Affine iterations on nonnegative vectors

V. Blondel, L. Ninove, P. Van Dooren
CESAME, Université catholique de Louvain
Av. G. Lemaître 4, B-1348, Louvain-la-Neuve, Belgium

1 Introduction

In this paper we consider three different iterations :

$$x_{k+1} = \frac{Ax_k}{\|Ax_k\|}, \tag{1}$$

$$x_{k+1} = \frac{Ax_k + b}{\|Ax_k + b\|}, \tag{2}$$

$$x_{k+1} = \frac{Ax_k + b}{y^T(Ax_k + b)}, \tag{3}$$

and analyze their fixed points and rate of convergence. The initial vector x_0 is positive, and the vectors b and y and the iteration matrix A are all nonnegative.

These iterations occur in the definition of the PageRank of a webpage [1] and in the calculation of the similarity matrix of two graphs, which was recently introduced by several authors [3, 6, 8]. In these applications, the iteration matrix A is nonnegative and very large but also very sparse because it is derived from the adjacency matrix of a very large and loosely connected graph (like the internet graph or a product graph of large databases). The iteration matrix A is either the adjacency matrix of a particular graph [1], or its symmetric part [3], or a weighted version of one of those [6, 8]. The fixed points in the positive quadrant provide answers to specific questions, but since the matrix A is very large one is also interested in understanding (and possibly influencing) the rate of convergence to these fixed points. In this short version of the paper we give the main ideas without proofs.

2 Eigenvector iteration

In the basic case of [1, 3, 8], the iteration being considered is the power method :

$$x_{k+1} = \frac{Ax_k}{\|Ax_k\|}, \quad x_0 \in \mathbb{R}^n_{>0}.$$

Since A is nonnegative the limiting vectors of this iteration are obviously in the closed positive quadrant $\mathbb{R}^n_{\geq 0}$ and must be eigenvectors of A. But if the starting vector $x_0 \in \mathbb{R}^n_{>0}$ then it can be shown that every limiting vector must have a component in the direction of the Perron space of A (with this we mean the eigenspace of the Perron root $\rho(A)$ of A). In other words, the reachable subspace of (A, x_0) contains eigenvectors of A that correspond to the Perron root $\rho(A)$, and since the starting vector has a component in this eigenspace, the only vectors one can hope to converge to are linear combinations of eigenvectors of A corresponding to eigenvalues of modulus equal to $\rho(A)$.

If A is symmetric, then there can only be two such eigenvalues, namely $\pm\rho(A)$, and it is shown in [3] that the even and odd iterates always converge to a fixed point. But the limit point of these subsequences depends on the initial vector x_0, and so this initial vector must be chosen appropriately. A good choice is the vector whose entries are all equal to 1, which guarantees that the limit point of the even iterates is the unique vector of largest 1-norm among the possible choices for x_0 (see [3]).

In the more general case the iteration will converge to the dominant subspace of the reachable subspace of A, and convergence will occur for every k-th step, where k is the greatest common divisor of the cycles in the dominant eigenspace of A.

It is well known that, when eigenvalues are distinct, the rate of convergence of the power method is given by $|\lambda_2/\lambda_1|$, the ratio between the module of the two first eigenvalues of A. For column-stochastic matrices, this rate of convergence can be improved by a low-rank correction

$$A_c := (1-c)A + c\,ve^T, \quad 0 < c < 1,$$

where v is a nonnegative vector whose 1-norm is equal to 1, and e is the vector of ones. This improved rate of convergence is important since the power method (or a Krylov-based variant) is typically used for computing the limiting vectors of (1) because of the size and sparsity pattern of the matrix A. Similarly, for an arbitrary matrix, we will consider in section 4 the scaled affine iteration (3) which is equivalent to a low rank correction of the matrix.

3 Normalized affine iteration

Since the power method (1) does not necessarily converge to a unique eigenvector when A has several dominant eigenvalues, it is useful to modify the algorithm in order to ensure a better convergence. In [8], the proposed variant is the affine normalized iteration (2) (where b is a positive vector), which is supposed to guarantee and speed up the convergence towards a fixed point not too far from an eigenvector of A. Unfortunately, there is no conclusive proof given in [8] about the convergence of this scheme.

Below, we use the notion of projective distance [9] to prove the existence of a unique fixed point and the global convergence of iteration (2) towards this point on the positive quadrant. For this we need to assume that $\|\cdot\|$ is a monotone norm, that the matrix A is *row-allowable* (i.e. has no zero rows) and the initial point x_0 is nonnegative with $\|x_0\| = 1$. We first recall Hilbert's projective distance between two positive vectors (see [9, chapter 3] for details).

Definition 1. The *projective distance* d_{pr} between positive vectors x and y is defined as

$$d_{\mathrm{pr}} \colon \mathbb{R}^n_{>0} \times \mathbb{R}^n_{>0} \to \mathbb{R} \colon (x, y) \mapsto \max_{i,j} \ln \frac{x_i/y_i}{x_j/y_j}.$$

For any vector norm $\|\cdot\|$, this defines a distance on the normalized positive quadrant $\left\{x \in \mathbb{R}^n_{>0} : \|x\| = 1\right\}$. Using this, we obtain the following result, due to Birkhoff [2, 9].

Lemma 1. *Let $A \in \mathbb{R}^{n\times n}_{\geq 0}$ be a row-allowable nonnegative matrix, then for all $x, y \in \mathbb{R}^n_{>0}$,*

$$0 \leq \tau_B(A) := \sup_{\substack{x,y\neq 0\\ x\neq \lambda y}} \frac{d_{\mathrm{pr}}(Ax, Ay)}{d_{\mathrm{pr}}(x, y)} \leq 1.$$

Moreover, if $A > 0$ then $\tau_B(A) < 1$.

The number $\tau_B(A)$ is often called *Birkhoff's contraction coefficient* of A.

Definition 2. A vector norm $\|\cdot\| : \mathbb{R}^n \to \mathbb{R}_{\geq 0}$ is said to be *monotone* if $\|x\| > \|y\|$ for all $x, y \in \mathbb{R}^n_{\geq 0}$ such that $x > y$.

For $A \in \mathbb{R}^{n\times n}_{\geq 0}$, $b \in \mathbb{R}^n_{\geq 0}$ and $\|\cdot\|$ a monotone norm on $\mathbb{R}^n$, we define the *hyperbolic contraction coefficient*

$$\tau_H(A, b, \|\cdot\|) = \frac{c_{\|\cdot\|}\, |||A|||}{c_{\|\cdot\|}\, |||A||| + \min_i b_i},$$

where $\mathbf{e}_i$ is the i^{th} column of the identity matrix, $c_{\|\cdot\|} = (\min_i \|\mathbf{e}_i\|)^{-1}$ is a coefficient dependent on the considered norm and $\|\|\cdot\|\| : \mathbb{R}^{n\times n} \to \mathbb{R} : A \mapsto \max_{\|x\|=1} \|Ax\|$ is the induced matrix norm.

Lemma 2. *Let $A \in \mathbb{R}^{n\times n}_{\geq 0}$, $b \in \mathbb{R}^n_{\geq 0}$ and $\|\cdot\|$ a monotone norm on $\mathbb{R}^n$. Then, for all $x \in \mathbb{R}^n_{\geq 0}$ such that $\|x\| = 1$ and $Ax + b > 0$,*

$$\max_i \frac{(Ax)_i}{(Ax)_i + b_i} \leq \tau_H(A, b, \|\cdot\|).$$

Obviously, $\tau_H(A, b, \|\cdot\|) \leq 1$. Moreover $\tau_H(A, b, \|\cdot\|) < 1$ if and only if $b > 0$.

The following theorem shows that, if $b > 0$ then $x \mapsto (Ax+b)/\|Ax+b\|$ is a contractive map on the positive quadrant.

Theorem 1. *Let $\|\cdot\|$ be a monotone norm on $\mathbb{R}^n$ and $\mathcal{S}$ be the open bounded set $\left\{x \in \mathbb{R}^n_{>0} : \|x\| = 1\right\}$ with the projective distance d_{pr}. Let $A \in \mathbb{R}^{n\times n}_{\geq 0}$ be a nonnegative row-allowable matrix and let $b \in \mathbb{R}^n_{>0}$ be a positive vector. Then*

$$f : \mathcal{S} \to \mathcal{S} : x \mapsto \frac{Ax + b}{\|Ax + b\|}$$

is a contractive map, i.e. $\sup_{x\neq y\in\mathcal{S}} \frac{d_{\text{pr}}(f(x), f(x))}{d_{\text{pr}}(x,y)} < 1$.

Banach's fixed point theorem then leads to existence and uniqueness of a fixed point.

Theorem 2. *Let $\|\cdot\|$ be a monotone norm and $\bar{\mathcal{S}} = \left\{x \in \mathbb{R}^n_{\geq 0} : \|x\| = 1\right\}$. Let $A \in \mathbb{R}^{n\times n}_{\geq 0}$ be a nonnegative row-allowable matrix and let $b \in \mathbb{R}^n_{>0}$ be a positive vector. Then*

$$f : \bar{\mathcal{S}} \to \bar{\mathcal{S}} : x \mapsto \frac{Ax + b}{\|Ax + b\|}$$

has one and only one fixed point x^* in $\bar{\mathcal{S}}$ and, whatever $x_0 \in \bar{\mathcal{S}}$, the sequence $f(x_0)$, $f(f(x_0))$, ..., $f^k(x_0)$, ..., converges towards x^* with respect to the projective distance d_{pr}. Moreover, x^* is a positive vector.

The following result is a direct consequence of theorem 2.

Theorem 3. *Let $\|\cdot\|$ be a monotone norm and $\bar{\mathcal{S}} = \left\{x \in \mathbb{R}^n_{\geq 0} : \|x\| = 1\right\}$. Let $A \in \mathbb{R}^{n\times n}_{\geq 0}$ be a nonnegative row-allowable matrix and let $b \in \mathbb{R}^n_{\geq 0}$ be a nonnegative and nonzero vector. Then, if A or b is positive,*

$$f : \bar{\mathcal{S}} \to \bar{\mathcal{S}} : x \mapsto \frac{A(x + b)}{\|A(x + b)\|}$$

has one and only one fixed point x^* in $\bar{\mathcal{S}}$ and, whatever $x_0 \in \bar{\mathcal{S}}$, the sequence $f(x_0)$, $f(f(x_0))$, ..., $f^k(x_0)$, ..., converges towards x^* with respect to the projective distance d_{pr}.

We conclude with some remarks on these theorems.

Remark 1. The fixed point of the iterations of theorems 2 and 3 are dependent on the chosen norm.

Remark 2. If A and b are nonnegative but are not positive, then f is no longer necessarily contractive. Take for example $A = \left(\begin{smallmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{smallmatrix}\right)$ and $b = \left(\begin{smallmatrix} 0 \\ 0 \\ 0.1 \end{smallmatrix}\right)$. Then, for $x = \frac{\sqrt{6}}{6}\left(\begin{smallmatrix} 1 \\ 2 \\ 1 \end{smallmatrix}\right)$ and $y = \frac{\sqrt{6}}{6}\left(\begin{smallmatrix} 2 \\ 1 \\ 1 \end{smallmatrix}\right)$,

$$d_{\mathrm{pr}}(Ax+b, Ay+b) = d_{\mathrm{pr}}(x,y) = \ln 4.$$

Remark 3. Let us see now that a monotone norm is necessary. Let $A = \left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$, $b = \left(\begin{smallmatrix} 0.1 \\ 1 \end{smallmatrix}\right)$ and

$$\|\cdot\| : \mathbb{R}^2 \to \mathbb{R}^2 \colon (x_1, x_2) \mapsto 4\,|x_1 - x_2| + |x_1 + x_2|\,.$$

For $x = \left(\begin{smallmatrix} 0.25 \\ 0.35 \end{smallmatrix}\right)$ and $y = \left(\begin{smallmatrix} 0.5 \\ 0.5 \end{smallmatrix}\right)$ ($\|x\| = \|y\| = 1$),

$$d_{\mathrm{pr}}(Ax+b, Ay+b) = \ln \frac{54}{35} > d_{\mathrm{pr}}(x,y) = \ln \frac{49}{35},$$

so the map is not contractive.

Remark 4. In general, we do not have $d_{\mathrm{pr}}(Ax+b, Ay+b) \leq d_{\mathrm{pr}}(Ax, Ay)$. For example, take $A = \left(\begin{smallmatrix} 10 & 1 \\ 10 & 1 \end{smallmatrix}\right) > 0$, $b = \left(\begin{smallmatrix} 0.1 \\ 1 \end{smallmatrix}\right) > 0$ and $\|\cdot\|_\infty$ which is a monotone norm. For $x = \left(\begin{smallmatrix} 1 \\ 0.1 \end{smallmatrix}\right)$ and $y = \left(\begin{smallmatrix} 0.1 \\ 1 \end{smallmatrix}\right)$,

$$d_{\mathrm{pr}}(Ax+b, Ay+b) = 0.2721 > d_{\mathrm{pr}}(Ax, Ay) = 0.$$

4 Scaled affine iteration

We now consider the following variant of the normalized affine iteration (2)

$$x_{k+1} = \frac{Ax_k + b}{y^T(Ax_k + b)}, \tag{4}$$

where $A \in \mathbb{R}^{n\times n}$, $b \in \mathbb{R}^n$ and where the scaling of the iterates are done according to a nonzero vector $y \in \mathbb{R}^n$, instead of a monotone norm.

Notice that if $y > 0$ then we can define a monotone norm $\|\cdot\|$ on $\mathbb{R}^n_{\geq 0}$ such that $\|x\| = y^T x$ for all vector $x \in \mathbb{R}^n_{\geq 0}$. If moreover $b > 0$ and A is a nonnegative row-allowable matrix, then theorem 2 ensures the convergence of the iterates of (4) towards the unique fixed point.

To analyze the general case, where A, b and y are not necessarily nonnegative, we express the iteration (4) as

$$r_{k+1}\begin{pmatrix} x_{k+1} \\ 1 \end{pmatrix} = \begin{pmatrix} A & b \\ y^T A & y^T b \end{pmatrix}\begin{pmatrix} x_{k+1} \\ 1 \end{pmatrix},$$

with $r_{k+1} = y^T(Ax_k + b)$. If we assume that $r_{k+1} \neq 0$ for all $k \in \mathbb{N}$, the $(n+1) \times (n+1)$ iteration matrix admits the factorization

$$\begin{pmatrix} A & b \\ y^T A & y^T b \end{pmatrix} = \begin{pmatrix} I_n & 0 \\ y^T & 1 \end{pmatrix}\begin{pmatrix} A + by^T & b \\ 0 & 0 \end{pmatrix}\begin{pmatrix} I_n & 0 \\ -y^T & 1 \end{pmatrix},$$

which implies that the spectrum of this matrix is that of $A + by^T$ plus the zero eigenvalue. With $y^T x_k = 1$ for all $k \in \mathbb{N}$, we find that

$$x_{k+1} = \frac{(A + by^T)x_k}{y^T(A + by^T)x_k},$$

which is essentially the power method on the matrix $A + by^T$. In order to analyze this correctly one needs to consider the unobservable subspace

$$\mathcal{O}_c(A, y^T) := \operatorname{Ker} \begin{pmatrix} y^T \\ y^T A \\ \vdots \\ y^T A^{n-1} \end{pmatrix}$$

which has dimension $d < n$. The iteration (4) can be rewritten in a coordinate system where

$$\begin{aligned} \hat{A} &= \begin{pmatrix} A_{11} & 0 \\ A_{21} & A_{22} \end{pmatrix} \quad \hat{b} = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} \\ \hat{y}^T &= \begin{pmatrix} y_1^T & 0 \end{pmatrix}, \end{aligned} \tag{5}$$

and where the subsystem (A_{11}, y_1^T) is observable. For convenience, we suppose that A, b and y already have the structure of (5). We then find the following result.

Proposition 3. *Let A, b and y have the block-decomposition* (5)*. Suppose that $A_{11} + b_1 y_1^T$ and A_{22} have no common eigenvalues, that $A + by^T$ has a basis of eigenvectors, that x_0 has a nonzero component in the direction of a* *dominant eigenvector of $A + by^T$, and that $y^T(Ax_k + b) \neq 0$ for all $k \in \mathbb{N}$. If $\rho(A_{22}) < \rho(A_{11} + b_1 y_1^T)$ (the observable part is dominant) then subsequences of the iterates* (4) *tend to vectors which are not in the space $\mathcal{O}_c(A, y^T)$. Moreover, if $A + by^T$ has only one dominant eigenvector, then the iterates converge to this eigenvector scaled according to the vector y. If $\rho(A_{22}) > \rho(A_{11} + b_1 y_1^T)$ (the unobservable part is dominant) then the iterates* (4) *diverge.*

We also have the following.

Proposition 4. *Let A, b and y have the block-decomposition* (5)*, and suppose that A_{11} and A_{22} have no common eigenvalues. If the iterates* (4) *diverge to infinity then x_0 or b are not orthogonal to the left invariant subspace of A corresponding to the spectrum $\Lambda(A_{22})$.*

Note that, in order to have $y^T(Ax_k + b) \neq 0$ for $k \in \mathbb{N}$, it is necessary that the vectors x_0 and b do not lie in $\mathcal{O}_c(A, y^T)$.

Remark 5. If we consider the iteration (4) on the nonnegative quadrant with $A, b, y, x_0 \geq 0$ and $y^T x_0 = 1$, then, if $y^T b \neq 0$, we will never have $y^T(Ax_k + b) = 0$ for some $k \in \mathbb{N}$. On the other hand, if $y^T b = 0$, then $y^T(Ax_k + b) = 0$ may occur. Consider for example $A = \left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right)$, $b = \left(\begin{smallmatrix} 0 \\ 1 \end{smallmatrix}\right)$ and $y = \left(\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right)$, and take $x_0 = \left(\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right)$. Then $y^T(Ax_0 + b) = 0$.

5 Linearized affine iteration

We now try to link sections 3 and 4. Let A, b and $\|\cdot\|$ satisfy the hypotheses of theorem 2, and let $x > 0$ be the unique fixed point on the nonnegative quadrant of the iteration

$$x_{k+1} = \frac{Ax_k + b}{\|Ax_k + b\|}.$$

For that x, there exists a dual vector $y_x \geq 0$ (not necessarily unique) such that $y_x^T x = 1$ and the dual norm of y_x is equal to 1 (see [5]). Let $r_x = \|Ax + b\|$, and let

$$\bar{A}_x = \begin{pmatrix} A & b \\ y_x^T A & y_x^T b \end{pmatrix}$$

then

$$r_x \begin{pmatrix} x \\ 1 \end{pmatrix} = \bar{A}_x \begin{pmatrix} x \\ 1 \end{pmatrix}$$

at the fixed point x. It then follows from the nonnegativity of the matrices and vectors that

$$\rho(A) < r_x = \rho(\bar{A}_x).$$

The speed of convergence of the affine normalized iteration near the fixed point can be characterized by the ratio $|\lambda_2| / r_x$, where λ_2 is the second eigenvalue of $A+by_x^T$, and where r_x is strictly bigger than the spectral radius of the matrix A.

The advantage of the normalized affine iteration (2) over the standard power method (1) is that convergence is guaranteed, even if A has several eigenvalues of maximum module. Moreover, since $\rho(\bar{A}_x) > \rho(A)$, it seems plausible that the ratio of the first two eigenvalues of $\bar{A}_x$ is often better than that of A. This was observed experimentally by the authors of [8].

However, convergence speed does not always improve with an affine term b, as shown in the following example. Let $A = \left(\begin{smallmatrix} 1 & 0 \\ 0.1 & 0 \end{smallmatrix}\right)$, $b = \left(\begin{smallmatrix} 0.2 \\ 1 \end{smallmatrix}\right)$, $x_0 = \left(\begin{smallmatrix} 0.1 \\ 1 \end{smallmatrix}\right)$ and let $\|\cdot\|$ be the 1-norm (with this norm, the dual vector is $y = \left(\begin{smallmatrix} 1 \\ 1 \end{smallmatrix}\right)$ for all point of $\mathbb{R}^n_{\geq 0}$). Then, the spectrum $\Lambda(A + by^T) = \{1.58; 0.62\}$ whereas $\Lambda(A) = \{1; 0\}$, and the affine normalized iteration needs several steps to give a good approximation of the fixed point even though the power method converges in only one iteration to the fixed point. Let us also notice that the fixed points obtained with these two methods can be quite different.

Remarks

After the submission of this paper, we found out that theorems of section 3 are consequences of more general results on nonlinear mappings on cones, also based on the Hilbert's projective distance, see [7].

Moreover, Yurii Nesterov proposed us a simple alternative proof of existence and uniqueness of the fixed point of theorem 2.

Acknowledgements

This paper presents research results of the Belgian Programme on Interuniversity Attraction Poles, initiated by the Belgian Federal Science Policy Office. The scientific responsibility rests with its authors.
The second author was supported by the Fonds National de la Recherche Scientifique.

References

[1] S. Brin, L. Page, *The Anatomy of a Large-Scale Hypertextual Web Search Engine*, Proceedings of the Seventh International Web Conference (WWW98), 1998.

[2] G. Birkhoff, *Extensions of Jentzsch's theorem*, Trans. Amer. Math. Soc., 85 (1957), 219-227.

[3] V. Blondel, A. Gajardo, M. Heymans, P. Senellart, P. Van Dooren, *A measure of similarity between graph vertices: Applications to synonym extraction and web searching*, SIAM Review, to appear.

[4] T. Haveliwala, S. Kamvar, *The second eigenvalue of the Google matrix*, Stanford University Technical Report, 2003.

[5] R. Horn, Ch. Johnson, *Matrix analysis*, Corrected reprint of the 1985 original, Cambridge University Press, Cambridge, 1990.

[6] G. Jeh, J. Widom, *SimRank: A measure of structural-context similarity*, Proc. of the KDD2002 Conf., Edmonton (2002)

[7] U. Krause, *A nonlinear extension of the Birkhoff-Jentzsch theorem*, J. Math. Anal. Appl., 114 (1986), 552-568.

[8] S. Melnik, H. Garcia-Molina and A. Rahm, *Similarity Flooding: A Versatile Graph Matching Algorithm and its Application to Schema Matching*, Proc. 18th ICDE Conf., San Jose (2002).

[9] E. Seneta, *Nonnegative matrices and Markov chains*, second ed., Springer Series in Statistics, Springer-Verlag, New York, 1981.