Wolfowitz's theorem and consensus algorithms in Hadamard spaces

P. Grohs

Wolfowitz's Theorem and Consensus Algorithms in Hadamard Spaces *

P. Grohs

August 29, 2012

Abstract

We provide a generalization of Wolfowitz's theorem on the products of stochastic, indecomposable and aperiodic (SIA) matrices to metric spaces with nonpositive curvature. As a result we show convergence for a wide class of distributed consensus algorithms operating on these spaces.

1 Introduction and Statement of the Result

Wolfowitz's Theorem. In the course of the present paper we fix a positive integer N and write $[k] := \{1, \ldots, k\}$ for $k \in \mathbb{N}$.

Definition 1.1. *A matrix $A = (a(i,j))_{i,j \in [N]} \in \mathbb{R}^{N \times N}$ is SIA (stochastic, indecomposable, aperiodic) if*

(i) $a(i,j) \geq 0$ for all $i, j \in [N]$

(ii) $\sum_{j \in [N]} a(i,j) = 1$ for all $i \in [N]$

(iii) The limit
$$Q = \lim_{n \to \infty} A^n$$
exists in $\mathbb{R}^{N \times N}$ and all the rows of Q are the same.

Definition 1.2. *A family $\mathcal{A} = \{A_1, \ldots, A_k\} \subset \mathbb{R}^{N \times N}$ is called SIA-family if every word*
$$A_\kappa := A_{\kappa_n} \cdot \ldots \cdot A_{\kappa_2} A_{\kappa_1}$$
is SIA, where $\kappa \in [k]^n$ and $n \in \mathbb{N}_0$.

Consider the quantity

$$\delta(A) := \max_{j \in [N]} \max_{k,l \in [N]} |A(k,j) - A(l,j)|, \tag{1}$$

defined for $A \in \mathbb{R}^{N \times N}$. We are now in a position to formulate the following well-known theorem by Wolfowitz [23].

Theorem 1.3 (Wolfowitz). *Let $\mathcal{A}$ be an SIA-family. Then for any $\varepsilon > 0$ there exists $\nu(\varepsilon)$ such that*

$$\delta(A_\kappa) < \varepsilon \quad \textit{for all} \quad \kappa \in \bigcup_{n \geq \nu(\varepsilon)} [k]^n.$$

Theorem 1.3 result comes with a number of important implications. For instance, picking $\kappa \in [k]^\mathbb{N}$ arbitrarily defines a nonhomogenous Markov chain with transition kernels

$$P_{n-1 \to n} := A_{\kappa_n}.$$

Wolfowitz's theorem implies that this Markov chain converges to a stationary state, whenever the underlying matrices belong to an SIA family.

Furthermore, Theorem 1.3 is of central importance in control theory and distributed processing, more precisely the study of convergence properties of *consensus algorithms* [21]. Consider N different agents $i = 1, \ldots, N$ each observing a common quantity $x(i) \in \mathbb{R}$. Due to several causes (noise, defective sensors, calibration errors, etc.) the measurements $x(i)$ will be different for different agents i. Several problems within the context of flocking and multiagent dynamics require the different agents to agree on a common value of the measurement. A natural choice would be to agree on the value $\frac{1}{M} \sum_{i \in [N]} x(i)$. However, often there are restrictions imposed on the ability for different agents to communicate with each other, and therefore it is necessary to reach a consensus value in a distributed fashion.

Let us see how Wolfowitz's theorem comes into play here. Pick an arbitrary SIA-family $\mathcal{A}$ and an arbitrary $\kappa \in [k]^\mathbb{N}$ such that each matrix A_{κ_k} is adapted to the underlying communication structure (meaning that averages are only computed among agents which can communicate with each other). Then one can iteratively compute averages

$$x_n := A_{\kappa_n} x_{n-1}, \quad x_0 := x. \tag{2}$$

Wolfowitz's theorem implies that this distributed averaging algorithm converges to a consensus value, i.e.

$$\lim_{n \to \infty} \max_{k,l \in [N]} |x_n(k) - x_n(l)| = 0. \tag{3}$$

It also gives quantitative information regarding the speed of convergence.

Observe that in the convergence result, the matrices A_{κ_n} can be chosen adaptively, based on the knowledge of $\prod_{i=1}^{n-1} A_{\kappa_i} x$.

A further application is in nonstationary subdivision schemes [7] whose convergence study leads to analogous problems as for consensus algorithms.

In view of these applications, the aim of the present paper is to generalize Theorem 1.3 to metric spaces.

Consensus in Metric Spaces. Assume that we are given a metric space (X, d). The goal of the present paper is to generalize Theorem 1.3 to the X-valued case. In order to formulate a meaningful result we observe that, due to assumption (i) in Definition 1.1, the consensus algorithm defined via the iteration (2) can be interpreted as a repeated averaging of the initial datum x.

Given weights $a : [N] \to \mathbb{R}_+$ such that $\sum_{i \in [N]} a(i) = 1$, a natural choice of a weighted average of points $x(i)$, $i \in [N]$, is the *barycenter* $\bar{x}^a$, which is defined by

$$\bar{x}^a := \operatorname*{argmin}_{z \in X} \sum_{i \in [N]} a(i)\mathrm{d}(x(i), z)^2.$$

In the literature, this definition is usually referred to as *Fréchet-mean* or *Karcher-mean* if X is a finite dimensional Riemannian manifold [14]. Provided that the barycenter operation is well-defined we can now formulate an iterative averaging algorithm generalizing the one defined above via (2):

Definition 1.4. *For an SIA matrix A and $x : [N] \to X$ we define the operator A^X by*

$$A^X x(i) := \operatorname*{argmin}_{z \in X} \sum_{j \in [N]} A(i, j)\mathrm{d}(x(j), z)^2. \tag{4}$$

For an SIA-family $\mathcal{A}$ and $\kappa \in [k]^{\mathbb{N}}$ define iteratively

$$x_n := A^X_{\kappa_n} x_{n-1}, \quad x_0 := x. \tag{5}$$

The natural question to ask is whether the iteration as defined in (5) converges to consensus, e.g., does

$$\lim_{n \to \infty} \max_{k,l \in [N]} \mathrm{d}\left(x_n(k), x_n(l)\right) = 0$$

hold?

Consensus algorithms in nonlinear spaces have been the subject of intense studies in recent years, see for instance [18] and the references therein. Closely related is the study of subdivision schemes operating in metric spaces [9, 10, 12, 22, 24].

Hadamard Spaces. We restrict our attention to a specific class of metric spaces in which barycenters are globally defined. These are so-called *Hadamard spaces*.

Definition 1.5. *A metric space (X, d) is called* Hadamard space *if for $x_1, x_1 \in X$ there exists $y \in X$ such that for all $z \in X$ one has*

$$\mathrm{d}(z, y)^2 \le \frac{1}{2}\mathrm{d}(z, x_0)^2 + \frac{1}{2}\mathrm{d}(z, x_1)^2 - \frac{1}{4}\mathrm{d}(x_0, x_1)^2 \tag{6}$$

The *Hadamard inequality* (6) can be interpreted as nonpositivity condition for the curvature, see [1, 2]. A basic fact about Hadamard spaces is that barycenters are globally well-defined [20]. Hadamard spaces include Cartan-Hadamard manifolds, Trees, Euclidean Bruhat-Tits buildings, spaces $L^2(M, N)$ with N Hadamard, certain spaces of Riemannian and Kähler metrics or spaces of connections [15]. Hadamard spaces also arise in several important applications, such as the study of phylogenetic trees [5], diffusion tensor MRI [17] or cost-minimizing networks [8].

Definition 1.6. *For* $x : [N] \to X$ *define the* diameter

$$D(x) := \max_{k,l \in [N]} \mathrm{d}(x(k), x(l)).$$

Our main theorem reads as follows.

Theorem 1.7. *Let* $\mathcal{A}$ *be an SIA-family and* (X, d) *Hadamard. Then there exists* $\nu \in \mathbb{N}$ *and* $0 < \gamma < 1$ *such that for all* $\kappa \in [k]^n$, $n \geq \nu$, *and* $x : [N] \to X$ *we have the contractivity property*

$$D\left(A^X_{\kappa_1} A^X_{\kappa_2} \dots A^X_{\kappa_n} x\right) \leq \gamma D(x). \tag{7}$$

It should be clear that Theorem 1.7 is a natural generalization of Theorem 1.3. In particular we have the following corollary, where we define for $\kappa \in [k]^n$ the operator

$$A^X_\kappa x := A^X_{\kappa_n} A^X_{\kappa_{n-1}} \dots A^X_{\kappa_1} x$$

acting on $x : [N] \to X$.

Corollary 1.8. *Let* $\mathcal{A}$ *be an SIA-family. Then for any* $\varepsilon > 0$ *there exists* $\nu(\varepsilon)$ *such that for all* $x : [N] \to X$

$$D(A^X_\kappa x) < \varepsilon D(x) \quad \text{for all} \quad \kappa \in \bigcup_{n \geq \nu(\varepsilon)} [k]^n.$$

Proof. This follows immediately from Theorem 1.7. $\qquad\square$

In view of nonlinear consensus in Hadamard spaces we have the following result.

Corollary 1.9. *The algorithm defined by (5) for a SIA-family converges geometrically to a consensus configuration: There exists* $\gamma < 1$, $C \geq 0$ *such that*

$$D(x_n) \leq C \gamma^n D(x_0).$$

Proof. This result is an immediate consequence of Theorem 1.7. $\qquad\square$

Section 2 below is devoted to the proof of Theorem 1.7. Finally, in Section 3 we treat several extensions of our main result and apply them to show convergence for a wide class of consensus algorithms in Hadamard spaces.

2 Proof of the Main Result

Our main result will be proven by utilizing results related to Markov chains in Hadamard spaces. This connection has been exploited in [9, 10] for related problems. Our work is inspired by the aforementioned articles.

Preliminaries on Markov Chains. We start with some basic facts related to Markov chains.

Definition 2.1. *A family* $p_{m \to n} : [N] \times [N] \to \mathbb{R}$, $m \leq n$ *is called a family of* transition kernels *if*

(i) $\sum_{j\in[N]} p_{m\to n}(i,j) = 1$ *for all* $i \in [N]$ *and* $p_{m\to n}(i,j) \geq 0$ *for all* $i,j \in [N]$

(ii) $p_{n\to n}(i,j) = \delta_{i,j}$

(iii) *Putting* $P_{m\to n} := (p_{m\to n}(i,j))_{i,j\in[N]}$ *we have the* Chapman-Kolmogorov *equations*

$$P_{m\to n} = P_{m\to l}P_{l\to n} \quad \text{for all} \quad m \leq l \leq n. \tag{8}$$

Remark 2.2. *Clearly, for* $\mathcal{A}$ *an SIA-family and* $\kappa \in [k]^{\mathbb{N}}$ *we can put*

$$P_{n-1\to n} := A_{\kappa_n}$$

and use (8) to define a family of transition kernels.

A family of transition kernels leads to an associated Markov chain as follows: Define the state space $\Omega := [N]^{\mathbb{N}_0}$ endowed with the sigma algebra $\mathfrak{F} := \bigotimes_{n\in\mathbb{N}_0} \mathfrak{P}([N])$, where $\mathfrak{P}$ denotes the power set.

Picking an initial probability distribution α on $[N]$ we define a probability distribution $\mathbb{P}$ on $(\Omega, \mathfrak{F})$ by its restriction to cylinder sets defined as

$$\mathbb{P}\left(\{i_0\} \times \{i_1\} \times \cdots \times \{i_n\} \times [N] \times \ldots\right) := \alpha(\{i_0\})p_{0\to 1}(i_0, i_1)p_{1\to 2}(i_1, i_2)\ldots p_{n-1\to n}(i_{n-1}, i_n)^1.$$

Consider the filtration

$$\mathfrak{F}_n := \sigma\left(\{i_0\} \times \{i_1\} \times \cdots \times \{i_n\} \times [N] \times \cdots : (i_0, \ldots, i_n) \in [N]^{n+1}\right), \ n \geq 0,$$

as well as the discrete stochastic process $(X_n)_{n\geq 0}$ defined by

$$X_n : \Omega \to [N]; \quad \omega \mapsto \omega_n,$$

which is adapted to the filtration $(\mathfrak{F}_n)_{n\geq 0}$. The process $(X_n)_{n\geq 0}$ is called *Markov chain* associated to the transition kernel $(P_{m\to n})_{m\leq n}$.

It is straightforward to verify the following *linear Markov property*

$$E\left(f(X_n)|\mathfrak{F}_m\right)(\omega) = \sum_{j\in[N]} p_{m\to n}(X_m(\omega), j)f(j), \quad \omega \in \Omega, \tag{9}$$

valid for any nonnegative function $f : [N] \to \mathbb{R}$.

Markov Chains in Hadamard Spaces. Here we collect some results concerning Markov processes in Hadamard spaces. It is remarkable that the convexity properties of the distance in a Hadamard space allow for far-reaching extensions of central concepts in probability theory to the Hadamard setting [19, 20]. We only collect the results needed for our purpose, see [4, 19, 20] for further information.

We start with the concept of conditional expectation in Hadamard spaces. Let $(\Omega, \mathfrak{F}, \mathbb{P})$ be a probability space, and $Y : \Omega \to X$ measurable and square-integrable, meaning that

$$\int_\Omega \mathrm{d}(Y(\omega), x)^2 \mathbb{P}(d\omega) < \infty$$

---

[1]The well-definedness of $\mathbb{P}$ follows by extension and the fact that cylinder sets span the sigma algebra.

for some (then all) $x \in X$. For $Y, Z : \Omega \to X$ we define the L_2-distance via

$$d_{L_2}(Y, Z) := \left(\int_\Omega d(Y(\omega), Z(\omega))^2 \mathbb{P}(d\omega) \right)^{1/2}.$$

Now we can define conditional expectations:

Definition 2.3. *Let $\mathfrak{G}$ be a subalgebra of $\mathfrak{F}$. Then for a square-integrable $Y : \Omega \to X$ with separable image, the* conditional expectation *of Y given $\mathfrak{G}$ is defined as*

$$E(Y|\mathfrak{G}) := \operatorname*{argmin}_{Z:\Omega\to X} E(d_{L_2}(Y, Z)).$$

Existence and uniqueness of the conditional expectation is shown in [20]. Nonlinear conditional expectations lack the usual associativity property enjoyed by the linear definition. Motivated by this, K.T. Sturm introduced the following concept of filtered conditional expectation in his article [20].

Definition 2.4. *Suppose $\mathfrak{F}_m \subset \mathfrak{F}_1 \subset \cdots \subset \mathfrak{F}_n$, $m \leq n$, is a sequence of subalgebras of $\mathfrak{F}$ and let $Y : \Omega \to X$ be $\mathfrak{F}_n$-measurable. Then the* filtered conditional expectation *of Y given $(\mathfrak{F}_k)_{m \leq k \leq n}$ is defined as*

$$E(Y|||\mathfrak{F}_m) := E(\ldots E(E(Y|\mathfrak{F}_{n-1})|\mathfrak{F}_{n-2})\ldots|\mathfrak{F}_0).$$

The filtered conditional expectation satisfies a conditional Jensen inequality as shown in [20]. The statement requires the notion of *convexity* of a function $\psi : X \to \mathbb{R}$, meaning that

$$\psi(x_t) \leq (1 - t)\psi(x_0) + t\psi(x_1),$$

where $(x_t)_{t \in [0,1]}$ denotes the unique geodesic connecting two points $x_0, \ x_1 \in X$. An important example of a convex function is the function

$$x \mapsto d(x, z_0)$$

for any $z_0 \in X$, as shown in [20].

The conditional Jensen inequality now reads as follows.

Lemma 2.5. *Suppose that $\psi : X \to \mathbb{R}$ is convex and lower semicontinuous. Moreover, suppose that $(\mathfrak{F}_n)_{n \geq 0}$ is a filtration on the probability space $(\Omega, \mathfrak{F}, \mathbb{P})$. Then for each bounded, $\mathfrak{F}_n$-measurable $Y : \Omega \to X$ we have*

$$\psi\left(E\left(Y|||(\mathfrak{F}_k)_{m \leq k \leq n}\right)\right) \leq E\left(\psi(Y)|\mathfrak{F}_m\right). \tag{10}$$

The following nonlinear Markov property holds for (X, d) a Hadamard space. It is shown in [10, Proposition 6].

Proposition 2.6. *Let $(X_n)_{n \geq 0}$ be a Markov chain as above and let $x : [N] \to X$. Then for any $m \leq n$ we have*

$$E\left(x(X_n)|\mathfrak{F}_m\right)(\omega) = \operatorname*{argmin}_{z \in X} \sum_{j \in [N]} p_{m \to n}(X_m(\omega), j) d(x(j), z)^2, \tag{11}$$

where E is the nonlinear conditional expectation for the probability space $(\Omega, \mathfrak{F}, \mathbb{P})$ associated with the Markov chain.

Proof of the Main Result. We have now assembled all the tools needed to establish our main result. First we state the following lemma.

Lemma 2.7. *Let $\mathcal{A}$ be an SIA-family and $\kappa \in [k]^{\mathbb{N}_0}$. Consider the Markov chain $(X_n)_{n \geq 0}$ defined as in Remark 2.2. Then we have*

$$A^X_{\kappa_1} A^X_{\kappa_2} \ldots A^X_{\kappa_n} x \circ X_0 = E\left(x \circ X_n \| \mathfrak{F}_0\right). \tag{12}$$

Proof. The proof uses Proposition 2.6, as well as the definition of the filtered conditional expectation. To this end, we first compute

$$
\begin{aligned}
E(x \circ X_n | \mathfrak{F}_{n-1})(\omega) &= \operatorname*{argmin}_{z \in X} \sum_{j \in [N]} p_{n-1 \to n}(X_{n-1}(\omega), j) \mathrm{d}(x(j), z)^2 \\
&= \operatorname*{argmin}_{z \in X} \sum_{j \in [N]} a_{\kappa_n}(X_{n-1}(\omega), j) \mathrm{d}(x(j), z)^2 \\
&= A^X_{\kappa_n} x \circ X_{n-1}(\omega).
\end{aligned}
$$

Consequently, by the same argument

$$E(E(x \circ X_n | \mathfrak{F}_{n-1}) | \mathfrak{F}_{n-2})(\omega) = E(A^X_{\kappa_n} x \circ X_{n-1} | \mathfrak{F}_{n-2})(\omega) = A^X_{\kappa_{n-1}} A^X_{\kappa_n} x \circ X_{n-2}(\omega).$$

Iterating this argument we arrive at the desired statement. $\qquad \square$

We proceed with the

proof of Theorem 1.7. Consider the function

$$X \to \mathbb{R}_+; \quad z \mapsto \mathrm{d}(z, z_0).$$

Clearly, this function is continuous and convex. We can therefore apply Jensen's inequality (10), as well as (12) to obtain

$$\mathrm{d}\left(A^X_{\kappa_1} A^X_{\kappa_2} \ldots A^X_{\kappa_n} x \circ X_0, z_0\right) = \mathrm{d}\left(E\left(x \circ X_n \| \mathfrak{F}_0\right), z_0\right) \leq E\left(\mathrm{d}(x \circ X_n, z_0) | \mathfrak{F}_0\right). \tag{13}$$

Using the linear Markov property (9), we see that the right-hand-side of (13) is equal to

$$E\left(\mathrm{d}(x \circ X_n, z_0) | \mathfrak{F}_0\right)(\omega) = \sum_{k \in [N]} p_{0 \to n}(X_0(\omega), k) \, \mathrm{d}(x(k), z_0),$$

and we arrive at the inequality

$$\mathrm{d}\left(A^X_{\kappa_1} A^X_{\kappa_2} \ldots A^X_{\kappa_n} x(i), z_0\right) \leq \sum_{k \in [N]} p_{0 \to n}(i, k) \, \mathrm{d}(x(k), z_0) \tag{14}$$

for all $i \in [N]$, $z_0 \in X$. Now we can put

$$z_0 := A^X_{\kappa_1} A^X_{\kappa_2} \ldots A^X_{\kappa_n} x(j)$$

and, by the same argument leading to (14), obtain the symmetric inequality

$$\mathrm{d}\left(A^X_{\kappa_1} A^X_{\kappa_2} \ldots A^X_{\kappa_n} x(i), A^X_{\kappa_1} A^X_{\kappa_2} \ldots A^X_{\kappa_n} x(j)\right) \leq \sum_{k,l \in [N]} p_{0 \to n}(i,k)\, p_{0 \to n}(j,l)\, \mathrm{d}\left(x(k), x(l)\right).$$

(15)

Denote the right-hand side of (15) by B. Then we have

$$\begin{aligned}
B &= \sum_{k,l \in [N]} p_{0 \to n}(1,k)\, p_{0 \to n}(1,l)\, \mathrm{d}\left(x(k), x(l)\right) \\
&\quad + \sum_{k,l \in [N]} \left(p_{0 \to n}(i,k) - p_{0 \to n}(1,k)\right) p_{0 \to n}(j,l)\, \mathrm{d}\left(x(k), x(l)\right) \\
&\quad + \sum_{k,l \in [N]} p_{0 \to n}(1,k) \left(p_{0 \to n}(j,l) - p_{0 \to n}(1,l)\right) \mathrm{d}\left(x(k), x(l)\right) \\
&\leq \sum_{k,l \in [N]} p_{0 \to n}(1,k)\, p_{0 \to n}(1,l)\, \mathrm{d}\left(x(k), x(l)\right) \\
&\quad + 2\delta(P_{0 \to n}) D(x) \\
&\leq \left(\sum_{k,l \in [N],\ k \neq l} p_{0 \to n}(1,k)\, p_{0 \to n}(1,l) + 2\delta(P_{0 \to n}) \right) D(x) \\
&= \left(1 - \sum_{k \in [N]} p_{0 \to n}(1,k)^2 + 2\delta(P_{0 \to n}) \right) D(x).
\end{aligned}$$

It remains to show that for n large enough, the constant

$$\gamma_n := 1 - \sum_{k \in [N]} p_{0 \to n}(1,k)^2 + 2\delta(P_{0 \to n}) < 1.$$

As a first step towards this goal we observe that

$$\sum_{k \in [N]} p_{0 \to n}(1,k)^2 \geq \frac{1}{\sqrt{N}} \sum_{k \in [N]} |p_{0 \to n}(1,k)| = \frac{1}{\sqrt{N}}.$$

(16)

Furthermore, we note that by definition we have

$$P_{0 \to n} = A_{\kappa_1} A_{\kappa_2} \ldots A_{\kappa_n} = A_{\overleftarrow{\kappa}},$$

where $\overleftarrow{\kappa}$ simply means the word κ with its indices reversed. By Wolfowitz's theorem 1.3, there exists $\nu \geq 0$ such that whenever $n \geq \nu$ we have

$$\delta(A_{\overleftarrow{\kappa}}) \leq \frac{1}{4\sqrt{N}}.$$

(17)

Putting together (16) and (17) we arrive at

$$\gamma_n \leq 1 - \frac{1}{2\sqrt{N}} < 1, \quad \text{for all} \quad n \geq \nu,$$

and this proves the result. $\qquad \square$

8

3 Some Extensions

We close with some extensions of our main result. In particular we will establish convergence of a class of consensus algorithms studied in [13] which is not covered by Theorem 1.7.

3.1 Block SIA-Families

We consider the following, more general case: Assume that the family $\mathcal{A}$ is *not* an SIA-family but that we can block together elements in $\mathcal{A}$ to get an SIA-family.

Definition 3.1. *Let $\mathcal{A} = \{A_1, \ldots, A_k\} \subset \mathbb{R}^{N \times N}$ and assume that there exists a family of words $\mathcal{K} = (\kappa^i)_{i \geq 0}$ of lengths $(L_i)_{i \geq 0}$ such that $L := \sup_i L_i < \infty$ and the family $\mathcal{A}_{\mathcal{K}} := \left\{ A_{\kappa^i} := \prod_{j=1}^{L_i} A_{\kappa_j^i} : i \geq 0 \right\}$ is an SIA-family (observe that $\mathcal{K}$ is a finite set by the assumption that all words κ^i are of uniformly bounded length). Then $\mathcal{A}$ is called a* block SIA-family.

In the Euclidean case, it is a direct consequence of Theorem 1.3 that for each $\varepsilon > 0$ there exists $\nu(\varepsilon)$ such that

$$\delta \left(\prod_{i=1}^{n} A_{\kappa^i} \right) \leq \varepsilon \quad \text{for all} \quad n \geq \nu(\varepsilon). \tag{18}$$

For a block SIA-family consider the consensus algorithm

$$x_{n_j^i} := A_{\kappa_j^i} x_{n_j^i - 1}, \quad x_0 = x, \tag{19}$$

where

$$n_j^i := \sum_{l=1}^{i-1} L_i + j, \quad j = 1, \ldots, L_i.$$

This consensus algorithm does not come from an SIA-family, but it consists of block matrices which form an SIA-family. In that sense the algorithm (19) is more general than (2). Nevertheless, it is easy to show that also the algorithm (19) converges, see e.g., [13].

We can establish the following result:

Theorem 3.2. *Let $\mathcal{A}$ be a block SIA-family. Then for each $\varepsilon > 0$ there exists $\nu(\varepsilon)$ such that*

$$D \left(\prod_{i=1}^{n} \prod_{j=1}^{L_i} A_{\kappa_j^i}^X \circ x \right) \leq \varepsilon D(x) \quad \text{for all} \quad n \geq \nu(\varepsilon), \ x : [N] \to X. \tag{20}$$

Here, the product symbol $\prod_{i=1}^{n} \prod_{j=1}^{L_i} A_{\kappa_j^i}^X$ denotes the composition operation of operators.

Proof. The proof follows the exact same arguments as the proof of Theorem 1.7, only with (18) replacing the application of Theorem 1.3. $\square$

We can also consider the nonlinear consensus algorithm defined for $x : [N] \to X$ via

$$x_{n_j^i} := A_{\kappa_j^i}^X x_{n_j^i - 1}, \quad x_0 = x, \tag{21}$$

and n_j^i as above. We have the following result.

Theorem 3.3. *The consensus algorithm defined by (21) converges for a block SIA-family $\mathcal{A}$. In particular we have*

$$\lim_{n \to \infty} D(x_n) \to 0 \quad \text{for all } x : [N] \to X.$$

Proof. As in the proof of Theorem 1.7 we can deduce that

$$D(x_{n_{L_i}^i}) \leq \varepsilon D(x)$$

for $i \geq \nu(\varepsilon)$ sufficiently large. To complete our theorem, we now show that

$$D(A_l^X x) \leq C D(x), \quad \text{for all } x : [N] \to X, \ l \in [k]. \tag{22}$$

This follows from continuity: By assumption there exists $z \in X$ such that

$$\sup_i \mathrm{d}(x(i), z) = O(D(x)).$$

We abuse notation and call z the function $[N] \to X$ which is constant and equal to z. Clearly for all $l \in [k]$ we have

$$A_l^X z = z.$$

Applying A_l^X and applying the same arguments as in the proof of Theorem 1.7 we get

$$\begin{aligned}
\mathrm{d}(A_l^X x(i), z) &= \mathrm{d}(A_l^X x(i), A_l^X z) \\
&\leq \sum_{k,m \in [N]} a_l(i, k) a_l(1, m) \mathrm{d}(x(k), z) \\
&\leq C D(x).
\end{aligned}$$

This implies (22) and we can deduce that

$$D(x_{n_{L_i}^i + l}) \leq C^L \varepsilon D(x), \quad l \in [L_{i+1}]$$

and consequently for any $\varepsilon > 0$ there exists $\nu'(\varepsilon)$ such that

$$D(x_n) \leq \varepsilon D(x) \quad \text{for all } n \geq \nu'(\varepsilon).$$

This proves the theorem. □

3.2   Examples

Leaderless Coordination.   To illustrate the usefulness of considering block SIA-families we briefly treat an algorithm studied in [13] under the name *leaderless coordination*. Consider an undirected, simple graph Γ with vertex set $[N]$. An edge (i, j) in Γ corresponds to the ability of the agents i, j to communicate with each other. Denoting $\mathcal{N}_\Gamma(i)$ the set of neighbors of i we can formulate the following algorithm for a family of undirected simple graphs $(\Gamma_n)_{n \geq 0}$

$$ x_n(i) := \frac{1}{1 + |\mathcal{N}_{\Gamma_{n-1}}(i)|} \left(x_{n-1}(i) + \sum_{j \in \mathcal{N}_{\Gamma_{n-1}}(i)} x_{n-1}(j) \right), \quad x_0 = x, \qquad (23) $$

which can be concisely written as

$$ x_n = A_{\Gamma_{n-1}} x_{n-1} $$

for a suitable matrix $A_{\Gamma_n} \in \mathbb{R}^{N \times N}$ in an obvious way.

Intuitively it is clear that in order to have a chance that (23) converges to a consensus value, at some point every two nodes need to communicate with each other at some point. The following definition takes this into account.

Definition 3.4. *A set $\{\Gamma_1, \ldots, \Gamma_p\}$ of simple graphs with vertex set $[N]$ is called jointly connected if the union graph $\bigcup_{j=1}^p \Gamma_j$ is connected.*

The next result establishes convergence of a large class of consensus algorithms whenever the underlying family of graphs can be partitioned into jointly connected families, i.e., any two agents can communicate every few steps of the algorithm.

Theorem 3.5. *Assume that $(\Gamma_n)_{n \geq 0}$ is a family of graphs such that there exists a partitioning $0 = n_0 < n_1 < n_2 < \cdots < \infty$, $n_i \in \mathbb{N}$ with $\sup_i |n_i - n_{i-1}| = L < \infty$ such that all sets $\kappa^i := \{\Gamma_{n_i}, \ldots, \Gamma_{n_{i+1}-1}\}$ are jointly connected. Let*

$$ \mathcal{A} = \{A_\Gamma : \ \Gamma \ simple, \ undirected \ graph \ on \ [N]\}. $$

Then $\mathcal{A}_\mathcal{K}$ is a block SIA-family where $\mathcal{K} = (\kappa^i)_i$. In particular, the corresponding consensus algorithm (23) converges on an arbitrary Hadamard space.

Proof. The fact that $\mathcal{A}_\mathcal{K}$ is a block SIA-family is Lemma 1 of [13]. The rest follows from Theorem 3.3. □

Binary Averages.   Given a consensus algorithm defined by an SIA-family $\mathcal{A}$, in applying the corresponding algorithm in a Hadamard space, one faces the question of solving the optimization problems in the averaging procedure in Definition 1.4. If X possesses a differentiable structure, this can be done with a gradient-descent scheme. However, the class of Hadamard spaces is much richer than that of Riemannian manifolds with nonpositive curvature. For instance, the space of phylogenetic trees, or more generally of $CAT(0)$-cubical complexes [5, 6]. Even though it is of considerable interest e.g., to compute averages of phylogenetic trees, it seems to difficult to solve (4) directly. On the other

hand, it is possible to compute weighted binary averages of two points in $CAT(0)$-cubical complexes [3, 6, 11, 16]. Also, it is well-known and not difficult to see that every linear averaging operation

$$\sum_{i\in[N]} a(i)x(i), \quad \sum_{i\in[N]} a(i) = 1$$

can be written as a sequence of binary averages, see e.g., [22]. Therefore, every consensus algorithm based on successive averages can be written as a consensus algorithm based on successive binary averages. The same can be done for the corresponding nonlinear algorithm operating in a Hadamard space. We have the following result.

Theorem 3.6. *Assume that $\mathcal{A}$ is a block SIA-family. Consider the corresponding X-valued consensus algorithm with all averaging operations factored into (nonlinear) binary averages. Then this algorithm converges in X.*

Proof. The theorem follows by observing that in the linear case, the family $\mathcal{A}'$ which consists of the binary factorizations of $\mathcal{A}$ is a block SIA-family. Therefore, the result follows from Theorem 3.3. $\square$

References

[1] A. Aleksandrov. A theorem on triangles in a metric space and some of its applications. *Trudy Matematicheskogo Instituta imeni V.A. Steklova*, 38:5–23, 1951.

[2] A. Aleksandrov. Über eine Verallgemeinerung der Riemannschen Geometrie. *Schriftenreihe des Forschungsinstituts für Mathematik bei der Deutschen Akademie der Wissenschaften zu Berlin*, 1:33–84, 1957.

[3] F. Ardila, M. Owen, and S. Sullivant. Geodesics in CAT(0) cubical complexes. *Advances in Applied Mathematics*, 2011.

[4] W. Ballmann. *Lectures on spaces of nonpositive curvature.* Birkhäuser Basel, 1995.

[5] L. Billera, S. Holmes, and K. Vogtmann. Geometry of the space of phylogenetic trees. *Advances in Applied Mathematics*, 27(4):733–767, 2001.

[6] V. Chepoi. Graphs of some CAT(0) complexes. *Advances in Applied Mathematics*, 24(2):125–179, 2000.

[7] A. Cohen and N. Dyn. Nonstationary subdivision schemes and multiresolution analysis. *SIAM Journal on Mathematical Analysis*, 27(6):1745–1769, 1996.

[8] J. Dahl. Steiner problems in optimal transport. *Transactions of the American Mathematical Society*, 363:1805–1819, 2011.

[9] O. Ebner. Convergence of refinement schemes on metric spaces. *Proceedings of the American Mathematical Society*, 2011. to appear.

[10] O. Ebner. Nonlinear Markov semigroups and refinement schemes on metric spaces. *Arxiv preprint arXiv:1112.6003*, 2011.

[11] A. Feragen, S. Hauberg, M. Nielsen, and F. Lauze. Means in spaces of tree-like shapes. *Submitted to ICCV*, 2011.

[12] P. Grohs. A general proximity analysis of nonlinear subdivision schemes. *SIAM Journal on Mathematical Analysis*, 42:729–750, 2010.

[13] A. Jadbabaie, J. Lin, and A. Morse. Coordination of groups of mobile autonomous agents using nearest neighbor rules. *IEEE Transactions on Automatic Control*, 48(6):988–1001, 2003.

[14] H. Karcher. Mollifier smoothing and Riemannian center of mass. *Communications on Pure and Applied Mathematics*, 30:509–541, 1977.

[15] S. Lang. *Fundamentals of differential geometry.* Springer Verlag, 1999.

[16] M. Owen and J. Provan. A fast algorithm for computing geodesic distances in tree space. *IEEE/ACM Transactions on Computational Biology and Bioinformatics (TCBB)*, 8(1):2–13, 2011.

[17] X. Pennec, P. Fillard, and N. Ayache. A Riemannian framework for tensor computing. *International Journal of Computer Vision*, 66(1):41–66, 2006.

[18] R. Sepulchre. Consensus on nonlinear spaces. *Annual Reviews in Control*, 2011.

[19] K. Sturm. Nonlinear martingale theory for processes with values in metric spaces of nonpositive curvature. *Annals of Probability*, 30:1195–1222, 2002.

[20] K. Sturm. Probability measures on metric spaces of nonpositive curvature. *Contemporary Mathematics*, 338:357–390, 2003.

[21] J. Tsitsiklis, D. Bertsekas, and M. Athans. Distributed asynchronous deterministic and stochastic gradient optimization algorithms. *IEEE Transactions on Automatic Control*, 31(9):803–812, 1986.

[22] J. Wallner and N. Dyn. Convergence and C^1 analysis of subdivision schemes on manifolds by proximity. *Computer Aided Geometric Design*, 22:593–622, 2005.

[23] J. Wolfowitz. Products of indecomposable, aperiodic, stochastic matrices. *Proceedings of the American Mathematical Society*, 15(5):733–737, 1963.

[24] G. Xie and T. P.-Y. Yu. Smoothness equivalence properties of general manifold-valued data subdivision schemes. *Multiscale Modeling and Simulation*, 7(3):1073–1100, 2008.

Research Reports