\geq\rho(\mathbf{N}))$ [8].
So, 1 is not an eigenvalue of $(\mathbf{M}-\mathbf{V}^{(K)})$.
This implies that $\sum_{k\in K}y_k\neq 0$.
Now, $\forall\alpha\neq 0$, we have the following:

$$\begin{array}{rlcl} & \overline{\mathbf{z}} & := & (\mathbf{M}-\mathbf{V}^{(K)})\overline{\mathbf{z}}+\alpha\overline{\mathbf{v}}\\ \Rightarrow & \mathbf{I}\overline{\mathbf{z}} & = & (\mathbf{M}-\mathbf{V}^{(K)})\overline{\mathbf{z}}+\alpha\overline{\mathbf{v}}\\ \Rightarrow & (\mathbf{I}-(\mathbf{M}-\mathbf{V}^{(K)}))\overline{\mathbf{z}} & = & \alpha\overline{\mathbf{v}}\\ \Rightarrow & \overline{\mathbf{z}} & = & \alpha(\mathbf{I}-(\mathbf{M}-\mathbf{V}^{(K)}))^{-1}\overline{\mathbf{v}}\end{array} \tag{8}$$

The last of the above implications is valid:

$$\begin{array}{rcl}\rho(\mathbf{M}-\mathbf{V}^{(K)})<1 & \Rightarrow & \det(\mathbf{I}-(\mathbf{M}-\mathbf{V}^{(K)}))\neq 0\\ & \Rightarrow & (\mathbf{I}-(\mathbf{M}-\mathbf{V}^{(K)}))\text{ is invertible.}\end{array} \tag{9}$$

Accordingly, the fixed-point $\overline{\mathbf{x}}$ of

$$\overline{\mathbf{x}}=(\mathbf{M}-\mathbf{V}^{(K)})\overline{\mathbf{x}}+\overline{\mathbf{v}} \tag{10}$$

is a scalar multiple of $\overline{\mathbf{y}}$, and thus $\overline{\mathbf{x}}$ is a right eigenvector of $\mathbf{M}$ corresponding to 1.
Further, given the non-singularity of $(\mathbf{M} - \mathbf{V}^{(K)})$, $\overline{\mathbf{x}}$ is the unique fixed-point of (9).
Hence, 1 is a simple eigenvalue of $\mathbf{M}$.
Finally, $\overline{\mathbf{x}}$ comprises no zero entries.
This is known because we may choose $\mathbf{V}^{(K)}$ to comprise exactly one non-zero entry, so the non-zero coefficient $(\sum_{k\in K} y_k)$ in (5) reduces to a non-zero coefficient y_k, and, further, by irreducibility, we may choose $\mathbf{V}^{(K)}$ such that the index k corresponding to y_k is any index from 1 through n. $\square$

The theorem is significant for at least four reasons.

The first reason is that it permits asynchronous iterative solution for $\overline{\mathbf{x}}$ in equation (1). This is potentially advantageous for a number of reasons: faster asymptotic convergence through a "Gauss-Seidel effect" [2, 4], simplified task-management and programming [9], and better use of the communication channel [9]. Below we will consider the primary motivation for asynchrony, namely circumventing the synchronisation penalty [2], and we will then, for the interested reader, consider a somewhat formal outline of linear asynchronous iterative methods.

In implementations over distributed or parallel computing systems, synchronous iterative solution methods for linear systems involve processors performing an iteration, then communicating their results, and then all processors moving on to the next iteration (consider, for example, Jacobi or Gauss–Seidel methods [2]). When processors perform iterations at different speeds or interprocessor communication times are inhomogeneous, then we expect significant processor idle time (during which processors wait for information before moving on to subsequent iterations). This becomes all the more troubling when iterations leave entries unchanged or very minimally changed—in synchronous methods, processors still need to communicate the fact that there has been little or no change before all processors can move on to the next iteration. Asynchronous methods are so named as they involve the removal of synchronisation points between communication and iteration. Each processor is allowed to keep iterating at its own pace, using whatever information is available at the time (possibly old information). Within an asynchronous set-up, it is not incumbent upon processors to communicate a particular piece of information, or indeed to communicate at any particular point. This often corresponds to faster solution (faster convergence, in computational time, to the fixed-point). [2]

The mathematical formulation of linear asynchronous iterative methods was originally given in [5]. We consider here a slight generalisation of this formulation. We start with fixed-point problem:

$$\overline{\mathbf{x}} = \mathbf{A}\overline{\mathbf{x}} + \overline{\mathbf{b}}, \text{ where } \mathbf{A} \in \mathbb{R}^{n\times n}, \overline{\mathbf{x}}, \overline{\mathbf{b}} \in \mathbb{R}^n. \tag{11}$$

To solve for the unique fixed-point $\overline{\mathbf{x}}$, we generate a convergent sequence of iterates from the corresponding *class of asynchronous iterations.*

The class of asynchronous iterations is given by the class of sequences, $<\overline{\mathbf{x}}^{(t)}>$, of column vectors in $\mathbb{R}^n$ recursively defined:

$$x_i^{(t+1)} := \begin{cases} \sum_{j=1}^n A_{ij} x_j^{(t-d(t,i,j))} + b_i & \text{if} \quad i \in U(t) \\ x_i^{(t)} & \text{if} \quad i \notin U(t) \end{cases} \tag{12}$$

where U, d are functions satisfying a set of conditions. Function $U : \mathbb{N} \to \mathbb{P}\{1,\ldots,n\}$, which we may refer to as the update function, gives the set of vector entries to be updated at each step (where $\mathbb{P}\{1,\ldots,n\}$ denotes the power set of $\{1,\ldots,\text{n}\}$). Function $d : \mathbb{N}\times\{1,\ldots,n\}\times\{1,\ldots,n\} \to \mathbb{N}$, which we may refer to as the delay function, gives the relative "age" of the entries used in the update of each vector entry at each step. These functions satisfy the following conditions:

$\mathcal{CM}$**1:** Each vector entry i $(1 \le i \le n)$ features in an infinite number of update sets.

$\mathcal{CM}$**2:** For each pair of vector entries i, j $(1 \le i, j \le n)$, we have that $\begin{cases} (t - d(t,i,j)) \to \infty \text{ as } t \to \infty; \\ \forall t (d(t,i,j) \le t). \end{cases}$

Generalising the key result from [5], we get the following convergence guarantee for generated sequences (the proof is a simple amendment to that offered in [5]). It allows us to know pre-generation whether sequences are convergent.

Theorem 2 *Consider the class of asynchronous iterations corresponding to equation (12). If* $\rho(|\mathbf{A}|) < 1$ *, then every sequence of iterates within the class converges to the unique fixed-point* $\overline{\mathbf{x}}$ *as* $t \to \infty$*; if* $\rho(|\mathbf{A}|) \geq 1$*, then some non-convergent sequence of iterates exists within the class.*

By this theorem, we know that we cannot guarantee convergence for all asynchronous iterations corresponding to equation (1), but, rather excitingly, we can guarantee convergence for asynchronous iterations corresponding to equation (2).

It is perhaps worth emphasising that (12) gives us a *class* of possible sequences—and, indeed, Gauss–Seidel and Jacobi sequences fall within this class. So, whilst we now know, by theorem 1, that we may employ any sequence within the class of asynchronous iterations to find the solution eigenvector of (1), it still remains for us to show which sequences (as defined by varying d and U) correspond to more efficient solution than Jacobi or Gauss-Seidel sequences.

The second noteworthy point regarding the theorem is that it opens the door to minimising spectral radius and/or optimising matrix sparsity patterns. This has knock-on implications for potential solution acceleration.

It is well-known that for equation (2), the smaller $\rho(|\mathbf{M} - \mathbf{V}^{(K)}|)$, the faster we may expect solution convergence—be this via synchronous methods [1] or asynchronous methods [3]. A worthwhile future research endeavour, building on theorem 1, would be to investigate strategies for choosing $K, \overline{\mathbf{v}}$ so as to minimise $\rho(|\mathbf{M} - \mathbf{V}^{(K)}|)$.

An equally interesting research pursuit concerns optimal sparsity patterns. With a view toward accelerating iterative linear system solution methods, much effort is currently being expended investigating graph and hypergraph partitioning strategies for matrices over several processes[12]. These strategies aim to balance computation work-load appropriately over the processes and to minimise inter-process communication. By suitably choosing $K, \overline{\mathbf{v}}$, we may maximise the sparsity $\rho(|\mathbf{M} - \mathbf{V}^{(K)}|)$, and, in conjunction with (or instead of) graph/hypergraph partitioning strategies we may optimise process work-load and minimise inter-process communication. These assignments need not just have $\mathbf{M} - \mathbf{V}^{(K)}$ comprise some extra zeros where $\mathbf{M}$ comprises non-zeroes . They may also have $\mathbf{M} - \mathbf{V}^{(K)}$ comprise some non-zeros where $\mathbf{M}$ comprises zeroes. For example, by reasoning similarly to the proof of $C1$ in theorem 1, it can be shown that any of the following $K, \overline{\mathbf{v}}$ assignments yield also yield equation (2) with requisite sub-unit spectral radius for the modulus matrix.

$A_1 := \{K, \overline{\mathbf{v}} : \max_j\{\sum_i |(\mathbf{M} - \mathbf{V}^{(K)})_{ij}|\} \leq 1$ and there is a non-empty set L such that firstly we have that $\forall l(l \in L \to \sum_i |(\mathbf{M} - \mathbf{V}^{(K)})_{il}| < 1)$ and also we may construct an irreducible matrix from $\mathbf{M} - \mathbf{V}^{(K)}$ by replacing some of the zeros which have column indices in L with positive reals.$\}$

$A_2 := \{K, \overline{\mathbf{v}} : \max_i\{\sum_j |(\mathbf{M} - \mathbf{V}^{(K)})_{ij}|\} \leq 1$ and there is a non-empty set L such that firstly we have that $\forall l(l \in L \to \sum_j |(\mathbf{M} - \mathbf{V}^{(K)})_{lj}| < 1)$ and also we may construct an irreducible matrix from $\mathbf{M} - \mathbf{V}^{(K)}$ by replacing some of the zeros which have row indices in L with positive reals.$\}$

Quite how best to choose $K, \overline{\mathbf{v}}$ for these ends is as yet unknown.

The third point to be noted as regards theorem 1 concerns the possibility of $\mathbf{M}$ comprising negative entries. By the Perron–Frobenius theorem, we know that if $\mathbf{M}$ is non-negative then there is necessarily a unique (up to scalar multiples) vector $\overline{\mathbf{x}}$ and this vector comprises only positive entries. Theorem 1, does not necessitate that there be a vector $\overline{\mathbf{x}}$ if $\mathbf{M}$ comprises negative entries. Consider for example,

$$\begin{pmatrix} \frac{-1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{-1}{2} \end{pmatrix}. \tag{13}$$

(This matrix has no eigenvalue 1, so no corresponding vector $\overline{\mathbf{x}}$.) But, what theorem 1 does give us is that if there is a $\overline{\mathbf{x}}$ when $\mathbf{M}$ comprises negative entries, e.g.

$$\begin{pmatrix} \frac{1}{2} & \frac{-1}{2} \\ \frac{-1}{2} & \frac{1}{2} \end{pmatrix} \tag{14}$$

then this vector $\overline{\mathbf{x}}$ is unique (up to scalar multiples), the vector comprises only non-zero entries, and it can be derived by way of equation (2).

The fourth point of significance concerns PageRank computation. Let there be n Webpages and let $\deg(j)$ denote the number of hyperlinks from a Webpage j. Let $\mathbf{P} \in \mathbb{R}^{n\times n}$ be defined by:

$$P_{ij} := \begin{cases} \frac{1}{deg(j)} & \text{if there is a hyperlink from } j \text{ to } i \\ 0 & \text{otherwise} \end{cases} \tag{15}$$

Let $\overline{\mathbf{v}} \in \mathbb{R}^{n\times n}$ be such that $\forall i(v_i > 0)$ and $\sum_i v_i = 1$.

Let $\mathbf{D} \in \mathbb{R}^{n\times n}$ be defined by:

$$D_{ij} := \begin{cases} v_i & \text{if } \deg(j){=}0 \\ 0 & \text{otherwise} \end{cases} \tag{16}$$

Let $\mathbf{E} \in \mathbb{R}^{n\times n}$ be defined by:

$$E_{ij} := v_i \tag{17}$$

Then, for some fixed α $(0 < \alpha < 1)$, we may define PageRank vector $\overline{\mathbf{x}}$ by:

$$\overline{\mathbf{x}} := (\alpha\mathbf{P} + \alpha\mathbf{D} + (1-\alpha)\mathbf{E})\overline{\mathbf{x}}. \tag{18}$$

where $\forall i(x_i > 0)$ and $\sum_i x_i = 1$.

Now, the traditional solution method for $\overline{\mathbf{x}}$ is the power method ($(\alpha\mathbf{P} + \alpha\mathbf{D} + (1-\alpha)\mathbf{E})$ is easily shown to be irreducible and aperiodic). However, the matrix $(\alpha\mathbf{P} + \alpha\mathbf{D} + (1-\alpha)\mathbf{E})$ is completely dense, so this traditional method is anything but memory and operation-count efficient.

To improve matters, we may choose instead to employ iterative methods based on the following (because $\overline{\mathbf{x}}$ is positive and entries sum to one):

$$\overline{\mathbf{x}} := (\alpha\mathbf{P} + \alpha\mathbf{D})\overline{\mathbf{x}} + \overline{\mathbf{v}}. \tag{19}$$

But, if there are many dangling Webpages (Webpages j for which $\deg(j){=}0$), then $(\alpha\mathbf{P} + \alpha\mathbf{D})$ may still be troublingly dense.

It is here that theorem 1 assists (see [6] for an alternative way to address this problem). Using equation (19) and theorem 1, we know that there is some positive scalar multiple $\overline{\mathbf{y}}$ of $\overline{\mathbf{x}}$ which uniquely satisfies the following:

$$\begin{aligned} \overline{\mathbf{y}} &= (\alpha\mathbf{P} + \alpha\mathbf{D} + (1-\alpha)\mathbf{E} - \alpha\mathbf{D})\overline{\mathbf{y}} + \alpha\overline{\mathbf{v}} \\ &= (\alpha\mathbf{P} + (1-\alpha)\mathbf{E})\overline{\mathbf{y}} + \alpha\overline{\mathbf{v}} \\ &= \alpha\mathbf{P}\overline{\mathbf{y}} + (1-\alpha)\mathbf{E}\overline{\mathbf{y}} + \alpha\overline{\mathbf{v}} \\ &= \alpha\mathbf{P}\overline{\mathbf{y}} + \beta\overline{\mathbf{v}} + \alpha\overline{\mathbf{v}} && \text{, where } \beta \text{ is some positive scalar} \\ &= \alpha\mathbf{P}\overline{\mathbf{y}} + \gamma\overline{\mathbf{v}} && \text{, where } \gamma \text{ is some positive scalar} \end{aligned} \tag{20}$$

From the proof of theorem 1, we know that this β serves only as a scaling factor, so for convenience we we simply set it to 1. And, so, rather happily, we may employ iterative solutions methods corresponding to equation (20) with $\beta = 1$, and these methods will be free of the density impositions originally relating to matrices $\mathbf{D}$ and $\mathbf{E}$. (Our concern is relative PageRank, not absolute values. But, even were we concerned with absolute values, we would simply normalise the vector $\overline{\mathbf{y}}$ in (20) to obtain these.)

References

[1] Owe Axelsson. *Iterative solution methods.* Cambridge University Press, New York, NY, USA, 1994.

[2] D. Bertsekas and J. Tsitsiklis. Some aspects of parallel and distributed iterative algorithms – a survey. *Automatica*, 27(1):3–21, 1991.

[3] Dimitri P. Bertsekas and John N. Tsitsiklis. Convergence rate and termination of asynchronous iterative algorithms. In *ICS '89: Proceedings of the 3rd international conference on Supercomputing*, pages 461–470, New York, NY, USA, 1989. ACM Press.

[4] J. Mark Bull and T. L. Freeman. Numerical performance of an asynchronous Jacobi iteration. In *Conference on Algorithms and Hardware for Parallel Processing*, pages 361–366, 1992.

[5] D. Chazan and W. L. Miranker. Chaotic relaxation. *Linear Algebra and Its Applications*, 2:199–222, 1969.

[6] Gianna M. Del Corso, Antonio Gull, and Francesco Romani. Fast pagerank computation via a sparse linear system (extended abstract). In *Proceedings of Algorithms and Models for the Web-Graph: Third International Workshop, WAW 2004 (LNCS, 3243)*, 2004.

[7] G. Frobenius. Über matrizen aus nicht negativen elementen. S.-B. Preuss. Akad. Wiss. Berlin, 1912.

[8] Roger A. Horn and Charles R. Johnson. *Matrix analysis.* Cambridge University Press, New York, NY, USA, 1986.

[9] Boris Lubachevsky and Debasis Mitra. A chaotic asynchronous algorithm for computing the fixed point of a nonnegative matrix of unit spectral radius. *Journal of the Association for Computing Machinery*, 33(1):130–150, January 1986.

[10] Raul Naulin and Carlos Pabst. The roots of a polynomial depend continuously on its coefficients. *Revista Colombiana de Matemáticas*, 28(1):35–37, 1994.

[11] O. Perron. Zur theorie der matrizen. *Math. Ann*, 64:248–263, 1907.

[12] A. Trifunovic. *Parallel Algorithms for Hypergraph Partitioning.* PhD thesis, Imperial College London, February 2006.