On the Solution of Linear Mean Recurrences

D. & J. Borwein and B. Sims

CARMA
University of Newcastle
http://carma.newcastle.edu.au/jon/meantalk.pdf

CARMA Colloquium

July 5, 2012 Revised 23-06-12

An obligatory irrelevant cartoon

Abstract

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

Problem I. *Determine the behaviour of the sequence:*

$$x_n := \frac{x_{n-1} + x_{n-2} + \cdots + x_{n-m}}{m} \qquad \textit{for } n \geq m+1 \quad (1)$$

and satisfying the initial conditions

$$x_k = a_k, \qquad \textit{for } k = 1, 2, \cdots, m, \qquad (2)$$

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

Abstract

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

Problem I. *Determine the behaviour of the sequence:*

$$x_n := \frac{x_{n-1} + x_{n-2} + \cdots + x_{n-m}}{m} \qquad \textit{for } n \geq m+1 \quad (1)$$

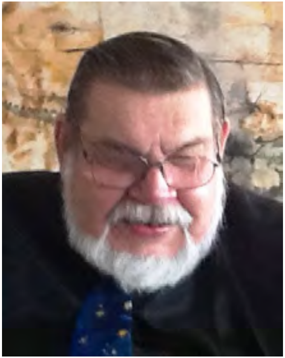
and satisfying the initial conditions

$$x_k = a_k, \qquad \textit{for } k = 1, 2, \cdots, m, \qquad (2)$$

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

My Coauthors

David Borwein and Bessie Borwein



Brailey Sims

Outline of Lecture

First attempts

Problem I. *Determine the behaviour of the sequence:*

$$x_n := \frac{x_{n-1} + x_{n-2} + \cdots + x_{n-m}}{m} \qquad \text{for } n \geq m+1 \tag{3}$$

and satisfying the initial conditions

$$x_k = a_k, \qquad \text{for } k = 1, 2, \cdots, m, \tag{4}$$

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

In light of questions posed in [1]—which encountered Problem I while computing zeroes of maximal monotone operators—we consider various approaches to addressing it.

We suspect that, like us, the first thing most readers do when shown an iteration is to try to find the limit, call it L, by taking the limit in (3).

First attempts

Problem I. *Determine the behaviour of the sequence:*

$$x_n := \frac{x_{n-1} + x_{n-2} + \cdots + x_{n-m}}{m} \qquad \textit{for } n \geq m+1 \tag{3}$$

and satisfying the initial conditions

$$x_k = a_k, \qquad \textit{for } k = 1, 2, \cdots, m, \tag{4}$$

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

In light of questions posed in [1]—which encountered Problem I while computing zeroes of maximal monotone operators—we consider various approaches to addressing it.
We suspect that, like us, the first thing most readers do when shown an iteration is to try to find the limit, call it L, by taking the limit in (3).

First attempts

Supposing the limit to exist we deduce

$$L = \frac{\overbrace{L + L + \cdots + L}^{m}}{m} = L, \tag{5}$$

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 3 sections, we present three distinct approaches.
- While at least one will be familiar to many, we suspect not all three will be.
- Each has its advantages, both as an example of more general techniques and since each opens up a beautiful corpus of mathematics.

First attempts

Supposing the limit to exist we deduce

$$L = \frac{\overbrace{L + L + \cdots + L}^{m}}{m} = L, \tag{5}$$

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 3 sections, we present three distinct approaches.
- While at least one will be familiar to many, we suspect not all three will be.
- Each has its advantages, both as an example of more general techniques and since each opens up a beautiful corpus of mathematics.

First attempts

Supposing the limit to exist we deduce

$$L = \frac{\overbrace{L + L + \cdots + L}^{m}}{m} = L, \tag{5}$$

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 3 sections, we present three distinct approaches.
- While at least one will be familiar to many, we suspect not all three will be.
- Each has its advantages, both as an example of more general techniques and since each opens up a beautiful corpus of mathematics.

First attempts

Supposing the limit to exist we deduce

$$L = \frac{\overbrace{L + L + \cdots + L}^{m}}{m} = L, \tag{5}$$

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 3 sections, we present three distinct approaches.
- While at least one will be familiar to many, we suspect not all three will be.
- Each has its advantages, both as an example of more general techniques and since each opens up a beautiful corpus of mathematics.

First attempts

Supposing the limit to exist we deduce

$$L = \frac{\overbrace{L + L + \cdots + L}^{m}}{m} = L, \tag{5}$$

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 3 sections, we present three distinct approaches.
- While at least one will be familiar to many, we suspect not all three will be.
- Each has its advantages, both as an example of more general techniques and since each opens up a beautiful corpus of mathematics.

Spectral solution

We start with the best known approach which turns up in most linear algebra courses along with the Fibonacci numbers:

$$F_n = F_{n-1} + F_{n-2} \qquad \text{with} \qquad F_0 = 0, F_1 = 1. \tag{6}$$

Equations (6) and (3) are examples of a *linear homogeneous recurrence relation of order* m with constant coefficients.

- Typically, elementary books only consider simple roots as suffices for (6). In *Maple*

$$solve(\{F(n) = F(n-1) + F(n-2), F(0) = 0, F(1) = 1\}, F(n))$$

returns $-1/5\sqrt{5}\left(1/2 - 1/2\sqrt{5}\right)^n + 1/5\sqrt{5}\left(1/2 + 1/2\sqrt{5}\right)^n$.

Spectral solution

We start with the best known approach which turns up in most linear algebra courses along with the Fibonacci numbers:

$$F_n = F_{n-1} + F_{n-2} \qquad \text{with} \qquad F_0 = 0, F_1 = 1. \tag{6}$$

Equations (6) and (3) are examples of a *linear homogeneous recurrence relation of order* m with constant coefficients.

- Typically, elementary books only consider simple roots as suffices for (6). In *Maple*

$$solve(\{F(n) = F(n-1) + F(n-2), F(0) = 0, F(1) = 1\}, F(n))$$

returns $-1/5\,\sqrt{5}\left(1/2 - 1/2\,\sqrt{5}\right)^n + 1/5\,\sqrt{5}\left(1/2 + 1/2\,\sqrt{5}\right)^n$.

Spectral solution

We start with the best known approach which turns up in most linear algebra courses along with the Fibonacci numbers:

$$F_n = F_{n-1} + F_{n-2} \qquad \text{with} \qquad F_0 = 0, F_1 = 1. \tag{6}$$

Equations (6) and (3) are examples of a *linear homogeneous recurrence relation of order* m with constant coefficients.

- Typically, elementary books only consider simple roots as suffices for (6). In *Maple*

$$solve(\{F(n) = F(n-1) + F(n-2), F(0) = 0, F(1) = 1\}, F(n))$$

returns $-1/5\sqrt{5}\left(1/2 - 1/2\sqrt{5}\right)^n + 1/5\sqrt{5}\left(1/2 + 1/2\sqrt{5}\right)^n$.

Theorem (General solution of a linear recurrence)

Standard theory [5, 9] runs as follows:

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

with constant coefficients, has the form

$$x_n = \sum_{k=1}^{l} q_k(n)\, r_k^n \tag{7}$$

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

$$p(r) := r^m - \sum_{k=1}^{m} \alpha_k r^{k-1}, \tag{8}$$

with multiplicity m_k *and polynomials* q_k *of degree less than* m_k*.*

Our equation analysed, I

Equation 3 has characteristic polynomial:

$$\begin{aligned} p(r) &:= r^m - \frac{1}{m}(r^{m-1} + r^{m-2} + \cdots + r + 1) \\ &= \frac{mr^{m+1} - (m+1)r^m + 1}{m(r-1)} \end{aligned} \tag{9}$$

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 one is simple.

We next show that if $p(r) = 0$ and $r \neq 1$, then $|r| < 1$. We argue as follows. From (9) we know $p(r) = 0$ if and only if

$$r + \frac{1}{mr^m} = 1 + \frac{1}{m}. \tag{10}$$

Our equation analysed, I

Equation 3 has characteristic polynomial:

$$p(r) := r^m - \frac{1}{m}(r^{m-1} + r^{m-2} + \cdots + r + 1) = \frac{mr^{m+1} - (m+1)r^m + 1}{m(r-1)} \quad (9)$$

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 one is simple.

We next show that if $p(r) = 0$ and $r \neq 1$, then $|r| < 1$. We argue as follows. From (9) we know $p(r) = 0$ if and only if

$$r + \frac{1}{mr^m} = 1 + \frac{1}{m}. \quad (10)$$

Our equation analysed, I

Equation 3 has characteristic polynomial:

$$p(r) := r^m - \frac{1}{m}(r^{m-1} + r^{m-2} + \cdots + r + 1) = \frac{mr^{m+1} - (m+1)r^m + 1}{m(r-1)} \tag{9}$$

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 one is simple.

We next show that if $p(r) = 0$ and $r \neq 1$, then $|r| < 1$. We argue as follows. From (9) we know $p(r) = 0$ if and only if

$$r + \frac{1}{mr^m} = 1 + \frac{1}{m}. \tag{10}$$

Our equation analysed, II

If $|r| > 1$, then

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

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 now that $p(r) = 0$ with $r = e^{i\theta}$, $0 \le \theta < 2\pi$. By (10)

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

which means $\theta = 0$. By (7) we have

$$x_n = c_1 + \sum_{k=2}^{r} q_k(n)\, r_k^n \tag{11}$$

where r_k lies in the open unit disc for $2 \le k \le m$. Thus, the limit in (11) exists and equals the coefficient c_1.

Our equation analysed, II

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 now that $p(r) = 0$ with $r = e^{i\theta}$, $0 \leq \theta < 2\pi$. By (10)

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

which means $\theta = 0$. By (7) we have

$$x_n = c_1 + \sum_{k=2}^{r} q_k(n)\, r_k^n \tag{11}$$

where r_k lies in the open unit disc for $2 \leq k \leq m$. Thus, the limit in (11) exists and equals the coefficient c_1.

Our equation analysed, II

If $|r| > 1$, then

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

since the function $f(x) := x + \frac{1}{m x^m}$ is strictly increasing for real $x > 1$ and $f(1) = 1 + \frac{1}{m}$. Thus $p(r) \neq 0$ when $|r| > 1$. Suppose now that $p(r) = 0$ with $r = e^{i\theta}$, $0 \le \theta < 2\pi$. By (10)

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

which means $\theta = 0$. By (7) we have

$$x_n = c_1 + \sum_{k=2}^{r} q_k(n)\, r_k^n \tag{11}$$

where r_k lies in the open unit disc for $2 \le k \le m$. Thus, the limit in (11) exists and equals the coefficient c_1.

Identifying the limit, I

Remark (The roots are simple)

In fact we may use (9) to see all roots are simple as follows: It follows from (9) that

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

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

$$x_n = c_1 + \sum_{k=2}^{m} c_k\, r_k^n, \tag{12}$$

as asserted (all the polynomials are constant).

Identifying the limit, I

Remark (The roots are simple)

In fact we may use (9) to see all roots are simple as follows: It follows from (9) that

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

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

$$x_n = c_1 + \sum_{k=2}^{m} c_k\, r_k^n, \tag{12}$$

as asserted (all the polynomials are constant). ◀

Identifying the limit, II

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

$$\sum_{n=1}^{m} nr^n = \frac{mr^{m+2} - (m+1)r^{m+1} + r}{(r-1)^2} = \frac{m\,rp(r)}{r-1} = 0, \qquad (13)$$

and summing (12) gives

$$c_1 = \frac{2}{m(m+1)} \sum_{n=1}^{m} na_n. \qquad (14)$$

Thence, we have convergence and a limit $L = c_1$ given by (14). □

The same analysis, works if in (3) we replace the arithmetic average 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 $\alpha_k > 0$ summing to one. ($W_{(1/m)} = A$ is the arithmetic mean of Problem I.)

- As often the analysis becomes easier when we generalize.

Identifying the limit, II

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

$$\sum_{n=1}^{m} nr^n = \frac{mr^{m+2} - (m+1)r^{m+1} + r}{(r-1)^2} = \frac{m\,rp(r)}{r-1} = 0, \qquad (13)$$

and summing (12) gives

$$c_1 = \frac{2}{m(m+1)} \sum_{n=1}^{m} na_n. \qquad (14)$$

Thence, we have convergence and a limit $L = c_1$ given by (14). □

The same analysis, works if in (3) we replace the arithmetic average 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 $\alpha_k > 0$ summing to one. ($W_{(1/m)} = A$ is the arithmetic mean of Problem I.)

- As often the analysis becomes easier when we generalize.

Identifying the limit, II

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

$$\sum_{n=1}^{m} nr^n = \frac{mr^{m+2} - (m+1)r^{m+1} + r}{(r-1)^2} = \frac{m\, rp(r)}{r-1} = 0, \tag{13}$$

and summing (12) gives

$$c_1 = \frac{2}{m(m+1)} \sum_{n=1}^{m} na_n. \tag{14}$$

Thence, we have convergence and a limit $L = c_1$ given by (14). □

The same analysis, works if in (3) we replace the arithmetic average 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 $\alpha_k > 0$ summing to one. ($W_{(1/m)} = A$ is the arithmetic mean of Problem I.)

- As often the analysis becomes easier when we generalize.

Example (The weighted mean)

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*

$$A_m := \begin{bmatrix} a_m & a_{m-1} & \cdots & a_2 & a_1 \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \cdots & \cdots & 1 & 0 & 0 \\ 0 & 0 & \cdots & 1 & 0 \end{bmatrix}. \quad (15)$$

The corresponding *characteristic polynomial of the recurrence* is

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

is also the characteristic polynomial of the matrix.

Example (The weighted mean)

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*

$$A_m := \begin{bmatrix} a_m & a_{m-1} & \cdots & a_2 & a_1 \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \cdots & \cdots & 1 & 0 & 0 \\ 0 & 0 & \cdots & 1 & 0 \end{bmatrix}. \tag{15}$$

The corresponding *characteristic polynomial of the recurrence* is

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

is also the characteristic polynomial of the matrix.

Example (Root behaviour for a weighted mean, I)

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

$$\rho^m \leq \sum_{k=1}^{m} \alpha_k \rho^{k-1} \leq \max_{1 \leq k \leq m} \rho^{k-1}, \tag{16}$$

and so $0 \leq \rho \leq 1$. If $\rho = 1$ but $r \neq 1$ then $r = e^{i\theta}$ for $0 < \theta < 2\pi$. Since $r^{-m}p(r) = 0$, on 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.
Thence, roots other than 1 have modulus strictly less than one. ◀

Example (Root behaviour for a weighted mean, I)

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

$$\rho^m \leq \sum_{k=1}^{m} \alpha_k \rho^{k-1} \leq \max_{1 \leq k \leq m} \rho^{k-1}, \tag{16}$$

and so $0 \leq \rho \leq 1$. If $\rho = 1$ but $r \neq 1$ then $r = e^{i\theta}$ for $0 < \theta < 2\pi$. Since $r^{-m} p(r) = 0$, on 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.

Thence, roots other than 1 have modulus strictly less than one. ◀

Example (Root behaviour for a weighted mean, I)

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

$$\rho^m \leq \sum_{k=1}^{m} \alpha_k \rho^{k-1} \leq \max_{1 \leq k \leq m} \rho^{k-1}, \tag{16}$$

and so $0 \leq \rho \leq 1$. If $\rho = 1$ but $r \neq 1$ then $r = e^{i\theta}$ for $0 < \theta < 2\pi$. Since $r^{-m} p(r) = 0$, on 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.
Thence, roots other than 1 have modulus strictly less than one. ◀

Example (Root behaviour for a weighted mean, I)

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

$$p(r) = (r-1)\sum_{k=1}^{m} \sigma_k r^{k-1}. \tag{17}$$

Hence, p has no other positive real root ($\sigma_k > 0$).
In particular, from (7) 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 one is simple while all other roots are strictly inside the unit disc—but need not be simple as illustrated in the next Example.

Example (Root behaviour for a weighted mean, I)

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

$$p(r) = (r-1)\sum_{k=1}^{m}\sigma_k r^{k-1}. \tag{17}$$

Hence, p has no other positive real root ($\sigma_k > 0$).

In particular, from (7) 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 one is simple while all other roots are strictly inside the unit disc—but need not be simple as illustrated in the next Example.

Example (Root behaviour for a weighted mean, I)

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

$$p(r) = (r-1)\sum_{k=1}^{m}\sigma_k r^{k-1}. \tag{17}$$

Hence, p has no other positive real root ($\sigma_k > 0$).
In particular, from (7) 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 one is simple while all other roots are strictly inside the unit disc—but need not be simple as illustrated in the next Example. ◀

Example (A weighted mean with multiple roots)

- p below has a root at 1 and a repeated pair of roots at $\pm\frac{i}{3}$:

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

$$= \frac{1}{162}\,(2\,r+1)\,(r-1)\,\left(1+9\,r^2\right)^2. \tag{19}$$

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 weighted mean Example. And

$$L := \frac{162\,a_6 + 161\,a_5 + 160\,a_4 + 144\,a_3 + 126\,a_2 + 81\,a_1}{834}. \tag{20}$$

is the limit.

Example (A weighted mean with multiple roots)

- p below has a root at 1 and a repeated pair of roots at $\pm\frac{i}{3}$:

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

$$= \frac{1}{162}\,(2\,r+1)\,(r-1)\left(1+9\,r^2\right)^2. \tag{19}$$

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 weighted mean Example. And

$$L := \frac{162\,a_6 + 161\,a_5 + 160\,a_4 + 144\,a_3 + 126\,a_2 + 81\,a_1}{834}. \tag{20}$$

is the limit. ◀

Example (A weighted mean with multiple roots)

- p below has a root at 1 and a repeated pair of roots at $\pm\frac{i}{3}$:

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

$$= \frac{1}{162}\,(2\,r+1)\,(r-1)\left(1+9\,r^2\right)^2. \tag{19}$$

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 weighted mean Example. And

$$L := \frac{162\,a_6 + 161\,a_5 + 160\,a_4 + 144\,a_3 + 126\,a_2 + 81\,a_1}{834}. \tag{20}$$

is the limit. ◀

Remark (How this recursion was found)

We examined how to place repeated roots on the imaginary axis while preserving increasing coefficients as required in (17).
One general potential form is then

$$p(\sigma,\tau) := (r-1)(r+\sigma)(r^2+\tau^2)^2$$

and we selected $p(\frac{1}{2},\frac{1}{3})$. In the same fashion

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

This has a zero coefficient of r^4, but the corresponding iteration remains well behaved, see below.

- L was found by computing A^{1000} to 14 places and rationalizing!

Remark (How this recursion was found)

We examined how to place repeated roots on the imaginary axis while preserving increasing coefficients as required in (17).
One general potential form is then

$$p(\sigma,\tau) := (r-1)(r+\sigma)(r^2+\tau^2)^2$$

and we selected $p(\frac{1}{2},\frac{1}{3})$. In the same fashion

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

This has a zero coefficient of r^4, but the corresponding iteration remains well behaved, see below. ◀

- L was found by computing A^{1000} to 14 places and rationalizing!

Remark (How this recursion was found)

We examined how to place repeated roots on the imaginary axis while preserving increasing coefficients as required in (17).
One general potential form is then

$$p(\sigma,\tau) := (r-1)(r+\sigma)(r^2+\tau^2)^2$$

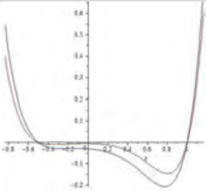
and we selected $p(\frac{1}{2},\frac{1}{3})$. In the same fashion

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

This has a zero coefficient of r^4, but the corresponding iteration remains well behaved, see below. ◀

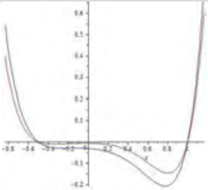
- L was found by computing A^{1000} to 14 places and rationalizing!

Further comments



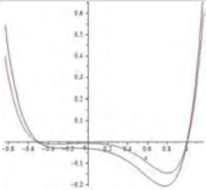
- The graphs are of $p(1/2, 1/3)$ and $p(1/2, 1/2)$. Is any such example of degree six or more?
- An analysis of the weighted mean Example shows it holds for non-negative weights if the highest-order term $\alpha_m > 0$.
  - We will see that the invariance principle below deals most efficiently with identifying limits for weighted linear means.
  - In fact, we shall discover that the numerator coefficients in (20) are the partial sums of those in (18).
  - The same method also provides a quick way to check the assertions about limits in the next Example.

Further comments



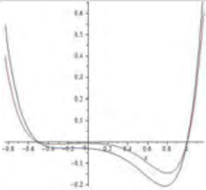
- The graphs are of $p(1/2, 1/3)$ and $p(1/2, 1/2)$. Is any such example of degree six or more?
- An analysis of the weighted mean Example shows it holds for non-negative weights if the highest-order term $\alpha_m > 0$.
  - We will see that the invariance principle below deals most efficiently with identifying limits for weighted linear means.
  - In fact, we shall discover that the numerator coefficients in (20) are the partial sums of those in (18).
  - The same method also provides a quick way to check the assertions about limits in the next Example.

Further comments



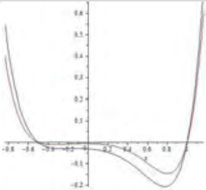
- The graphs are of $p(1/2, 1/3)$ and $p(1/2, 1/2)$. Is any such example of degree six or more?
- An analysis of the weighted mean Example shows it holds for non-negative weights if the highest-order term $\alpha_m > 0$.
  - We will see that the invariance principle below deals most efficiently with identifying limits for weighted linear means.
  - In fact, we shall discover that the numerator coefficients in (20) are the partial sums of those in (18).
  - The same method also provides a quick way to check the assertions about limits in the next Example.

Further comments



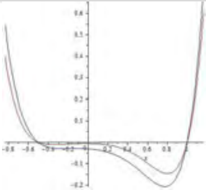
- The graphs are of $p(1/2, 1/3)$ and $p(1/2, 1/2)$. Is any such example of degree six or more?
- An analysis of the weighted mean Example shows it holds for non-negative weights if the highest-order term $\alpha_m > 0$.
  - We will see that the invariance principle below deals most efficiently with identifying limits for weighted linear means.
  - In fact, we shall discover that the numerator coefficients in (20) are the partial sums of those in (18).
  - The same method also provides a quick way to check the assertions about limits in the next Example.

Further comments



- The graphs are of $p(1/2, 1/3)$ and $p(1/2, 1/2)$. Is any such example of degree six or more?
- An analysis of the weighted mean Example shows it holds for non-negative weights if the highest-order term $\alpha_m > 0$.
  - We will see that the invariance principle below deals most efficiently with identifying limits for weighted linear means.
  - In fact, we shall discover that the numerator coefficients in (20) are the partial sums of those in (18).
  - The same method also provides a quick way to check the assertions about limits in the next Example.

Further comments



- The graphs are of $p(1/2, 1/3)$ and $p(1/2, 1/2)$. Is any such example of degree six or more?
- An analysis of the weighted mean Example shows it holds for non-negative weights if the highest-order term $\alpha_m > 0$.
  - We will see that the invariance principle below deals most efficiently with identifying limits for weighted linear means.
  - In fact, we shall discover that the numerator coefficients in (20) are the partial sums of those in (18).
  - The same method also provides a quick way to check the assertions about limits in the next Example.

Example (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 is in Problem I with $m = 2$ on ignoring row and column 3. ◀

Example (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 is in Problem I with $m = 2$ on ignoring row and column 3. ◀

Example (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 is in Problem I with $m = 2$ on ignoring row and column 3. ◀

Example (Limiting examples I)

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 $x_n = x_{n-3}$.

Example (Limiting examples I)

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 $x_n = x_{n-3}$. ◀

Another irrelevant cartoon

Mean iteration solution

The second approach, based on [3, Section 8.7], deals very efficiently with equation 3.

- As a bonus, our convergence proof holds for nonlinear means given positive starting values.

Definition (Strict mean)

We say M is a *strict m-variable mean* if always

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

with equality if and only if all variables are equal.

- While nonlinear means —such as $G := (x_1 x_2 \cdots x_m)^{1/m}$—are defined only for positive input, linear means are defined for all variables.

Mean iteration solution

The second approach, based on [3, Section 8.7], deals very efficiently with equation 3.

- As a bonus, our convergence proof holds for nonlinear means given positive starting values.

Definition (Strict mean)

We say M is a *strict m-variable mean* if always

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

with equality if and only if all variables are equal.

- While nonlinear means —such as $G := (x_1 x_2 \cdots x_m)^{1/m}$—are defined only for positive input, linear means are defined for all variables.

Mean iteration solution

The second approach, based on [3, Section 8.7], deals very efficiently with equation 3.

- As a bonus, our convergence proof holds for nonlinear means given positive starting values.

Definition (Strict mean)

We say M is a *strict m-variable mean* if always

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

with equality if and only if all variables are equal.

- While nonlinear means —such as $G := (x_1 x_2 \cdots x_m)^{1/m}$—are defined only for positive input, linear means are defined for all variables.

Mean iteration solution

The second approach, based on [3, Section 8.7], deals very efficiently with equation 3.

- As a bonus, our convergence proof holds for nonlinear means given positive starting values.

Definition (Strict mean)

We say M is a *strict m-variable mean* if always

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

with equality if and only if all variables are equal.

- While nonlinear means —such as $G := (x_1 x_2 \cdots x_m)^{1/m}$—are defined only for positive input, linear means are defined for all variables.

Convergence of mean iterations

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

Theorem (Convergence of a mean iteration)

Let M be any strict mean in m variables and consider the iteration

$$x_n := M(x_{n-m}, x_{n-m+1}, \cdots, x_{n-1}) \tag{21}$$

so that with $M = A$ we recover the iteration in (3)*. Then x_n converges to a finite limit $L(x_1, x_2, \ldots, x_m)$.*

- Specialization of [3, Exercise 7 of Section 8.7] showa convergence for an arbitrary strict mean. We shall make this explicit below.
- For general means we need to restrict the variables to non-negative values, but for linear means no such restriction is needed.

Convergence of mean iterations

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

Theorem (Convergence of a mean iteration)

Let M be any strict mean in m variables and consider the iteration

$$x_n := M(x_{n-m}, x_{n-m+1}, \cdots, x_{n-1}) \tag{21}$$

so that with $M = A$ we recover the iteration in (3). Then x_n converges to a finite limit $L(x_1, x_2, \ldots, x_m)$.

- Specialization of [3, Exercise 7 of Section 8.7] showa convergence for an arbitrary strict mean. We shall make this explicit below.
- For general means we need to restrict the variables to non-negative values, but for linear means no such restriction is needed.

Convergence of mean iterations

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

Theorem (Convergence of a mean iteration)

Let M be any strict mean in m variables and consider the iteration

$$x_n := M(x_{n-m}, x_{n-m+1}, \cdots, x_{n-1}) \tag{21}$$

so that with $M = A$ we recover the iteration in (3)*. Then x_n converges to a finite limit $L(x_1, x_2, \ldots, x_m)$.*

- Specialization of [3, Exercise 7 of Section 8.7] showa convergence for an arbitrary strict mean. We shall make this explicit below.
- For general means we need to restrict the variables to non-negative values, but for linear means no such restriction is needed.

Proof.

Let $\overline{x}_n := (x_n, x_{n-1}, \cdots, x_{n-m+1})$ and let

$$a_n := \max \overline{x}_n, \qquad b_n := \min \overline{x}_n.$$

For all n, the mean property shows

$$a_{n-1} \geq a_n \geq b_n \geq b_{n-1}. \tag{22}$$

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 $\overline{x}_{n_k} \to \overline{x}$. Thence

$$b \leq \min \overline{x} \leq \max \overline{x} \leq a \tag{23}$$

while

$$b = \min M(\overline{x}) \quad \text{and} \quad \max M(\overline{x}) = a. \tag{24}$$

Since M is a strict mean, we have $a = b$ and convergence. □

Proof.

Let $\overline{x}_n := (x_n, x_{n-1}, \cdots, x_{n-m+1})$ and let

$$a_n := \max \overline{x}_n, \qquad b_n := \min \overline{x}_n.$$

For all n, the mean property shows

$$a_{n-1} \geq a_n \geq b_n \geq b_{n-1}. \tag{22}$$

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 $\overline{x}_{n_k} \to \overline{x}$. Thence

$$b \leq \min \overline{x} \leq \max \overline{x} \leq a \tag{23}$$

while

$$b = \min M(\overline{x}) \quad \text{and} \quad \max M(\overline{x}) = a. \tag{24}$$

Since M is a strict mean, we have $a = b$ and convergence. □

Proof.

Let $\overline{x}_n := (x_n, x_{n-1}, \cdots, x_{n-m+1})$ and let

$$a_n := \max \overline{x}_n, \qquad b_n := \min \overline{x}_n.$$

For all n, the mean property shows

$$a_{n-1} \geq a_n \geq b_n \geq b_{n-1}. \tag{22}$$

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 $\overline{x}_{n_k} \to \overline{x}$. Thence

$$b \leq \min \overline{x} \leq \max \overline{x} \leq a \tag{23}$$

while

$$b = \min M(\overline{x}) \quad \text{and} \quad \max M(\overline{x}) = a. \tag{24}$$

Since M is a strict mean, we have $a = b$ and convergence. □

Proof.

Let $\overline{x}_n := (x_n, x_{n-1}, \cdots, x_{n-m+1})$ and let

$$a_n := \max \overline{x}_n, \qquad b_n := \min \overline{x}_n.$$

For all n, the mean property shows

$$a_{n-1} \geq a_n \geq b_n \geq b_{n-1}. \tag{22}$$

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 $\overline{x}_{n_k} \to \overline{x}$. Thence

$$b \leq \min \overline{x} \leq \max \overline{x} \leq a \tag{23}$$

while

$$b = \min M(\overline{x}) \quad \text{and} \quad \max M(\overline{x}) = a. \tag{24}$$

Since M is a strict mean, we have $a = b$ and convergence. □

Proof.

Let $\overline{x}_n := (x_n, x_{n-1}, \cdots, x_{n-m+1})$ and let

$$a_n := \max \overline{x}_n, \qquad b_n := \min \overline{x}_n.$$

For all n, the mean property shows

$$a_{n-1} \geq a_n \geq b_n \geq b_{n-1}. \tag{22}$$

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 $\overline{x}_{n_k} \to \overline{x}$. Thence

$$b \leq \min \overline{x} \leq \max \overline{x} \leq a \tag{23}$$

while

$$b = \min M(\overline{x}) \quad \text{and} \quad \max M(\overline{x}) = a. \tag{24}$$

Since M is a strict mean, we have $a = b$ and convergence. □

Determining the limit

Both the Limit theorem above and the Invariance principle below show the power of identifying (3) as a mean iteration.

Theorem (Invariance principle, see ref. 3.)

For any convergent mean iteration M, the limit L is necessarily a mean and is the unique diagonal mapping satisfying the Invariance principle:

$$\begin{aligned} & L\left(x_{n-m}, x_{n-m+1}, \ldots, x_{n-1}\right) \\ = \; & L\left(x_{n-m+1}, \ldots, x_{n-1}, M\left(x_{n-m}, x_{n-m+1}, \ldots, x_{n-1}\right)\right). \end{aligned} \tag{25}$$

Moreover, L is linear as soon as M is.

Determining the limit

Both the Limit theorem above and the Invariance principle below show the power of identifying (3) as a mean iteration.

Theorem (Invariance principle, see ref. 3.)

For any convergent mean iteration M, the limit L is necessarily a mean and is the unique diagonal mapping satisfying the Invariance principle:

$$\begin{aligned} & L\left(x_{n-m}, x_{n-m+1}, \ldots, x_{n-1}\right) \\ = \quad & L\left(x_{n-m+1}, \ldots, x_{n-1}, M\left(x_{n-m}, x_{n-m+1}, \ldots, x_{n-1}\right)\right). \end{aligned} \tag{25}$$

Moreover, L is linear as soon as M is.

Determining the limit

We sketch the important direction leaving the other to the reader. Details are again in [3, Section 8.7].

Proof.

One first checks that the limit is a mean (as a point-wise limit of means) and so is continuous on the diagonal. The principle says

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

as required. □

- The proof just quantifies the shift invariance of the limit.
- We can mix-and-match arguments—if we have used the ideas of the previous section to convince ourselves the limit exists, the invariance principle is ready to finish the job.

Determining the limit

We sketch the important direction leaving the other to the reader. Details are again in [3, Section 8.7].

Proof.

One first checks that the limit is a mean (as a point-wise limit of means) and so is continuous on the diagonal. The principle says

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

as required. □

- The proof just quantifies the shift invariance of the limit.
- We can mix-and-match arguments—if we have used the ideas of the previous section to convince ourselves the limit exists, the invariance principle is ready to finish the job.

Determining the limit

We sketch the important direction leaving the other to the reader. Details are again in [3, Section 8.7].

Proof.

One first checks that the limit is a mean (as a point-wise limit of means) and so is continuous on the diagonal. The principle says

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

as required. □

- The proof just quantifies the shift invariance of the limit.
- We can mix-and-match arguments—if we have used the ideas of the previous section to convince ourselves the limit exists, the invariance principle is ready to finish the job.

Example (A general strict linear mean)

Suppose that $M(y_1,\ldots,y_m) = \sum_{i=1}^m \alpha_i y_i$, with all $\alpha_i > 0$, and $L(y_1,\ldots,y_m) = \sum_{i=1}^m \lambda_i y_i$ are both linear. We may solve (25) to determine that for $k = 1, 2, \ldots m-1$ we have

$$\lambda_{k+1} = \lambda_k + \lambda_m \alpha_{k+1}. \tag{26}$$

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

$$\lambda_k / \lambda_m = \sigma_k. \tag{27}$$

Since L is a mean we have $L(1,1,\ldots,1) = 1$ and so

$$\lambda_k = \frac{\sigma_k}{\sum_{k=1}^m \sigma_k}. \tag{28}$$

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 (14).

Example (A general strict linear mean)

Suppose that $M(y_1, \ldots, y_m) = \sum_{i=1}^{m} \alpha_i y_i$, with all $\alpha_i > 0$, and $L(y_1, \ldots, y_m) = \sum_{i=1}^{m} \lambda_i y_i$ are both linear. We may solve (25) to determine that for $k = 1, 2, \ldots m-1$ we have

$$\lambda_{k+1} = \lambda_k + \lambda_m \alpha_{k+1}. \tag{26}$$

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

$$\lambda_k / \lambda_m = \sigma_k. \tag{27}$$

Since L is a mean we have $L(1, 1, \ldots, 1) = 1$ and so

$$\lambda_k = \frac{\sigma_k}{\sum_{k=1}^{m} \sigma_k}. \tag{28}$$

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 (14).

Example (A general strict linear mean)

Suppose that $M(y_1,\ldots,y_m) = \sum_{i=1}^m \alpha_i y_i$, with all $\alpha_i > 0$, and $L(y_1,\ldots,y_m) = \sum_{i=1}^m \lambda_i y_i$ are both linear. We may solve (25) to determine that for $k = 1, 2, \ldots m-1$ we have

$$\lambda_{k+1} = \lambda_k + \lambda_m \alpha_{k+1}. \tag{26}$$

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

$$\lambda_k/\lambda_m = \sigma_k. \tag{27}$$

Since L is a mean we have $L(1,1,\ldots,1) = 1$ and so

$$\lambda_k = \frac{\sigma_k}{\sum_{k=1}^m \sigma_k}. \tag{28}$$

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 (14).

Example (A general strict linear mean)

Suppose that $M(y_1, \ldots, y_m) = \sum_{i=1}^m \alpha_i y_i$, with all $\alpha_i > 0$, and $L(y_1, \ldots, y_m) = \sum_{i=1}^m \lambda_i y_i$ are both linear. We may solve (25) to determine that for $k = 1, 2, \ldots m-1$ we have

$$\lambda_{k+1} = \lambda_k + \lambda_m \alpha_{k+1}. \tag{26}$$

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

$$\lambda_k/\lambda_m = \sigma_k. \tag{27}$$

Since L is a mean we have $L(1, 1, \ldots, 1) = 1$ and so

$$\lambda_k = \frac{\sigma_k}{\sum_{k=1}^m \sigma_k}. \tag{28}$$

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 (14). ◀

Example (A nonlinear mean)

We may replace A by the *Hölder mean*

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

for $-\infty < p < \infty$. The limit is $\left(\sum_{k=1}^m \lambda_k a_k^p\right)^{1/p}$ with λ_k from (28).

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 may also consider weighted Hölder means. ◀

- We end this section with an especially neat application of the Invariance principle to an example of Carlson [3, Section 8.7].
- One can similarly analyse Archimedes's method for π.

Example (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 is $\left(\sum_{k=1}^m \lambda_k a_k^p\right)^{1/p}$ with λ_k from (28). 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 may also consider weighted Hölder means.

- We end this section with an especially neat application of the Invariance principle to an example of Carlson [3, Section 8.7].
- One can similarly analyse Archimedes's method for π.

Example (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 is $\left(\sum_{k=1}^{m} \lambda_k a_k^p\right)^{1/p}$ with λ_k from (28). 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 may also consider weighted Hölder means. ◀

- We end this section with an especially neat application of the Invariance principle to an example of Carlson [3, Section 8.7].
- One can similarly analyse Archimedes's method for π.

Example (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 is $\left(\sum_{k=1}^{m} \lambda_k a_k^p\right)^{1/p}$ with λ_k from (28). 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 may also consider weighted Hölder means. ◀

- We end this section with an especially neat application of the Invariance principle to an example of Carlson [3, Section 8.7].
- One can similarly analyse Archimedes's method for π.

Example (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 is $\left(\sum_{k=1}^m \lambda_k a_k^p\right)^{1/p}$ with λ_k from (28). 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 may also consider weighted Hölder means. ◀

- We end this section with an especially neat application of the Invariance principle to an example of Carlson [3, Section 8.7].
- One can similarly analyse Archimedes's method for π.

Example (Carlson's logarithmic mean)

Consider the iteration 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

$$|a_{n+1} - b_{n+1}| = \frac{1}{2} |a_n - b_n|.$$

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

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

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

Example (Carlson's logarithmic mean)

Consider the iteration 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

$$|a_{n+1} - b_{n+1}| = \frac{1}{2}\,|a_n - b_n|.$$

If asked for the limit, you might make little progress.

But suppose you are told the answer is

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

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

Example (Carlson's logarithmic mean)

Consider the iteration 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

$$|a_{n+1} - b_{n+1}| = \frac{1}{2} |a_n - b_n|.$$

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

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

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

Example (Carlson's logarithmic mean)

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 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.

Example (Carlson's logarithmic mean)

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 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.

Example (Carlson's logarithmic mean)

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

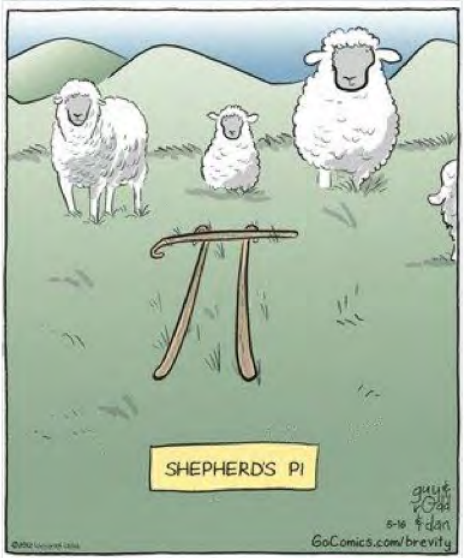
$$2\log\frac{\sqrt{a_n}}{\sqrt{b_n}} = \log\frac{a_n}{b_n}.$$

The Invariance principle 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. ◀

Another irrelevant cartoon



Nonnegative matrix solution

A third approach directly exploits non-negativity of the entries of the matrix A_m. This is best organized as a case of the *Perron-Frobenius theorem* [2], [6, Theorem 8.8.1] or [8].

- A is *row stochastic* if all entries are non-negative and each row sums to one.
- A is *irreducible* if for every pair of indices i, j, there is a natural number k with $(A^k)_{ij} \neq 0$.
- The *spectral radius* [6, p. 177] is

$$\rho(A) := \sup\{|\lambda| : \lambda \text{ is an eigenvalue of } A\}.$$

- Since A is not assumed symmetric, we may have distinct eigenvectors for A and its transpose with the same non-zero eigenvalue. We call the later *left eigenvectors*.

Below we view l as a column with highest order entry at the top.

Nonnegative matrix solution

A third approach directly exploits non-negativity of the entries of the matrix A_m. This is best organized as a case of the *Perron-Frobenius theorem* [2], [6, Theorem 8.8.1] or [8].

- A is *row stochastic* if all entries are non-negative and each row sums to one.
- A is *irreducible* if for every pair of indices $i, j,$ there is a natural number k with $(A^k)_{ij} \neq 0$.
- The *spectral radius* [6, p. 177] is

$$\rho(A) := \sup\{|\lambda| : \lambda \text{ is an eigenvalue of } A\}.$$

- Since A is not assumed symmetric, we may have distinct eigenvectors for A and its transpose with the same non-zero eigenvalue. We call the later *left eigenvectors*.

Below we view l as a column with highest order entry at the top.

Nonnegative matrix solution

A third approach directly exploits non-negativity of the entries of the matrix A_m. This is best organized as a case of the *Perron-Frobenius theorem* [2], [6, Theorem 8.8.1] or [8].

- A is *row stochastic* if all entries are non-negative and each row sums to one.
- A is *irreducible* if for every pair of indices $i, j,$ there is a natural number k with $(A^k)_{ij} \neq 0$.
- The *spectral radius* [6, p. 177] is

$$\rho(A) := \sup\{|\lambda| : \lambda \text{ is an eigenvalue of } A\}.$$

- Since A is not assumed symmetric, we may have distinct eigenvectors for A and its transpose with the same non-zero eigenvalue. We call the later *left eigenvectors*.

Below we view l as a column with highest order entry at the top.

Nonnegative matrix solution

A third approach directly exploits non-negativity of the entries of the matrix A_m. This is best organized as a case of the *Perron-Frobenius theorem* [2], [6, Theorem 8.8.1] or [8].

- A is *row stochastic* if all entries are non-negative and each row sums to one.
- A is *irreducible* if for every pair of indices i, j, there is a natural number k with $(A^k)_{ij} \neq 0$.
- The *spectral radius* [6, p. 177] is

$$\rho(A) := \sup\{|\lambda| : \lambda \text{ is an eigenvalue of } A\}.$$

- Since A is not assumed symmetric, we may have distinct eigenvectors for A and its transpose with the same non-zero eigenvalue. We call the later *left eigenvectors*.

Below we view l as a column with highest order entry at the top.

Theorem (Perron Frobenius, Utility grade)

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

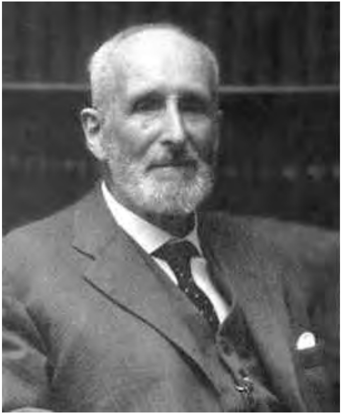
$$\lim_{k\to\infty} A^k = \begin{bmatrix} l_m & l_{m-1} & \cdots & l_2 & l_1 \\ l_m & l_{m-1} & \cdots & l_2 & l_1 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ l_m & l_{m-1} & \cdots & l_2 & l_1 \\ l_m & l_{m-1} & \cdots & l_2 & l_1 \end{bmatrix}. \tag{29}$$

Theorem (Perron Frobenius, Utility grade)

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

$$\lim_{k\to\infty} A^k = \begin{bmatrix} l_m & l_{m-1} & \cdots & l_2 & l_1 \\ l_m & l_{m-1} & \cdots & l_2 & l_1 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ l_m & l_{m-1} & \cdots & l_2 & l_1 \\ l_m & l_{m-1} & \cdots & l_2 & l_1 \end{bmatrix}. \tag{29}$$

Perron (1907) and Frobenius (1912)



Oskar Perron (1880-1975) and **Georg Frobenius** (1849-1917)

Perron-Frobenius theory

The full version of the Perron-Frobenius theorem treats arbitrary irreducible matrices with non-negative entries.

- Even in our setting, not all eigenvalues are simple: this is equivalent to A being similar to a diagonal matrix D, with entries are the eigenvalues in decreasing order, say. Then

$$A^n = U^{-1} D^n U \to U^{-1} D^\infty U$$

where the diagonal of D^∞ is $[1, 0, \cdots, 0_m]$.

- The *Jordan normal form* [7] shows (29) still follows.
- See [11] for a very nice reprise of general Perron-Frobenius theory and its multi-fold applications (and indeed *Wikipedia*).

Perron-Frobenius theory

The full version of the Perron-Frobenius theorem treats arbitrary irreducible matrices with non-negative entries.

- Even in our setting, not all eigenvalues are simple: this is equivalent to A being similar to a diagonal matrix D, with entries are the eigenvalues in decreasing order, say. Then

$$A^n = U^{-1} D^n U \to U^{-1} D^\infty U$$

where the diagonal of D^∞ is $[1, 0, \cdots, 0_m]$.

- The *Jordan normal form* [7] shows (29) still follows.
- See [11] for a very nice reprise of general Perron-Frobenius theory and its multi-fold applications (and indeed *Wikipedia*).

Perron-Frobenius theory

The full version of the Perron-Frobenius theorem treats arbitrary irreducible matrices with non-negative entries.

- Even in our setting, not all eigenvalues are simple: this is equivalent to A being similar to a diagonal matrix D, with entries are the eigenvalues in decreasing order, say. Then

$$A^n = U^{-1} D^n U \to U^{-1} D^\infty U$$

  where the diagonal of D^∞ is $[1, 0, \cdots, 0_m]$.
- The *Jordan normal form* [7] shows (29) still follows.
- See [11] for a very nice reprise of general Perron-Frobenius theory and its multi-fold applications (and indeed *Wikipedia*).

- In particular [11, §4] gives Karlin's resolvent-based proof of Perron-Frobenius.

Remark (Collatz and Wielandt (ref. 10.))

An attractive proof of the Perron-Frobenius theorem, originating with Collatz [4] and before him Perron, is to consider

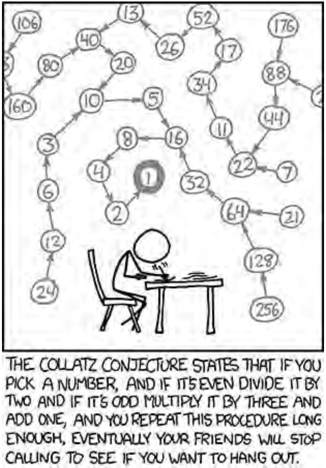
$$g(x_1, x_2, \cdots, x_m) := \min_{1 \le k \le m} \left\{ \frac{\sum_{j=1}^{m} a_{j,k} x_j}{x_k} \right\}.$$

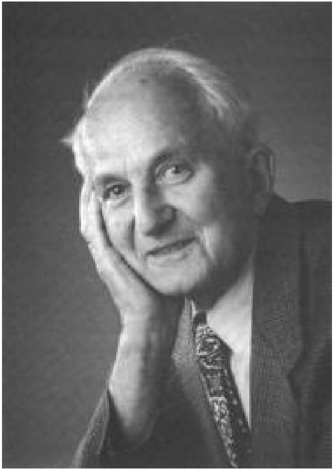
Then the maximum,

$$\max_{\sum x_j = 1, x_j \ge 0} g(x) = g(v) = 1$$

exists and yields uniquely the Perron-Frobenius vector v (which in our case is the constant vector e). ◀

The same Collatz





Lothar Collatz (1910-1990)

Example (The closed form for l)

The recursion we study is expressible as

$$\overline{x}_{n+1} = A\overline{x}_n$$

where A has k-th row A_k for m strict arithmetic means A_k. Hence A is row stochastic and strictly positive; so its *Perron eigenvalue* is 1, while $A^* l = l$ shows the limit l is the adjoint eigenvector.

- Equivalently, this is a so called *compound iteration*

$$L := \bigotimes A_k$$

as in [3, Section 8.7] and mean arguments much as in the previous section also establish convergence.

- Here we identify the eigenvector l with the corresponding linear function L since $L(x) = \langle l, x \rangle$ ◀.

Example (The closed form for l)

The recursion we study is expressible as

$$\overline{x}_{n+1} = A\overline{x}_n$$

where A has k-th row A_k for m strict arithmetic means A_k. Hence A is row stochastic and strictly positive; so its *Perron eigenvalue* is 1, while $A^* l = l$ shows the limit l is the adjoint eigenvector.

- Equivalently, this is a so called *compound iteration*

$$L := \bigotimes A_k$$

  as in [3, Section 8.7] and 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 (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 Invariance principle we gave and a specialization of the general Invariance principle in [3, Section 8.7]. ◀

The case used in (3) 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 $\sum_{k=1}^m a_k = 1$.

Remark (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 Invariance principle we gave and a specialization of the general Invariance principle in [3, Section 8.7]. ◀

The case used in (3) 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 (Weighted means revisited)

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

Proof.

We *induct* on k. If the first $k < m$ rows of A_m^k are strictly positive:

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

It follows that $(A_m^{k+1})_{1j} = \sum_{r=1}^m (A_m)_{1r}(A_m^k)_{rj} > 0$, and that, for $2 \le i \le k+1 \le m$, $(A_m^{k+1})_{ij} = \sum_{r=1}^m (A_m)_{ir}(A_m^k)_{rj} = (A_m^k)_{i-1,j} > 0$. Thus, the first $k+1$ rows of A_m^{k+1} have strictly positive entries, and we are done. □

Proposition (Weighted means revisited)

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

Proof.

We *induct* on k. If the first $k < m$ rows of A_m^k are strictly positive:

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

It follows that $(A_m^{k+1})_{1j} = \sum_{r=1}^m (A_m)_{1r}(A_m^k)_{rj} > 0$, and that, for $2 \le i \le k+1 \le m$, $(A_m^{k+1})_{ij} = \sum_{r=1}^m (A_m)_{ir}(A_m^k)_{rj} = (A_m^k)_{i-1,j} > 0$. Thus, the first $k+1$ rows of A_m^{k+1} have strictly positive entries, and we are done. □

Proposition (Weighted means revisited)

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

Proof.

We *induct* on k. If the first $k < m$ rows of A_m^k are strictly positive:

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

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. □

Proposition (Weighted means revisited)

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

Proof.

We *induct* on k. If the first $k < m$ rows of A_m^k are strictly positive:

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

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. □

Irreducibility of matrices

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 fairly 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.*

Remark (A picture is often worth a thousand words)

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 $(A')_{ij} = 1$. Also, 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*. ◀

Irreducibility of matrices

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 fairly 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.

Remark (A picture is often worth a thousand words)

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 $(A')_{ij} = 1$. Also, 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*.

Irreducibility of matrices

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 fairly 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.*

Remark (A picture is often worth a thousand words)

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 $(A')_{ij} = 1$. Also, 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*.

Irreducibility of matrices

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 fairly 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.*

Remark (A picture is often worth a thousand words)

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 $(A')_{ij} = 1$. Also, 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*. ◀

Remark (A picture is often worth a thousand words)

For our particular matrix A_m, as given in (15), the associated graph G_m is depicted in the Figure below. 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 moment's 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.

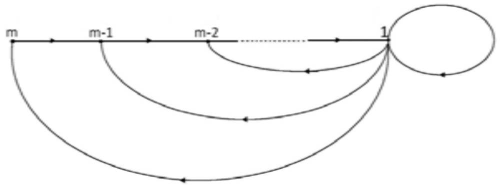


Figure: The graph G_m with adjacency matrix A_m'.

Remark (A picture is often worth a thousand words)

For our particular matrix A_m, as given in (15), the associated graph G_m is depicted in the Figure below. 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 moment's 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. ◀

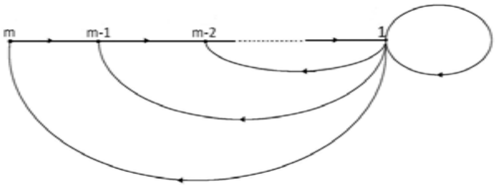


Figure: The graph G_m with adjacency matrix A_m'.

Example (Limiting examples, II)

We return to the matrices of Limiting Examples I. First

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

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.

Example (Limiting examples, II)

We return to the matrices of Limiting Examples I. First

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

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.

Example (Limiting examples, II)

Next we consider

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

Now A_3 is reducible and the limit eigenvector $[2/3, 1/3, 0]$ exists but is not strictly positive. The corresponding iteration is $x_n = (x_{n-1} + x_{n-2})/2$ with limit $(a_1 + 2a_2)/3$. (Consider our starting case in with $m = 2$ and ignore 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 $x_n = (x_{n-2} + x_{n-3})/2$.

Example (Limiting examples, II)

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 $x_n = x_{n-3}$. ◀

Example (Limiting examples, II)

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 $x_n = x_{n-3}$. ◀

Conclusion (and a Gaussian bonus)

- All three approaches have their delights and advantages.
- For the original problem, analysis as a mean iteration—while least well known—is by far the most efficient and also most elementary.
- Moreover, each approach provides lovely examples for any linear algebra class, or any introduction to computer algebra.
- Indeed, they offer different flavours of algorithmics, analysis, combinatorics, algebra and graph theory.

Carl Friedrich Gauss (1777-1855)

Conclusion (and a Gaussian bonus)

- All three approaches have their delights and advantages.
- For the original problem, analysis as a mean iteration—while least well known—is by far the most efficient and also most elementary.
- Moreover, each approach provides lovely examples for any linear algebra class, or any introduction to computer algebra.
- Indeed, they offer different flavours of algorithmics, analysis, combinatorics, algebra and graph theory.

Carl Friedrich Gauss (1777-1855)

Conclusion (and a Gaussian bonus)

- All three approaches have their delights and advantages.
- For the original problem, analysis as a mean iteration—while least well known—is by far the most efficient and also most elementary.
- Moreover, each approach provides lovely examples for any linear algebra class, or any introduction to computer algebra.
- Indeed, they offer different flavours of algorithmics, analysis, combinatorics, algebra and graph theory.

Carl Friedrich Gauss (1777-1855)

Conclusion (and a Gaussian bonus)

- All three approaches have their delights and advantages.
- For the original problem, analysis as a mean iteration—while least well known—is by far the most efficient and also most elementary.
- Moreover, each approach provides lovely examples for any linear algebra class, or any introduction to computer algebra.
- Indeed, they offer different flavours of algorithmics, analysis, combinatorics, algebra and graph theory.

Carl Friedrich Gauss (1777-1855)

Conclusion (and a Gaussian bonus)

- All three approaches have their delights and advantages.
- For the original problem, analysis as a mean iteration—while least well known—is by far the most efficient and also most elementary.
- Moreover, each approach provides lovely examples for any linear algebra class, or any introduction to computer algebra.
- Indeed, they offer different flavours of algorithmics, analysis, combinatorics, algebra and graph theory.

Carl Friedrich Gauss (1777-1855)

Example (Gauss's arithmetic-geometric mean, see ref. 3)

Consider the iteration with $a_0 := a > 0, b_0 := b > 0$ and for $n \geq 0$

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

Convergence is easy and quadratic. If asked the limit, you might again make little progress. For $a, b > 0$ let

$$\mathcal{I}(a,b) := \int_0^{\pi/2} \frac{\mathrm{d}\theta}{\sqrt{a^2\cos^2(\theta) + b^2\sin^2(\theta)}}.$$

A young Gauss discovered—and proved as *Maple* now can—that the *elliptic integral* $\mathcal{I}$ satisfies

$$\mathcal{I}(a_{n+1}, b_{n+1}) = \mathcal{I}(a_n, b_n).$$

The *Invariance principle* then confirms that $\frac{\pi/2}{\mathcal{I}(a,b)}$ is the limit.

Example (Gauss's arithmetic-geometric mean, see ref. 3)

Consider the iteration with $a_0 := a > 0, b_0 := b > 0$ and for $n \geq 0$

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

Convergence is easy and quadratic. If asked the limit, you might again make little progress. For $a, b > 0$ let

$$\mathcal{I}(a,b) := \int_0^{\pi/2} \frac{\mathrm{d}\theta}{\sqrt{a^2\cos^2(\theta) + b^2\sin^2(\theta)}}.$$

A young Gauss discovered—and proved as *Maple* now can—that the *elliptic integral* $\mathcal{I}$ satisfies

$$\mathcal{I}(a_{n+1}, b_{n+1}) = \mathcal{I}(a_n, b_n).$$

The *Invariance principle* then confirms that $\frac{\pi/2}{\mathcal{I}(a,b)}$ is the limit.

Example (Gauss's arithmetic-geometric mean, see ref. 3)

Consider the iteration with $a_0 := a > 0, b_0 := b > 0$ and for $n \geq 0$

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

Convergence is easy and quadratic. If asked the limit, you might again make little progress. For $a, b > 0$ let

$$\mathcal{I}(a,b) := \int_0^{\pi/2} \frac{\mathrm{d}\theta}{\sqrt{a^2\cos^2(\theta) + b^2\sin^2(\theta)}}.$$

A young Gauss discovered—and proved as *Maple* now can—that the *elliptic integral* $\mathcal{I}$ satisfies

$$\mathcal{I}(a_{n+1}, b_{n+1}) = \mathcal{I}(a_n, b_n).$$

The *Invariance principle* then confirms that $\frac{\pi/2}{\mathcal{I}(a,b)}$ is the limit.

Example (Gauss's arithmetic-geometric mean, see ref. 3)

Consider the iteration with $a_0 := a > 0, b_0 := b > 0$ and for $n \geq 0$

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

Convergence is easy and quadratic. If asked the limit, you might again make little progress. For $a, b > 0$ let

$$\mathcal{I}(a,b) := \int_0^{\pi/2} \frac{\mathrm{d}\theta}{\sqrt{a^2\cos^2(\theta) + b^2\sin^2(\theta)}}.$$

A young Gauss discovered—and proved as *Maple* now can—that the *elliptic integral* $\mathcal{I}$ satisfies

$$\mathcal{I}(a_{n+1}, b_{n+1}) = \mathcal{I}(a_n, b_n).$$

The *Invariance principle* then confirms that $\frac{\pi/2}{\mathcal{I}(a,b)}$ is the limit.

Example (Gauss's arithmetic-geometric mean, see ref. 3)

Consider the iteration with $a_0 := a > 0, b_0 := b > 0$ and for $n \geq 0$

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

Convergence is easy and quadratic. If asked the limit, you might again make little progress. For $a, b > 0$ let

$$\mathcal{I}(a,b) := \int_0^{\pi/2} \frac{\mathrm{d}\theta}{\sqrt{a^2\cos^2(\theta) + b^2\sin^2(\theta)}}.$$

A young Gauss discovered—and proved as *Maple* now can—that the *elliptic integral* $\mathcal{I}$ satisfies

$$\mathcal{I}(a_{n+1}, b_{n+1}) = \mathcal{I}(a_n, b_n).$$

The *Invariance principle* then confirms that $\frac{\pi/2}{\mathcal{I}(a,b)}$ is the limit. ◀

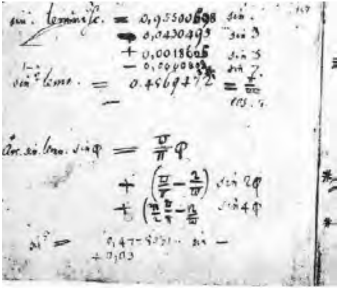


Figure 1.1. Gauss on the lemniscate.

Here is another example of Gauss's prowess at "mental experimental mathematics." One day in 1799, while examining tables of integrals provided originally by James Stirling, he noticed that the reciprocal of the integral

$$\frac{2}{\pi}\int_0^1 \frac{dt}{\sqrt{1-t^4}},$$

agreed numerically with the limit of the rapidly convergent arithmetic-geometric mean iteration: $a_0 = 1$, $b_0 = \sqrt{2}$;

$$a_{n+1} = \frac{a_n + b_n}{2}, \quad b_{n+1} = \sqrt{a_n b_n}. \tag{1.1}$$

The sequences (a_n) and (b_n) have the limit 1.1981402347355922074... in common. Based on this purely computational observation, Gauss was able to conjecture and subsequently prove that the integral is indeed equal to this common limit. It was a remarkable result, of which he wrote in his diary (see [74, pg. 5] and below) "[the result] will surely open up a whole new field of analysis." He was right. It led to the entire vista of 19th century elliptic and modular function theory.

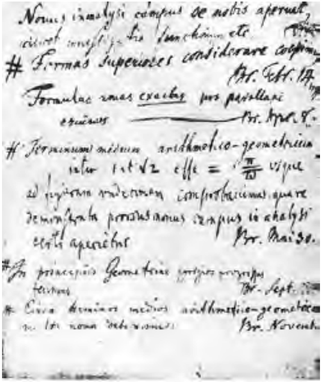


Figure 1.2. Gauss on the arithmetic-geometric mean.

In Figure 1.2, an excited Gauss writes:

> *Novus in analysi campus se nobis aperuit, scilicet investigatio functionum etc.* (October 1798) [A new field of analysis has appeared to us, evidently in the study of functions etc.]

And in May 1799 (a little further down the page), he writes:

> *Terminum medium arithmetico-geometricum inter 1 et (root 2) esse pi/omega usque ad figuram undecimam comprobavimus, qua re demonstrata prorsus novus campus in analysi certo aperietur.* [We have shown the limit of the arithmetical-geometric mean between 1 and root 2 to be pi/omega up to eleven figures, which on having been demonstrated, a whole new field in analysis is certain to be opened up.]

Example (Archimedes method, see ref. 3)

Take the slightly different iteration with $a_0 := a > 0, b_0 := b > 0$ and for $n \geq 0$

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

Convergence is easy and linear. The *Invariance principle* establishes that the limit is:

$$\mathcal{A}(a,b) := \begin{cases} \frac{\sqrt{b^2-a^2}}{\arccos(a/b)}, & 0 \leq a < b; \\ a, & a = b; \\ \frac{\sqrt{a^2-b^2}}{\operatorname{arccosh}(a/b)}, & 0 < b < a. \end{cases}.$$

Updating $1/a_n$ and $1/b_n$ tracks circumscribed and inscribed perimeters as number of sides doubles.

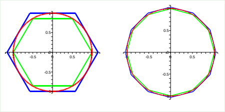


Figure 5: Archimedes' method of computing π with 6- and 12-gons

Example (Archimedes method, see ref. 3)

Take the slightly different iteration with $a_0 := a > 0, b_0 := b > 0$ and for $n \geq 0$

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

Convergence is easy and linear. The *Invariance principle* establishes that the limit is:

$$\mathcal{A}(a,b) := \begin{cases} \frac{\sqrt{b^2-a^2}}{\arccos(a/b)}, & 0 \leq a < b; \\ a, & a = b; \\ \frac{\sqrt{a^2-b^2}}{\operatorname{arccosh}(a/b)}, & 0 < b < a. \end{cases}.$$

Updating $1/a_n$ and $1/b_n$ tracks circumscribed and inscribed perimeters as number of sides doubles.

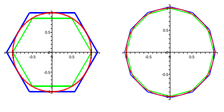


Figure 5: Archimedes' method of computing π with 6- and 12-gons

Example (Archimedes method, see ref. 3)

Take the slightly different iteration with $a_0 := a > 0, b_0 := b > 0$ and for $n \geq 0$

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

Convergence is easy and linear. The *Invariance principle* establishes that the limit is:

$$\mathcal{A}(a,b) := \begin{cases} \frac{\sqrt{b^2-a^2}}{\arccos(a/b)}, & 0 \leq a < b; \\ a, & a = b; \\ \frac{\sqrt{a^2-b^2}}{\operatorname{arccosh}(a/b)}, & 0 < b < a. \end{cases} .$$

Updating $1/a_n$ and $1/b_n$ tracks circumscribed and inscribed perimeters as number of sides doubles.

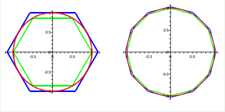


Figure 5: Archimedes' method of computing π with 6- and 12-gons

References

1. H. Bauschke, J. Sarada & S. Wang, "On moving averages." Preprint.
2. A. Berman and R. J. Plemmons, *Nonnegative Matrices in the Math Sciences*, SIAM, 1994.
3. J.M. and P. B. Borwein, *Pi and the AGM,* John Wiley, 1987.
4. Lothar Collatz, "Einschlieungssatz für die charakteristischen Zahlen von Matrize," *Math Zeit.* **48** (1) (1942), 221–226.
5. C.-E. Froberg, *Introduction to Numerical Analysis*, 2nd ed., Addison-Wesley, 1969.
6. C. Godsil and G.F. Royle, *Algebraic Graph Theory* Springer, 2001.
7. G.H. Golub and C.F. Van Loan, *Matrix Computations*, 3rd ed., Johns Hopkins, 1996.
8. C. R. MacCluer, "The Many Proofs and Applications of Perron's Theorem," *SIAM Review*, **42**, (2000), 487–498.
9. A.M. Ostrowski, *Solution of Equations in Euclidean and Banach Spaces*, Academic Press, 1973.