Preprints of the 19th World Congress
The International Federation of Automatic Control
Cape Town, South Africa. August 24-29, 2014

Distributed Nonlinear Consensus in the Space of Probability Measures

Adrian N. Bishop * **and Arnaud Doucet** **

* *Australian National University and NICTA* (adrianb@kth.se)
** *University of Oxford, Dept. of Statistics* (doucet@stats.ox.ac.uk)

Abstract: Distributed consensus in the Wasserstein metric space of probability measures is introduced for the first time in this work. It is shown that convergence of the individual agents' measures to a common measure value is guaranteed so long as a weak network connectivity condition is satisfied asymptotically. The common measure achieved asymptotically at each agent is the one closest simultaneously to all initial agent measures in the sense that it minimises a weighted sum of Wasserstein distances between it and all the initial measures. This algorithm has applicability in the field of distributed estimation.

1. INTRODUCTION

The problem of distributed consensus concerns a group of dynamic agents that seek to develop a distributed agreement upon certain state variables of interest by exchanging information across a network.

For example, the problem of average-consensus concerns a group of agents that interact via some (possibly time-varying) interaction network. Each agent has an initial state value (say on the real line) and shares this value with their local neighbours in the network. Each agent then has an update rule for their state value that takes, as input, their own state value at the previous iteration and their received neighbour agent states. The goal of the update rule is to drive each agent (say asymptotically) to a state that corresponds to the average value of all initial state values. The agents are then said to have reached an average-consensus.

Typically the agents are connected via a network that changes with time due to link failures, node failure, packet drops etc. In distributed sensor networks the individual nodes (or some subset of such) may be mobile etc. in which case the interaction topology may change due to communication constraints etc. All such variations in topology can happen randomly and often the network is disconnected for some time. Studies on the convergence of consensus algorithms are often motivated by such complex time-varying network communication constraints.

1.1 Background

The consensus problem has a long history [Berger (1981); Borkar and Varaiya (1982); DeGroot (1974); Tsitsiklis and Athans (1984); Tsitsiklis et al. (1986)] which is too broad to cover here. We highlight Jadbabaie et al. (2003); Moreau (2005); Olfati-Saber et al. (2007); Olfati-Saber and Murray (2004); Ren and Beard (2005); Tsitsiklis et al. (1986) for further history, background and novel extensions.

⋆ A.N. Bishop is supported by NICTA, Canberra Research Lab and the Australian Research Council (ARC) via a Discovery Early Career Researcher Award (DE-120102873). The authors are also supported by AFOSR/AOARD via AOARD-144042.

Consensus algorithms as studied in Cao et al. (2008); Olfati-Saber and Murray (2004); Ren and Beard (2005); Tsitsiklis et al. (1986) focus on linear update rules (at each agent) and typically concern average-consensus or consensus about some linear function of all the agent's initial state values in Euclidean space. The average-consensus problem has a natural relationship with distributed linear least squares or distributed (linear) maximum likelihood estimation [Xiao et al. (2005)] and distributed Kalman filtering [Carli et al. (2008); Cattivelli and Sayed (2010); Olfati-Saber (2007); Spanos and Murray (2005)].

Consensus via nonlinear update rules is studied in Ajorlou et al. (2011); Hui and Haddad (2008); Moreau (2005); Yu et al. (2011). Here consensus to general functions (e.g. the maximum or minimum etc.) of all initial agent states may be sought as in Cortés (2008); Wang and Hong (2010) and even finite-time convergence may be achievable [Cortés (2006); Wang and Hong (2010)]. One may also want to achieve consensus to some time-varying reference signal (e.g. at some leader etc.) as in Hong et al. (2006); Zhu and Martínez (2010).

We note here that much of the existing literature on consensus concerns agreement in Euclidean space; e.g. the seminal papers of Moreau (2005); Olfati-Saber et al. (2007); Olfati-Saber and Murray (2004); Ren and Beard (2005) etc. all restrict themselves to Euclidean spaces.

The problem of synchronisation is closely related to consensus but typically deals with the problem of driving a network of oscillators to a common frequency/phase etc. This work typically concerns nonlinear manifolds such as the circle etc. A survey on synchronisation is given in Strogatz (2000) while consensus and synchronisation are related in Li et al. (2010). Some other notable exceptions of consensus in spaces other than Euclidean are Grohs (2012); Matei and Baras (2010); Sarlette and Sepulchre (2009); Sepulchre (2011). In particular, Sarlette and Sepulchre (2009); Sepulchre (2011) consider general nonlinear consensus on manifolds by embedding such manifolds in a suitably high-dimensional Euclidean space. In Sarlette and Sepulchre (2009); Sepulchre (2011) consensus on the special orthogonal group and on Grassmann manifolds are explored using this embedding approach. Separate work in Matei and Baras (2010) considers consensus in convex metric spaces while Grohs (2012) develops an analogue of Wolfowtiz's theorem [Wolfowitz (1963)] for a special class of metric spaces with non-positive curvature which leads to a notion of consensus in such spaces.

1.2 Contribution

The contribution of this paper is a novel algorithm and convergence results for distributed consensus in the space of probability measures with time-varying interaction networks. We introduce a well-studied metric known as the Wasserstein distance which allows us to consider an important set of probability measures as a metric space [Givens and Shortt (1984)]. The introduced consensus algorithm is based on iteratively updating each agent's probability measure by finding a measure that is a minimal (weighted) Wasserstein distance from the agent's own previous measure plus all neighbour agents' measures. We show that convergence of the individual agents measures to a common probability measure is guaranteed as long as a weak network connectivity condition is satisfied. The common measure that is achieved asymptotically at each agent is the one that is closest simultaneously to all initial agent measure values in the sense of the Wasserstein distance.

This work has wide applicability in the field of distributed estimation [Doucet et al. (2001)] and distributed information fusion [Liggins et al. (1997); Mahler (2007)]. For example, suppose each agent starts with a probability measure conditioned on some common underlying event of interest. Then one would like to combine all these measures (which amount to each agents estimate and/or belief of the underlying event) into a common probability measure that captures all the agents beliefs. The proposed consensus algorithm can do this in a very general distributed setting. Related work in Olfati-Saber et al. (2005) considers the application of consensus to the problem of distributed Bayesian information fusion. Unlike Olfati-Saber et al. (2005), the proposed method can deal with singular probability measures and the connectivity condition proposed here is more relaxed than in Olfati-Saber et al. (2005).

1.3 Paper Organization

The paper is organised as follows. In the next section we provide the general problem setup and a short background on linear consensus via a study on the convergence of infinite matrix products. In Section 3 we introduce a notion of consensus in Euclidean space under a broad class of metrics in order to gently introduce the idea of consensus in general metric spaces. In Section 4 we introduce the main contribution of this paper which is an algorithm and convergence analysis for consensus in the Wasserstein metric space of probability measures. Concluding remarks are given in 5.

2. SETUP AND A QUICK BACKGROUND ON CONSENSUS IN $\mathbb{R}$

2.1 Common Setup

Consider a group of agents indexed in $\mathcal{V} = \{1, \dots, n\}$ and a set of possible time-varying undirected links $\mathcal{E}(t) \subset \mathcal{V}\times\mathcal{V}$ defining a network graph $\mathcal{G}(t)(\mathcal{V}, \mathcal{E}(t))$. The neighbor set at agent i is denoted by $\mathcal{N}_i(t) = \{j \in \mathcal{V} : (i, j) \in \mathcal{E}(t)\}$ and $j \in \mathcal{N}_i(t) \Leftrightarrow i \in \mathcal{N}_j(t)$ for undirected topologies.

The graph adjacency matrix $\mathbf{A}(t) \in \mathbb{R}^{n\times n}$ obeys $\mathbf{A} = \mathbf{A}^\top = [a_{ij}(t)]$ where $a_{ij}(t) = 1 \Leftrightarrow \{i, j\} \in \mathcal{E}(t)$ and $a_{ij}(t) = 0$ otherwise. A weighted adjacency matrix is denoted by $\mathbf{W}(t) = [w_{ij}(t)] \in \mathbb{R}^{n\times n}$. Throughout, we will restrict $w_{ij}(t) \in (0, 1)$ and require $(w_{ii}(t) + \sum_{j\in\mathcal{N}_i(t)} w_{ij}(t)) = 1$ (although we allow without further reference $w_{ii}(t) = 1$ in which case we require $w_{ij}(t) = 0$ for all $j \neq i$). Time is indexed in $\mathbb{N}$ and $0 \in \mathbb{N}$.

Consider the sequence of graphs $\mathcal{G}(t_k), \mathcal{G}(t_{k_1}), \dots, \mathcal{G}(t_{k+1})$ on the same vertex set $\mathcal{V}$. The union of this sequence of graphs is the graph $\mathfrak{G}(t_k, t_{k+1})(\mathcal{V}, \cup_{t\in[t_k,t_{k+1})}\mathcal{E}(t))$. The sequence is said to be jointly connected if $\mathfrak{G}$ is connected.

2.2 A Quick Background on Consensus in $\mathbb{R}$

A matrix with nonnegative elements is denoted by $\mathbf{P} \succeq 0$ or by $\mathbf{P} \succ 0$ when all elements are strictly positive. Similar notation is also used for nonnegative and positive vectors. See Bremaud (1999); Seneta (1973) for a definition of stochastic, irreducible and primitive matrices.

Theorem 1 (Perron; see Bremaud (1999); Seneta (1973)). *Let $\mathbf{P} \in \mathbb{R}^{n\times n}$ be nonnegative and primitive. Then*

$$\lim_{k\to\infty} \left[\frac{\mathbf{P}}{\max(|\lambda(\mathbf{P})|)}\right]^k = \mathbf{v}\mathbf{u}^\top$$

where $\mathbf{P}\mathbf{v} = \max(|\lambda(\mathbf{P})|)\mathbf{v}$, $\mathbf{u}^\top\mathbf{P} = \max(|\lambda(\mathbf{P})|)\mathbf{u}^\top$ and $\mathbf{v} \succ 0$, $\mathbf{u} \succ 0$ and $\mathbf{v}^\top\mathbf{u} = 1$.

If $\mathbf{P} \in \mathbb{R}^{n\times n}$ is also row-stochastic then $\mathbf{P}\mathbf{1} = \mathbf{1}$ and thus $\lim_{k\to\infty} \mathbf{P}^k = \mathbf{1}\mathbf{u}^\top$. If $\mathbf{P} \in \mathbb{R}^{n\times n}$ is doubly stochastic then also $\mathbf{1}^\top\mathbf{P} = \mathbf{1}^\top$ and thus $\lim_{k\to\infty} \mathbf{P}^k = \frac{1}{n}\mathbf{1}\mathbf{1}^\top$.

A generalisation of the above theorem due to Wolfowitz can be stated as follows.

Theorem 2 (Wolfowitz (1963)). *Consider a finite set $\mathcal{P} = \{\mathbf{P}_1, \dots, \mathbf{P}_m\}$, $m \geq 1$, of non-negative, row-stochastic, primitive matrices $\mathbf{P} \in \mathbb{R}^{n\times n}$ with the property that for any sequence s_n, $n \geq 1$ the product $\mathbf{P}_{s_k} \cdots \mathbf{P}_{s_2}\mathbf{P}_{s_1}$ is also primitive for any $k > 1$. Then*

$$\lim_{k\to\infty} \mathbf{P}_{s_k} \cdots \mathbf{P}_{s_2}\mathbf{P}_{s_1} = \mathbf{1}\mathbf{u}^\top$$

where $\mathbf{u}$ is dependent on the particular sequence.

Suppose the state of agent i at time $t = 0$ is given by a point $x_i(0) \in \mathbb{R}$. Then suppose that agent i updates its state iteratively according to

$$x_i(t + 1) = w_{ii}(t)x_i(t) + \sum_{j\in\mathcal{N}_i(t)} w_{ij}(t)x_j(t) \qquad (1)$$

where we assume $(w_{ii}(t) + \sum_{j\in\mathcal{N}_i(t)} w_{ij})(t) = 1$. The evolution of the group can be written in vector form as

$$\mathbf{x}(t + 1) = \mathbf{W}(t)\mathbf{x}(t)$$

where $\mathbf{W}(t) \in \mathbb{R}^{n\times n}$ is a non-negative row-stochastic matrix and $\mathbf{x}(t) \in \mathbb{R}^n$.

Theorem 3 (Jadbabaie et al. (2003)). *Consider a group of agents $\mathcal{V}$ and network $\mathcal{G}(t)(\mathcal{V}, \mathcal{E}(t))$. Suppose the state of each agent is $x_i(t) \in \mathbb{R}$ and that each agent applies (1) where $\mathbf{W}(t) \in \mathbb{R}^{n\times n}$ is a non-negative, row-stochastic, primitive matrix for each $t \geq 0$. If there exists an infinite* *sequence of contiguous, nonempty, bounded, time-intervals $[t_k, t_{k+1})$, starting at $t = 0$, with the property that across each such interval the graph union $\mathfrak{G}(t_k, t_{k+1})$ is connected and $\prod_{t\in[t_k,t_{k+1})} \mathbf{W}(t)$ is primitive, then*

$$\lim_{t\to\infty} \mathbf{x}(t) = (\mathbf{u}^\top \mathbf{x}(0))\mathbf{1}$$

for some constant $\mathbf{u} \succ 0$ with $\mathbf{u}^\top \mathbf{1} = 1$.

The proof of this theorem is a simple application of Wolfowitz's theorem. In practice, for a particular class of matrices $\mathbf{W}(t)$ being row-stochastic and primitive one must establish $\prod_{t\in[t_k,t_{k+1})} \mathbf{W}(t)$ is also primitive in order to make use of the theorem. For example if $w_{ii}(t) > 0$ for all i and t then $\prod_{t\in[t_k,t_{k+1})} \mathbf{W}(t)$ is primitive [Jadbabaie et al. (2003)]. If $\mathbf{W}(t)$ is constant in t and doubly stochastic then each agent converges to $\frac{1}{n}\sum_{i\in\mathcal{V}} x_i(0)$.

3. A DIFFERENT VIEW OF CONSENSUS IN $\mathbb{R}$ AND IN ARBITRARY METRIC SPACES $(\mathbb{R}, d)$

The motivation for this section is just to introduce the idea of consensus in metric spaces in the simplest fashion. Suppose we consider an arbitrary metric $d : \mathbb{R}\times\mathbb{R} \to [0,\infty)$ on $\mathbb{R}$ and define an update by

$$x_i(t+1) = \operatorname*{argmin}_z \sum_{j\in\mathcal{N}_i(t)\cup\{i\}} w_{ij}(t) d(z, x_j(t))^2$$

Let $\mathbf{x}(t) = [x_1(t) \ \ldots \ x_n(t)]^\top$. Then we can define

$$\begin{aligned} x_i(t+1) &= W_i^\star(t)\mathbf{x}(t) \\ &= \operatorname*{argmin}_z \sum_{j\in\mathcal{N}_i(t)\cup\{i\}} w_{ij}(t) d(z, x_j(t))^2 \end{aligned} \quad (2)$$

and the dynamics of the entire group of agents by

$$\mathbf{x}(t+1) = \mathbf{W}^\star(t)\mathbf{x}(t) = [W_1^\star(t)\mathbf{x}^\top(t) \ \ldots \ W_n^\star(t)\mathbf{x}^\top(t)]^\top$$

where in this case $\mathbf{W}^\star(t)$ (and each $W_i^\star(t)$) represents a nonlinear operator with parameters w_{ij}. Clearly this is a generalisation of the consensus problem considered in the previous section (where (1) is actually a special case of (2) with d given by the usual Euclidean metric).

Note we derive the results here for the space $(\mathbb{R}, d)$ for comparison with, and inline with, the results of the previous section. However, the results presented in this section can straightforwardly be extended to $(\mathbb{R}^m, d)$ with $1 \le m < \infty$.

All metrics are continuous (in the topology induced by the metric) and, for metrics defined on $\mathbb{R}$, we say d is strictly increasing if $d(x,y) \le c\cdot d(x,z)$ for all $x < y < z$ in $\mathbb{R}$ and some $c > 0$ that may depend on all x, y, z. Similarly, a metric d defined on $\mathbb{R}$ is said to have uniform growth if $d(x, x+c) = d(y, y+c)$ for all $x, y, c \in \mathbb{R}$.

Lemma 1. *Define a closed set including those states $x_j(t)$ for all $j \in \{\mathcal{N}_i(t)\cup\{i\}\}$ by the line segment in $\mathbb{R}$ given by $\{x \in \mathbb{R} : \min_j[x_j(t)] \le x \le \max_j[x_j(t)]\}$. Then, for any strictly increasing metric $d : \mathbb{R}\times\mathbb{R} \to [0,\infty)$ with uniform growth, the result*

$$x_i(t+1) = W_i^\star(t)\mathbf{x}(t) = \operatorname*{argmin}_z \sum_{j\in\mathcal{N}_i(t)\cup\{i\}} w_{ij}(t) d(z, x_j(t))^2$$

is strictly within this closed set of states whenever the cardinality of $\{\mathcal{N}_i(t)\cup\{i\}\}$ is greater than two and at least two of the state values are distinct.

Proof. Let $x_{max} = \max_j[x_j(t)]$ and q denote the index j of the maximum element $\max_j[x_j(t)]$. Suppose that $W_i^\star(t)\mathbf{x}(t)$ is strictly greater than x_{max}. Then clearly one must find (by the triangle inequality)

$$\operatorname*{argmin}_z \sum_{j\in\{\mathcal{N}_i(t)\cup\{i\}\}\setminus\{q\}} w_{ij}(t) d(z, x_j(t))^2 > \sum_{j\in\mathcal{N}_i(t)\cup\{i\}} w_{ij}(t) d(x_{max}, x_j(t))^2$$

which is a contradiction. Thus, we can restrict $W_i^\star(t)\mathbf{x}(t) \le x_{max}$. A similar argument can be made about $\min_j[x_j(t)]$ implying that $W_i^\star(t)\mathbf{x}(t)$ can be readily restricted to within or on the boundary of the closed set of interest for any metric d. Now suppose there are only two agents (the proof to this point allows this) with states $x_1 < x_2$. For $i \in \{1,2\}$ it follows that

$$\begin{aligned} W_i^\star(t)\mathbf{x}(t) &= \operatorname*{argmin}_z \ w_{i1}(t) d(z, x_1(t))^2 + (1 - w_{i1}(t)) d(z, x_2(t))^2 \\ &= \operatorname*{argmin}_z \ w_{i1}(t)\left(d(z, x_1(t))^2 - d(z, x_2(t))^2\right) + d(z, x_2(t))^2 \end{aligned}$$

where $w_{i1}(t)\left(d(z, x_1(t))^2 - d(z, x_2(t))^2\right)$ is strictly negative initially and strictly increasing as z moves from $x_1(t)$ to $x_2(t)$ (crossing zero at some point on the interval) and conversely $d(z, x_2(t))^2$ is strictly positive and strictly decreasing to zero as z moves from $x_1(t)$ to $x_2(t)$ (obtaining zero only at $z = x_2(t)$).

Then for any $w_{i1} \in (0,1)$ and because d is continuous it follows that there exists some $\epsilon > 0$ such that

$$\begin{aligned} w_{i1}(t)\left(d(z, x_1(t))^2 - d(z, x_2(t))^2\right) &< 0 \\ |w_{i1}(t)\left(d(z, x_1(t))^2 - d(z, x_2(t))^2\right)| &< d(z, x_2(t))^2 \end{aligned}$$

on $z \in [x_1(t), x_1(t)+\epsilon]$. Consequently, $W_i^\star(t)\mathbf{x}(t)$ is strictly decreasing on $z \in [x_1(t), x_1(t)+\epsilon]$ for some $\epsilon > 0$. Hence for any $w_{i1} \in (0,1)$ the point $x_1(t)$ cannot be a minimum. The same argument can be applied for the point $x_2(t)$ by readily swapping w_{i1} and $1 - w_{i1}$ for $1 - w_{i2}$ and w_{i2} respectively. □

Note that uniform growth is not necessary for $W_i^\star(t)\mathbf{x}(t)$ to be within the strict interior of the desired closed set and nor is the requirement that d be strictly increasing.

Theorem 4. *Consider a group of agents $\mathcal{V}$ and network $\mathcal{G}(t)(\mathcal{V}, \mathcal{E}(t))$. Suppose the state of each agent is $x_i(t) \in \mathbb{R}$ and that each agent applies (2) where $d : \mathbb{R}\times\mathbb{R} \to [0,\infty)$ is a strictly increasing metric with uniform growth. If for all $t_0 \in \mathbb{N}$ the graph union $\mathfrak{G}(t_0, \infty)$ is connected then*

$$\lim_{t\to\infty} d(x_i(t), x_j(t))^2 = 0$$

for all $i, j \in \mathcal{V}$ and $x_i(t)$ converges to some constant in $\{x \in \mathbb{R} : \min_j[x_j(0)] \le x \le \max_j[x_j(0)]\}$ as $t \to \infty$.

Proof. Proof of this result follows from Lemma 1 and the main result of Moreau (2005). □

Note that the requirement on the connectivity property of the network over time is actually less restrictive than the requirement given in Theorem 3.

4. CONSENSUS IN THE SPACE OF PROBABILITY MEASURES (THE WASSERSTEIN METRIC SPACE)

The main contribution of this work is given in this section where we introduce a consensus protocol in the Wasserstein metric space of probability measures.

Consider the same setup as before but suppose the state of agent i is given by a Radon probability measure μ_i defined on the Borel sets of $(\mathbb{R}^m, d)$ with $0 < m < \infty$ where in this section we restrict $d : \mathbb{R}^m \times \mathbb{R}^m \to [0, \infty)$ to be the usual Euclidean distance.

Define the space of all such measures on $(\mathbb{R}^m, d)$ by $\mathfrak{U}(\mathbb{R}^m)$ and the subset of all such measures with bounded, finite, p^{th} moment by $\mathfrak{U}_p(\mathbb{R}^m)$ for some suitably small $1 \le p < \infty$. That is, $\mathfrak{U}_p$ is the collection of probability measures on the Borel sets of $(\mathbb{R}^m, d)$ such that

$$\int_{\mathbb{R}^m} d(\mathbf{x}, \mathbf{x}_0)^p \, \mathrm{d}\mu_i(\mathbf{x}) < \infty$$

for all bounded $\mathbf{x} \in \mathbb{R}^m$ and a given $\mathbf{x}_0 \in \mathbb{R}^m$. We typically won't note the dependence on $\mathbb{R}^m$ unless it is a special case.

One can associate the so-called Wasserstein metric $\ell_p : \mathfrak{U}_p \times \mathfrak{U}_p \to [0, \infty)$ with $\mathfrak{U}_p$. This metric is defined by

$$\ell_p(\mu_i, \mu_j) = \left(\inf_{\gamma \in \Gamma(\mu_i, \mu_j)} \int_{\mathbb{R}^m \times \mathbb{R}^m} d(\mathbf{x}_i, \mathbf{x}_j)^p \, \mathrm{d}\gamma(\mathbf{x}_i, \mathbf{x}_j) \right)^{1/p}$$

where $\Gamma(\mu_i, \mu_j)$ denotes the collection of all measures on $\mathbb{R}^m \times \mathbb{R}^m$ with marginals μ_i and μ_j on the first and second factors; see Ambrosio et al. (2005); Villani (2003).

Lets collect some facts about the Wasserstein metric space $(\mathfrak{U}_p, \ell_p)$ with $p \ge 2$ from Ambrosio et al. (2005); Bertrand and Kloeckner (2012); Kloeckner (2010); Villani (2003).

(1) $(\mathfrak{U}_p, \ell_p)$ is a complete and separable metric space.
(2) Convergence $\lim_{k\to\infty} \ell_p(\mu_k, \mu) = 0$ is equivalent to weak convergence and convergence in the first p moments.
(3) Considering two measures $\mu_i, \mu_j \in \mathfrak{U}_p$ then $\ell_p(\mu_i, \mu_j) = \ell_p(\mu_i, \mu) + \ell_p(\mu_j, \mu)$ for some $\mu \in \mathfrak{U}_p$.
(4) More generally, there exists a continuously parameterised constant speed path $\mu_s \in \mathfrak{U}_p$, $s \in [0, 1]$ such that for $\mu_i, \mu_j \in \mathfrak{U}_p$ we have $\mu_{s=0} = \mu_i$ and $\mu_{s=1} = \mu_j$ and $\ell_p(\mu_i, \mu_j) = \ell_p(\mu_i, \mu_s) + \ell_p(\mu_j, \mu_s)$ for all $s \in [0, 1]$. The measure μ_s is known as the interpolant measure; see McCann (1997).
(5) The interpolant measure defines a geodesic and consequently $(\mathfrak{U}_p, \ell_p)$ is geodesic.
(6) $(\mathfrak{U}_p(\mathbb{R}), \ell_p)$ has vanishing curvature in the sense of Alexandrov (a subset of CAT(0)) when $\mu_i \in \mathfrak{U}_p$ is defined on the Borel sets of $(\mathbb{R}, d)$ with d the usual Euclidean metric on $\mathbb{R}$; see Kloeckner (2010).
(7) $(\mathfrak{U}_p(\mathbb{R}), \ell_p)$ is simply connected; see Bertrand and Kloeckner (2012); Kloeckner (2010).
(8) $(\mathfrak{U}_p(\mathbb{R}^m), \ell_p)$ with $1 < m < \infty$ has positive curvature in the sense of Alexandrov; see Ambrosio et al. (2005).

All metrics are continuous and a constant speed geodesic in $(\mathfrak{U}_p, \ell_p)$ is a curve $\mu_s : \mathbb{I} \to \mathfrak{U}_p$ parameterised on some interval $s \in \mathbb{I} \subset \mathbb{R}$ that satisfies $\ell_p(\mu_{s_i}, \mu_{s_j}) = v|s_i - s_j|$ for some constant $v > 0$ and for all $s_i, s_j \in \mathbb{I}$.

Suppose the measure at agent i is updated by

$$\boxed{\mu_i(t+1) = \inf_{z \in \mathfrak{U}_p} \sum_{j \in \mathcal{N}_i(t) \cup \{i\}} w_{ij}(t) \ell_p(z, \mu_j(t))^p} \quad (3)$$

for all $i \in \mathcal{V}$ where as before we restrict $w_{ij}(t) \in (0, 1)$ and require $(w_{ii}(t) + \sum_{j \in \mathcal{N}_i(t)} w_{ij}(t)) = 1$ (again allowing the special case $w_{ii}(t) = 1$ and $w_{ij}(t) = 0$ for all $j \ne i$).

Application of (3) at each agent $i \in \mathcal{V}$ corresponds to the proposed nonlinear (distributed) consensus algorithm in the Wasserstein metric space of probability measures.

Proposition 1. *Consider a group of agents $\mathcal{V}$ and network $\mathcal{G}(t)(\mathcal{V}, \mathcal{E}(t))$. Suppose $\mu_i(t) \in \mathfrak{U}_p$ and that each agent applies (3) where ℓ_p is the Wasserstein metric. If for all $t_0 \in \mathbb{N}$ the graph union $\mathfrak{G}(t_0, \infty)$ is connected then*

$$\lim_{t\to\infty} \ell_p(\mu_i, \mu_j)^p = 0$$

for all $i, j \in \mathcal{V}$ and $\lim_{t\to\infty} \mu_i(t) = \text{some constant}$ in $\mathfrak{U}_p$.

Corollary 1. *Consider a group of agents $\mathcal{V}$ and network $\mathcal{G}(t)(\mathcal{V}, \mathcal{E}(t))$. Suppose the state of each agent is $\mu_i(t) \in \mathfrak{U}_p(\mathbb{R})$ and that each agent applies (3) where ℓ_p is the Wasserstein metric. If for all $t_0 \in \mathbb{N}$ the graph union $\mathfrak{G}(t_0, \infty)$ is connected then*

$$\lim_{t\to\infty} \ell_p(\mu_i, \mu_j)^p = 0$$

$\forall i, j \in \mathcal{V}$ and $\lim_{t\to\infty} \mu_i(t) = \text{some constant}$ in $\mathfrak{U}_p(\mathbb{R})$.

Here we will concentrate on proof of the stated corollary.

A subset $\mathfrak{X} \subset \mathfrak{U}_p(\mathbb{R})$ is convex if every geodesic segment whose endpoints are in $\mathfrak{X}$ lies entirely in $\mathfrak{X}$. The (closed) convex hull $\mathrm{co}(\mathfrak{Y})$ of a subset $\mathfrak{Y} \subset \mathfrak{U}_p$ is the intersection of all (closed) convex subsets of $\mathfrak{U}_p$ that contain $\mathfrak{Y}$.

Lemma 2. *Suppose $\mu_i(t) \in \mathfrak{U}_p$ is defined on the Borel sets of $(\mathbb{R}, d)$ and thus $(\mathfrak{U}_p(\mathbb{R}), \ell_p)$ is CAT(0) in addition to being uniquely geodesic, complete and separable. Then the operation (3) at each agent is well-defined in the sense that it has a solution and this solution is unique.*

This lemma follows from the fact that $(\mathfrak{U}_p(\mathbb{R}), \ell_p)$ is Hadamard and Fréchet averages such as defined by operations of the form (3) are well defined in such spaces; see Burago et al. (2001).

Lemma 3. *Consider a collection $\{\mu_i\}$, $i \in \widetilde{\mathcal{V}} \subseteq \mathcal{V}$ of distinct points in $(\mathfrak{U}_p(\mathbb{R}), \ell_p)$. The convex hull of $\{\mu_i\}$ is $\mathrm{co}(\{\mu_i\}) \subset \mathfrak{U}_p(\mathbb{R})$ and is isometric to a l-sided convex polygon in $\mathbb{R}^2$ with $2 \le l \le |\{\mu_i\}|$.*

Before proceeding with the proof we point to Bridson and Haefliger (1999) for background on comparison triangles and Alexandrov curvature of metric spaces. We also note that in a general geodesic CAT(0) space, i.e. some arbitrary geodesic space with non-positive curvature, the preceding lemma is not true and the convex hull of a 'geodesic triangle' [Bridson and Haefliger (1999)] defined by three points in such spaces may be of dimension greater than two [Bridson and Haefliger (1999)]. Thus, our Euclidean intuition is generally incorrect when it suggests the existence of a two-dimensional convex hull for a triangle defined by three points and the geodesics connecting them (albeit this is hard to visualise of course).

Proof. Lemma 3 is really just a straightforward consequence of the vanishing curvature property of $(\mathfrak{U}_p(\mathbb{R}), \ell_p)$. We elaborate for completeness.

$(\mathfrak{U}_p(\mathbb{R}), \ell_p)$ has vanishing curvature in the sense of Alexandrov [Kloeckner (2010)] which formally means that for any triangle of points $\{\mu_i\}$, $i \in \{i_1, i_2, i_3\}$ and any point on the geodesic $\mu_s \in \mathfrak{U}_p(\mathbb{R})$, $s \in [0,1]$ such that, for example, $\mu_{s=0} = \mu_{i_1}$ and $\mu_{s=1} = \mu_{i_2}$ then the ℓ_p distance between μ_{i_3} and μ_s, $s \in [0,1]$ is the same as the corresponding distance $d : \mathbb{R}^2 \times \mathbb{R}^2 \to [0,\infty)$ in a comparison triangle in $\mathbb{R}^2$. Consider also any pair of points μ_j and μ_k with μ_j on the geodesic connecting μ_{i_1} and μ_{i_2} and μ_k on the geodesic connecting μ_{i_1} and μ_{i_3} with $\{\mu_j, \mu_k\} \cap \{\mu_i\} = \emptyset$, $i \in \{i_1, i_2, i_3\}$. Then vanishing curvature also implies $\angle_{\mu_{i_1}}(\mu_j, \mu_k)$ is equal to the usual interior Euclidean angle at the corresponding vertex in the comparison triangle in $\mathbb{R}^2$. Here the angle $\angle_{\mu_{i_1}}(\mu_j, \mu_k)$ is the Alexandrov angle in arbitrary metric spaces; see Bridson and Haefliger (1999). It now follows that the convex hull of any triangle of points $\{\mu_i\}$ in $(\mathfrak{U}_p(\mathbb{R}), \ell_p)$ is isometric to a triangle in $\mathbb{R}^2$; e.g. see Proposition 2.9 (Flat Triangle Lemma) in Bridson and Haefliger (1999).

Now define $\mathfrak{C} = \{\Delta_j\}$ to be the collection of geodesic triangles in $(\mathfrak{U}_p(\mathbb{R}), \ell_p)$ defined by every combination of three points in $\{\mu_i\}$, $i \in \widetilde{\mathcal{V}} \subseteq \mathcal{V}$. Clearly $\text{co}(\{\mu_i\}) = \cup_j \Delta_j$. Consider also the corresponding collection $\mathfrak{C}^+ = \{\Delta_j^+\}$ of comparison triangles in $\mathbb{R}^2$. The Flat Triangle Lemma implies that this collection can be arranged in $\mathbb{R}^2$ such that each angle $\angle_{\mu_i}(\mu_j, \mu_k)$ and each distance $\ell_p(\mu_i, \mu_j)$ for all $i, j, k \in \widetilde{\mathcal{V}}$ in $(\mathfrak{U}_p(\mathbb{R}), \ell_p)$ equals exactly the corresponding angle or distance in the comparison configuration of points in $\mathbb{R}^2$. Obviously, the convex hull of the comparison configuration is a l-sided convex polygon in $\mathbb{R}^2$ with $2 \leq l \leq |\{\mu_i\}|$ and equal to $\cup_j \Delta_j^+$.

Define the following map

$$f_{\ell_p,d} : \text{co}(\{\mu_i\}) \to \mathbb{R}^2, \quad i \in \widetilde{\mathcal{V}} \tag{4}$$

so the restriction $f_{\ell_p,d}(\Delta_j) = f_{\ell_p,d}(\text{co}(\{\mu_{j_1}, \mu_{j_2}, \mu_{j_3}\})) = \text{co}(\{f_{\ell_p,d}(\mu_{j_1}), f_{\ell_p,d}(\mu_{j_2}), f_{\ell_p,d}(\mu_{j_3})\}) = \Delta_j^+$, $\forall j \in \mathfrak{C} = \{\Delta_j\}$ is an isometry. Then

$$f_{\ell_p,d}(\text{co}(\{\mu_i\})) = f_{\ell_p,d}(\cup_j \Delta_j) = \cup_j f_{\ell_p,d}(\Delta_j) = \cup_j \Delta_j^+$$

from the Flat Triangle Lemma and the property of vanishing curvature. For any two points in $\text{co}(\{\mu_i\})$ there exists a particular $\Delta_j \in \mathfrak{C}$ that contains them and the restriction $f_{\ell_p,d}(\Delta_j)$ is an isometry to a convex subset of $\cup_j \Delta_j^+$. Thus, $f_{\ell_p,d}$ is an isometry and this completes the proof. □

More precisely, the convex hull in $(\mathfrak{U}_p(\mathbb{R}), \ell_p)$ is defined by

$$\text{co}(\{\mu_i\}) = \{\inf_{z \in \mathfrak{U}_p} \sum_{i \in \widetilde{\mathcal{V}}} w_i \ell_p(z, \mu_i)^p | w_i \in [0,1], \sum_i w_i = 1\}$$

on the set of points $\{\mu_i\}$, $i \in \widetilde{\mathcal{V}} \subseteq \mathcal{V}$.

Lemma 4. *Consider the convex hull* $\text{co}(\{\mu_i(0)\})$ *of all initial agent states in* $(\mathfrak{U}_p(\mathbb{R}), \ell_p)$. *If each agent applies (3) it follows that* $\text{co}(\{\mu_i(t)\}) \subseteq \text{co}(\{\mu_i(0)\})$ *for all* t.

Actually, this is really a consequence of the following, stronger result.

Lemma 5. *Consider the convex hull* $\text{co}(\{\mu_j(t)\})$, *with* $j \in \mathcal{N}_i(t) \cup \{i\}$ *at time* t. *If agent* i *applies (3) it follows that* $\mu_i(t+1)$ *is strictly within the convex hull* $\text{co}(\{\mu_j(t)\})$ *whenever* $|\{\mu_j(t)\}| \geq 2$ *and two agent states are distinct.*

Proof. It is enough to consider two agents $i, j \in \mathcal{V}$ with (3) then given by

$$\mu_i(t+1) = \inf_{z \in \mathfrak{U}_p} w_{ii}(t)\left(\ell_p(z, \mu_i(t))^p - \ell_p(z, \mu_j(t))^p\right) + \ell_p(z, \mu_j(t))^p$$

and to note that z must lie on a geodesic $\mu_s : \mathbb{I} \to \mathfrak{U}_p$. The proof relies on showing that $\mu_i(t+1) \notin \{\mu_i(t), \mu_j(t)\}$ when $w_{ii}, w_{ij} \in (0,1)$.

The first term

$$w_{ii}(t)\left(\ell_p(z, \mu_i(t))^p - \ell_p(z, \mu_j(t))^p\right)$$

is strictly negative at $z = \mu_i(t)$ and strictly increasing as z moves from $\mu_i(t)$ to $\mu_j(t)$ and conversely $\ell_p(z, \mu_j(t))^p$ is strictly positive at $z = \mu_i(t)$ and strictly decreasing to zero as z moves from $\mu_i(t)$ to $\mu_j(t)$. Then for any $w_{ii} \in (0,1)$ and because ℓ_p is continuous it follows that there exists some μ_ϵ on μ_s with $\epsilon > 0$ such that

$$w_{ii}(t)\left(\ell_p(z, \mu_i(t))^p - \ell_p(z, \mu_j(t))^p\right) < 0$$
$$|w_{ii}(t)\left(\ell_p(z, \mu_i(t))^p - \ell_p(z, \mu_j(t))^p\right)| < \ell_p(z, \mu_j(t))^p$$

on $z \in \mu_s$, $s \in [0, \epsilon]$. Consequently, $\mu_i(t+1)$ is strictly decreasing on $z \in \mu_s$, $s \in [0, \epsilon]$. Hence for any $w_{i1} \in (0,1)$ the point $\mu_i(t)$ cannot be a minimum. The same argument can be applied for the point $\mu_j(t)$. □

Lemma 6. *Suppose* $\mathcal{G}(\mathcal{V}, \mathcal{E})$ *is time-invariant, connected, but not (necessarily) complete. Suppose the state of each agent is* $\mu_i(t) \in \mathfrak{U}_p(\mathbb{R})$ *and that each agent applies (3) where* ℓ_p *is the Wasserstein metric. Then*

$$\lim_{t \to \infty} \ell_p(\mu_i, \mu_j)^p = 0$$

$\forall i, j \in \mathcal{V}$ *and* $\lim_{t\to\infty} \mu_i(t) = \text{some constant}$ *in* $\mathfrak{U}_p(\mathbb{R})$.

Proof. It almost goes without saying that $\ell_2(\mu_i, \mu_j)^p = 0$, $\forall i, j \in \mathcal{V}$ and $\mu_i = \text{some constant in } \mathfrak{U}_p(\mathbb{R})$ is an equilibrium state of (3). Now consider a Lyapunov-like function $\nu(\mu) : \mathfrak{U}_p \to \mathbb{R}$ given by

$$\nu(\mu) = \sup_{x, y \in \{\mu_i(t)\}_{i \in \mathcal{V}}} \ell_p(x, y)^p \tag{5}$$

and note that $\nu(\mu) \geq 0$ with $\nu(\mu) = 0$ if and only if $\mu_i = \mu_j$ for all $i, j \in \mathcal{V}$. By Lemma 4 it follows that $\nu(\mu)$ is non-increasing along trajectories of (3). It suffices to show $\nu(\mu(t + n - 1)) < \nu(\mu(t))$ for each t.

Firstly, pick a $t_0 \geq 0$ and note $\text{co}(\{\mu_i(t_0)\}) \subseteq \text{co}(\{\mu_i(0)\})$ and thus $f_{\ell_p,d}(\text{co}(\{\mu_i(t_0)\})) \subseteq f_{\ell_p,d}(\text{co}(\{\mu_i(0)\}))$ from Lemma 4 and where $f_{\ell_p,d}$ is an isometry given by (4). Without loss of generality, via Lemma 3, suppose that $f_{\ell_p,d}(\text{co}(\{\mu_i(t_0)\}))$ is a l-sided polygon in $\mathbb{R}^2$ with $2 \leq l \leq |\mathcal{V}|$ on the collection of vertices $\{\mathbf{x}_j(t_0)\}$, $j \in \{1, \ldots, l\}$ with $\mathbf{x}_j(t_0) \in \mathbb{R}^2$. If we chose a t_0 such that $l = 1$ then we would be done. Define the following set-valued function

$$h_j(t) = \{i \in \mathcal{V} : f_{\ell_p,d}(\mu_i(t)) = \mathbf{x}_j(t_0)\}\,, \quad \forall j \in \{1,\ldots,l\} \tag{6}$$

for each time $t \geq t_0$. It is immediate from Lemma 5 that $h_j(t+1) \subseteq h_j(t)$ for all $j \in \{1,\ldots,l\}$; i.e. more generally, no agent state $f_{\ell_p,d}(\mu_i(t))$ which is not on the boundary of the l-sided polygon at time t can ever reach this same boundary at $t+1$ as a consequence of Lemma 5. Note that $|h_j(t_0)| \leq n-1$ for all $j \in \{1,\ldots,l\}$ with $l \geq 2$ at t_0.

Recall the neighbour set at agent i is given by $\mathcal{N}_i(t)$. Because the network is connected, for each $k \in h_j(t_0)$ the neighbour set obeys $\mathcal{N}_k(t_0) \neq \emptyset$ for each $j \in \{1,\ldots,l\}$. Then by Lemma 5 it follows that $h_j(t_0+1) \subset h_j(t_0)$ since at least one $k \in h_j(t_0)$ must be connected to an agent outside $h_j(t_0)$ and this agent's state must change $\mu_k(t_0) \neq \mu_k(t_0+1)$ as a consequence of Lemma 5 such that $f_{\ell_p,d}(\mu_k(t_0+1)) \neq \mathbf{x}_j$ (indeed $f_{\ell_p,d}(\mu_k(t_0+1))$ cannot even lie on the same convex hull). At the next time t_0+1 it holds again that for each $k \in h_j(t_0+1)$ (assuming $h_j(t_0+1) \neq \emptyset$) the neighbour set obeys $\mathcal{N}_k(t_0+1) \neq \emptyset$ for each $j \in \{1,\ldots,l\}$. Then by application of Lemma 5 it follows again that $h_j(t_0+2) \subset h_j(t_0+1) \subset h_j(t_0)$. Thus, $h_j(t+1) \subset h_j(t)$ is a strictly decreasing set-valued function unless $h_j(t) = \emptyset$. By at most time t_0+n-1 it follows that $h_j(t_0+n-1) = \emptyset$ and the argument can reset by redefining t_0. It thus follows that $f_{\ell_p,d}(\mathrm{co}(\{\mu_i(t_0+n-1)\})) \subset f_{\ell_p,d}(\mathrm{co}(\{\mu_i(t_0)\}))$ for all $t_0 \geq 0$. Following the proof of Lemma 3 we know $\mathrm{co}(\{\mu_i(t_0+n-1)\}) \subset \mathrm{co}(\{\mu_i(t_0)\})$ and thus because we chose t_0 arbitrarily $\nu(\mu(t+n-1)) < \nu(\mu(t))$ for each $t \in \mathbb{N}$ unless $\mu_i(t+n-1) = \mu_i(t)$, $\forall i$, as desired. The strictly decreasing Lyapunov-like function completes proof. □

The preceding lemma specialises the corollary to the case in which the network topology is connected and time-invariant (but otherwise arbitrary). This lemma is of interest on its own in many applications in which the network topology is static or otherwise changes very slowly. Proof of this lemma, given Lemmas 3-5, follows roughly the analysis of Moreau (2005) on nonlinear consensus in the usual Euclidean metric space.

Proof. (**of Corollary 1**) The proof here relies on extending the previous lemma to the case in which $\mathcal{G}(t)(\mathcal{V},\mathcal{E}(t))$ is time-varying and for all $t_0 \in \mathbb{N}$ the graph union $\mathfrak{G}(t_0,\infty)$ is connected. Recall the same Lyapunov function (5) as used in the proof of Lemma 6 (we assume familiarity with the proof of Lemma 6 going forward).

We note that it suffices to show that there is a countably infinite number of finite time intervals $t \in [t_0^q, \widehat{t}_0^q]$, $q \in \mathbb{N}$ such that $\nu(\mu(t_0^q + \widehat{t}_0^q)) < \nu(\mu(t_0^q))$.

Pick $t_0^q \geq 0$, $q \in \mathbb{N}$ so $f_{\ell_p,d}(\mathrm{co}(\{\mu_i(t_0^q)\}))$ is a l-sided polygon in $\mathbb{R}^2$ with $2 \leq l \leq |\mathcal{V}|$ on the collection of vertices $\{\mathbf{x}_j(t_0^q)\}$, $j \in \{1,\ldots,l\}$ with $\mathbf{x}_j(t_0^q) \in \mathbb{R}^2$. Recall (6). Then define a sequence of times $\{t_{s(j)}^q\}$, $s(j) \in \mathbb{N}$ each greater than t_0^q for each $j \in \{1,\ldots,l\}$ with $l \geq 2$. The connectivity condition implies the existence of such a sequence for each j with the property that, if $h_j(t_{s(j)}^q) \neq \emptyset$, there exists a $k \in h_j(t_{s(j)}^q)$ that is connected to an agent outside $h_j(t_{s(j)}^q)$. Then, this agent's state must change $\mu_k(t_{s(j)}^q) \neq \mu_k(t_{s(j)}^q+1)$ as a consequence of Lemma 5 and $f_{\ell_p,d}(\mu_k(t_{s(j)}^q+1)) \neq \mathbf{x}_j(t_0^q)$. Then $h_j(t_{s(j)}^q+1) \subset h_j(t_{s(j)}^q)$ for all $j \in \{1,\ldots,l\}$ unless obviously $h_j(t_{s(j)}^q) = \emptyset$. As in the proof of Lemma 6 it holds that $s(j) \geq n-1$ implies $h_j(t_{s(j)+1}^q) = \emptyset$ for all j. Let $\widehat{t}_0^q = \min\{t \in \mathbb{N} : t > t_0^q,\ s(j) \geq n-1, \forall j\}$ and note then that the interval $t \in [t_0^q, \widehat{t}_0^q]$ is finite owing to the connectivity condition. Moreover, as in the proof of Lemma 6 one can then show that $\nu(\mu(t_0^q + \widehat{t}_0^q)) < \nu(\mu(t_0^q))$. Restart the argument by picking t_0^{q+1} to be equal or sufficiently close to $\widehat{t}_0^q$ and note that the connectivity condition then implies the number of such (finite) intervals $t \in [t_0^q, \widehat{t}_0^q]$ is countably infinite on $q \in \mathbb{N}$. We thus have a strictly decreasing Lyapunov function $\nu(\mu(t_0^q + \widehat{t}_0^q)) < \nu(\mu(t_0^q))$ on the sequence of finite intervals $t \in [t_0^q, \widehat{t}_0^q]$, $q \in \mathbb{N}$ and this completes proof. □

Finally, given Proposition 1 and Corollary 1, it is worth noting the following result.

Proposition 2. *Consider a group of agents $\mathcal{V}$ and network $\mathcal{G}(t)(\mathcal{V},\mathcal{E}(t))$. Suppose the initial state of each agent is $\mu_i(0) \in \mathfrak{U}_p$ and that each agent applies (3). Suppose for all $t_0 \in \mathbb{N}$ the graph union $\mathfrak{G}(t_0,\infty)$ is connected. Define $\overline{\mu} = \lim_{t\to\infty} \mu_i(t)$, $\forall i \in \mathcal{V}$. Then there exists some symmetric weight matrix $\overline{\mathbf{W}} = [\overline{w}_{ij}] \in \mathbb{R}^{n\times n}$ with $\overline{w}_{ij} \in (0,1)$ and $\sum_{j\in\mathcal{V}} \overline{w}_{ij} = 1$ for all i such that*

$$\overline{\mu} = \inf_{z\in\mathfrak{U}_p} \sum_{j\in\mathcal{V}} \overline{w}_{ij}\ell_p(z,\mu_j(0))^p, \quad \forall i \in \mathcal{V}$$

where we note that $\overline{\mathbf{W}}$ is not the weighting matrix $\mathbf{W}(t)$ associated with the network and the update (3) but it is solely dependent on the sequence $\mathbf{W}(t)$, $t \in \mathbb{N}$ and the initial measures $\mu_i(0) \in \mathfrak{U}_p$, $i \in \mathcal{V}$.

Proof of this proposition is straightforward given the actual convergence result is stated in Proposition 1 and Corollary 1. This result states that the common measure which all agent states converge to must be strictly within the convex hull of all initial agent measures in $\mathfrak{U}_p$.

An interesting open problem is how one can design (or at least restrict) the evolution of $\mathbf{W}(t)$, $t \in \mathbb{N}$ such that for a set of measures $\mu_i(0) \in \mathfrak{U}_p$, $i \in \mathcal{V}$ the final weighting matrix $\overline{\mathbf{W}}$ specifies a limit $\overline{\mu} = \lim_{t\to\infty} \mu_i(t)$, $\forall i \in \mathcal{V}$ that is optimal in some desired sense (e.g. minimum variance over all possible $\overline{\mathbf{W}}$ given $\mu_i(0) \in \mathfrak{U}_p$, $i \in \mathcal{V}$). Some performance criteria may be completely independent of the initial measures $\mu_i(0) \in \mathfrak{U}_p$, $i \in \mathcal{V}$; i.e. some criteria may be achievable for any arbitrary set of initial measures by simply restricting $\mathbf{W}(t)$, $t \in \mathbb{N}$. We conjecture that this *might* be possible if one wants to achieve $\overline{w}_{ij} = 1/n$, so that $\overline{\mu}$ is an equal distance in the Wasserstein metric from all initial agent measures $\mu_i(0) \in \mathfrak{U}_p$, $i \in \mathcal{V}$.

We have not explored the idea of designing/constraining the evolution of $\mathbf{W}(t)$ so as to achieve some desired consensus value. However, such work would prove valuable in distributed information fusion, where one wants to combine numerous conditional beliefs (i.e. measures) about some underlying event into a single belief that should somehow be 'better' than each individual belief. We do not go into defining the notion of 'better' here but note that one example may be a reduction in variance as one may often want to reduce the variance of one's belief on some underlying event by combining multiple, additional, sources of information concerning that event.

5. CONCLUDING REMARKS

Distributed consensus in the Wasserstein metric space of probability measures was introduced in this paper. It is shown that convergence of the individual agents' measures to a common measure value is guaranteed if a relatively weak network connectivity condition is satisfied. This common measure value that is achieved asymptotically at each agent is the one that is closest simultaneously to all the initial agent measure values in the sense that it will minimise some weighted sum of Wasserstein distances between it and all initial measures.

REFERENCES

Ajorlou, A., Momeni, A., and Aghdam, A. (2011). Sufficient conditions for the convergence of a class of nonlinear distributed consensus algorithms. *Automatica*, 47(3), 625–629.

Ambrosio, L., Gigli, N., and Savaré, G. (2005). *Gradient Flows in Metric Spaces and in the Space of Probability Measures.* Birkhauser Verlag, Basel, Switzerland.

Berger, R. (1981). A necessary and sufficient condition for reaching a consensus using DeGroot's method. *Journal of the American Statistical Association*, 76(374), 415–418.

Bertrand, J. and Kloeckner, B. (2012). A geometric study of Wasserstein spaces: Hadamard spaces. *Journal of Topology and Analysis*, 4(4), 515–542.

Borkar, V. and Varaiya, P. (1982). Asymptotic agreement in distributed estimation. *IEEE Transactions on Automatic Control*, 27(3), 650–655.

Bremaud, P. (1999). *Markov Chains: Gibbs Fields, Monte Carlo Simulation and Queues.* Springer-Verlag, New York, NY.

Bridson, M. and Haefliger, A. (1999). *Metric Spaces of Non-Positive Curvature.* Springer, Berlin, Germany.

Burago, D., Burago, Y., and Ivanov, S. (2001). *A Course in Metric Geometry.* American Mathematical Society, Providence, R.I.

Cao, M., Morse, A., and Anderson, B. (2008). Reaching a consensus in a dynamically changing environment: A graphical approach. *SIAM Journal on Control and Optimization*, 47(2), 575–600.

Carli, R., Chiuso, A., Schenato, L., and Zampieri, S. (2008). Distributed Kalman filtering based on consensus strategies. *IEEE Journal on Selected Areas in Communications*, 26(4), 622–633.

Cattivelli, F. and Sayed, A. (2010). Diffusion strategies for distributed Kalman filtering and smoothing. *IEEE Transactions on Automatic Control*, 55(9), 2069–2084.

Cortés, J. (2006). Finite-time convergent gradient flows with applications to network consensus. *Automatica*, 42(11), 1993–2000.

Cortés, J. (2008). Distributed algorithms for reaching consensus on general functions. *Automatica*, 44(3), 726–737.

DeGroot, M. (1974). Reaching a consensus. *Journal of the American Statistical Association*, 69(345), 118–121.

Doucet, A., de Freitas, N., and Gordon, N. (eds.) (2001). *Sequential Monte Carlo Methods in Practice.* Springer.

Givens, C. and Shortt, R. (1984). A class of Wasserstein metrics for probability distributions. *The Michigan Mathematical Journal*, 31(2), 231–240.

Grohs, P. (2012). Wolfowitz's theorem and convergence of consensus algorithms in hadamard spaces. Technical Report SAM Report 2012-27, ETH, Zurich, Switzerland.

Hong, Y., Hu, J., and Gao, L. (2006). Tracking control for multi-agent consensus with an active leader and variable topology. *Automatica*, 42(7), 1177–1182.

Hui, Q. and Haddad, W. (2008). Distributed nonlinear control algorithms for network consensus. *Automatica*, 44(9), 2375–2381.

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

Kloeckner, B. (2010). A geometric study of Wasserstein spaces: Euclidean spaces. *Annali della Scuola Normale Superiore di Pisa*, IX(2), 297–323.

Li, Z., Duan, Z., Chen, G., and Huang, L. (2010). Consensus of multiagent systems and synchronization of complex networks: A unified viewpoint. *IEEE Trans. on Circuits and Systems I*, 57(1), 213–224.

Liggins, M., Chong, C.Y., Kadar, I., Alford, M., Vannicola, V., and Thomopoulos, S. (1997). Distributed fusion architectures and algorithms for target tracking. *Proceedings of the IEEE*, 85(1), 95–107.

Mahler, R. (2007). *Statistical Multisource-Multitarget Information Fusion.* Artech House.

Matei, I. and Baras, J. (2010). The asymptotic consensus problem on convex metric spaces. In *Proc. of the 2nd IFAC Workshop on Distributed Estimation and Control in Networked Systems.* Annecy, France.

McCann, R. (1997). A convexity principle for interacting gases. *Advances in Mathematics*, 128(1), 153–179.

Moreau, L. (2005). Stability of multiagent systems with time-dependent communication links. *IEEE Transactions on Automatic Control*, 50(2), 169–182.

Olfati-Saber, R. (2007). Distributed kalman filtering for sensor networks. In *Proc. of the 46th IEEE Conference on Decision and Control*, 5492–5498. New Orleans, Louisiana.

Olfati-Saber, R., Fax, J., and Murray, R. (2007). Consensus and cooperation in networked multi-agent systems. *Proceedings of the IEEE*, 95(1), 215–223.

Olfati-Saber, R., Franco, E., Frazzoli, E., and Shamma, J. (2005). Belief consensus and distributed hypothesis testing in sensor networks. In *Proc. of the Workshop on Network Embedded Sensing and Control.* Notre Dame University, South Bend, Indiana.

Olfati-Saber, R. and Murray, R. (2004). Consensus problems in networks of agents with switching topology and time-delays. *IEEE Transactions on Automatic Control*, 49(9), 1520–1533.

Ren, W. and Beard, R. (2005). Consensus seeking in multiagent systems under dynamically changing interaction topologies. *IEEE Transactions on Automatic Control*, 50(5), 655–661.

Sarlette, A. and Sepulchre, R. (2009). Consensus optimization on manifolds. *SIAM J. on Control and Optimization*, 48(1), 56–76.

Seneta, E. (1973). *Non-Negative Matrices: An Introduction to Theory and Applications.* George Allen & Unwin, London, U.K.

Sepulchre, R. (2011). Consensus on nonlinear spaces. *Annual Reviews in Control*, 35(1), 56–64.

Spanos, D. and Murray, R. (2005). Distributed sensor fusion using dynamic consensus. In *Proc. of the 16th IFAC World Congress.* Prague, Czech Republic.

Strogatz, S. (2000). From Kuramoto to Crawford: Exploring the onset of synchronization in populations of coupled oscillators. *Physica D: Nonlinear Phenomena*, 143(1), 1–20.

Tsitsiklis, J. and Athans, M. (1984). Convergence and asymptotic agreement in distributed decision problems. *IEEE Transactions on Automatic Control*, 29(1), 42–50.

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

Villani, C. (2003). *Topics in Optimal Transportation.* American Mathematical Society.

Wang, X. and Hong, Y. (2010). Distributed finite-time χ-consensus algorithms for multi-agent systems with variable coupling topology. *Journal of Systems Science and Complexity*, 23(2), 209–218.

Wolfowitz, J. (1963). Products of indecomposable, aperiodic, stochastic matrices. *Proceedings of the American Mathematical Society*, 14(5), 733–737.

Xiao, L., Boyd, S., and Lall, S. (2005). A scheme for robust distributed sensor fusion based on average consensus. In *Proc. of the 4th International Symposium Information Processing in Sensor Networks*, 63–70. Los Angeles, California.

Yu, W., Chen, G., and Cao, M. (2011). Consensus in directed networks of agents with nonlinear dynamics. *IEEE Transactions on Automatic Control*, 56(6), 1436–1441.

Zhu, M. and Martínez, S. (2010). Discrete-time dynamic average consensus. *Automatica*, 46(2), 322–329.