On the Solution of Linear Mean Recurrences

David Borwein* Jonathan M. Borwein† Brailey Sims ‡

October 26, 2012

Abstract

Motivated by questions of algorithm analysis, we provide several distinct approaches to determining convergence and limit values for a class of linear iterations.

1 Introduction

Problem I. *Determine the behaviour of the sequence defined recursively by,*

$$x_n := \frac{x_{n-1} + x_{n-2} + \cdots + x_{n-m}}{m} \qquad \text{for } n \geq m+1 \tag{1}$$

and satisfying the initial conditions

$$x_k = a_k, \qquad \text{for } k = 1, 2, \cdots, m, \tag{2}$$

where $a_1, a_2, \cdots, a_m$ *are given real numbers.*

This problem was encountered by Bauschke, Sarada and Wang [1] while examining algorithms to compute zeroes of maximal monotone operators in optimization. Questions they raised concerning its resolution motivated our ensuing consideration of various approaches whereby it might be addressed.

*Department of Mathematics, University of Western Ontario, London, ON, Canada. Email: dborwein@uwo.ca.

†Centre for Computer-assisted Research Mathematics and its Applications (CARMA), School of Mathematical and Physical Sciences, University of Newcastle, Callaghan, NSW 2308, Australia. Email: jonathan.borwein@newcastle.edu.au, jborwein@gmail.com

‡Centre for Computer-assisted Research Mathematics and its Applications (CARMA), School of Mathematical and Physical Sciences, University of Newcastle, Callaghan, NSW 2308, Australia. Email: brailey.sims@newcastle.edu.au.

We suspect that, like us, the first thing most readers do when presented with a discrete iteration is to try to solve for the limit, call it L, by taking the limit in (1). Supposing the limit to exist we deduce

$$L = \frac{\overbrace{L + L + \cdots + L}^{m}}{m} = L, \tag{3}$$

and learn nothing—at least not about the limit. There is a clue in that the result is vacuous in large part because it involves an average, or *mean*.

In the next three sections, we present three quite distinct approaches. While at least one will be familiar to many readers, we suspect not all three will be. Each has its advantages, both as an example of more general techniques and as a doorway to a beautiful corpus of mathematics.

2 Spectral solution

We start with what may well be the best known approach. It may be found in many linear algebra courses often along with a discussion of the Fibonacci numbers: $F_n = F_{n-1} + F_{n-2}$ with $F_0 = 0, F_1 = 1$.

Equation (1) is an example of a *linear homogeneous recurrence relation of order* m with constant coefficients. Standard theory, see for example [5, Chapter 13.2, p. 252] or [9, Section 12.5 on page 90], runs as follows.

Theorem 2.1 (Linear recurrences). *The general solution of a linear recurrence*

$$x_n = \sum_{k=1}^{m} \alpha_k x_{n-k}$$

with constant coeffients, has the form

$$x_n = \sum_{k=1}^{l} q_k(n)\, r_k^n \tag{4}$$

where the r_k *are the* l *distinct roots of the characteristic polynomial*

$$p(r) := r^m - \sum_{k=1}^{m} \alpha_k r^{k-1}, \tag{5}$$

with algebraic multiplicity m_k *and* q_k *are polynomials of degree at most* $m_k - 1$*.*

Typically, elementary books only consider simple roots but we shall use a little more.

2.1 Our equation analysed

The linear recurrence relation specified by equation 1 has characteristic polynomial:

$$\begin{aligned} p(r) &:= r^m - \frac{1}{m}(r^{m-1} + r^{m-2} + \cdots + r + 1) \\ &= \frac{mr^{m+1} - (m+1)r^m + 1}{m(r-1)} \end{aligned} \tag{6}$$

with roots $r_1 = 1, r_2, r_3, \ldots, r_m$. Since

$$p'(1) = m - \frac{1}{m}\sum_{n=1}^{m-1} n = m - \frac{m-1}{2} = \frac{m+1}{2}$$

the root at 1 is simple.

We next show that if $p(r) = 0$ and $r \neq 1$, then $|r| < 1$. We argue as follows. We know from (6) that $p(r) = 0$ if and only if

$$r + \frac{1}{mr^m} = 1 + \frac{1}{m}. \tag{7}$$

If $|r| > 1$, then

$$\left| r + \frac{1}{mr^m} \right| \leq |r| + \frac{1}{m|r|^m} < 1 + \frac{1}{m},$$

since the function $f(x) := x + \frac{1}{mx^m}$ is strictly increasing for real $x > 1$ and $f(1) = 1 + \frac{1}{m}$. Thus $p(r) \neq 0$ when $|r| > 1$. Suppose therefore that $p(r) = 0$ with $r = e^{i\theta}$, $0 \leq \theta < 2\pi$. Then by (7) we must have

$$\cos(\theta) + \frac{\cos(-m\theta)}{m} = 1 + \frac{1}{m},$$

which is only possible when $\theta = 0$.

By (4) we must have

$$x_n = c_1 + \sum_{k=2}^{r} q_k(n)\, r_k^n \tag{8}$$

where r_k lies in the open unit disc for $2 \leq k \leq m$. Thus, the limit in (8) exists and equals $c_1 = q_k(1)$, the constant polynomial coefficient of the eigenvalue 1.

2.2 Identifying the limit

In fact we may use (6) to see all roots are simple. It follows from (6) that

$$((1-r)p(r))' = (m+1)r^{m-1}(1-r),$$

and hence that the only possible multiple root of p is $r_1 = 1$ which we have already shown to be simple, and so the solution is actually of the form

$$x_n = c_1 + \sum_{k=2}^{m} c_k\, r_k^n, \quad \text{with } c_1, \cdots, c_m \text{ constants.} \tag{9}$$

Observe now that if r is any of the roots $r_2, r_3, \ldots, r_m$, then

$$\sum_{n=1}^{m} n r^n = \frac{m r^{m+2} - (m+1) r^{m+1} + r}{(r-1)^2} = \frac{m\, r p(r)}{r-1} = 0, \tag{10}$$

and so multiplying (9) by n and summing from $n = 1$ to m we obtain

$$c_1 = \frac{2}{m(m+1)} \sum_{n=1}^{m} n a_n. \tag{11}$$

Thence, we do have convergence and the limit $L = c_1$ is given by (11).

Example 2.2 (The weighted mean)**.** We may perform the same analysis, if the arithmetic average in (1) is replaced by any weighted arithmetic mean

$$W_{(\alpha)}(x_1, x_2, \cdots, x_m) := \alpha_1 x_1 + \alpha_2 x_2 + \cdots + \alpha_m x_m$$

for strictly positive weights $\alpha_k > 0$ with $\sum_{k=1}^{m} \alpha_k = 1$. Then $W_{(1/m)} = A$ is the arithmetic mean of Problem I. As is often the case, the analysis becomes easier when we generalize. The recurrence relation in this case is

$$x_n = \alpha_m x_{n-1} + \alpha_{m-1} x_{n-2} + \cdots + \alpha_1 x_{n-m}$$

for $n \geq m + 1$, with *companion matrix*

$$A_m := \begin{bmatrix} \alpha_m & \alpha_{m-1} & \cdots & \alpha_2 & \alpha_1 \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \cdots & \cdots & 1 & 0 & 0 \\ 0 & 0 & \cdots & 1 & 0 \end{bmatrix}. \tag{12}$$

The corresponding characteristic polynomial of the recurrence relation

$$p(r) := r^m - \left(\alpha_m r^{m-1} + \alpha_{m-1} r^{m-2} + \cdots + \alpha_2 r^1 + \alpha_1\right)$$

is also the characteristic polynomial of the matrix.

Clearly $p(1) = 0$. Now suppose r is a root of p and set $\rho := |r|$. Then the triangle inequality and the mean property of $W_{(\alpha)}$ imply that

$$\rho^m \leq \sum_{k=1}^{m} \alpha_k \rho^{k-1} \leq \max_{1\leq k\leq m} \rho^{k-1}, \tag{13}$$

and so $0 \leq \rho \leq 1$.

If $\rho = 1$ but $r \neq 1$ then $r = e^{i\theta}$ for $0 < \theta < 2\pi$ and, on observing that $r^{-m}p(r) = 0$ and equating real parts, we get

$$1 = \sum_{k=1}^{m} \alpha_k e^{i(k-m-1)\theta} = \sum_{k=1}^{m-1} \alpha_k \cos((m+1-k)\theta) + \alpha_m \cos(\theta)$$

whence $\cos(\theta) = 1$ which is a contradiction. [Alternatively we may note that the modulus is a strictly convex function, whence $\exp(i\theta) = 1$ which is again a contradiction.] Thence, all roots other than 1 have modulus strictly less than 1.

Finally, since $p'(1) = m - \sum_{k=1}^{m}(k-1)\alpha_k \geq m - (m-1)\sum_{k=1}^{m} \alpha_k = 1$ the root at 1 is still simple. Moreover, if $\sigma_k := \alpha_1 + \alpha_2 + \cdots + \alpha_k$, then

$$p(r) = (r-1) \sum_{k=1}^{m} \sigma_k r^{k-1}. \tag{14}$$

Hence, p has no other positive real root. In particular, from (4) we again have

$$x_n = L + \sum_{k=2}^{r} q_k(n)\, r_k^n = L + \varepsilon_n$$

where $\varepsilon_n \to 0$ since the root at 1 is simple while all other roots are strictly inside the unit disc—but need not be simple as illustrated in Example 2.4. $\triangleleft$

Remark 2.3. An analysis of the proof in Example 2.2 shows that the conclusions continue to hold for non-negative weights as long as the highest-order term $\alpha_m > 0$. $\triangleleft$

Example 2.4 (A weighted mean with multiple roots)**.** The polynomial

$$p(r) = r^6 - \frac{r^5 + r^4 + 16\,r^3 + 18\,r^2 + 45\,r + 81}{162} \tag{15}$$

$$= \frac{1}{162}\,(2\,r+1)\,(r-1)\left(1+9\,r^2\right)^2, \tag{16}$$

has a root at one and a repeated pair of conjugate roots at $\pm\frac{i}{3}$. Nonetheless, the weighted mean iteration

$$x_n = \frac{81\,x_{n-6} + 45\,x_{n-5} + 18\,x_{n-4} + 16\,x_{n-3} + x_{n-2} + x_{n-1}}{162}$$

is covered by the development of Example 2.2. The limit is

$$L := \frac{162\,a_6 + 161\,a_5 + 160\,a_4 + 144\,a_3 + 126\,a_2 + 81\,a_1}{834}. \tag{17}$$

Once found, this is easily checked (in a computer algebra system) from the Invariance principle of the next section. In fact the coefficients were found by looking in *Maple* at the thousandth power of the corresponding matrix and converting the entries to be rational.

The polynomial was constructed by examining how to place repeated roots on the imaginary axis while preserving increasing coefficients as required in (14). One general potential form is then $p(\sigma,\tau) := (r-1)(r+\sigma)(r^2+\tau^2)^2$ and we selected $p(1/2,1/3)$. In the same fashion

$$p\left(\frac{1}{2},\frac{1}{2}\right) = r^6 - \frac{16\,r^5 + 8\,r^3 + 6\,r^2 + r + 1}{32},$$

in which r^4 has a zero coefficient, but the corresponding iteration remains well behaved, see Remark 2.3. ◁

We will show in Example 3.3 that the approach of the next section provides the most efficient way of identifying the limit in this generalization. (In fact, we shall discover that the numerator coefficients in (17) are the partial sums of those in (15).) Example 3.3 also provides a quick way to check the assertions about limits in the next example.

Example 2.5 (Limiting examples I)**.** Consider first

$$A_3 := \begin{bmatrix} \frac{1}{2} & 0 & \frac{1}{2} \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}.$$

The corresponding iteration is $x_n = (x_{n-1} + x_{n-3})/2$ with limit $a_1/4 + a_2/4 + a_3/2$. By comparison, for

$$A_3 := \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}.$$

the corresponding iteration is $x_n = (x_{n-1} + x_{n-2})/2$ with limit $(a_1 + 2a_2)/3$. This can also be deduced by considering Problem I with $m = 2$ and ignoring the third row and column. The third permutation

$$A_3 := \begin{bmatrix} 0 & \frac{1}{2} & \frac{1}{2} \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}.$$

corresponding to the iteration $x_n = (x_{n-2} + x_{n-3})/2$ has limit $(a_1 + 2a_2 + 2a_3)/5$.

Finally,

$$A_3 := \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$$

has $A_3^3 = I$ and so is A_3^k is periodic of period three as is obvious from the iteration $x_n = x_{n-3}$.

We return to these matrices in Example 4.7 of the penultimate section. $\triangleleft$

3 Mean iteration solution

The second approach, based on [3, Section 8.7], deals very efficiently with equation 1; as a bonus the proof we give below of convergence holds for nonlinear means given positive starting values.

We say a real valued function of M is a *strict m-variable mean* if

$$\min(x_1, x_2, \cdots, x_m) \leq M(x_1, x_2, \cdots, x_m) \leq \max(x_1, x_2, \cdots, x_m)$$

with equality only when all variables are equal.

We observe that when M is a weighted arithmetic mean we may take its domain to be $\mathbb{R}^m$, however certain nonlinear means —such as $G := (x_1 x_2 \cdots x_m)^{1/m}$—are defined only for positive values of the variables.

3.1 Convergence of mean iterations

In the language of [3, Section 8.7], we have the following:

Theorem 3.1 (Convergence of a mean iteration)**.** *Let M be any* strict m-variable mean *and consider the iteration*

$$x_n := M(x_{n-m}, x_{n-m+1}, \cdots, x_{n-1}), \tag{18}$$

so when $M = A$ we recover the iteration in (1)*. Then x_n converges to a finite limit $L(x_1, x_2, \ldots, x_m)$.*

Proof. Indeed, specialization of [3, Exercise 7 of Section 8.7] actually establishes convergence for an arbitrary strict mean; but let us make this explicit for this case.

Let $\overline{x}_n := (x_n, x_{n-1}, \cdots, x_{n-m+1})$ and let $a_n := \max \overline{x}_n, b_n := \min \overline{x}_n$. As noted above, for general means we need to restrict the variables to non-negative values, but for linear means no such restriction is needed. Then for all n, the mean property implies that

$$a_{n-1} \geq a_n \geq b_n \geq b_{n-1}. \tag{19}$$

Thus, $a := \lim_n a_n$ and $b := \lim_n b_n$ exist with $a \geq b$. In particular $\overline{x}_n$ remains bounded. Select a subsequence $\overline{x}_{n_k}$ with $\overline{x}_{n_k} \to \overline{x}$. It follows that

$$b \leq \min \overline{x} \leq \max \overline{x} \leq a \tag{20}$$

while

$$b = \min M(\overline{x}) \quad \text{and} \quad \max M(\overline{x}) = a. \tag{21}$$

Since M is a strict mean we must have $a = b$ and the iteration converges. □

It is both here and in Theorem 3.2 that we see the power of identifying the iteration as a mean iteration.

3.2 Determining the limit

In what follows a mapping $L : \mathbb{D}^n \to \mathbb{R}$, where $\mathbb{D} \subseteq \mathbb{R}$, is said to be a diagonal mapping if $L(x, x, \cdots, x) = x$ for all $x \in D$.

Theorem 3.2 (Invariance principle [3])**.** *For any mean iteration, the limit L is necessarily a mean and is the unique diagonal mapping satisfying the* Invariance principle*:*

$$L\left(x_{n-m}, x_{n-m+1}, \ldots, x_{n-1}\right) = L\left(x_{n-m+1}, \ldots, x_{n-1}, M(x_{n-m}, x_{n-m+1}, \ldots, x_{n-1})\right). \tag{22}$$

Moreover, L is linear whenever M is.

Proof. We sketch the proof (details may again be found in [3, Section 8.7]). One first checks that the limit, being a pointwise limit of means is itself a mean and so is continuous on the diagonal.

The principle follows since,

$$L(\overline{x}_m) = \cdots = L(\overline{x}_n) = L(\overline{x}_{n+1}) = L(\lim_n \overline{x}_n) = \lim_n (x_n).$$

We leave it to the reader to show that L is linear whenever M is. □

We note that we can mix-and-match arguments—if we have used the ideas of the previous section to convince ourselves that the limit exists, the Invariance principle is ready to finish the job.

Example 3.3 (A general strict linear mean)**.** If we suppose that $M(y_1, \ldots, y_m) = \sum_{i=1}^m \alpha_i y_i$, with all $\alpha_i > 0$, and that $L(y_1, \ldots, y_m) = \sum_{i=1}^m \lambda_i y_i$ are both linear, we may solve (22) to determine that for $k = 1, 2, \ldots m-1$ we have

$$\lambda_{k+1} = \lambda_k + \lambda_m \alpha_{k+1}. \tag{23}$$

Whence, on setting $\sigma_k := \alpha_1 + \cdots + \alpha_k$, we obtain

$$\frac{\lambda_k}{\lambda_m} = \sigma_k. \tag{24}$$

Further, since L is a linear mean we have $1 = L(1, 1, \ldots, 1) = \Sigma_{k=1}^m \lambda_k$, whence, summing (3.3) from $k = 1$ to m yields, $\frac{1}{\lambda_m} = \Sigma_{k=1}^m \sigma_k$ and so becomes,

$$\lambda_k = \frac{\sigma_k}{\sum_{k=1}^m \sigma_k}. \tag{25}$$

In particular, setting $\alpha_k \equiv \frac{1}{m}$ we compute that $\sigma_k = \frac{k}{m}$ and so $\lambda_k = \frac{2k}{m(m+1)}$ as was already determined in (11) of the previous section. $\triangleleft$

Example 3.4 (A nonlinear mean)**.** We may replace A by the *Hölder mean*

$$H_p(x_1, x_2, \ldots, x_m) := \left(\frac{1}{m}\sum_{i=1}^m x_i^p\right)^{1/p}$$

for $-\infty < p < \infty$. The limit will be $\left(\sum_{k=1}^m \lambda_k a_k^p\right)^{1/p}$, with λ_k as in (25). In particular with $p = 0$ (taken as a limit) we obtain in the limit the weighted geometric mean $G(a_1, a_2, \cdots, a_m) = \prod_{k=1}^m a_k^{\lambda_k}$. We also apply the same considerations to weighted Hölder means. $\triangleleft$

We conclude this section with an especially neat application of the arithmetic Invariance principle to an example by Carlson [3, Section 8.7].

Example 3.5 (Carlson's logarithmic mean)**.** Consider the iterations with $a_0 := a > 0, b_0 := b > a$ and

$$a_{n+1} = \frac{a_n + \sqrt{a_n b_n}}{2}, \qquad b_{n+1} = \frac{b_n + \sqrt{a_n b_n}}{2},$$

for $n \geq 0$. In this case convergence is immediate since $|a_{n+1} - b_{n+1}| = |a_n - b_n|/2$.

If asked for the limit, you might make little progress. But suppose you are told the answer is the *logarithmic mean*

$$\mathcal{L}(a, b) := \frac{a - b}{\log a - \log b},$$

for $a \neq b$ and a (the limit as $a \to b$) when $a = b > 0$. We check that

$$\mathcal{L}(a_{n+1}, b_{n+1}) = \frac{a_n - b_n}{2\,\log \frac{a_n + \sqrt{b_n a_n}}{b_n + \sqrt{b_n a_n}}} = \mathcal{L}(a_n, b_n),$$

since $2\log\frac{\sqrt{a_n}}{\sqrt{b_n}} = \log\frac{a_n}{b_n}$. The invariance principle of Theorem 3.2 then confirms that $\mathcal{L}(a,b)$ is the limit. In particular, for $a > 1$,

$$\mathcal{L}\left(\frac{a}{a-1}, \frac{1}{a-1}\right) = \frac{1}{\log a},$$

which quite neatly computes the logarithm (slowly) using only arithmetic operations and square roots. ◁

4 Nonnegative matrix solution

A third approach is to directly exploit the non-negativity of the entries of the matrix A_m. This seems best organized as a case of the *Perron-Frobenius theorem* [6, Theorem 8.8.1].

Recall that a matrix A is *row stochastic* if all entries are non-negative and each row sums to one and is *irreducible* if for every pair of indices i and j, there exists a natural number k such that $(A^k)_{ij}$ is not equal to zero. Recall also that the *spectral radius* $\rho(A) := \sup\{|\lambda|\colon \lambda \text{ is an eigenvalue of } A\}$ [6, p. 177]. Since A is not assumed symmetric, we may have distinct eigenvectors for A and its transpose corresponding to the same non-zero eigenvalue. We call the later *left eigenvectors*.

Theorem 4.1 (Perron Frobenius, Utility grade [2, 6, 8])**.** *Let A be a row-stochastic irreducible square matrix. Then the spectral radius $\rho(A) = 1$ and 1 is a simple eigenvalue. Moreover, the right eigenvector $e := [1, 1, \cdots, 1_m]$ and the left eigenvector $l = [l_m, l_{m-1}, \ldots, l_1]$ are necessarily both strictly positive and hence one-dimensional.*

In consequence

$$\lim_{k\to\infty} A^k = \begin{bmatrix} l_m & l_{m-1} & \cdots & l_2 & l_1 \\ l_m & l_{m-1} & \cdots & l_2 & l_1 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ l_m & l_{m-1} & \cdots & l_2 & l_1 \\ l_m & l_{m-1} & \cdots & l_2 & l_1 \end{bmatrix}. \tag{26}$$

[We choose to consider l as a column vector with the highest order entry at the top.]

The full version of Theorem 4.1 treats arbitrary matrices with non-negative entries. Even in our setting, we do not know that the other eigenvalues are simple but we may observe that this is equivalent to the matrix A being similar to a diagonal matrix D—whose entries are the eigenvalues in decreasing order say. Then $A^n = U^{-1}D^nU \to U^{-1}D^\infty U$ where the diagonal of $D^\infty = [1, 0, \cdots, 0_m]$. More generally, the *Jordan form* [7] suffices to show that (26) still follows. See [8] for a very nice reprises of the general Perron-Frobenius theory and its multi-fold applications (and indeed [11]). In particular [8, §4] gives Karlin's resolvent proof of of Theorem 4.1.

Remark 4.2 (Collatz and Wielandt, [4, 10])**.** An attractive proof of Theorem 4.1, originating with Collatz and before him Perron, is to consider

$$g(x_1, x_2, \cdots, x_m) := \min_{1\le k\le m} \left\{ \frac{\sum_{j=1}^m a_{j,k} x_j}{x_k} \right\}.$$

Then the maximum, $\max_{\sum x_j=1, x_j\ge 0} g(x) = g(v) = 1$, exists and yields uniquely the Perron-Frobenius vector v (which in our case is e). ◁

Example 4.3 (The closed form for l)**.** The recursion we study is $\overline{x}_{n+1} = A\overline{x}_n$ where the matrix A has k-th row A_k for m strict arithmetic means A_k. Hence A is row stochastic and strictly positive and so its *Perron eigenvalue* is 1, while $A^* l = l$ shows the limit l is the left or adjoint eigenvector. Equivalently, this is also a so called *compound iteration* $L := \bigotimes A_k$ as in [3, Section 8.7] and so mean arguments much as in the previous section also establish convergence. Here we identify the eigenvector l with the corresponding linear function L since $L(x) = \langle l, x\rangle$. ◁

Remark 4.4 (The closed form for L)**.** Again we can solve for the right eigenvector $l = A^* l$, either numerically (using a linear algebra package or direct iteration) or symbolically. Note that this closed form is simultaneously a generalisation of Theorem 3.2 and a specialization of the general Invariance principle in [3, Section 8.7]. ◁

The case originating in (1) again has A being the companion matrix

$$A_m := \begin{bmatrix} a_m & a_{m-1} & \cdots & a_2 & a_1 \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \cdots & \cdots & 1 & 0 & 0 \\ 0 & 0 & \cdots & 1 & 0 \end{bmatrix}$$

with $a_k > 0$ and $\sum_{k=1}^m a_k = 1$.

Proposition 4.5. *Suppose for all $1 \le k \le m$ we have $a_k > 0$ then the matrix A_m^m has all entries strictly positive.*

Proof. We induct on k. Suppose that the first $k < m$ rows of A_m^k have strictly positive entries. Since

$$A_m^{k+1} = \begin{bmatrix} a_m & a_{m-1} & \cdots & a_2 & a_1 \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \cdots & \cdots & 1 & 0 & 0 \\ 0 & 0 & \cdots & 1 & 0 \end{bmatrix} A_m^k,$$

it follows that

$$(A_m^{k+1})_{1j} = \sum_{r=1}^{m} (A_m)_{1r} (A_m^k)_{rj} > 0,$$

and that, for $2 \leq i \leq k+1 \leq m$,

$$(A_m^{k+1})_{ij} = \sum_{r=1}^{m} (A_m)_{ir} (A_m^k)_{rj} = (A_m^k)_{i-1,j} > 0.$$

Thus, the first $k+1$ rows of A_m^{k+1} have strictly positive entries, and we are done. □

Remark 4.6 (A picture may be worth a thousand words)**.** The last theorem ensures the irreducibility of A_m by establishing the stronger condition that A_m^m is a strictly positive matrix.

Both the irreducibility of A_m and the stronger condition obtained above may be observed in the following alternative way. There are many equivalent conditions for the irreducibility of A. One obvious condition is that *an $m \times m$ matrix A with non-negative entries is irreducible if (and only if) A' is irreducible, where A' is A with each of its non-zero entries replaced by 1.*

Now, A' may be interpreted as the *adjacency matrix*, see [6, Chapter 8], for the directed graph G with vertices labeled $1, 2, \cdots, m$ and an edge from i to j precisely when $(A')_{ij} = 1$. In which case, the ij entry in the k'th power of A' equals the number of paths of length k from i to j in G. Thus, irreducibility of A corresponds to G being strongly connected.

For our particular matrix A_m, as given in (12), the associated graph G_m is depicted in Figure 1.

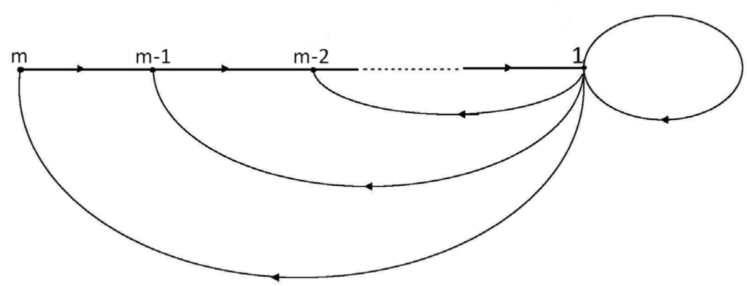

Figure 1: The graph G_m with adjacency matrix $A_m^{'}$.

The presence of the cycle $m \to m-1 \to m-2 \to \cdots \to 1 \to m$ shows that G_m is connected and hence that A_m is irreducible.

A moments' checking also reveals that in G_m any vertex i is connected to any other j by a path of length m (when forming such paths, the loop at 1 may be traced as many times as necessary), thus, also establishing the strict positivity of A_m^m. ◁

Example 4.7 (Limiting examples, II)**.** We return to the matrices of Example 2.5.

First we look again at

$$A_3 := \begin{bmatrix} \frac{1}{2} & 0 & \frac{1}{2} \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}.$$

Then A_3^4 is coordinate-wise strictly positive (but A_3^3 is not). Thus, A_3 is irreducible despite the first row not being strictly positive. The limit eigenvector is $[1/2, 1/4, 1/4]$ and the corresponding iteration is $x_n = (x_{n-1} + x_{n-3})/2$ with limit $a_1/4 + a_2/4 + a_3/2$, where the a_i are the given initial values.

Next we consider

$$A_3 := \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}.$$

In this case A_3 is reducible and the limit eigenvector $[2/3, 1/3, 0]$ exists but is not strictly positive (see Remark 2.3). The corresponding iteration is $x_n = (x_{n-1} + x_{n-2})/2$ with limit $(a_1 + 2a_2)/3$. This is also deducible by considering our starting case in with $m = 2$ and ignoring the third row and column.

The third case

$$A_3 := \begin{bmatrix} 0 & \frac{1}{2} & \frac{1}{2} \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}.$$

corresponds to the iteration $x_n = (x_{n-2} + x_{n-3})/2$. It, like the first, is irreducible with limit $(a_1 + 2a_2 + 2a_3)/5$.

Finally,

$$A_3 := \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$$

has $A_3^3 = I$ and so A_3^k is periodic of period three—and does not converge—as is obvious from the iteration $x_n = x_{n-3}$. ◁

5 Conclusion

All three approaches that we have shown have their delights and advantages. It seems fairly clear, however, that for the original problem, analysis as a mean iteration—while the least well known—is by far the most efficient and also the most elementary. Moreover, all three approaches provide for lovely examples in any linear algebra class, or any introduction to computer algebra. Indeed, they offer different flavours of algorithmics, combinatorics, analysis, algebra and graph theory.

References

[1] Heinz H. Bauschke, Joshua Sarada and Xianfu Wang, "On moving averages." Preprint, June 15, 2012.

[2] Abraham Berman and Robert J. Plemmons, *Nonnegative Matrices in the Mathematical Sciences*, 1994, SIAM.

[3] J.M. Borwein and P. B. Borwein, *Pi and the AGM: A Study in Analytic Number Theory and Computational Complexity,* John Wiley, New York, 1987, paperback 1998)

[4] Lothar Collatz, "Einschlieungssatz für die charakteristischen Zahlen von Matrize," *Mathematische Zeitschrift.* **48** (1) (1942), 221-226, doi:10.1007/BF01180013

[5] Carl-Erik Froberg, *Introduction to Numerical Analysis*, 2nd edition, Addison-Wesley, 1969.

[6] Chris Godsil and Gordon F. Royle, *Algebraic Graph Theory* Springer, New York, 2001.

[7] Gene H. Golub and Charles F. Van Loan, *Matrix Computations* (3rd ed.), Johns Hopkins University Press, Baltimore, 1996.

[8] C. R. MacCluer, "The Many Proofs and Applications of Perron's Theorem," *SIAM Review*, Vol. 42, No. 3 (Sep., 2000), 487–498.

[9] Alexander M. Ostrowski, *Solution of Equations in Euclidean and Banach Spaces.* Academic Press, 1973.

[10] Helmut Wielandt, "Unzerlegbare, nicht negative Matrizen," *Mathematische Zeitschrift* 52 (1) (150), 642-648, doi:10.1007/BF02230720

[11] Wikipedia entry on the Perron-Frobenius theorem:
http://en.wikipedia.org/wiki/Perron%E2%80%93Frobenius_theorem.