On moving averages

Heinz H. Bauschke[*], Joshua Sarada[†], and Xianfu Wang[‡]

April 1, 2013

Abstract

We show that the moving arithmetic average is closely connected to a Gauss–Seidel type fixed point method studied by Bauschke, Wang and Wylie, and which was observed to converge only numerically. Our analysis establishes a rigorous proof of convergence of their algorithm in a special case; moreover, the limit is explicitly identified. Moving averages in Banach spaces and Kolmogorov means are also studied. Furthermore, we consider moving proximal averages and epi-averages of convex functions.

2010 Mathematics Subject Classification: Primary 15B51, 26E60, 47H10; Secondary 39A06, 65H04, 47J25, 49J53.

Keywords: Arithmetic mean, difference equation, epi-average, Kolmogorov mean, linear recurrence relation, means, moving average, proximal average, stochastic matrix.

1 Introduction

Throughout this paper, we assume that $m \in \{2, 3, \ldots\}$ and that

(1) $X = \mathbb{R}^m$ is an m-dimension real Euclidean space with standard inner product $\langle \cdot, \cdot \rangle$

[*]Mathematics, University of British Columbia, Kelowna, B.C. V1V 1V7, Canada. E-mail: heinz.bauschke@ubc.ca.

[†]59 Tuscany Valley Hts NW, Calgary, Alberta T3L 2E7, Canada. E-mail: jshsarada@gmail.com.

[‡]Mathematics, University of British Columbia, Kelowna, B.C. V1V 1V7, Canada. E-mail: shawn.wang@ubc.ca.

and induced norm $\|\cdot\|$. We also assume that

$$\alpha_0, \ldots, \alpha_{m-1} \text{ are nonnegative real numbers such that } \sum_{i=0}^{m-1} \alpha_i = 1, \text{ and } \alpha_m := \alpha_0. \tag{2}$$

It will be convenient to introduce the following notation for the partial sums and associated weights:

$$(\forall k \in \{0,1,\ldots,m-1\}) \quad a_k = \sum_{i=0}^{k} \alpha_i \ \text{ and } \ \mathbf{a} = (a_0, a_1, \ldots, a_{m-1})^*. \tag{3}$$

and

$$(\forall k \in \{0,1,\ldots,m-1\}) \quad \lambda_k = \frac{\mathbf{a}^*\mathbf{e}_k}{\mathbf{a}^*\mathbf{e}} = \frac{a_k}{\sum_{i=0}^{m-1} a_i} = \frac{\sum_{i=0}^{k} \alpha_i}{\sum_{j=0}^{m-1}\sum_{i=0}^{j} \alpha_i} \text{ and } \boldsymbol{\lambda} = \begin{pmatrix} \lambda_0 \\ \vdots \\ \lambda_{m-1} \end{pmatrix}. \tag{4}$$

We will study the homogeneous linear difference equation

$$(\forall n \geq m) \quad y_n = \alpha_{m-1} y_{n-1} + \cdots + \alpha_0 y_{n-m}, \ \text{ where } \ (y_0, \ldots, y_{m-1}) \in \mathbb{R}^m. \tag{5}$$

In the literature, this is called the *moving average* (and it is also known as the rolling average, rolling mean or running average). It says that given a series of numbers and a fixed subset size, the first element of the moving average is obtained by taking the average of the initial fixed subset of the number series. Then the subset is modified by "shifting forward", excluding the first number of the series and including the next number following the original subset in the series. This creates a new subset of numbers, which is averaged. This process is repeated over the entire data series. The moving average is widely applied in statistics, signal processing, econometrics and mathematical finance; see, e.g., also [9, 10, 15, 13, 26].

In [4, 5], we observed the numerical convergence of a Gauss–Seidel type fixed point iteration numerically, but we were unable to provide a rigorous proof. In this note, we present a connection between the moving average and this fixed point recursion; this allows us to give an analytical proof for the case when all monotone operators are zero.

Moreover, the limit is identified. However this approach is unlike to generalize to the general fixed point iteration due to the interlaced nonlinear resolvents.

While the results rest primarily on results from linear algebra, we consider in the second half of the paper several related highly nonlinear moving averages that exhibit a "hidden linearity" and thus allow to be rigorously studied.

The paper is organized as follows. In Section 2, we present various facts and auxiliary results rooted ultimately in Linear Algebra. In Section 3 we establish the connection between the moving average and the Gauss–Seidel type iteration scheme studied by Bauschke, Wang and Wylie, which leads to a rigorous proof for the convergence of their algorithm in the aforementioned special case. In Section 4 we show that the iteration matrix has a closed form when $\alpha_0 = 0$ and $\alpha_1 = \cdots = \alpha_{m-1} = 1/(m-1)$. Moving Kolmogorov means (also known as f-means) are considered in Section 5. In particular, various known means such as arithmetic mean, harmonic mean, resolvent mean, etc. [10, 7] all turn out to be special cases of this general framework. The final Section 6 concerns moving proximal averages and epi-averages of convex functions.

Our notation follows [3, 19, 24, 25]. The *identity operator* Id is defined by $X \to X \colon x \mapsto x$. A mapping $T \colon X \to X$ is *nonexpansive* (Lipschitz-1) if $(\forall x \in X)(\forall y \in X)$ $\|Tx - Ty\| \leq \|x - y\|$. The set of *fixed points* of T is denoted by $\operatorname{Fix} T = \{x \in X \mid x = Tx\}$. The following matrix plays a central role in this paper

$$\mathbf{A} = \begin{pmatrix} 0 & 1 & 0 & \ldots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ \alpha_0 & \alpha_1 & \alpha_2 & \cdots & \alpha_{m-1} \end{pmatrix}. \tag{6}$$

Note that

$$\operatorname{Fix} \mathbf{A} = \mathbf{D} = \left\{ \mathbf{x} = (x, x, \ldots, x) \in X \;\middle|\; x \in \mathbb{R} \right\}, \tag{7}$$

where $\mathbf{D}$ is also called the *diagonal* in X. By [19, page 648], the characteristic polynomial of $\mathbf{A}$ is

$$p(x) = x^m - \alpha_{m-1}x^{m-1} - \cdots - \alpha_1 x - \alpha_0, \tag{8}$$

and, in turn, $\mathbf{A}$ is the (transpose of the) companion matrix of $p(x)$.

As usual, we use $\mathbb{C}$ for the field of complex numbers; $\mathbb{R}_+$ ($\mathbb{R}_{++}$) for the nonnegative real numbers (positive real numbers); $\mathbb{N} = \{0, 1, 2, \ldots\}$ for the natural numbers. For

$\mathbf{B} \in \mathbb{C}^{m\times m}$, $\rho(\mathbf{B}) = \max_{\lambda\in\sigma(\mathbf{B})} |\lambda|$ is the spectral radius of $\mathbf{B}$ and $\sigma(\mathbf{B})$ denotes the *spectrum* of $\mathbf{B}$, i.e., the set of eigenvalues of $\mathbf{B}$. Given $\lambda \in \sigma(\mathbf{B})$, we say that λ is *simple* if its algebraic multiplicity is 1 and that it is *semisimple* if its algebraic and geometric multiplicities coincide. A simple eigenvalue is always semisimple, see [19, pages 510–511]. A vector $\mathbf{y} \in \mathbb{C}^m \smallsetminus \{0\}$ satisfying $\mathbf{B}\mathbf{y} = \lambda\mathbf{y}$ ($\mathbf{y}^*\mathbf{B} = \lambda\mathbf{y}^*$) is called a *right-hand (left-hand) eigenvector* of $\mathbf{B}$ with respect to the eigenvalue λ. (The * denotes the conjugate transpose of a vector or a matrix.) Then *range* of an operator $\mathbf{B}$ is denoted by $\operatorname{ran}(\mathbf{B})$; if $\mathbf{B}$ is linear, we denote its *kernel* by $\ker(\mathbf{B})$. It will be convenient to denote the ith standard unit column vector by $\mathbf{e}_i$ and to also set $\mathbf{e} = \sum_{i=1}^m \mathbf{e}_i = (1, 1, \dots, 1)^* \in X$. We denote by $\mathbb{S}^{N\times N}$ the space of all $N \times N$ real symmetric matrices, while $\mathbb{S}_+^{N\times N}$ and $\mathbb{S}_{++}^{N\times N}$ stand, respectively, for the set of $N \times N$ positive semidefinite and positive definite matrices. The *greatest common divisor* of a set of integers S is denoted by $\gcd(S)$. Turning to functions, we let $\Gamma(X)$ be the set of all functions that are convex, lower semicontinuous, and proper. We denote the quadratic energy function by $\mathfrak{q} = \frac{1}{2}\|\cdot\|^2$. Finally, the *Fenchel conjugate* g^* of a function g is given by $g^*(x) = \sup_{y\in X}(\langle x, y\rangle - g(y))$.

2 Linear algebraic results

To make our analysis self-contained, let us gather in this section some useful facts and auxiliary results on stochastic matrices, linear recurrence relations and the convergence of matrix powers.

Basic properties of stochastic matrices

Recall that a matrix $\mathbf{B}$ with nonnegative entries is called *stochastic* (or row-stochastic) if each row sum is equal to 1. (See [8, Chapter 8] or [19, Section 8.4] for more on stochastic matrices.)

Fact 2.1 (See [19, pages 689 and 696].) *Let $\mathbf{B} \in \mathbb{R}^{m\times m}$ be stochastic. Then $\rho(\mathbf{B}) = 1$ and 1 is a semisimple eigenvalue of $\mathbf{B}$ with eigenvector $\mathbf{e}$.*

The proof of the next result is a simple verification and hence omitted.

Proposition 2.2 *The vectors $\mathbf{a}^*$ and $\mathbf{e}$ are, respectively, left-hand and right-hand eigenvectors of $\mathbf{A}$ associated with the eigenvalue 1.*

Linear recurrence relations and polynomials

Consider the linear recurrence relation

$$(9) \qquad (\forall n \geq m) \quad y_n = \alpha_{m-1} y_{n-1} + \cdots + \alpha_0 y_{n-m}, \quad \text{where} \quad (y_0, \ldots, y_{m-1}) \in \mathbb{R}^m.$$

Setting $(\forall n \in \mathbb{N})$ $\mathbf{y}^{[n]} = (y_n, y_{n+1}, \ldots, y_{n+m-1})^*$, we see that we can rewrite (9) as

$$(10) \qquad (\forall n \in \mathbb{N}) \quad \mathbf{y}^{[n+1]} = \mathbf{A}\mathbf{y}^{[n]} = \cdots = \mathbf{A}^{n+1}\mathbf{y}^{[0]}.$$

This explains our interest in understanding the limiting behaviour of powers of $\mathbf{A}$, which yield information about the limiting behaviour of $(\mathbf{y}^{[n]})_{n \in \mathbb{N}}$ and hence of $(y_n)_{n \in \mathbb{N}}$. The solution to (9) depends on the roots of characteristic polynomial

$$(11) \qquad p(x) = x^m - \alpha_{m-1} x^{m-1} - \cdots - \alpha_1 x - \alpha_0$$

of $\mathbf{A}$.

Fact 2.3 (See [22, page 90] or [21, pages 74–75]) *Let k be the number of distinct roots $u_1, \ldots, u_k$ of* (11) *with multiplicities $m_1, \ldots, m_k$, respectively. Then for each $i \in \{1, \ldots, k\}$, there exists a polynomial q_i of degree $m_i - 1$ such that the general solution of* (9) *is*

$$(12) \qquad (\forall n \in \mathbb{N}) \quad y_n = \sum_{i=1}^{k} q_i(n) u_i^n.$$

Consequently, if $k = m$, i.e., all roots are distinct, then there exists $(\nu_1, \ldots, \nu_m) \in \mathbb{C}^m$ such that

$$(13) \qquad (\forall n \in \mathbb{N}) \quad y_n = \sum_{i=1}^{m} \nu_i u_i^n.$$

Fact 2.4 (Ostrowski) (See [22, Theorem 12.2].) *Let $(b_1, \ldots, b_m) \in \mathbb{R}_+^m$ and set*

$$(14) \qquad q(x) = x^m - b_1 x^{m-1} - \cdots - b_m.$$

Suppose that $\gcd \big\{ k \in \{1, \ldots, m\} \;\big|\; b_k > 0 \big\} = 1$. *Then q has a unique positive root r. Moreover, r is a simple root of q, and the modulus of every other root of q is strictly less than r.*

See also [12, 23] for further results on polynomials. Let us now state a basic assumption that we will impose repeatedly.

Basic Hypothesis 2.5 The polynomial $p(x)$ given by (11) has a unique positive root, which is simple and equal to 1, and each other root has modulus strictly less than 1. This happens if one of the following holds:

(i) $\gcd\big\{i \in \{1,\ldots,m\} \;\big|\; \alpha_{m-i} > 0\big\} = 1$.

(ii) $\alpha_{m-1} > 0$.

(iii) $(\exists\, i \in \{1,\ldots,m-1\})\ \alpha_{m-i}\alpha_{m-(i+1)} > 0$.

(iv) $(\exists\, \{i,j\} \subseteq \{1,\ldots,m-1\})\ \gcd\{i,j\} = 1$ and $\alpha_{m-i}\alpha_{m-j} > 0$.

Proof. (i): This is clear from Fact 2.4 and the definition of $p(x)$. (ii)–(iv): Each condition implies (i). ■

Limits of matrix powers and linear recurrence relations

Fact 2.6 (See [19, pages 383–386, 518 and 630].) *Let* $\mathbf{B} \in \mathbb{C}^{m\times m}$*. Then the following hold:*

(i) $(\mathbf{B}^n)_{n\in\mathbb{N}}$ *converges to a nonzero matrix if and only if* 1 *is a semisimple eigenvalue of* $\mathbf{B}$ *and every other eigenvalue of* $\mathbf{B}$ *has modulus strictly less than* 1.

(ii) *If* $\lim_{n\in\mathbb{N}} \mathbf{B}^n$ *exists, then it is the projector onto* $\ker(\operatorname{Id} - \mathbf{B})$ *along* $\operatorname{ran}(\operatorname{Id} - \mathbf{B})$.

(iii) *If* $\rho(\mathbf{B}) = 1$ *is a simple eigenvalue with right-hand and left-hand eigenvectors* x *and* y^* *respectively, then* $\displaystyle\lim_{n\in\mathbb{N}} \mathbf{B}^n = \frac{xy^*}{y^*x}$.

Corollary 2.7 *Suppose that the* **Basic Hypothesis 2.5** *holds. Then*

$$\lim_{n\in\mathbb{N}} \mathbf{A}^n = \frac{\mathbf{e}\mathbf{a}^*}{\mathbf{a}^*\mathbf{e}} = \frac{1}{\sum_{k=0}^{m-1}\sum_{i=0}^{k}\alpha_i}\begin{pmatrix} \alpha_0 & \sum_{i=0}^{1}\alpha_i & \cdots & \sum_{i=0}^{m-2}\alpha_i & 1 \\ \alpha_0 & \sum_{i=0}^{1}\alpha_i & \cdots & \sum_{i=0}^{m-2}\alpha_i & 1 \\ \vdots & & & & \\ \alpha_0 & \sum_{i=0}^{1}\alpha_i & \cdots & \sum_{i=0}^{m-2}\alpha_i & 1 \end{pmatrix} = \mathbf{e}\boldsymbol{\lambda}^*. \tag{15}$$

Proof. Clear from Proposition 2.2, Fact 2.6(iii) and the definitions. ■

Corollary 2.8 *Suppose that the* **Basic Hypothesis 2.5** *holds and consider the sequence* $(y_n)_{n\in\mathbb{N}}$ *generated by the linear recurrence relation*

$$(\forall n \geq m) \quad y_n = \alpha_{m-1}y_{n-1} + \cdots + \alpha_0 y_{n-m}, \quad \text{where} \quad \mathbf{y} = (y_0,\ldots,y_{m-1}) \in \mathbb{R}^m. \tag{16}$$

Then

$$\lim_{n\in\mathbb{N}} y_n = \frac{\mathbf{a}^*\mathbf{y}}{\mathbf{a}^*\mathbf{e}} = \frac{\sum_{k=0}^{m-1} a_k y_k}{\sum_{k=0}^{m-1} a_k} = \frac{\sum_{k=0}^{m-1}\sum_{i=0}^{k}\alpha_i y_k}{\sum_{k=0}^{m-1}\sum_{i=0}^{k}\alpha_i} = \sum_{k=0}^{m-1}\lambda_k y_k. \tag{17}$$

Proof. This follows from Corollary 2.7 and (10). ■

Definition 2.9 *Let Y be a real Banach space and consider the sequence $(y_n)_{n\in\mathbb{N}}$ in Y generated by the linear recurrence relation*

$$(18)\qquad (\forall n \geq m) \quad y_n = \alpha_{m-1}y_{n-1} + \cdots + \alpha_0 y_{n-m}, \ \ \textit{where} \ \ \mathbf{y} = (y_0, \ldots, y_{m-1}) \in Y^m.$$

If $\lim_{n\in\mathbb{N}} y_n$ *exists, then we set it equal to* $\ell(\mathbf{y})$ *and we write* $\mathbf{y} \in \operatorname{dom} \ell$.

Remark 2.10 Note that $\operatorname{dom}\ell$ is a linear subspace of Y^m, that $\{\mathbf{y} = (y, \ldots, y) \in Y^m \mid y \in Y\} \subseteq \operatorname{dom}\ell$, and that $\ell\colon \operatorname{dom}\ell \to Y$ is a linear operator. Furthermore,

$$(19)\quad \big(\forall (y_0, \ldots, y_{m-1}) \in Y^m\big) \quad \{y_n\}_{n\in\mathbb{N}} \subseteq \operatorname{conv}\{y_0, \ldots, y_{m-1}\} \subseteq \operatorname{span}\{y_0, \ldots, y_{m-1}\}.$$

This implies

$$(20)\quad (\forall \mathbf{y} = (y_0, \ldots, y_{m-1}) \in \operatorname{dom}\ell) \quad \ell(\mathbf{y}) \in \operatorname{conv}\{y_0, \ldots, y_{m-1}\} \subseteq \operatorname{span}\{y_0, \ldots, y_{m-1}\}$$

because the convex hull of finitely many points is compact (see, e.g., [1, Corollary 5.30]).

We are now ready to lift Corollary 2.8 to general Banach spaces.

Corollary 2.11 *Suppose that the* **Basic Hypothesis 2.5** *holds and consider sequence $(y_n)_{n\in\mathbb{N}}$ generated by the linear recurrence relation*

$$(21)\qquad (\forall n \geq m) \quad y_n = \alpha_{m-1}y_{n-1} + \cdots + \alpha_0 y_{n-m}, \ \ \textit{where} \ \ \mathbf{y} = (y_0, \ldots, y_{m-1}) \in Y^m,$$

where Y is a real Banach space. Then

$$(22)\qquad \lim_{n\in\mathbb{N}} y_n = \ell(y_0, \ldots, y_{m-1}) = \frac{\sum_{k=0}^{m-1} a_k y_k}{\sum_{k=0}^{m-1} a_k} = \frac{\sum_{k=0}^{m-1}\sum_{i=0}^{k} \alpha_i y_k}{\sum_{k=0}^{m-1}\sum_{i=0}^{k} \alpha_i} = \sum_{k=0}^{m-1} \lambda_k y_k.$$

Proof. Denote the right-hand side of (22) by z, let $y^* \in Y^*$ and define $(\forall n \in \mathbb{N})$ $\eta_n = y^*(y_n)$. Applying y^* to (21) gives rise to the linear recurrence relation

$$(23a)\qquad (\forall n \geq m) \quad \eta_n = \alpha_{m-1}\eta_{n-1} + \cdots + \alpha_0\eta_{n-m},$$

where

$$(23b)\qquad (\eta_0, \ldots, \eta_{m-1}) = \big(y^*(y_0), \ldots, y^*(y_{m-1})\big) \in \mathbb{R}^m.$$

Corollary 2.8 and the linearity and continuity of y^* imply that

$$(24)\qquad y^*(y_n) = \eta_n \to \frac{\sum_{k=0}^{m-1} a_k \eta_k}{\sum_{k=0}^{m-1} a_k} = y^*\left(\frac{\sum_{k=0}^{m-1} a_k y_k}{\sum_{k=0}^{m-1} a_k}\right) = y^*(z).$$

It follows that $(y_n)_{n\in\mathbb{N}}$ converges weakly to z. On the other hand, $(y_n)_{n\in\mathbb{N}}$ lies not only in a compact subset but also in a finite-dimensional subspace of Y (see Remark 2.10). Altogether, we deduce that $(y_n)_{n\in\mathbb{N}}$ strongly converges to z. ■

Reducible matrices

Recall (see, e.g., [19, page 671]) that $\mathbf{B} \in \mathbb{C}^{m\times m}$ is *reducible* if there exists a permutation matrix P such that

$$P^*\mathbf{B}P = \begin{pmatrix} U & W \\ 0 & V \end{pmatrix}, \tag{25}$$

where U and V are nontrivial square matrices; otherwise, $\mathbf{B}$ is *irreducible*.

Proposition 2.12 *Let $\mathbf{B} \in \mathbb{C}^{m\times m}$ have at least one zero column. Then $\mathbf{B}$ is reducible.*

Proof. By assumption, there exists a permutation matrix P such that the first column of $\mathbf{B}P$ is equal to the zero vector. Then

$$P^*\mathbf{B}P = \begin{pmatrix} U & W \\ 0 & V \end{pmatrix}, \tag{26}$$

where $U = 0 \in \mathbb{C}^{1\times 1}$ and $V \in \mathbb{C}^{(m-1)\times(m-1)}$ are nontrivial square matrices. ■

Circulant matrices

It is interesting to compare the above results to circulant matrices. To this end, we set

$$\mathbf{C} = \begin{pmatrix} \alpha_0 & \alpha_1 & \alpha_2 & \cdots & \alpha_{m-1} \\ \alpha_{m-1} & \alpha_0 & \alpha_1 & \cdots & \alpha_{m-2} \\ \alpha_{m-2} & \alpha_{m-1} & \alpha_0 & \cdots & \alpha_{m-3} \\ & \cdots & & \vdots & \\ \alpha_1 & \alpha_2 & \alpha_3 & \cdots & \alpha_0 \end{pmatrix}. \tag{27}$$

Fact 2.13 (See [18, Theorems 1 and 2].) *Set $U = \big\{k \in \{1,\ldots,m\} \;\big|\; \alpha_k > 0\big\}$. Then*

$$\mathbf{L} = \lim_{n\in\mathbb{N}} \mathbf{C}^n \text{ exists} \tag{28}$$

if and only if $\gamma = \gcd(U \cup \{m\}) = \gcd((U-U)\cup\{m\})$, in which case the entries of $\mathbf{L}$ satisfy

$$\mathbf{L}_{i,j} = \begin{cases} \gamma/m, & \text{if } i \equiv j \pmod{\gamma}; \\ 0, & \text{otherwise.} \end{cases} \tag{29}$$

Remark 2.14 Note that $\mathbf{C}$ is *bistochastic*, i.e., both $\mathbf{C}$ and $\mathbf{C}^*$ are stochastic. Since permutation matrices are clearly nonexpansive, it follows from Birkhoff's theorem (see, e.g., [17, Theorem 8.7.1]) that $\mathbf{C}$ is convex combination of nonexpansive matrices. Hence, $\mathbf{C}$ is nonexpansive as well. Moreover, one verifies readily that $\operatorname{Fix}\mathbf{C} = \mathbf{D}$.

Example 2.15 Suppose there exists $i \in \{1, \ldots, m-1\}$ such that $\{i, i+1\} \subseteq U$. Then $(\mathbf{C}^n)_{n\in\mathbb{N}}$ converges to the orthogonal projector onto $\mathbf{D}$, i.e.,

$$\lim_{n\in\mathbb{N}} \mathbf{C}^n = \frac{1}{m}\mathbf{e}\mathbf{e}^* = \frac{1}{m}\begin{pmatrix} 1 & 1 & \cdots & 1 \\ 1 & 1 & \cdots & 1 \\ \vdots & & & \\ 1 & 1 & \cdots & 1 \end{pmatrix}. \tag{30}$$

For further results on circulant matrices, see [14] and in particular [16].

3 Main results

In this section, we study the convergence of the Gauss–Seidel type fixed point iteration scheme proposed in [4, 5]. Recalling (2), we set

$$\mathbf{T}_1 = \begin{pmatrix} \alpha_0 & \alpha_1 & \alpha_2 & \cdots & \alpha_{m-1} \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & & & & \\ 0 & 0 & 0 & \cdots & 1 \end{pmatrix}, \quad \mathbf{T}_2 = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ \alpha_{m-1} & \alpha_0 & \alpha_1 & \cdots & \alpha_{m-2} \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & & & & \\ 0 & 0 & 0 & \cdots & 1 \end{pmatrix}, \tag{31a}$$

$$\vdots$$

$$\mathbf{T}_m = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & & & & \\ \alpha_1 & \alpha_2 & \alpha_3 & \cdots & \alpha_0 \end{pmatrix}; \tag{31b}$$

or entrywise

$$\big(\mathbf{T}_k\big)_{i,j} = \begin{cases} \alpha_{m-k+j}, & \text{if } i = k \text{ and } 1 \leq j < k; \\ \alpha_{j-k}, & \text{if } i = k \text{ and } m \geq j \geq k; \\ 1, & \text{if } i \neq k \text{ and } j = i; \\ 0, & \text{otherwise.} \end{cases} \tag{32}$$

Furthermore, we define

$$\mathbf{T} = \mathbf{T}_m \cdots \mathbf{T}_1. \tag{33}$$

When $\alpha_0 = 0, \alpha_1 = \cdots = \alpha_{m-1} = \frac{1}{m-1}$, iterating $\mathbf{T}$ exactly corresponds to the algorithm investigated in [4, 5] when all monotone operators are zeros. The authors there observed the numerical convergence but were not able to provide a rigorous proofs. We will present a rigorous proof by connecting $\mathbf{T}$ and $\mathbf{A}$. We start with some basic properties.

Proposition 3.1 *Let $k \in \{1, \ldots, m\}$. Then the following hold:*

(i) *$\mathbf{T}_k$ and $\mathbf{T}$ are stochastic.*

(ii) *If $\alpha_0 = 0$, then neither $\mathbf{T}_k$ nor $\mathbf{T}$ is nonexpansive.*

(iii) *If $\alpha_0 = 0$, then neither $\mathbf{T}_k$ nor $\mathbf{T}$ is irreducible.*

Proof. (i): By (2), $\mathbf{T}_k$ is stochastic and so is therefore $\mathbf{T}$ as a product of stochastic matrices.

(ii): Indeed, in this case $\mathbf{T}_k(\mathbf{e} - \mathbf{e}_k) = \mathbf{e}$ and thus $\|\mathbf{T}_k(\mathbf{e} - \mathbf{e}_k)\|^2 = m > m - 1 = \|\mathbf{e} - \mathbf{e}_k\|^2$. Similarly, $\mathbf{T}(\mathbf{e} - \mathbf{e}_1) = \mathbf{e}$ and thus $\|\mathbf{T}(\mathbf{e} - \mathbf{e}_1)\|^2 = m > m - 1 = \|\mathbf{e} - \mathbf{e}_1\|^2$.

(iii): In this case, the $\mathbf{T}_1$ and $\mathbf{T}_k$ have a zero column, hence so does $\mathbf{T}$. Now apply Proposition 2.12. ■

Proposition 3.1 illustrates neither the theory of nonexpansive mappings nor that of irreducible matrices (as is done in, e.g., [9, 19]) is applicable to study limiting properties of $(\mathbf{T}^n)_{n\in\mathbb{N}}$. Fortunately, we are able to base our analysis on the right-shift and left-shift operators which are respectively given by

$$R = \begin{pmatrix} 0 & 0 & 0 & \cdots & 1 \\ 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \vdots & & \ddots & \ddots & \\ 0 & \cdots & 0 & 1 & 0 \end{pmatrix}, \qquad L = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ 1 & 0 & 0 & \cdots & 0 \end{pmatrix}. \tag{34}$$

Note that R and L are permutation matrices which satisfy the following properties which we will use repeatedly:

$$R^m = L^m = \mathrm{Id},\ R^* = R^{-1} = L,\ L^* = L^{-1} = R,\ R^{m-k} = L^k,\ \text{where } k \in \{0, \ldots, m\}. \tag{35}$$

A key observation is the following result which connects $\mathbf{T}_k$ and the companion matrix $\mathbf{A}$.

Proposition 3.2 *For every $k \in \{1, \ldots, m\}$, we have*

$$\mathbf{T}_k = R^k \mathbf{A} L^{k-1}, \tag{36a}$$

$$\mathbf{T}_k\mathbf{T}_{k-1}\cdots\mathbf{T}_1 = R^k\mathbf{A}^k, \tag{36b}$$

$$\mathbf{T} = \mathbf{A}^m. \tag{36c}$$

Proof. We prove this by induction on k. Clearly, (36a) and (36b) hold when $k = 1$. Now assume that (36a) and (36b) hold for some $k \in \{1, \ldots, m-1\}$. Then $\mathbf{T}_{k+1} = R\mathbf{T}_k L = R(R^k A L^{k-1})L = R^{k+1}AL^k$, which is (36a) for $k+1$. Hence $\mathbf{T}_{k+1}\cdots\mathbf{T}_1 = \mathbf{T}_{k+1}(\mathbf{T}_k\cdots\mathbf{T}_1) = (R^{k+1}\mathbf{A}L^k)(R^k\mathbf{A}^k) = R^{k+1}\mathbf{A}^{k+1}$, which is (36b) for $k+1$. Finally, (36c) follows from (33) and (36b) with $k = m$. ■

We are now able to derive our main results which resolves a special case of an open problem posed in [5].

Theorem 3.3 (main result) *Suppose that the* **Basic Hypothesis 2.5** *holds. Then*

$$\lim_{n\in\mathbb{N}} \mathbf{T}^n = \frac{\mathbf{e}\mathbf{a}^*}{\mathbf{a}^*\mathbf{e}} = \frac{1}{\sum_{k=0}^{m-1}\sum_{i=0}^{k}\alpha_i}\begin{pmatrix} \alpha_0 & \sum_{i=0}^{1}\alpha_i & \cdots & \sum_{i=0}^{m-2}\alpha_i & 1 \\ \alpha_0 & \sum_{i=0}^{1}\alpha_i & \cdots & \sum_{i=0}^{m-2}\alpha_i & 1 \\ \vdots & & & & \\ \alpha_0 & \sum_{i=0}^{1}\alpha_i & \cdots & \sum_{i=0}^{m-2}\alpha_i & 1 \end{pmatrix} = \mathbf{e}\boldsymbol{\lambda}^*. \tag{37}$$

and $\operatorname{Fix}\mathbf{T} = \mathbf{D} = \operatorname{Fix}\mathbf{T}_1 \cap \cdots \cap \operatorname{Fix}\mathbf{T}_m$.

Proof. In view of (36c), it is clear that $(\mathbf{T}^n)_{n\in\mathbb{N}}$ is a subsequence of $(\mathbf{A}^n)_{n\in\mathbb{N}}$. Hence (37) follows from Corollary 2.7.

Now set $\mathbf{L} = \lim_{n\in\mathbb{N}}\mathbf{T}^n = \mathbf{a}^*\mathbf{e}/(\mathbf{e}^*\mathbf{a})$. On the one hand, $\mathbf{D} = \operatorname{Fix}\mathbf{A} \subseteq \operatorname{Fix}\mathbf{T}$ by (7) and (36c). On the other hand, $\operatorname{Fix}\mathbf{T} \subseteq \operatorname{ran}\mathbf{L} \subseteq \mathbf{D}$. Altogether, $\operatorname{Fix}\mathbf{T} = \mathbf{D}$. Finally, clearly $\operatorname{Fix}\mathbf{T}_1 \cap \cdots \cap \operatorname{Fix}\mathbf{T}_m \subseteq \operatorname{Fix}\mathbf{T} = \mathbf{D} \subseteq \operatorname{Fix}\mathbf{T}_1 \cap \cdots \cap \operatorname{Fix}\mathbf{T}_m$. ■

Remark 3.4 Theorem 3.3, with the choice $(\alpha_0, \ldots, \alpha_{m-1}) = (m-1)^{-1}(0,1,1,\ldots,1)$, settles [5, Questions **Q1** and **Q2** on p. 36] when all resolvents coincide with Id. Furthermore, (37) gives an *explicit formula* for the limit of $(\mathbf{T}^n)_{n\in\mathbb{N}}$.

4 A special case: $\alpha_0 = 0$ and $\alpha_1 = \cdots = \alpha_{m-1} = 1/(m-1)$

This assignment of parameters was the setting of [5]. In this case, even closed forms for $\mathbf{T}$, and for the partial products $\mathbf{T}_k\mathbf{T}_{k-1}\cdots\mathbf{T}_1$ and A^k, are available as we demonstrate in this section. To this end, we abbreviate

$$\gamma = \frac{1}{m-1}. \tag{38}$$

Then (31) and (6) turn into

$$\mathbf{T}_1 = \begin{pmatrix} 0 & \gamma & \gamma & \cdots & \gamma \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & & & & \\ 0 & 0 & 0 & \cdots & 1 \end{pmatrix}, \qquad \mathbf{T}_2 = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ \gamma & 0 & \gamma & \cdots & \gamma \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & & & & \\ 0 & 0 & 0 & \cdots & 1 \end{pmatrix}, \tag{39a}$$

$$\ldots,$$

$$\mathbf{T}_m = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & & & & \\ \gamma & \gamma & \gamma & \cdots & 0 \end{pmatrix}, \text{ and } \mathbf{A} = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 \\ \vdots & & & & \\ 0 & \gamma & \gamma & \cdots & \gamma \end{pmatrix}. \tag{39b}$$

Proposition 4.1 *Define* $\widetilde{\mathbf{T}} \in \mathbb{R}^{m \times m}$ *entrywise by*

$$\widetilde{\mathbf{T}}_{i,j} := \begin{cases} \gamma(1+\gamma)^{i-1}, & \text{if } i < j; \\ \\ \gamma(1+\gamma)^{i-1} - \gamma(1+\gamma)^{i-j}, & \text{if } i \geq j. \end{cases} \tag{40}$$

Let $k \in \{1, \ldots, m\}$. *Then*

$$\mathbf{T}_k \mathbf{T}_{k-1} \cdots \mathbf{T}_1 = \begin{pmatrix} \text{first } k \text{ rows of } \widetilde{\mathbf{T}} \\ 0_{(m-k)\times k} \quad I_{m-k} \end{pmatrix}; \tag{41}$$

or entrywise

$$\left(\mathbf{T}_k \mathbf{T}_{k-1} \cdots \mathbf{T}_1\right)_{i,j} := \begin{cases} \widetilde{\mathbf{T}}_{i,j}, & \text{if } 1 \leq i \leq k; \\ 1, & \text{if } k+1 \leq i \leq m \text{ and } j = i; \\ 0, & \text{if } k+1 \leq i \leq m \text{ and } j \neq i. \end{cases} \tag{42}$$

In particular,

$$\mathbf{T} = \widetilde{\mathbf{T}}. \tag{43}$$

Proof. Set

$$\mathbf{S}_k = \mathbf{T}_k \mathbf{T}_{k-1} \cdots \mathbf{T}_1. \tag{44}$$

We prove (41) by induction on $k \in \{1, \ldots, m\}$; the rest follows readily.

$k = 1$: On the one hand, by definition,

$$\mathbf{T}_1 = \begin{pmatrix} 0 & \gamma & \gamma & \cdots & \gamma \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & & & & \\ 0 & 0 & 0 & \cdots & 1 \end{pmatrix}. \tag{45}$$

On the other hand, the first row of $\widetilde{\mathbf{T}}$, using (40), is $(0, \gamma, \ldots, \gamma)$. Hence the statement is true for $k = 1$.

$k - 1 \rightsquigarrow k$: Assume the statement is true for $k \in \{2, \ldots, m\}$, i.e.,

$$\mathbf{T}_{k-1} \cdots \mathbf{T}_1 = \begin{pmatrix} \multicolumn{2}{c}{\text{first } k-1 \text{ rows of } \widetilde{\mathbf{T}}} \\ 0_{(m-(k-1))\times(k-1)} & I_{m-(k-1)} \end{pmatrix} = \begin{pmatrix} \multicolumn{2}{c}{\text{first } k-1 \text{ rows of } \widetilde{\mathbf{T}}} \\ 0_{(m-k+1))\times(k-1)} & I_{m-k+1}; \end{pmatrix} \tag{46}$$

or entrywise, this matrix $\mathbf{S}$ satisfies

$$(\forall i)(\forall j) \quad \mathbf{S}_{i,j} := \begin{cases} \widetilde{\mathbf{T}}_{i,j}, & \text{if } 1 \le i \le k-1; \\ 1, & \text{if } k \le i \le m \text{ and } j = i; \\ 0, & \text{if } k \le i \le m \text{ and } j \ne i. \end{cases} \tag{47}$$

Recall that

$$(\forall i)(\forall j) \quad \left(\mathbf{T}_k\right)_{i,j} = \begin{cases} \gamma, & \text{if } i = k \text{ and } j \ne k; \\ 1, & \text{if } i \ne k \text{ and } j = i; \\ 0, & \text{otherwise.} \end{cases} \tag{48}$$

It is clear that $\mathbf{T}_k\mathbf{S} = \mathbf{T}_k\mathbf{T}_{k-1} \cdots \mathbf{T}_1$. We must show that

$$\mathbf{T}_k\mathbf{S} \stackrel{?}{=} \begin{pmatrix} \multicolumn{2}{c}{\text{first } k \text{ rows of } \widetilde{\mathbf{T}}} \\ 0_{(m-k)\times k} & I_{m-k} \end{pmatrix}. \tag{49}$$

We do this entrywise, and thus fix i and j in $\{1, \ldots, m\}$.

Case 1: $1 \le i \le k-1$.
Then $(\mathbf{T}_k\mathbf{S})_{i,j} = \sum_{l=1}^{m} (\mathbf{T}_k)_{i,l}\mathbf{S}_{l,j} = (\mathbf{T}_k)_{i,i}\mathbf{S}_{i,j} = \mathbf{S}_{i,j}$, which shows that the first $k-1$ rows of $\mathbf{T}_k\mathbf{S}$ are the same as the first $k-1$ rows of $\mathbf{S}$, which in turn are the same as the first $k-1$ rows of $\widetilde{\mathbf{T}}$, as required.

Case 2: $i = k$.
Then $(\mathbf{T}_k\mathbf{S})_{i,j} = (\mathbf{T}_k\mathbf{S})_{k,j} = \sum_{l=1}^{m} (\mathbf{T}_k)_{k,l}\mathbf{S}_{l,j} = \sum_{l=1}^{k-1} (\mathbf{T}_k)_{k,l}\mathbf{S}_{l,j} + \sum_{l=k+1}^{m} (\mathbf{T}_k)_{k,l}\mathbf{S}_{l,j} = \sum_{l=1}^{k-1} \gamma\widetilde{\mathbf{T}}_{l,j} + \sum_{l=k+1}^{m} \gamma\mathbf{S}_{l,j}$.

Subcase 2.1: $j \leq k-1$. Then $(\mathbf{T}_k\mathbf{S})_{i,j} = (\mathbf{T}_k\mathbf{S})_{k,j} = \sum_{l=1}^{k-1} \gamma\widetilde{\mathbf{T}}_{l,j} = \gamma\sum_{l=1}^{j-1}\widetilde{\mathbf{T}}_{l,j} + \gamma\sum_{l=j}^{k-1}\widetilde{\mathbf{T}}_{l,j} = \gamma\sum_{l=1}^{j-1}\gamma(1+\gamma)^{l-1} + \gamma\sum_{l=j}^{k-1}\left(\gamma(1+\gamma)^{l-1} - \gamma(1+\gamma)^{l-j}\right) = \gamma^2\sum_{l=1}^{k-1}(1+\gamma)^{l-1} - \gamma\sum_{l=j}^{k-1}\gamma(1+\gamma)^{l-j} = \gamma(1+\gamma)^{k-1} - \gamma(1+\gamma)^{k-j} = \widetilde{\mathbf{T}}_{k,j} = \widetilde{\mathbf{T}}_{i,j}$.

Subcase 2.2: $j = k$. Then $(\mathbf{T}_k\mathbf{S})_{i,j} = (\mathbf{T}_k\mathbf{S})_{k,k} = \sum_{l=1}^{k-1}\gamma\widetilde{\mathbf{T}}_{l,k} = \gamma\sum_{l=1}^{k-1}\gamma(1+\gamma)^{l-1} = \gamma(1+\gamma)^{k-1} - \gamma = \widetilde{\mathbf{T}}_{k,k} = \widetilde{\mathbf{T}}_{i,j}$.

Subcase 2.3: $k+1 \leq j$. Then $(\mathbf{T}_k\mathbf{S})_{i,j} = (\mathbf{T}_k\mathbf{S})_{k,j} = \sum_{l=1}^{k-1}\gamma\widetilde{\mathbf{T}}_{l,j} + \gamma = \gamma\sum_{l=1}^{k-1}\gamma(1+\gamma)^{l-1} + \gamma = \gamma(1+\gamma)^{k-1} = \widetilde{\mathbf{T}}_{k,j} = \widetilde{\mathbf{T}}_{i,j}$.

Case 3: $i \geq k+1$.
Then $(\mathbf{T}_k\mathbf{S})_{i,j} = \sum_{l=1}^{m}(\mathbf{T}_k)_{i,l}\mathbf{S}_{l,j} = (\mathbf{T}_k)_{i,i}\mathbf{S}_{i,j} = \mathbf{S}_{i,j}$.

Subcase 3.1: $j = i$. Then $(\mathbf{T}_k\mathbf{S})_{i,j} = \mathbf{S}_{i,i} = 1$.

Subcase 3.2: $j \neq i$. Then $(\mathbf{T}_k\mathbf{S})_{i,j} = \mathbf{S}_{i,j} = 0$.

Altogether, we have shown that

$$\mathbf{T}_k\mathbf{S} = \begin{pmatrix} \text{first } k \text{ rows of } \widetilde{\mathbf{T}} \\ 0_{(m-k)\times k} \quad I_{m-k} \end{pmatrix}. \tag{50}$$

The "In particular" part is the case when $k = m$. ■

Corollary 4.2 *Let $k \in \{1, \ldots, m\}$. Then*

$$\mathbf{A}^k = \begin{pmatrix} 0_{(m-k)\times k} \quad I_{m-k} \\ \textit{first } k \textit{ rows of } \mathbf{T} \end{pmatrix}; \tag{51}$$

or entrywise

$$\left(\mathbf{A}^k\right)_{i,j} = \begin{cases} 1, & \text{if } 1 \leq i \leq m-k \text{ and } j = i+k; \\ 0, & \text{if } 1 \leq i \leq m-k \text{ and } j \neq i+k; \\ \mathbf{T}_{i+k-m,j}, & \text{if } m-k+1 \leq i \leq m. \end{cases} \tag{52}$$

Proof. Proposition 3.2 yields $\mathbf{A}^k = L^k(\mathbf{T}_k\mathbf{T}_{k-1}\cdots\mathbf{T}_1)$ and the result now follows from Proposition 4.1. ■

Remark 4.3 We do not know whether it is possible to obtain a closed form for the powers of $\mathbf{T}$ that does not rely on the eigenvalue analysis from Section 2. If such a formula exists, one may be able to construct a different proof of Theorem 3.3 in this setting.

5 Kolmogorov means

We now turn to moving Kolmogorov means, which are sometimes also called f-means.

Theorem 5.1 (moving Kolmogorov means) *Let D be a nonempty topological Hausdorff space, let Z be a real Banach space, and let $f\colon D \to Z$ be an injective mapping such that* $\operatorname{ran} f$ *is convex. Consider the linear recurrence relation*

(53a) $$(\forall n \geq m) \quad y_n = f^{-1}\big(\alpha_{m-1} f(y_{n-1}) + \cdots + \alpha_0 f(y_{n-m})\big),$$

where

(53b) $$(y_0, \ldots, y_{m-1}) \in D^m.$$

Then the following hold:

(i) *The sequence $(y_n)_{n\in\mathbb{N}}$ is well defined and the sequence $(f(y_n))_{n\in\mathbb{N}}$ lies in the compact convex set* $\operatorname{conv}\{f(y_0), \ldots, f(y_{m-1})\}$.

(ii) *Suppose that the* **Basic Hypothesis 2.5** *holds and that f^{-1} is continuous from $f(D)$ to D. Then*

(54) $$\lim_{n\in\mathbb{N}} y_n = f^{-1}\bigg(\sum_{k=0}^{m-1} \lambda_k f(y_k)\bigg) = f^{-1}\bigg(\frac{\sum_{k=0}^{m-1} a_k f(y_k)}{\sum_{k=0}^{m-1} a_k}\bigg) = f^{-1}\bigg(\frac{\sum_{k=0}^{m-1}\sum_{i=0}^{k} \alpha_i f(y_k)}{\sum_{k=0}^{m-1}\sum_{i=0}^{k} \alpha_i}\bigg).$$

Proof. (i): This is similar to Remark 2.10 and proved inductively.

(ii): By Corollary 2.11, $(f(y_n))_{n\in\mathbb{N}}$ converges to $\ell(f(y_0), \ldots, f(y_{m-1}))$, which belongs to $\operatorname{conv}\{f(y_0), \ldots, f(y_{m-1})\} \subseteq \operatorname{conv} f(D) = f(D) = \operatorname{ran} f$ by assumption and (i). Since f^{-1} is continuous, the result follows from (22). ■

Theorem 5.1(ii) allows for a universe of examples by appropriately choosing f and $(\alpha_0, \ldots, \alpha_{m-1})$. Let us present some classical moving means on (subsets) of the real line with the equal weights $\alpha_0 = \cdots = \alpha_{m-1} = 1/m$. Note that the **Basic Hypothesis 2.5** is satisfied.

Corollary 5.2 (some classical means) *Let D be a nonempty subset of $\mathbb{R}$, let $(y_0, \ldots, y_{m-1}) \in D^m$. Here is a list of choices for f in Theorem 5.1, along with the limits obtained by* (54)*:*

(i) *If $D = \mathbb{R}$ and $f = \operatorname{Id}$, then the moving* arithmetic mean *sequence satisfies*

(55) $$y_n = \frac{1}{m} y_{n-1} + \cdots + \frac{1}{m} y_{n-m} \to \frac{2}{m(m+1)} \sum_{j=0}^{m-1} (j+1) y_j.$$

(ii) *If* $D = \mathbb{R}_{++}$ *and* $f = \ln$*, then the moving* geometric mean *sequence satisfies*

$$(56) \qquad y_{n+m} = \sqrt[m]{y_{n+m-1}\cdots y_n} \to \left(\prod_{j=0}^{m-1} y_j^{j+1}\right)^{\frac{2}{m(m+1)}}.$$

(iii) *If* $D = \mathbb{R}_{+}$ *and* $f\colon x \mapsto x^p$*, then the moving* Hölder mean *sequence satisfies*

$$(57) \qquad y_{n+m} = \left(\frac{1}{m}y_{n+m-1}^p + \cdots + \frac{1}{m}y_n^p\right)^{1/p} \to \left(\frac{2}{m(m+1)}\sum_{j=0}^{m-1}(j+1)y_j^p\right)^{1/p}.$$

(iv) *If* $D = \mathbb{R}_{++}$ *and* $f\colon x \mapsto 1/x$*, then the moving* harmonic mean *sequence satisfies*

$$(58) \qquad y_{n+m} = \left(\frac{1}{m}\frac{1}{y_{n+m-1}} + \cdots + \frac{1}{m}\frac{1}{y_n}\right)^{-1} \to \left(\frac{2}{m(m+1)}\sum_{j=0}^{m-1}(j+1)\frac{1}{y_j}\right)^{-1}.$$

Let us provide some means situated in a space of matrices.

Corollary 5.3 (some matrix means) *Let* $(Y_0, \ldots, Y_{m-1}) \in \mathbb{S}_{++}^{N\times N}$*. Then the following hold:*

(i) *The moving arithmetic mean satisfies*

$$(59) \qquad Y_n = \frac{1}{m}Y_{n-1} + \cdots + \frac{1}{m}Y_{n-m} \to \frac{2}{m(m+1)}\sum_{j=0}^{m-1}(j+1)Y_j.$$

(ii) *The moving harmonic mean satisfies*

$$(60) \qquad Y_n = \left(\frac{1}{m}Y_{n-1}^{-1} + \cdots + \frac{1}{m}Y_{n-m}^{-1}\right)^{-1} \to \left(\frac{2}{m(m+1)}\sum_{j=0}^{m-1}(j+1)Y_j^{-1}\right)^{-1}.$$

(iii) *The moving resolvent mean (see also* [7] *) satisfies*

$$(61) \quad Y_n = \left(\frac{1}{m}(Y_{n-1}+\mathrm{Id})^{-1} + \cdots + \frac{1}{m}(Y_{n-m}+\mathrm{Id})^{-1}\right)^{-1} - \mathrm{Id}$$
$$\to \left(\frac{2}{m(m+1)}\sum_{j=0}^{m-1}(j+1)(Y_j+\mathrm{Id})^{-1}\right)^{-1} - \mathrm{Id}\,.$$

Proof. This follows from Theorem 5.1 with $D = \mathbb{S}_{++}^{N\times N}$ and $f = \mathrm{Id}$, $f = Z \mapsto Z^{-1}$, and $f = Z \mapsto (Z+\mathrm{Id})^{-1}$, respectively. Note that matrix inversion is continuous; see, e.g., [19, Example 6.2.7]. ∎

6 Moving proximal and epi averages for functions

In this last section, we apply the moving average (in particular Corollary 2.8) to functions in the context of proximal and epi averages. We set

$$Z = \mathbb{R}^l, \text{ where } l \in \{1,2,\ldots\}. \tag{62}$$

Let us start by reviewing a key notion (see also [2, 11, 25]).

Definition 6.1 (epi-convergence) (See [25, Proposition 7.2].) *Let g and $(g_n)_{n\in\mathbb{N}}$ be functions from Z to $\left]-\infty,+\infty\right]$. Then*

(i) $(g_n)_{n\in\mathbb{N}}$ pointwise converges *to g, in symbols $g_n \overset{\mathrm{p}}{\to} g$, if for every $z \in Z$ we have $g(z) = \lim_{n\in\mathbb{N}} g_n(z)$;*

(ii) $(g_n)_{n\in\mathbb{N}}$ epi-converges *to g, in symbols $g_n \overset{\mathrm{e}}{\to} g$, if for every $z \in Z$ one has*

  (a) $\big(\forall\,(z_n)_{n\in\mathbb{N}}\big)\ z_n \to z \Rightarrow g(z) \leq \varliminf g_n(z_n)$, *and*

  (b) $\big(\exists\,(y_n)_{n\in\mathbb{N}}\big)\ y_n \to z$ *and* $\varlimsup g_n(y_n) \leq g(z)$.

Let $g : Z \to \left]-\infty,+\infty\right]$ be an extended real valued function. Recall that

$$\big(g\,\Box\,\mathfrak{q}\big)\colon z \mapsto \inf_{y\in Z}\Big(g(y) + \tfrac{1}{2}\|z-y\|^2\Big) \tag{63}$$

is the *Moreau envelope* [20] of g. By [25, Theorem 2.26] (see also [20]), if $g \in \Gamma(Z)$, then $g\,\Box\,\mathfrak{q}$ is convex and continuously differentiable on Z. Especially important are the following connections among epi-convergence pointwise convergence of envelopes, and the epi-continuity of Fenchel conjugation.

Fact 6.2 (See [25, Theorems 7.37 and 11.34].) *Let $(g_n)_{n\in\mathbb{N}}$ and g be in $\Gamma(Z)$. Then*

$$g_n \overset{\mathrm{e}}{\to} g \quad\Leftrightarrow\quad g_n\,\Box\,\mathfrak{q} \overset{\mathrm{p}}{\to} g\,\Box\,\mathfrak{q} \quad\Leftrightarrow\quad g_n^* \overset{\mathrm{e}}{\to} g^*. \tag{64}$$

We now turn to the moving proximal average (see [6] for further information on this operation).

Theorem 6.3 (moving proximal average) *Let $(g_0,\ldots,g_{m-1}) \in (\Gamma(Z))^m$ and define*

$$(\forall n \geq m)\quad g_n = \big(\alpha_{m-1}(g_{n-1}+\mathfrak{q})^* + \cdots + \alpha_0(g_{n-m}+\mathfrak{q})^*\big)^* - \mathfrak{q}. \tag{65}$$

Then the following hold:

(i) *The sequence* $(g_n)_{n\in\mathbb{N}}$ *lies in* $\Gamma(Z)$.

(ii) $(\forall n \geq m)$ $g_n \Box \mathfrak{q} = \alpha_{m-1}(g_{n-1}\Box \mathfrak{q}) + \cdots + \alpha_0(g_{n-m}\Box \mathfrak{q})$.

(iii) *Suppose that the* **Basic Hypothesis 2.5** *holds. Then*

$$g_n \overset{e}{\to} g = \big(\lambda_{m-1}(g_{m-1} + \mathfrak{q})^* + \cdots + \lambda_0(g_0 + \mathfrak{q})^*\big)^* - \mathfrak{q}. \tag{66a}$$

and

$$g_n^* \overset{e}{\to} g^* = \big(\lambda_{m-1}(g_{m-1}^* + \mathfrak{q})^* + \cdots + \lambda_0(g_0^* + \mathfrak{q})^*\big)^* - \mathfrak{q}. \tag{66b}$$

Proof. (i): Combine [6, Propositions 4.3 and 5.2].

(ii): See [6, Theorem 6.2].

(iii): It follows from (ii) and Corollary 2.8 that $g_n\Box \mathfrak{q} \overset{p}{\to} \sum_{k=0}^{m-1}\lambda_k(g_k\Box \mathfrak{q}) = g\Box \mathfrak{q}$. The result now follows from Fact 6.2 and [6, Theorem 5.1]. ■

We conclude the paper with the moving epi-average. Recall that for $\alpha > 0$ and $g \in \Gamma(Z)$, $\alpha \star g = \alpha g \circ \alpha^{-1}\mathrm{Id}$ and that $0 \star g = \iota_{\{0\}}$, where $\iota_{\{0\}}(0) = 0$ and $\iota_{\{0\}}(z) = +\infty$ if $z \in Z \smallsetminus \{0\}$.

Theorem 6.4 (moving epi-average) *Let* $(g_0,\ldots,g_{m-1}) \in (\Gamma(Z))^m$ *and define*

$$(\forall n \geq m) \quad g_n = (\alpha_{m-1}\star g_{n-1})\Box\cdots\Box(\alpha_0\star g_{n-m}). \tag{67}$$

Suppose that for every $k \in \{0,\ldots,m-1\}$, g_k *is cofinite and that the* **Basic Hypothesis 2.5** *holds. Then*

$$g_n \overset{e}{\to} g = (\lambda_{m-1}\star g_{m-1})\Box\cdots\Box(\lambda_0\star g_0). \tag{68}$$

Proof. It follows from [3, Proposition 15.7(iv)] that $(\forall n \in \mathbb{N})$ $g_n \in \Gamma(Z)$, g_n is cofinite, and g_n^* is continuous everywhere. Furthermore, taking the Fenchel conjugate of (67) yields

$$(\forall n \geq m) \quad g_n^* = \alpha_{m-1}g_{n-1}^* + \cdots + \alpha_0 g_{n-m}^*. \tag{69}$$

This and Corollary 2.8 imply not only that $g_n^* \overset{p}{\to} g^*$ but also that the sequence $(g_n^*)_{n\in\mathbb{N}}$ is equi-lsc everywhere (see [25, pages 248f]). In turn, [25, Theorem 7.10] implies that $g_n^* \overset{e}{\to} g^*$. The conclusion therefore follows from Fact 6.2. ■

Acknowledgments

The authors are grateful to an anonymous referee, Jon Borwein and Tamás Erdélyi for some helpful comments. Heinz Bauschke was partially supported by the Natural Sciences and Engineering Research Council of Canada and by the Canada Research Chair Program. Joshua Sarada was partially supported by the Irving K. Barber Endowment Fund. Xianfu Wang was partially supported by the Natural Sciences and Engineering Research Council of Canada.

References

[1] C.D. Aliprantis and K.C. Border, *Infinite Dimensional Analysis*, third edition, Springer, 2006.

[2] H. Attouch, *Variational Convergence for Functions and Operators*, Applicable Mathematics Series, Pitman Advanced Publishing Program, Boston, MA, 1984.

[3] H.H. Bauschke and P.L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, Springer, 2011.

[4] H.H. Bauschke, X. Wang, and C.J.S. Wylie, Fixed points of averages of resolvents: geometry and algorithms, `http://arxiv.org/pdf/1102.1478v1`, February 2011.

[5] H.H. Bauschke, X. Wang, and C.J.S. Wylie, Fixed points of averages of resolvents: geometry and algorithms, *SIAM Journal on Optimization* 22 (2012), 24–40.

[6] H.H. Bauschke, R. Goebel, Y. Lucet, and X. Wang, The proximal average: basic theory, *SIAM Journal on Optimization* 19 (2008), no. 2, 766–785.

[7] H.H. Bauschke, S.M. Moffat and X. Wang, The resolvent average for positive semidefinite matrices, *Linear Algebra Appl.* 432 (2010), no. 7, 1757-1771.

[8] A. Berman and R.J. Plemmons, *Nonnegative Matrices in the Mathematical Sciences*, SIAM, 1994.

[9] D. Borwein, J.M. Borwein, and B. Sims, On the solution of linear mean recurrences, *American Mathematical Monthly*, to appear.

[10] J.M. Borwein and P.B. Borwein, *Pi and the AGM*, Wiley, New York, 1987.

[11] J.M. Borwein and J.D. Vanderwerff, *Convex Functions*, Cambridge University Press, 2010.

[12] P. Borwein and T. Erdélyi, *Polynomials and Polynomial Inequalites*, Springer-Verlag, 1995.

[13] G. Box, G. Jenkins, and G. Reinsel, *Time Series Analysis: Forecasting and Control*, third edition, Prentice Hall, Englewood Cliffs, NJ, 1994.

[14] W.S. Chou, B.S. Du, and P.J.-S. Shiue, A note on circulant transition matrices in Markov chains, *Linear Algebra and its Applications* 429 (2008), 1699–1704.

[15] P.L. Combettes and T. Pennanen, Generalized Mann iterates for constructing fixed points in Hilbert spaces, *Journal of Mathematical Analysis and Applications* 275 (2002), no. 2, 521–536.

[16] P.J. Davis, *Circulant Matrices*, Wiley, 1979.

[17] R.A. Horn and C.R. Johnson, *Matrix Analysis*, Cambridge University Press, Cambridge, 1985.

[18] O. Krafft and M. Schaefer, Convergence of the powers of a circulant stochastic matrix, *Linear Algebra and its Applications* 127 (1990), 59–69.

[19] C.D. Meyer, *Matrix Analysis and Applied Linear Algebra*, SIAM, 2000.

[20] J.-J. Moreau, Proximité et dualité dans un espace hilbertien, *Bulletin de la Société Mathématique de France* 93 (1965), 273–299.

[21] J.M. Ortega, *Numerical Analysis:* A Second Course, second edition, SIAM, 1990.

[22] A.M. Ostrowski, *Solution of Equations and Systems of Equations*, second edition, Academic Press, New York and London, 1966.

[23] V.V. Prasolov, *Polynomials*, Springer-Verlag, Berlin, 2004.

[24] R.T. Rockafellar, *Convex Analysis*, Princeton University Press, Princeton, 1970.

[25] R.T. Rockafellar and R.J-B Wets, *Variational Analysis*, Springer, corrected third printing, 2009.

[26] C.P. Tsokos, K-th moving, weighted and exponential moving average for time series forecasting models, *European Journal of Pure and Applied Mathematics* 3 (2010), 406–416.