New Convergence and Exact Performance Results for Linear Consensus Algorithms Using Relative Entropy and Lossless Passivity Properties

Herbert Mangesius

***Abstract*— Despite the importance of the linear consensus algorithm for networked systems, yet, there is no agreement on the intrinsic mathematical structure that supports the observed exponential averaging behavior among n agents for any initial condition. Here we add to this discussion in linear consensus theory by introducing relative entropy as a novel Lyapunov function. We show that the configuration space of consensus systems is isometrically embedded into a statistical manifold. On projective n-1-space relative entropy is a common time-invariant Lyapunov function along solutions of the time-varying algorithm. For cases of scaled symmetry of the update law, we expose a gradient flow structure underlying the dynamics that evolve relative entropy in a steepest descent gradient scheme. On that basis we provide exact performance rates and upper bounds based on spectral properties of the update law governing the behavior on the statistical manifold. The condition of scaled symmetry allows to exhibit gradient flow structures for cases where the original update law is neither doubly stochastic, nor self-adjoint. The results related to the gradient flow structure are obtained by exploiting lossless passivity properties. We show that lossless passivity of a dynamical system implies a gradient flow structure on a manifold and vice versa. Exploiting lossless passivity amounts to constructing the combination of dissipation (pseudo)metric with Lyapunov function.**

I. INTRODUCTION

In recent years it has been seen that the linear consensus algorithm is a core element in the theory of interconnected systems. Applications range from distributed computation, and information diffusion, to the analysis of chemical reaction networks, and novel methods for the analysis of electric power system dynamics. Despite its prominent role in networked systems theory and applications, yet, there is no agreement on the intrinsic mathematical structure that produces the observed averaging behavior among n scalar subsystems.

With this paper we contribute to ongoing discussions by using a geometric and information theoretic approach to linear consensus systems as contraction mappings. This highlights the nonlinear aspects despite the linearity of the dynamics. We show that a natural configuration space for linear consensus algorithms is a statistical manifold up to a constant scaling that depends on the initial condition. By relating lossless passivity with gradient flows on a manifold, we establish a gradient flow property for linear consensus systems evolving on the space of discrete probability densities, based on relative entropy as Lyapunov function. This Lyapunov function bears the intuitive notion of each subsystem forgetting its specific initial condition over time. Moreover, on projective n-1-space relative entropy provides a common, time-invariant Lyapunov function for the time-varying algorithm.

Our results are novel contributions to linear consensus theory as follows. First, relative entropy is established as time-varying, and common time-invariant Lyapunov function. In its second role it is a dissipation measure that bases on the information of all n state values, rather than the two extremal state values used in the most general Lyapunov function known so far, which is set up by the diameter of the convex hull of all states. Second, we establish a gradient flow property for situations where the linear update law of consensus algorithm is not a doubly stochastic or symmetric matrix, and for that situation we give novel convergence bounds. This yields a direct extension of a Theorem given in [1] on the convergence under time-varying communication structures. Third, we verify a conjecture given in [2] that the underlying state space may be not linear. In this context we build relations to Markov chains.

Notation: The notation grad denotes gradient, and it corresponds to ∇, (containing partial derivatives), only in Euclidean space.

II. LINEAR CONSENSUS SYSTEMS - BASICS & SETTING

A. Representation and General Stability Result

The classical linear consensus algorithm is a time-varying system of n agents embedded in a directed graph $\mathcal{G}(t) = (\mathcal{V}, \mathcal{E}(t))$ with $t \geq 0$ denoting time. The embedding is such that each node $i \in \mathcal{V} = \{1, \dots, n\}$ indexes one agent, and whenever an ordered pair (i, j) is an element of the set of edges $\mathcal{E}(t) \subseteq \mathcal{V} \times \mathcal{V}$ there is a communication link from agent j to agent i at time t.

The algorithm is initialized by assigning a real scalar to each agent's state $x_i(t = 0)$. Then, the system evolves by communicating scalar state values $x_j(t)$ to agents i, whenever $(i, j) \in \mathcal{E}(t)$, and performing the update

$$x_i(t+1) = \sum_{j=1}^{n} a_{ij}(t) x_j(t), \quad i = 1, \dots, n, \qquad (1)$$

where $a_{ij}(t) \geq \alpha > 0$, ($\alpha$ some threshold value), if $(i, j) \in \mathcal{E}(t)$ and $\sum_{j=1}^{n} a_{ij} = 1, \forall j \in \mathcal{V}$; else, if $(i, j) \notin \mathcal{E}(t)$ we have $a_{ij}(t) = 0$. The local update law (1) yields the time-varying system $\boldsymbol{x}(t+1) = \boldsymbol{A}(t)\boldsymbol{x}(t)$, with $\boldsymbol{x}(t) = (x_1(t), \cdots, x_n(t))' \in \mathbb{R}^n$, and $\boldsymbol{A}(t)$ being a (right-) stochastic matrix of dimension $n \times n$.

H. Mangesius is with the Institute for Advanced Study (IAS), Technische Universität München (TUM), Lichtenbergstrasse 2a, D-85748 Garching. mangesius@tum.de, With the support of the TUM-IAS, funded by the German Excellence Initiative.

It is well-known that linear consensus algorithms converge under the mild assumption of uniform connectedness, see for instance [3], [4].

Definition 1 (Uniform connectedness): For a given time interval $[t_1, t_1+t_2]$, $t_1 \in [0,\infty)$, $t_1 \le t_2 < \infty$, consider the matrix $\bar{\boldsymbol{A}} := \sum_{t=t_1}^{t_1+t_2} \boldsymbol{A}(t)$ with associated graph $\mathscr{G}_{[t_1,t_1+t_2]} := (\mathcal{V}, \cup_{t\in[t_1,t_1+t_2]}\mathcal{E}(t))$. The time-varying graph $\mathcal{G}(t)$ is said to be uniformly connected over bounded intercommunication intervals, if for any t_1 there exists a bounded interval $[t_1, t_1+t_2]$ and at least one node $k \in \mathcal{V}$ such that $\mathscr{G}_{[t_1,t_1+t_2]}$ contains a directed path satisfying $\bar{a}_{ij} > \alpha$ from k to any other node $i \in \mathcal{V}$.

Theorem 1 (General convergence result [3], [4]): Consider the linear time-varying system (1). The consensus set $\mathcal{C} := \{c^\infty \mathbf{1}, c^\infty \in \mathbb{R}\}$ is uniformly exponentially stable if and only if $\mathcal{G}(t)$ is uniformly connected. In particular, for any initial condition $\boldsymbol{x}(t=0) \in \mathbb{R}^n$, solutions of (1) asymptotically convergence to a common consensus value $c^\infty = x_1(t\to\infty) = \cdots = x_n(t\to\infty)$.

This observation relies on the fact that over sufficiently large intercommunication intervals the product of matrices $\boldsymbol{\Phi} := \boldsymbol{A}(t_1+t_2)\circ\cdots\circ\boldsymbol{A}(t_1)$ is again a stochastic matrix and a contraction map, see [3], [4]; that is, for any $\boldsymbol{x},\boldsymbol{y} \in \mathcal{U} \subseteq \mathbb{R}^n$, $\mathcal{U} \not\subseteq \mathcal{C}$ application of $\boldsymbol{\Phi}$ yields

$$d(\boldsymbol{\Phi x}, \boldsymbol{\Phi y}) \le \lambda d(\boldsymbol{x},\boldsymbol{y}), \quad \lambda < 1 \tag{2}$$

where $d(\cdot,\cdot)$ is a metric function on $\mathcal{U}$, see [5] Ch. 1.

Associated to the discrete-time algorithm (1) is the differential algorithm in continuous time

$$\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{x}(t) = -\boldsymbol{L}(t)\boldsymbol{x}(t), \quad \text{with} \quad \boldsymbol{A} \equiv e^{-\boldsymbol{L}}. \tag{3}$$

The negative semidefinit $n\times n$ matrix $-\boldsymbol{L}(\mathcal{G}(t))$ generates $\boldsymbol{A}(\mathcal{G}(t))$ infinitesimally, and has the property of vanishing row sums, i.e. $\sum_{j\in\mathcal{V}} l_{ij} = 0$, in particular, $l_{ii} + \sum_{j\neq i} l_{ij} = 0,\ i \in \mathcal{V}$. In accordance to (1) a similar threshold value needs to be assumed for the weights l_{ij}.

B. Lyapunov Functions & Performance (In-)Equalities

Uniform convergence of algorithm (1) is classically established by means of the time-invariant Lyapunov function

$$V_T := \max_{i\in\mathcal{V}} x_i(t) - \min_{i\in\mathcal{V}} x_i(t), \tag{4}$$

which is nonincreasing along solutions of (1) and strictly decreasing, (i.e. it is a Lyapunov function), over sufficiently large, finite time intervals satisfying uniform connectedness. This observation has been made in [6], and in [4] this set-valued Lyapunov function is related to the contracting diameter of the convex hull $\mathrm{co}\{x_1(t),\cdots,x_n(t)\}$.

Despite its generality, convergence analysis using (4) suffers from the fact that it provides overly conservative estimates for the convergence speed, as stated for instance in [7]. Alternatively, it is well-known from linear systems theory, that the existence of a (common) quadratic Lyapunov function yields exponential convergence speed, with constant rate usually being related to spectral properties of the system.

Theorem 2 (Quadratic behavior, [1]): Consider the system (1) under the assumption that $\mathcal{G}(t)$ is strongly connected and balanced at each time instant. Define $\boldsymbol{e}(t) := \boldsymbol{x}(t) - c^\infty \mathbf{1}$, $c^\infty = \mathbf{1}'\boldsymbol{x}(t=0)$. The quadratic function $V_Q := \frac{1}{2}||\boldsymbol{e}(t)||^2$ is a common Lyapunov function. It satisfies the inequality

$$\frac{\mathrm{d}}{\mathrm{d}t}V_Q = -\boldsymbol{e}(t)'\left(\boldsymbol{L}(t)+\boldsymbol{L}'(t)\right)(\mathcal{G}(t))\boldsymbol{e}(t) \le -\kappa_Q V_Q, \tag{5}$$

where the rate $-\kappa_Q$ denotes the second-largest eigenvalue that is smallest within all second-largest eigenvalues corresponding to matrices $-\left(\boldsymbol{L}(\mathcal{G}(t)) + \boldsymbol{L}'(\mathcal{G}(t))\right),\ t \ge 0$.

Clearly, in this situation we have exponential convergence with $V_Q(\boldsymbol{x}(t_2)) \le V_Q(\boldsymbol{x}(t_1))e^{-\kappa_Q(t_2-t_1)},\ 0 \le t_1 < t_2$. The result relies on the fact that each $\boldsymbol{L}(t)$ has $\mathbf{1}$ as valid left eigenvector associated to the zero eigenvalue, and by that one can show that c^∞ remains invariant along solutions.

For situations where $\boldsymbol{L}(t)$ is symmetric, the Lyapunov function $V_\pi := \frac{1}{2}\boldsymbol{x}^T\boldsymbol{L}(t)\boldsymbol{x}$ is a time-varying potential and (3) evolves as gradient-descent algorithm in this scalar potential field according to

$$\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{x}(t) = -\nabla V_\pi(\boldsymbol{x};t) \tag{6}$$

Recently, in [8], the Lyapunov function (4) has been derived in logarithmic coordinates. The authors proposed the Lyapunov function, (over sufficiently large time intervals),

$$V_B := \max_{i\in\mathcal{V}} \log x_i(t) - \min_{i\in\mathcal{V}} \log x_i(t) = \log\frac{\max_i x_i(t)}{\min_i x_i(t)}, \tag{7}$$

on the basis of G. Birkhoff's work [9] on positive mappings that are contractions on the interior of the positive orthant $\mathbb{K}_+ = \{\boldsymbol{x}\in\mathbb{R}^n, x_i > 0, \forall i \in \mathcal{V}\}$ relative to the projective Hilbert metric. In the context of (2), V_B measures the projective distance between rays $[\boldsymbol{x}] := \{c\boldsymbol{x}, c > 0\}$ to the ray of consensus states $[\mathbf{1}] := \{c\mathbf{1}, c>0\}$, what we denote by writing $V_B(\boldsymbol{x}||\mathbf{1})$. Thus, while (7) defined on $\mathbb{K}_+$ is a pseudometric, it is a metric on the projective n-1-space $\mathbb{P}_+^{n-1} := \mathbb{K}_+/\sim$, where $\sim$ defines the equivalence classes of rays $[\cdot]$. The function V_B also measures the diameter of $\mathrm{co}\{x_1(t),\cdots,x_n(t)\}$, see [8]. Furthermore, using [9], a performance measure is given in [8] in terms of the contraction ratio

$$\begin{aligned}\kappa_B :&= \inf\left\{\kappa : V_B(\boldsymbol{\Phi x}||\mathbf{1}) \le \kappa V_B(\boldsymbol{x}||\mathbf{1}),\ \forall \boldsymbol{x}\in\mathcal{K}^+\right\}\\ &= \tanh\left(\frac{1}{4}\Delta(\boldsymbol{\Phi})\right),\end{aligned} \tag{8}$$

where $\Delta(\boldsymbol{\Phi})$ is the projective diameter of $\boldsymbol{\Phi}$; it is defined as

$$\begin{aligned}\Delta(\boldsymbol{\Phi}) :&= \sup\left\{V_B(\boldsymbol{\Phi x}||\mathbf{1}),\ \boldsymbol{x}\in\mathcal{K}^+\right\}\\ &= \sup\left\{\log\left(\frac{\varphi_{ij}\varphi_{pq}}{\varphi_{iq}\varphi_{pj}}\right),\ i,j,p,q\in\mathcal{V}\right\}.\end{aligned} \tag{9}$$

A performance measure for the convergence speed based on (8) can only be given when the projective diameter (9) takes finite values. This however requires irreducibility of $\boldsymbol{\Phi}$, cf. [10], otherwise $\Delta(\boldsymbol{\Phi}) = \infty$ and $\kappa_B = 1$. Thus, a finite diameter is only a sufficient condition to prove strict decay of V_B over finite time intervals and does not support the generality in Th. 1.

III. Contraction Behavior & Lossless Passivity - Novel Methodology

In the following we propose a method that allows to exploit passivity properties of lossless dissipative dynamical system by choice of a storage function, (that serves as Lyapunov function), and by construction of an appropriate (pseudo)metric that renders the autonomous system lossless. Such systems exhibit exponential convergence and evolve as gradient flow.

Our starting point is the dissipation inequality in integral and differential form

$$V(\boldsymbol{x}(t_1)) + \int_{t_1}^{t_2} w(\tau)\mathrm{d}\tau \geq V(\boldsymbol{x}(t_2)) \;\Leftrightarrow\; \dot{V}(\boldsymbol{x}(t)) \leq w(t) \tag{10}$$

with $0 \leq t_1 < t_2$, as presented in the framework of dissipative dynamical systems established in the work [11]. The sufficiently smooth function $V : \mathcal{X} \to \mathbb{R}_0^+$, $\mathcal{X} \subseteq \mathbb{R}^n$ the state space, is called storage, and the locally integrable function w, depending on external inputs and outputs, is called supply rate.

Remark 1: Here, as in [11], we share the general understanding of a dynamical system as being a semigroup. Hence, we do not restrict to a special class of dynamical systems.

A dynamical system for which a storage and supply can be found such that inequality (10) holds is called dissipative. Instances of dissipativeness are losslessness and passivity.

Definition 2 (Lossless passivity, [11], [12] Ch. 2): A dissipative dynamical system is called lossless if for all $0 \leq t_1 < t_2$ the equality

$$V(\boldsymbol{x}(t_1)) + \int_{t_1}^{t_2} w(\tau)\mathrm{d}\tau = V(\boldsymbol{x}(t_2)), \tag{11}$$

holds along any possible solution. A dissipative dynamical system is said to be passive if it has bilinear supply rate. If a dissipative dynamical system is both lossless and passive we say it has the property of lossless passivity.

In [11] it is already stated that any dissipative dynamical system with inputs and outputs can be made lossless w.r.t. an appropriate supply rate taking the form $\tilde{w}(t) = w(t) - d(t)$, where d is a nonnegative (internal) dissipation (rate) function. Appropriate means, that the internal dissipation function is known exactly.

Remark 2: While in [11] the choice $\tilde{w}(t) = w(t) + d(t)$ is made, here we subtract the dissipation rate, because internally dissipated energy is removed from the stored energy.

In the following we consider dissipative dynamical systems in the sense of its definition as autonomous systems. Hence, we have $w(t) = 0, \forall t$, (because external inputs and outputs are not present), and the remaining part of the supply rate $\tilde{w}$ relates to the internal dissipation. The original inequality together with the corresponding differential form become

$$V(\boldsymbol{x}(t_2)) - V(\boldsymbol{x}(t_1)) \leq -\int_{t_1}^{t_2} d(\tau)\mathrm{d}\tau \;\Leftrightarrow\; \dot{V}(t) \leq -d(t), \tag{12}$$

with equality whenever the system is lossless. Clearly, nonnegativity of the dissipation rate implies dissipativeness and positivity implies asymptotic stability with V being a Lyapunov function.

Lemma 1 (Gradient flows and lossless passivity): A dissipative dynamical system evolving on $\mathcal{X}$ has the property of lossless passivity if and only if there is at least a pseudometric given by an inner product function $g : T_{\boldsymbol{x}}\mathcal{X} \times T_{\boldsymbol{x}}\mathcal{X} \to \mathbb{R}_0^+$ such that the state dynamics and the storage have the gradient flow property

$$\dot{\boldsymbol{x}} = -\mathrm{grad}V(\boldsymbol{x}) \;\text{ and }\; \dot{V}(\boldsymbol{x}) = -g(\nabla V(\boldsymbol{x}), \nabla V(\boldsymbol{x})) \tag{13}$$

for any possible configuration $\boldsymbol{x} \in \mathcal{X}$.

Proof: [(13) $\Rightarrow$ Lossless passivity]: The choice $d(t) = g(\nabla V(\boldsymbol{x}), \nabla V(\boldsymbol{x}))$ yields relations (12) with equality, and hence losslessness by Def. 2. The gradient of $V(\boldsymbol{x})$, (on a possibly nonlinear space), is defined as the element in the tangent space $T_{\boldsymbol{x}}\mathcal{X}$ given by $\boldsymbol{G}_{\boldsymbol{x}}^{-1}\nabla V(\boldsymbol{x})$, where $\boldsymbol{G}_{\boldsymbol{x}} = \{g_{ij}(\boldsymbol{x})\}$, $i, j \in \mathcal{V}$, is a symmetric, positive semidefinit matrix valued function with inner products as elements, see for instance [13] Ch. 3 and [14]. We denote its inverse elements by g_{ij}^{-1}. Using this fact and the component inner product functions we can rewrite (13) such that

$$\begin{aligned}\dot{V} &= -\,g(\nabla V, \nabla V) = -||\nabla V||^2_{G_{\boldsymbol{x}}^{-1}} = -\nabla V' \boldsymbol{G}_{\boldsymbol{x}}^{-1} \nabla V \\ &= -\sum_{i=1}^{n}\sum_{j=1}^{n} g_{ij}^{-1}(\boldsymbol{x}) \frac{\partial V}{\partial x_i}\frac{\partial V}{\partial x_j} = -\sum_{i=1}^{n}\sum_{j=1}^{n} g_{ij}(\boldsymbol{x}) \dot{x}_i \dot{x}_j .\end{aligned} \tag{14}$$

The last equality comes from substitution using the identities $\sum_j g_{ij}^{-1} \nabla_{x_j} V = \dot{x}_i$ and $\nabla_{x_i} V = \sum_j g_{ij}\dot{x}_j$. By that, the dissipation rate $d = ||\nabla V||^2_{\boldsymbol{G}_{\boldsymbol{x}}^{-1}}$ is a metric having the bilinear form $\mathrm{d}d = \sum_i \sum_j g_{ij}(\boldsymbol{x})\mathrm{d}x_i \mathrm{d}x_j$, and hence passivity holds, cf. [5] Ch. 1. We only require a pseudometric, since any two solutions that are equivalent in the sense of having identical dynamics should be indistinguishable when measured in this metric.

[lossless passivity $\Rightarrow$ (13)]: follows from taking the reverse arguments: Passivity requires a bilinear form and losslessness a steepest descent gradient structure to obtain equality in (12). Thus, the dissipation rate needs to take the bilinear form $\mathrm{d}d$ being at least a pseudometric. ■

Theorem 3 (Performance and lossless passivity): Consider a dissipative dynamical system on $\mathcal{X}$ that satisfies lossless passivity, and a storage $V : \boldsymbol{x} \mapsto \mathbb{R}_0^+$, $\boldsymbol{x} \in \mathcal{X}$ with appropriate dissipation rate $d : T_{\boldsymbol{x}}\mathcal{X} \times T_{\boldsymbol{x}}\mathcal{X} \to \mathbb{R}_0^+$. For any $\boldsymbol{x} \in \mathcal{X}$ and $0 \leq t_1 < t_2$, the equality

$$V(\boldsymbol{x}(t_2)) = e^{-d(t_2 - t_1)} V(\boldsymbol{x}(t_1)) \tag{15}$$

holds. Denote by Γ the set of all possible solutions starting in a configuration $\boldsymbol{x}(t_1) \in \mathcal{X}$, and terminating in $\boldsymbol{x}(t_2) \in \mathcal{X}$. The dissipation rate d has the variational characterization

$$\int_{t_1}^{t_2} d(\tau)\mathrm{d}\tau = \sup_{\Gamma} \left\{ V(\boldsymbol{x}(t_1)) - V(\boldsymbol{x}(t_2)) \right\}, \tag{16}$$

$$\text{and } -d = \inf_{\Gamma} \left\{ \kappa : V(\boldsymbol{x}(t_2)) \leq e^{\kappa(t_2 - t_1)} V(\boldsymbol{x}(t_1)) \right\}, \tag{17}$$

and the square root of (16) is a (directed) distance (to steady state).

Proof: It is well-known that gradient flows have exponential convergence to equilibrium for any configuration, see for instance [5] Ch. 1. Then, the connection of lossless passivity with gradient flows as in Lem. 1 leads to (15).
The variational characterization of d is also an immediate consequence of the gradient flow property, see for instance [13] Ch. 3, because (16) corresponds to the squared length of the geodesic connecting $\boldsymbol{x}(t_1)$ and $\boldsymbol{x}(t_2)$. Then, (17) follows from noting that the dynamics evolve as differential steepest gradient descent algorithm of the storage V. The length of a geodesic is a distance, see for instance [13] Ch. 3, and the square root of the integral (16) is a directed distance, because the flow is directed in time. ■

Remark 3: The variational characterization in Thm. 3 is identical with the infimum characterization (8) when going to discrete time. The dissipation rate d being a constant according to (17) implies that the system evolves as constant speed geodesic, cf. [5] Ch. 5 and [14] Ch. 2 .

A method that exploits lossless passivity properties is to construct the combination of storage and dissipation "metric", e.g. by fixing a storage and trying to find, (for instance by reverse engineering), a suitable inner product g that brings forth the appropriate dissipation rate, (and by that a gradient flow structure).

IV. Dissipation in Linear Consensus Algorithms - Statement of Main Results

The main result establishes isometric and equivalence relations between the linear consensus algorithm (1), resp. (3), and the associated (time-varying) Markov chain with same update law, hereby denoted by Σ_M. In this context we consider the space of positive probability vectors

$$\mathcal{P} := \left\{ \boldsymbol{p} = (p_1, \cdots, p_n)' \in (0,1)^n : \sum_{i \in \mathcal{V}} p_i = 1 \right\}. \quad (18)$$

The evolution of any $\boldsymbol{p} \in \mathcal{P}$ under Σ_M is governed by the update law $\boldsymbol{p}'(t+1) = \boldsymbol{p}'(t)\boldsymbol{A}(t)$. A weaker representation of the state and its evolution can be accomplished in terms of probability density vectors. We define the space of positive probability densities as

$$\mathcal{M} := \left\{ \boldsymbol{\rho} = (\rho_1, \cdots, \rho_n)' \in \mathcal{K}^+ : \sum_{i \in \mathcal{V}} \nu_i \rho_i = 1 \right\}. \quad (19)$$

where $\nu : i \in \mathcal{V} \mapsto [0,1]$ is a reference probability function. A Lyapunov function for Σ_M is the relative entropy (Kullback-Leibler divergence) of $\boldsymbol{p}$ w.r.t. the the asymptotic probability vector $\boldsymbol{q}$, (this probability state is only reached when the update law does not switch, else it is time-varying), where $\boldsymbol{q}$ is the left Perron vector of $\boldsymbol{A}(t)$. The relative entropy of $\boldsymbol{p}$ w.r.t $\boldsymbol{q}$ is defined as

$$V_P(\boldsymbol{p}||\boldsymbol{q}) := \sum_{i \in \mathcal{V}} p_i \log \frac{p_i}{q_i}. \quad (20)$$

We consider linear consensus algorithms and make the following hypotheses:

(H1) Each graph $\mathcal{G}(t)$ is strongly connected.
(H2) The storage function is of strictly additive form such that $V(\boldsymbol{x}(t)) = \sum_{i \in \mathcal{V}} V_i(x_i)$.
(H3) The algorithm is defined on $\mathcal{X} = \mathbb{K}_+$.

Remark 4: Under (H1), (irreducibility of $\boldsymbol{A}$), uniqueness and positivity of $\boldsymbol{q}$ is guaranteed.

Remark 5: Within the context of interconnected, physical, and process systems (H2) is standard, see for instance [11], [15] Ch. 1, or [16] Ch 3.

Remark 6: Without loss of generality we can set (H3). That is, we can always find a constant $c \in \mathbb{R}$ such that the shifted state $\boldsymbol{x}(t) - c\mathbf{1}$ remains in $\mathbb{K}_+$ for all time. This does not alter the dynamics, because $c\mathbf{1}$ lies in the right kernel of $\boldsymbol{L}$, i.e. $\frac{\mathrm{d}}{\mathrm{d}t}(\boldsymbol{x} - c\mathbf{1}) = -\boldsymbol{L}\boldsymbol{x} + c\boldsymbol{L}\mathbf{1} = -\boldsymbol{L}\boldsymbol{x} = \frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{x}$.

Theorem 4 (Relative entropy & lossless passivity): Under (H1)-(H3), the following statements are true :

(i) The function

$$V_H(\boldsymbol{x}) := \sum_{i \in \mathcal{V}} q_i x_i \log x_i, \quad (21)$$

is a Lyapunov function along solutions of (1). Define $\boldsymbol{x}^s(t) := \frac{1}{c^\infty(t)}\boldsymbol{x}(t),\ c^\infty(t) := \boldsymbol{q}'(t)\boldsymbol{x}(t)$. Then, for all t inbetween two switching times $\boldsymbol{x}^s(t) \in \mathcal{M}$.

(ii) The space of solutions $\{\boldsymbol{x}(t)\}_{0<t<\infty}$ of (1) is isometrically embedded into the space of solutions $\{\boldsymbol{\rho}(t)\}_{0<t<\infty}$ of Σ_M. The isometric embedding is time-varying, and given by $\boldsymbol{\Pi}_{\mathcal{M}} : \mathcal{X} \to \mathcal{M}(\mathcal{X})$, with $\boldsymbol{x} \mapsto \boldsymbol{\Pi}_{\mathcal{M}}(\boldsymbol{x}) = \boldsymbol{x}/c^\infty(t)$. In particular, on $\mathbb{P}^{n-1}_+$, the relative entropy V_P is given by $V_H(\boldsymbol{x}(t)||c^\infty(t)\mathbf{1}) = V_M(\boldsymbol{\rho}(t)||\mathbf{1})$, and it is a time-invariant common Lyapunov function for all $t \in [0, \infty)$.

(iii) Within two switching times, if the time-invariant update law $-\boldsymbol{L}$, is scaled symmetric so that $q_i l_{ij} = q_j l_{ji},\ \forall i, j \in \mathcal{V}$, then $V_H(\boldsymbol{x}(t))$, together with $V_M(\boldsymbol{x}^s(t))$ and $V_P(\boldsymbol{p}(t))$ are gradient flows. In particular, we have the proportional evolution $V_H = c^\infty V_M$, and $\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{x}^s = \frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{p} = \boldsymbol{G}_{\boldsymbol{x}}^{-1}\nabla V_P(\boldsymbol{p})$. The matrices $\boldsymbol{G}_{\boldsymbol{x}}^{-1}$ define a pseudometric, and components are given by

$$g_{ij}^{-1} := \begin{cases} l_{ij}\frac{x_j^s - x_i^s}{\log x_j^s - \log x_i^s}, & \text{if } i \neq j \\ -\sum_{(i,j)\in\mathcal{E}(t)} l_{ij} q_i \frac{x_j^s - x_i^s}{\log x_j^s - \log x_i^s}, & \text{if } i = j \end{cases} \quad (22)$$

Define $\boldsymbol{e}^s(t) := \boldsymbol{x}^s - \mathbf{1}$. The function $V_Q^s := \frac{1}{2}||\boldsymbol{e}^s(t)||^2$ is a Lyapunov function, convergence is exponential, and

$$\frac{\mathrm{d}}{\mathrm{d}t}V_Q^s(t) = -\langle \boldsymbol{x}^s, \boldsymbol{Q}\boldsymbol{L}\boldsymbol{x}^s \rangle \leq \lambda_2(-\boldsymbol{Q}\boldsymbol{L})V_Q^s. \quad (23)$$

Remark 7: The result in part (iii) provides an immediate extension to Thm. 2 by allowing balanced weighted matrices that are scaled symmetric, so that the left-eigenvector differs from $\mathbf{1}$. Further, part (iii) provides an extension to the known gradient flow property in Euclidean space using V_π. All results can be relaxed to cases where $\mathcal{G}(t)$ is not always strongly connected, but over sufficiently large time intervals. Then, $\boldsymbol{\Phi}$ replaces $\boldsymbol{A}(t)$.

Remark 8: While V_T and V_B are measures of only two extremal state values, the function V_H uses all n state values. Observe that the logarithmic mean in g_{ij}^{-1} provides a means to convert between V_T and V_B. It is well-known that contraction rates for (relative) entropy measures are related to contractions of phase space volume, see for instance [17].

V. Proof of Theorem 4

A. *Proof of Part* (i)

The following Lemma bases on the work [18].

Lemma 2 (Extensive storage): Let $I \subseteq \mathbb{R}$ be an interval, and $f : I \to \mathbb{R}$ be a convex function. Consider the function

$$V(\boldsymbol{x}) = \sum_{i=1}^{n} V_i(x_i) = \sum_{i=1}^{n} \vartheta_i f(x_i), \quad \vartheta_i \geq 0, \tag{24}$$

The function (24) is nonincreasing and a storage along solutions of (1), if an only if

$$\sum_{i=1}^{n} \vartheta_i a_{ij}(t) = \vartheta_j, \quad j \in \mathcal{V}. \tag{25}$$

Proof: [V is storage $\Rightarrow$ (25)]: Define the output $\boldsymbol{y} := \boldsymbol{A}(t)\boldsymbol{x}, \boldsymbol{x} \in \mathcal{X}$. Then, dissipativity implies $V(\boldsymbol{y}) = \sum_{i=1}^{n} \vartheta_i f(y_i) \leq \sum_{i=1}^{n} \vartheta_i f(x_i) = V(\boldsymbol{x})$. This is the case, because

$$\begin{aligned}\sum_{i=1}^{n} \vartheta_i f(y_i) &= \sum_{i=1}^{n} \vartheta_i f(\sum_{j=1}^{n} a_{ij}(t) x_j) \\ &\leq \sum_{i=1}^{n} \sum_{j=1}^{n} \vartheta_i a_{ij}(t) f(x_j) \stackrel{(25)}{=} \sum_{j=1}^{n} \vartheta_j f(x_j).\end{aligned} \tag{26}$$

Here we used Jensen's inequality for convex functions on convex sets. Also, as a consequence of Jensen's theorem, the inequality (26) is strict, when f is strictly convex.

[V is storage $\Rightarrow$ (25)]: follows from taking any norm as extrem case of a convex function. Then, only equality can be obtained and this implies (25). ■

Corollary 1: Under $(H1)$ and $(H3)$ the function $V = \sum_{i=1}^{n} q_i x_i \log x_i$ is a Lyapunov function.

Proof: Condition (25) yields in vector matrix notation $\boldsymbol{\vartheta}'\boldsymbol{A}(t) = \boldsymbol{\vartheta}'$, which is precisely the definition of the first left eigenvector of the stochastic matrix $\boldsymbol{A}$, (that corresponds to the eigenvalue $\lambda = 1$). Since the left Perron vector is given as $\boldsymbol{q} \triangleq \boldsymbol{\vartheta}, ||\boldsymbol{\vartheta}||_1 = 1$, it is sufficient to set $\vartheta_i = q_i, i \in \mathcal{V}$. The function $f(x_i) = x_i \log x_i$ is strictly convex on the positive real line, and with $\boldsymbol{x} \in \mathbb{K}_+$, Lem. 2 implies Cor. 1. ■

Proposition 1: Under the assmptions and definitions made in Thm. 4, $c^\infty(t) > 0$, and $\frac{1}{c^\infty(t)} \sum_{i \in \mathcal{V}} q_i x_i = 1$.

Proof: The quantity $\sum_{i \in \mathcal{V}} q_i x_i(t) = \boldsymbol{q}'\boldsymbol{x}(t)$ is a system invariant along dynamics with time-invariant update law, i.e.

$$const. = \boldsymbol{q}'\boldsymbol{x}(t+1) = \boldsymbol{q}'\boldsymbol{A}(t)\boldsymbol{x}(t) = \boldsymbol{q}'\boldsymbol{x}(t) \triangleq c^\infty(t). \tag{27}$$

By definition and using (H1) we have $q_i, x_i > 0, \forall i \in \mathcal{V}$, cf. Rem. 4. This implies $c^\infty(t) > 0$. A scaling of $\boldsymbol{x}(t)$ with the inverse of $c^\infty(t)$ results in $\frac{1}{c^\infty(t)}\boldsymbol{q}'\boldsymbol{x}(t) = 1$. ■

Clearly, $\boldsymbol{x}^s = 1/c^\infty(t)\boldsymbol{x} \in \mathcal{M}$ by definition of $\mathcal{M}$. The scaling with $1/c^\infty(t)$ remains constant when the update law is time-invariant, due to the invariance of $c^\infty(t)$ along consensus dynamics. This completes the proof of part (i). ■

B. *Proof of Part* (ii)

Given two metric spaces $(\mathcal{S}_2, d_{S_2})$, $(\mathcal{S}_1, d_{S_1})$, and a function $s : \mathcal{S}_1 \to \mathcal{S}_2$, such that for all $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{S}_1$, the relation $d_{S_2}(s(\boldsymbol{x}), s(\boldsymbol{y})) = d_{S_1}(\boldsymbol{x}, \boldsymbol{y}))$ holds, then s is an isometry from $\mathcal{S}_1$ to $\mathcal{S}_2$.

Proposition 2: Define $V_M(\boldsymbol{\rho}||\mathbf{1}) := \sum_{i \in \mathcal{V}} q_i \rho_i \log \rho_i$. Then $V_P = V_M$, and along a trajectory produced by a time-invariant update law, the relative entropy V_M is a distance from $\boldsymbol{\rho}$ to the stationary density $\mathbf{1}$.

Proof: From definitions of $\mathcal{M}$ and $\mathcal{P}$ observe that $p_i = \rho_i q_i$. Substitution into relative entropy yields

$$V_P(\boldsymbol{p}||\boldsymbol{q}) = \sum_{i \in \mathcal{V}} p_i \log \frac{p_i}{q_i} = \sum_{i \in \mathcal{V}} q_i \rho_i \log \rho_i = V_M(\boldsymbol{\rho}||\mathbf{1}).$$

It is well-known that along a solution, the Kullback-Leibler divergence is positive, strictly decreasing and it vanishes if and only if $\boldsymbol{p} = \boldsymbol{q}$, or $\boldsymbol{\rho} = \mathbf{1}$. Thus, it induces a partial ordering among points lying on a trajectory along time, and this allows to use $V_M(\boldsymbol{\rho}||\mathbf{1})$ as a distance to equilibrium along a trajectory. ■

From Thm. 4 part (i) we know that any solution $\boldsymbol{x}(t), 0 \leq t < \infty$ is related to a density solution via time-varying scaling as $\boldsymbol{\rho}(t) = \frac{\boldsymbol{x}(t)}{c^\infty(t)}, 0 \leq t < \infty$.

Lemma 3: The time-varying function $\boldsymbol{\Pi}_{\mathcal{M}}$ as in Thm. 4 part (ii) is an isometric embedding from $\mathcal{X}$ to $\mathcal{M}$.

Proof: Set $(\mathcal{S}_2, d_{S_2}) = (\mathcal{M}, V_M)$, and fix $\boldsymbol{y} = \mathbf{1}$. According to Prop. 2, we can use $V_M(\boldsymbol{\rho}||\mathbf{1})$ as distance to equilibrium along a solution. Then, $V_H(\boldsymbol{x}(t)||c^\infty(t)\mathbf{1}) = V_M(\boldsymbol{\Pi}(\boldsymbol{x}), \boldsymbol{\Pi}(c^\infty(t)\mathbf{1})$, which is identical to $V_M(\boldsymbol{\rho}||\mathbf{1})$. ■

It remains to show that on $\mathbb{P}_+^{n-1}$, V_M is equivalent to V_H, and relative entropy becomes a common, time-invariant Lyapunov function.

By definition of $\mathbb{P}_+^{n-1}$, $\boldsymbol{x} \sim \frac{1}{c^\infty(t)}\boldsymbol{x} = \boldsymbol{\rho}$, and they are members of the ray $[\boldsymbol{x}] = \{\beta\boldsymbol{x},\ \beta > 0\}$. Accordingly, $c^\infty(t)\mathbf{1} \sim \mathbf{1}$ belong to the same equivalence class (ray) $[\mathbf{1}] = \{\beta\mathbf{1}, \beta > 0\}$. Thus, on projective n-1-space, functions $V_M(\boldsymbol{x}^s)$ and $V_H(\boldsymbol{x})$ are indistinguishable and therefore equivalent, since they are related by constant scaling of their arguments.

Lemma 4: Consider two reference vectors $\boldsymbol{q}^1, \boldsymbol{q}^2 \in \mathcal{P}$, with associated density vectors $\boldsymbol{\rho}^1, \boldsymbol{\rho}^2 \in \mathcal{M}$. Then, there exists a real scalar $\beta > 0$, such that $\sum_{i \in \mathcal{V}} q_i^2 \rho_i^2 = \sum_{i \in \mathcal{V}} q_i^2 \beta \rho_i^1 = 1$, and $V_M(\boldsymbol{\rho}^1)|_{\boldsymbol{q}^1} = V_M(\boldsymbol{\rho}^1\beta)|_{\boldsymbol{q}^2}$.

Proof: Set $\sum_{i \in \mathcal{V}} q_i^2 \rho_i^1 = \beta^{-1}$. Then, globally total probability must be preserved, i.e. $\boldsymbol{q}(t)'\boldsymbol{\rho}(t) = 1$ is an invariant, and the choice $\boldsymbol{\rho}^2 = \beta\boldsymbol{\rho}^1$ satisfies this conservation law. Further, because $\sum_{i \in \mathcal{V}} q_i^1 \rho_i^1 \log \rho^1 = \sum_{i \in \mathcal{V}} q_i^2 \rho_i^2 \log \rho^2$ the replacement $\rho_i^2 = \rho_i^1 \beta$ yields $V_M(\boldsymbol{\rho}^1)|_{\boldsymbol{q}^1} = V_M(\boldsymbol{\rho}^1\beta)|_{\boldsymbol{q}^2}$. ■

While $V_P(\boldsymbol{p}(t)||\boldsymbol{q}(t))$ is time-varying in both arguments, the function $V_M(\boldsymbol{\rho}(t)||\mathbf{1})$ is time-varying only in the first argument, and according to Lem. 4, this time-variance can be captured by scalings with a positive real number, without altering the value of the relative entropy in the system. Because a positively scaled density can always be chosen such that it belongs to the same equivalence class (ray) as the state, the relative entropy measured on $\mathbb{P}_+^{n-1}$ is a time-invariant quantity. This completes the proof of part (ii). ∎

C. *Proof of Part* (iii)

Remember that $x_i^s = \rho_i = p_i/q_i$. Further, componentwise we have $\nabla_{p_i} V_P(\boldsymbol{p}) = \log \frac{p_i}{q_i} + p_i \frac{q_i}{p_i}\frac{1}{q_i} = \log \frac{p_i}{q_i} + 1$. Denote by LM_{ij} the logarithmic mean between x_j^s and x_i^s. Then, we can write $\boldsymbol{G}_{\boldsymbol{x}}^{-1}\nabla V_P(\boldsymbol{p})$ in the i-th component as

$$\begin{aligned}
&\sum_{j\in\mathcal{V}} g_{ij}^{-1}\nabla_{p_j} V_P = \sum_{j\in\mathcal{V}} l_{ij} q_i \mathrm{LM}_{ij}(\log x_j^s + 1)\\
&= -\sum_{\substack{j\in\mathcal{V}\\ j\neq i}} l_{ij} q_i \mathrm{LM}_{ij}(\log x_i^s + 1) + \sum_{\substack{j\in\mathcal{V}\\ j\neq i}} l_{ij} q_i \mathrm{LM}_{ij}(\log x_j^s + 1)\\
&= \sum_{j\in\mathcal{V}, j\neq i} l_{ij} q_i \frac{x_j^s - x_i^s}{\log x_j^s - \log x_i^s}(\log x_j^s - \log x_i^s)\\
&= \sum_{j\in\mathcal{V}, j\neq i} l_{ij} q_i (x_j^s - x_i^s) \qquad (28)
\end{aligned}$$

Hence, we obtain in vector matrix form

$$\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{x}^s = -\boldsymbol{Q}\boldsymbol{L}\boldsymbol{x}^s = -\boldsymbol{L}\boldsymbol{Q}\boldsymbol{x}^s = -\boldsymbol{L}\boldsymbol{p} = \frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{p}. \qquad (29)$$

The matrix $\boldsymbol{G}_{\boldsymbol{x}}^{-1}$ is symmetric and positive semidefinite, because the logarithmic mean is positive and symmetric, and $\boldsymbol{Q}\boldsymbol{L}$ is symmetric. By that, $\boldsymbol{G}_{\boldsymbol{x}}^{-1}$ defines a pseudometric.

From Thm. 4 part (i) and (ii) we know that trajectories $\boldsymbol{x}(t)$ evolve proportional to $\boldsymbol{x}^s(t)$ and $V_P = V_M$ for all t. Clearly, $\boldsymbol{p}$ and V_P evolve as gradient flow, V_M is identical to V_P, and V_H evolves proportional to V_M because it is related via constant scaling $V_H = c^\infty V_M$, cf. Lem. 4. Since $\boldsymbol{x}^s$ corresponds to a density and is a gradient flow of relative entropy, it converges exponentially fast to $\mathbf{1}$, and thus, V_Q^s is a Lyapunov function. The dynamics are governed by the symmetric differential update law $-\boldsymbol{Q}\boldsymbol{L}$, which is a negative semidefinit matrix, and hence, the convergence rate is upper bounded by the second-largest eigenvalue of $-\boldsymbol{Q}\boldsymbol{L}$. This completes the proof of part (iii). ∎

VI. Conclusion

In this paper we show that under the assumption of strong connectedness the relative entropy of the associated time-varying Markov chain is a Lyapunov function for the time-varying consensus system. On projective n-1-space relative entropy is a common time-invariant Lyapunov function. We show that the configuration space of the linear consensus system is isometrically embedded into a statistical manifold. Moreover, under the assumption of scaled symmetry, the consensus system evolves isometrically to the gradient flow of relative entropy associated to the corresponding Markov chain. Scaled symmetry includes cases where the update law is not doubly stochastic or symmetric. Exact exponential convergence rates are given for the evolution of the consensus system considered on the statistical manifold. The results related to the gradient flow structure base on lossless passivity properties inherent to the consensus system. We show that lossless passivity implies a gradient flow structure on some nonlinear space and vice versa, and we provide a method to exploit this property by constructing the combination of a Lyapunov function with associated dissipation metric.

An upcoming work is in progress where we generalize the convergence results to classes of combinations that consist of storage functions with dissipation (pseudo)metrics.

VII. Acknowledgement

I would like to thank Professor Sanjoy K. Mitter for the stimulating discussions during his visits to Munich, that brought forth the main lines of thought. Further, I thank Prof. Sandra Hirche for her critical comments that have lead to an improved presentation.

References

[1] R. Olfati-Saber and R. M. Murray, "Consensus problems in networks of agents with switching topology and time-delays," *IEEE Transactions on Automatic Control*, vol. 49, p. 9, 2004.

[2] R. Sepulchre, "Consensus on nonlinear space," *Annual Reviews in Control*, vol. 35, pp. 56–64, 2011.

[3] L. Moreau, "Stability of continuous-time distributed consensus algorithms," in *43rd IEEE Conference on Decision and Control*, pp. 3998–4003, 2004.

[4] L. Moreau, "Stability of multiagent systems with time-dependent communication links," *IEEE Transactions on Automatic Control*, vol. 50, pp. 169–182, 2005.

[5] A. Katok and B. Hasselblatt, *Introduction to the Modern Theory of Dynamical Systems*. Cambridge University Press, 1999.

[6] J. N. Tsitsiklis, D. P. Bertsekas, and M. Athans, "Distributed asynchronous deterministic and stochastic gradient optimization algorithms," *IEEE Transactions on Automatic Control*, vol. 31, pp. 803–812, 1986.

[7] A. Olshevsky and J. N. Tsitsiklis, "On the nonexistence of quadratic lyapunov functions for consensus algorithms," *IEEE Transactions on Automatic Control*, vol. 53, pp. 2642–2645, 2008.

[8] R. Sepulchre, A. Sarlette, and P. Rouchon, "Consensus in non-commutative spaces," in *IEEE Conference on Decision and Control*, 2010.

[9] G. Birkhoff, "Extensions of jentzsch's theorem," *Transactions of the American Mathematical Society*, vol. 85, pp. 219–227, 1957.

[10] P. J. Bushell, "Hilbert's metric and positive contraction mappings in a banach space," *Archive for Rational Mechanics and Analysis*, vol. 52, no. 4, pp. 330–338, 1973.

[11] J. C. Willems, "Dissipative dynamical systems part 1: General theory," *Archive for Rational Mechanics and Analysis*, vol. 45, pp. 321– 351, 1972.

[12] R. Sepulchre, M. Jankovic, and P. Kokotovic, *Constructive Nonlinear Control*. Springer, 1996.

[13] P.-A. Absil, R. Mahony, and R. Sepulchre, *Optimization on Matrix Manifolds*. Princeton University Press, 2008.

[14] V. Trovimov and A. Fomenko, "Riemannian geometry," *Journal of Mathematical Sciences*, vol. 109, pp. 1345–1501, 2002.

[15] E. Cohen, T. Cvitas, J. Frey, B. Holmström, K. Kuchitsu, R. Marquardt, I. Mills, F. Pavese, M. Quack, J. Stohner, H. Strauss, M. Takami, and A. Thor, *Quantities, Units and Symbols in Physical Chemistry*. IUPAC Green Book, 3rd Edition, 2nd Printing, IUPAC & RSC Publishing, Cambridge, 2008.

[16] K. Hangos and I. Cameron, *Process Modelling and Model Analysis*, vol. 4 of *Process systems engineering*. Academic Press, 2001.

[17] D. Daems and G. Nicolis, "Entropy production and phase volume contraction," *Physical Review E*, vol. 59, pp. 4000–4006, 1999.

[18] J. C. Willems, "Lyapunov functions for diagonally dominant systems," *Automatica*, vol. 12, pp. 519–523, 1976.