Extending robustness and randomization from consensus to symmetrization algorithms

Luca Mazzarella, Francesco Ticozzi, Alain Sarlette

Extending robustness & randomization from Consensus to Symmetrization Algorithms

Luca Mazzarella* Francesco Ticozzi† Alain Sarlette‡

September 15, 2015

Abstract

This work interprets and generalizes consensus-type algorithms as switching dynamics leading to symmetrization of some vector variables with respect to the actions of a finite group. We show how the symmetrization framework we develop covers applications as diverse as consensus on probability distributions (either classical or quantum), uniform random state generation, and open-loop disturbance rejection by quantum dynamical decoupling. Robust convergence results are explicitly provided in a group-theoretic formulation, both for deterministic and for randomized dynamics. This indicates a way to directly extend the robustness and randomization properties of consensus-type algorithms to more fields of application.

1 Introduction

The investigation of randomized and robust algorithmic procedures has been a prominent development of applied mathematics and dynamical systems theory in the last decades [9, 23]. Among these, one of the most studied class of algorithms in the automatic control literature are those designed to reach *average consensus* [32, 26, 22, 5]: the most basic form entails a linear algorithm based on local iterations that reach agreement on a mean value among network nodes. Linear consensus, despite its simplicity, is used as a subtask for several distributed tasks like distributed estimation [36, 8], motion coordination [17], clock synchronization [6], optimization [24], and of course load balancing; an example is presented later in the paper, while more applications are covered in [26]. A universally appreciated feature of linear consensus is its robustness to parameter values and perfect behavior under time-varying network structure.

In the present paper, inspired by linear consensus, we present an abstract framework that produces linear procedures to solve a variety of (a priori apparently unrelated) *symmetrization* problems with respect to the action of finite groups. The main practical contribution of this unified framework is a systematic approach to prove effectiveness and robustness of a whole class of switching algorithms where iterations are associated to convex combinations of linear actions of a finite group. Our results prove asymptotic convergence to symmetrization by focusing only on the way the iteration

*Dipartimento di Ingegneria dell'Informazione, Università di Padova, via Gradenigo 6/B, 35131 Padova, Italy (mazzarella@dei.unipd.it).

†Dipartimento di Ingegneria dell'Informazione, Università di Padova, via Gradenigo 6/B, 35131 Padova, Italy and Department of Physics and Astronomy, Dartmouth College, 6127 Wilder Laboratory, Hanover, NH 03755, USA (ticozzi@dei.unipd.it).

‡INRIA Paris-Rocquencourt, QUANTIC project team, PSL Research University, France; and SYSTeMS, Ghent University, Technologiepark Zwijnaarde 914, 9052 Zwijnaarde(Gent), Belgium (alain.sarlette@inria.fr).

steps are selected, by studying a *lifted* dynamics. To this aim, only weak assumptions on the choice of possibly randomized actions applied at each iteration, and on the values of mixing parameters, are needed. Hence, the algorithms that converge in the proposed framework offer the same desirable features of linear consensus algorithms, including robustness and potential implementation in a randomized/unsupervised fashion.

In the second half of the paper, we show that our linear symmetrization framework covers a diversified set of previously proposed algorithms, and can suggest some new ones for suitable problems: the only requirement is that they can be recast as a symmetrization problem. This naturally includes only a subset, comprising linear consensus, of distributed algorithms while many other relevant ones, like belief propagation [28, 21], distributed pagerank [11], computations of other graph properties [35, 2], or various algorithms for distributed data fusion in sensor networks do not directly belong to this class. On the other hand, our framework does directly cover a set of tasks and procedures which do not even involve a distributed network, but just have a common group-theoretic structure with consensus. For instance, we show how our framework unveils the robustness of *quantum Dynamical Decoupling* (DD) [33] protocols which are used for open-loop disturbance rejection in quantum control. Circuits generating random states, or gates for quantum information processing, can also be viewed in this light. In fact, symmetric and invariant states are ubiquitous in classical and quantum physics, and symmetry-breaking or -preserving dynamics are sought for a variety of tasks. In particular, in quantum control, symmetries are known to be associated to uncontrollable sectors of the space [1] or to subsystems that are protected from noise [37, 16]; this seems to open the possibility for various future applications of our framework.

The paper is organized as follows. Section 2 outlines the main features of standard gossip consensus algorithms, that will serve as an inspirational and guiding example. Sections 3 and 4 develop our general framework, first relying on specific group actions and then moving to a general abstract framework. While the present paper mostly focuses on discrete-time dynamics, a natural continuous-time counterpart is introduced in Section 4.1, generalizing the idea first introduced in our paper [31] for a specific example. Section 5 proves convergence of general symmetrizing algorithms in deterministic and randomized settings. Finally, Section 6 presents a diverse set of problems and existing algorithms that are covered by our general framework, *and for which we can claim the same robustness features of gossip-type algorithms.* In the appendix, an alternative proof of convergence of the lifted dynamics using relative entropy is proposed.

Notation: Throughout the paper, we call a vector whose elements are nonnegative and sum to 1 a *vector of convex weights*. We denote by $|S|$ the cardinality of a set S (i.e. the number of elements it contains).

2 Guiding example: gossip iterations as randomized symmetrization

Consensus-type problems are formalized by assigning local agents (subsystems) to vertices $1, 2, ..., m \in V$ of a graph and association a state $x_k(t)$ to each vertex $k \in V$. The possibility of an interaction between agent pairs (j, k) at time t is modeled by the edges $E(t) \subset \{(j, k) : j, k \in V\}$ of the graph. We restrict ourselves to an undirected interaction graph, which identifies (j, k) with (k, j). The goal of *consensus algorithms* is, by iterating interactions between subsystems starting from an arbitrary initial state $x_1(0), x_2(0), ..., x_m(0)$, to reach a final state where $x_1 = x_2 = ... = x_m$ at a value that reflects a given function of the initial values, e.g. their mean.

There are many variants of consensus algorithms, and here as an example we consider *linear gossip* [5], with x_k belonging to $\mathbb{R}^n$ for $k = 1, 2, ..., m$. At each iteration, a single edge (j, k) is selected

from the set $E(t)$ of available edges at that time; the agents then update their state according to:

$$\begin{aligned} x_j(t+1) &= x_j(t) + \alpha(t)(x_k(t) - x_j(t)) \\ x_k(t+1) &= x_k(t) + \alpha(t)(x_j(t) - x_k(t)) \\ x_\ell(t+1) &= x_\ell(t) \quad \text{for all } \ell \notin \{j,k\}\,, \end{aligned} \tag{1}$$

where $\alpha(t) \in [\underline{\alpha}, \overline{\alpha}] \subset (0,1)$. If $\alpha = 1/2$, agents j and k move to the same point that is the average of their states. By iterating this rule, one hopes that all $x_j(t)$ asymptotically converge to the average of the $x_j(0)$.

The way in which the edges are selected over time leads to different evolutions for the whole system. We consider the following situations:

- *Cyclic interaction:* at each time t one link $(j(t), k(t))$ is selected deterministically by cycling through the elements of a time-invariant edge set E.
- *Random interaction:* at each time t one link $(j(t), k(t))$ is selected at random, $(j(t), k(t))$ being a single-valued random variable onto the edge set $E(t)$.

A well-known result in the consensus literature is that gossip iterations — both random and cyclic — lead to consensus under sufficient graph connectivity assumptions. In addition, gossip evolutions preserve the total average $\bar{x} = \frac{1}{m}\sum_{k=1}^{m} x_k$, so the state of each agent k converges to $x_k = \bar{x}(0) = \bar{x}(t)$ for all t.

Proposition 2.1. *[5, 22] If there exists some $B > 0$ (and $\delta > 0$) such that the union of edges selected during $[t, t+B]$ form a connected graph for all t (with probability $\geq \delta$), then iteration of (1) asymptotically leads to $x_k(t) = \bar{x}(0)$ for all k (with probability 1).*

Summing up, gossip iterations thus perform a distributed asynchronous computation of the mean, in a robust way with respect to the network size and structure and to parameter α, as long as the graph is not completely disconnected.

It is possible, however, to look at this gossip algorithm from another perspective. The evolution associated to (1) can be interpreted as a convex combination of two permutations, namely the trivial one (identity) and the transposition of the j and k state values:

$$\begin{aligned} \Big(x_j(t+1),\ x_k(t+1)\Big) &= (1\text{-}\alpha(t))\,\Big(x_j(t),\ x_k(t)\Big) + \alpha(t)\,\Big(x_k(t),\ x_j(t)\Big) \\ x_\ell(t+1) &= x_\ell(t) \quad \text{for all } \ell \notin \{j,k\}\,. \end{aligned} \tag{2}$$

Let $\mathfrak{P}$ denote the group of all permutations of the integers $1, 2, ..., m$ and for $\pi \in \mathfrak{P}$ let P_π be the unique linear operator such that $P_\pi\,(x_1, x_2, ..., x_m) = (x_{\pi(1)}, x_{\pi(2)}, ..., x_{\pi(m)})$ for any $x_1, x_2, ..., x_m$. It is easy to show that connectedness of a graph is equivalent to the property that the pairwise swaps associated to its edges generate the whole permutation group [10]. By using linearity of (1) and basic group properties, it is also possible to show that the evolution up to time t of the full state vector $x(t) = (x_1(t), \ldots, x_m(t))$ can always be written — although maybe not uniquely — as a convex combination of permutation operators on the initial states[1]:

$$x(t) = \sum_{\pi \in \mathfrak{P}} w_\pi(t) P_\pi\, x(0) \qquad \text{with} \quad w_\pi(t) \geq 0,\ \ \textstyle\sum_\pi w_\pi(t) = 1 \quad \forall t\,.$$

[1] This basic result will be proved in a more general setting later.

Any map of this form obviously preserves the average $\bar{x}(t)$. The reformulation in terms of permutations defines consensus as being any state in the set

$$C = \{x \in \mathcal{X} = \mathbb{R}^{mn} : \; P_\pi \, x = x \;\; \text{for all } \pi \in \mathfrak{P}\} \, . \tag{3}$$

Hence, consensus can be equivalently described as reaching a state that is invariant under (the action P_π on $\mathcal{X}$ of) any element of the permutation group.

We call this *symmetrization with respect to the permutation group.* In the next sections we develop a general framework to tackle symmetrization tasks by iterative, distributed algorithms. This allows for direct extension of the gossip consensus example to different state spaces, to networks that are more general than graphs, and to computational or control tasks not directly related to networks and consensus.

3 Symmetrization from group actions

This section presents the key definitions and algorithmic elements of finite-group symmetrization on vector spaces. In particular, linear gossip can be seen as a particular case of this class of symmetrizing iterations. Further examples are developed in Section 6.

3.1 Notation and Symmetrization Task

Let $\mathcal{G}$ be a finite group, with number of elements $|\mathcal{G}|$. Let $\mathcal{X}$ be a vector space over a field $\mathbb{R}$ or $\mathbb{C}$, endowed with an inner product $\langle \; , \; \rangle : \mathcal{X} \times \mathcal{X} \longrightarrow \mathbb{C}$.

We will consider a *linear action* of $\mathcal{G}$ on $\mathcal{X}$, that is a linear map $a : \mathcal{G} \times \mathcal{X} \to \mathcal{X}$ such that $a(hg, x) = a(h, a(g, x))$ and $a(\mathrm{e}_\mathcal{G}, x) = x$ for all $x \in \mathcal{X}$ and $g, h \in \mathcal{G}$, where $\mathrm{e}_\mathcal{G}$ is the identity of $\mathcal{G}$. Note that this implies among others $a(g^{-1}, a(g, x)) = x$. Although every linear action is associated to a representation[2] of $\mathcal{G}$ on $\mathcal{X}$, we maintain the action notation to make it directly applicable without re-parametrization, e.g. when considering the conjugate action of the unitary group on quantum operators. From the inner product, we can define the adjoint of $a(g, \cdot)$ as the unique operator $a^\dagger(g, \cdot)$ that satisfies: $\langle y \, , \, a(g, x) \rangle = \langle a^\dagger(g, y) \, , \, x \rangle \quad \forall \, x, y \, \in \mathcal{X}$.

An element $\bar{x} \in \mathcal{X}$ is a fixed point of the action of $\mathcal{G}$ if

$$a(g, \bar{x}) = \bar{x} \qquad \forall \, g \in \mathcal{G}. \tag{4}$$

We denote the set of such fixed points as $C^\mathcal{G} \subseteq \mathcal{X}$. Since the action is linear, $C^\mathcal{G}$ is a vector space. Our main goal is the *symmetrization* of any initial condition $x \in \mathcal{X}$ with respect to the action of $\mathcal{G}$, that is, construct an algorithm or a dynamical system that (asymptotically, with probability 1) drives any $x \in \mathcal{X}$ to some related $\bar{x} \in C^\mathcal{G}$.

Consider any time-varying discrete-time dynamics $x(t+1) = \mathcal{E}_t(\, x(t) \,)$ on $\mathcal{X}$. We denote $\mathcal{E}_{t,0}(\cdot)$ the map associated to the evolution from time 0 up to time t, such that $x(t) = \mathcal{E}_{t,0}(\, x(0) \,)$. Let $\| \cdot \|$ be a norm associated to the inner product in $\mathcal{X}$.

Definition 3.1. *The algorithm associated to iterations $\{\mathcal{E}_t\}_{t \geq 0}$ attains asymptotic symmetrization if for all $x \in \mathcal{X}$ it holds:*

$$\lim_{t \to \infty} \| a(g, \mathcal{E}_{t,0}(x)) - \mathcal{E}_{t,0}(x) \| = 0 \quad \forall \, g \in \mathcal{G}. \tag{5}$$

[2] Given a group $\mathcal{G}$, let $\mathcal{X}$ be a vector space and let us denote the set of bijective linear transformations on $\mathcal{X}$ as $\mathrm{GL}(\mathcal{X})$. A representation of $\mathcal{G}$ is an homomorphism from $\mathcal{G}$ to $\mathrm{GL}(\mathcal{X})$, i.e. a map $\gamma : \mathcal{G} \longrightarrow \mathrm{GL}(\mathcal{X})$ such that $\gamma(gh) = \gamma(g)\gamma(h) \;\; \forall \, g, h \in \mathcal{G}$.

We will also consider sequences of maps $\{\mathcal{E}_t\}_{t\geq 0}$ that can be randomized; in this case, the above definition applies but convergence with probability one is understood. Note that for finite-dimensional $\mathcal{X}$, by linearity this implies uniform convergence. For infinite-dimensional $\mathcal{X}$, it would indicate a weak type of convergence.

3.2 A Class of Algorithms

For a given group $\mathcal{G}$, vector space $\mathcal{X}$ and linear action $a : \mathcal{G} \times \mathcal{X} \to \mathcal{X}$, we will be interested in linear maps $\mathcal{F}$ of the form:

$$\mathcal{F}(x) \quad = \quad \sum_{g\in\mathcal{G}} \mathsf{s}_g \, a(g, x) \qquad \text{with} \quad \textstyle\sum_{g\in\mathcal{G}} \mathsf{s}_g = 1 \ , \ \mathsf{s}_g \geq 0 \ \forall g \, . \tag{6}$$

Such a map is completely specified by the choice of convex weights s_g. From here on, we shall call a vector whose elements are nonnegative and sum to 1 a *vector of convex weights*. We construct discrete-time dynamics on $\mathcal{X}$ by selecting at each time step t a vector of convex weights $\mathsf{s}(t) = (\mathsf{s}_{g_1}(t), \mathsf{s}_{g_2}(t), \ldots, \mathsf{s}_{g_{|\mathcal{G}|}}(t)) \in \mathbb{R}^{|\mathcal{G}|}$ and mapping $x(t)$ to $x(t+1)$ through the corresponding map of type $\mathcal{F}(x)$, i.e.

$$x(t+1) = \mathcal{E}_t(x(t)) := \sum_{g\in\mathcal{G}} \mathsf{s}_g(t) \, a(g, x(t)) \, . \tag{7}$$

We assume that $\mathsf{s}(t)$ is selected deterministically or randomly from some possibly infinite set $\mathcal{S}$. Typically any $\mathsf{s} \in \mathcal{S}$ assigns nonzero weights only to a restricted set of $g \in \mathcal{G}$. From a dynamical systems perspective, we can interpret (7) as a discrete-time *switching* system, whose generator is chosen at each time between a set of maps of the form (6), according to the switching signal $\mathsf{s}(t)$. The resulting $\mathcal{E}_{t,0}(\cdot)$ is also a convex combination of group actions, i.e. of the form $\mathcal{F}(\cdot)$ given in (6).

Lemma 3.1. *If the iterations have the form (7), then there exists a (possibly not unique) vector* $\mathsf{p}(t) = (\mathsf{p}_{g_1}(t), \mathsf{p}_{g_2}(t), \ldots, \mathsf{p}_{g_{|\mathcal{G}|}}(t)) \in \mathbb{R}^{|\mathcal{G}|}$ *such that for any t we can write:*

$$x(t) = \mathcal{E}_{t,0}(x(0)) = \sum_{g\in\mathcal{G}} \mathsf{p}_g(t) \, a(g, x(0)) \tag{8}$$

for any $x(0)$*, with*
- *at* $t = 0$, $\mathsf{p}_{e_\mathcal{G}}(0) = 1$ *and* $\mathsf{p}_g(0) = 0$ *for all* $g \neq 0$
- *for all t,* $\sum_{g\in\mathcal{G}} \mathsf{p}_g(t) = 1$ *and* $\mathsf{p}_g(t) \geq 0 \ \forall g$.

Proof. Proceed by inductive reasoning on t. For $t = 1$, (8) trivially holds because $\mathcal{E}_{1,0}(x) = \mathcal{E}_0(x)$ is given by (7). Now assume (8) holds for some t. Then

$$\begin{aligned}
\mathcal{E}_{t+1,0}(x) \quad &= \quad \mathcal{E}_t \circ \mathcal{E}_{t,0}(x) \\
\text{(def.}\mathcal{E}\text{)} \quad &= \quad \sum_{h\in\mathcal{G}} \mathsf{s}_h(t) a(h, \sum_{g\in\mathcal{G}} \mathsf{p}_g(t) a(g, x)) \\
\text{(linearity)} \quad &= \quad \sum_{h,g\in\mathcal{G}} \mathsf{s}_h(t)\mathsf{p}_g(t) \, a(h, a(g, x)) \\
\text{(def.action)} \quad &= \quad \sum_{h,g\in\mathcal{G}} \mathsf{s}_h(t)\mathsf{p}_g(t) \, a(hg, x)) \\
\text{(var.change)} \quad &= \quad \sum_{h,g'\in\mathcal{G}} \mathsf{s}_h(t)\mathsf{p}_{h^{-1}g'}(t) \, a(g', x)) \\
&= \quad \sum_{g'\in\mathcal{G}} \mathsf{p}_{g'}(t+1) \, a(g', x)) \quad ,
\end{aligned}$$

where we have defined $\mathsf{p}_{g'}(t+1) = \sum_{h\in\mathcal{G}} \mathsf{s}_h(t)\mathsf{p}_{h^{-1}g'}(t)$. Noting that $g' \mapsto h^{-1}g'$ is a group automorphism such that $\sum_{g'\in\mathcal{G}} \mathsf{p}_{h^{-1}g'}(t) = 1$ for each fixed h, one easily checks that $\mathsf{p}(t+1)$ satisfies the requirements of a vector of convex weights. Hence the statement holds for $t+1$ and we get the conclusion by induction. □

3.3 The symmetrizing map

A general time-varying map might achieve symmetrization according to (5) without ever converging to a fixed point. However, for dynamics of the form (7) we have the following result.

Proposition 3.1. *An evolution defined by $\mathcal{E}_t$ of the form (7) attains asymptotic symmetrization if and only if $\mathcal{E}_{t,0}(\cdot)$ converges to the fixed map*

$$\bar{\mathcal{F}}(\cdot) \;:=\; \frac{1}{|\mathcal{G}|}\sum_{g\in\mathcal{G}} a(g,\cdot) \tag{9}$$

pointwise for all $x \in \mathcal{X}$.

Proof. Assume symmetrization is attained. Taking the (finite) sum of (5) over all $g \in \mathcal{G}$, dividing by $|\mathcal{G}|$ and using the triangle inequality gives:

$$\begin{aligned}
0 \;&=\; \lim_{t\to+\infty} \left\| \tfrac{1}{|\mathcal{G}|} \sum_{g\in\mathcal{G}} a\left(g, \sum_{h\in\mathcal{G}} \mathsf{p}_h(t) a(h,x) \right) \;-\; \mathcal{E}_{t,0}(x) \right\| \\
\text{(linearity)} \;&=\; \lim_{t\to+\infty} \left\| \tfrac{1}{|\mathcal{G}|} \sum_{g\in\mathcal{G},\, h\in\mathcal{G}} \mathsf{p}_h(t)\, a\big(\, g,\, a(h,x)\,\big) \;-\; \mathcal{E}_{t,0}(x) \right\| \\
\text{(def.action)} \;&=\; \lim_{t\to+\infty} \left\| \tfrac{1}{|\mathcal{G}|} \sum_{g\in\mathcal{G},\, h\in\mathcal{G}} \mathsf{p}_h(t)\, a(gh,x) \;-\; \mathcal{E}_{t,0}(x) \right\| \\
\text{(var.change)} \;&=\; \lim_{t\to+\infty} \left\| \tfrac{1}{|\mathcal{G}|} \sum_{g\in\mathcal{G},\, h'\in\mathcal{G}} \mathsf{p}_{g^{-1}h'}(t)\, a(h',x) \;-\; \mathcal{E}_{t,0}(x) \right\| \qquad (10) \\
\text{(see below)} \;&=\; \lim_{t\to+\infty} \left\| \tfrac{1}{|\mathcal{G}|} \sum_{h'\in\mathcal{G}} a(h',x) \;-\; \mathcal{E}_{t,0}(x) \right\| \qquad (11)
\end{aligned}$$

for all $x \in \mathcal{X}$, which would imply that $\mathcal{E}_{t,0}$ converges to $\bar{\mathcal{F}}$. To go from (10) to (11), we sum on g for each fixed h': that yields $\sum_{g\in\mathcal{G}} \mathsf{p}_{g^{-1}h'}(t) = \sum_{g'\in\mathcal{G}} \mathsf{p}_{g'}(t) = 1$ for all h', thanks to the facts that $g \mapsto g^{-1}$, and $g \mapsto gh$ (for fixed h), are group automorphisms.

For the converse: Since both Definition 3.1 and the present Proposition 3.1 concern pointwise convergence, we can as well assume a fixed x and define $b_g = a(g,x) \in \mathcal{X}$ for all $g \in \mathcal{G}$, that is a finite number of points in $\mathcal{X}$. Then any action just maps a b_{g_1} to some other b_{g_2}, so the future evolution of the system can be restricted to the finite-dimensional linear subspace $\mathcal{B}$ of $\mathcal{X}$ spanned by the b_g. Then

we have, since $a(h, \bar{\mathcal{F}}(\cdot)) = \bar{F} i(\cdot)$ for all $h \in \mathcal{G}$ by definition, by linearity of the actions:

$$\begin{aligned} \|a(h, \mathcal{E}_{t,0}(x)) - \mathcal{E}_{t,0}(x)\| &= \|a(h,\ \mathcal{E}_{t,0}(x) - \bar{\mathcal{F}}(x)) + \bar{\mathcal{F}}(x) - \mathcal{E}_{t,0}(x)\| \\ &\leq \|a(h,\ \mathcal{E}_{t,0}(x) - \bar{\mathcal{F}}(x))\| + \|\bar{\mathcal{F}}(x) - \mathcal{E}_{t,0}(x)\| \\ &\leq (1 + \bar{b}(x))\, \|\bar{\mathcal{F}}(x) - \mathcal{E}_{t,0}(x)\| \end{aligned}$$

where $\bar{b}(x)$ is an upper bound on the norm of the linear operator resulting from the restriction of $a(g, .)$ to the finite-dimensional vector space $\mathcal{B}$. □

The proof builds on the finite cardinality of $\mathcal{G}$ and remains valid if $\mathcal{X}$ is infinite-dimensional. Notice however that if the actions associated to different $g \in \mathcal{G}$ are not all linearly independent, there will be more than one vector p corresponding to the same map $\bar{\mathcal{F}}$ (see the next section).

Lemma 3.2. *If there exists a group automorphism $g \mapsto h(g)$ such that*

$$a^\dagger(g, \cdot) = a(h(g), \cdot) \quad \forall g \in \mathcal{G}\,, \tag{12}$$

then $\bar{\mathcal{F}}$ is an orthogonal projection.

Proof. Eq. (9) readily yields that $\bar{\mathcal{F}} = \bar{\mathcal{F}}^2$ and that (12) ensures $\bar{\mathcal{F}} = \bar{\mathcal{F}}^\dagger$. □ Property (12) holds

e.g. for any action that is a unitary representation of $\mathcal{G}$. Another advantage of a self-adjoint actions set is that it allows to easily determine a set of preserved quantities, depending only on the initial $x(0)$, as is the case for the mean in the gossip example.

Lemma 3.3. *If there exists a map (not necessarily an automorphism) $g \in \mathcal{G} \mapsto h(g) \in \mathcal{G}$ such that (12) holds, then for any $\bar{z} \in C^{\mathcal{G}}$ we have*

$$\langle \bar{z}, x(t) \rangle = \langle \bar{z}, x(0) \rangle \quad \forall\, t\,. \tag{13}$$

Proof. For any t it holds that:

$$\begin{aligned} \langle \bar{z}, x(t) \rangle &= \langle \bar{z}\,, \sum_{g \in \mathcal{G}} \mathsf{p}_g(t)\, a(g, x_0) \rangle = \sum_{g \in \mathcal{G}} \mathsf{p}_g(t)\, \langle a^\dagger(g, \bar{z})\,,\ x_0 \rangle \\ &= \sum_{g \in \mathcal{G}} \mathsf{p}_g(t)\, \langle a(h(g), \bar{z})\,,\ x_0 \rangle = \sum_{g \in \mathcal{G}} \mathsf{p}_g(t)\ \langle \bar{z}, x_0 \rangle = \langle \bar{z}, x_0 \rangle\,. \end{aligned}$$

□

3.4 Example: linear gossip

Consider the gossip algorithm described in Section 2. To recast it in our framework, we choose $\mathcal{X} = \mathbb{R}^{mn}$ and $\mathcal{G} = \mathfrak{P}$ the group of all permutations of m elements. We can think of any $x \in \mathcal{X}$ as a column vector that stacks the n-dimensional state vectors of the m subsystems. With the linear permutation operator P_π defined Section 2, the action of the group is simply $a(\pi, x) = P_\pi x$. Notice that this action is self-adjoint. We have already established that consensus corresponds to the fixed points of this action, i.e. $C = C^{\mathfrak{P}}$. From Proposition 3.1 and Lemma 3.2 (with the trivial automorphism $h(g) = g$), the map $\bar{\mathcal{F}} = \frac{1}{m!} \sum_\pi P_\pi$ is the orthogonal projection onto the consensus set.

Next we turn to the evolution model. For linear gossip, the $m!$-dimensional vector $\mathsf{s}(t)$ has only two nonzero entries at any time: $(1-\alpha(t))$ on the component corresponding to the group identity, and $\alpha(t)$ associated to swapping j and k. If α and the graph with $|E|$ edges are constant, then $\mathsf{s}(t)$ can switch between $|E|$ values. Let P_{e} and $P_{(j,k)}$ denote the linear operators P_π that respectively implement the identity and the swapping of subsystems j and k. These can be represented as $nm \times nm$ matrices: $P_{\mathrm{e}} = I_{nm}$, the identity, and $P_{(j,k)} = Q_{(j,k)} \otimes I_n$, the Kronecker product between the identity on $\mathbb{R}^n$ and $Q_{(j,k)}$ the $m \times m$ matrix that swaps the coordinates j and k of a vector of length m. Then the elementary evolution step associated to the selection of edge (j,k) at time t writes:

$$x(t+1) \;=\; \sum_{\pi} \mathsf{s}_\pi(t)\, a(\pi, x(t)) \;=\; (1-\alpha(t))\, P_{\mathrm{e}} x(t) + \alpha(t)\, P_{(j,k)} x(t)\,.$$

Finally, let us look at preserved quantities. Denoting z_c the value on row c of vector $z \in \mathcal{X} = \mathbb{R}^{mn}$, the set $\mathcal{C} = \mathcal{C}^{\mathfrak{P}}$ consists of all $z \in \mathcal{X}$ such that $z_{jn-d+1} = z_{kn-d+1}$ for all subsystems $j,k \in \{1,2,...,m\}$ and all components $d \in \{1,2,...,n\}$. This vector space is spanned in particular by the vectors $z^d \in X$, $d = 1,2,...,n$, defined by:

$$z^d_{jn-d+1} = 1/m \text{ for all } j\,, \text{ other components } 0\,.$$

Hence by Lemma 3.3, we get as conserved quantities any linear functional of the form

$$\langle \bar{z}, x\rangle \;=\; \sum_{d=1}^{n} f_d \,\langle z^d, x\rangle \;=\; \sum_{d=1}^{n} f_d \;\; \mathrm{avg}(x)_d$$

with arbitrary $f_1, f_2, ..., f_n \in \mathbb{R}$, where $\mathrm{avg}(x)_d$ denotes the average of the d^{th} component of the subsystem states.

4 Action-independent dynamics

This section discusses *sufficient conditions* for obtaining symmetrization, that are *independent of the actions* but depend only on $\mathcal{G}$ and on the selected sequence of convex weights $\mathsf{s}(t)$ at each step. These conditions are also *necessary* if the particular actions associated to all elements of $\mathcal{G}$ are linearly independent. Since such actions exist for any finite group $\mathcal{G}$, the following conditions can be viewed as *necessary and sufficient for obtaining symmetrization on all possible actions associated to a given group dynamics*[3]. In other words, we ensure asymptotic symmetrization for a general group-based algorithm in the form (7) based only on the group properties and the selection rules for the convex vectors $\mathsf{s}(t)$, for *any* underlying vector spaces and action. This frees us from the need to prove convergence for each specific application. Section 6 provides a series of examples obtained by extending in this way the gossip-type algorithm.

More explicitly, Lemma 3.1 suggests that for studying the dynamics on $\mathcal{X}$ according to (7), it is sufficient to look at the evolution of the convex weights $\mathsf{p}(t)$. The proof of the Lemma proposes the dynamics

$$\mathsf{p}_g(t+1) \;=\; \sum_{h\in\mathcal{G}} \mathsf{s}_h(t)\mathsf{p}_{h^{-1}g}(t) \tag{14}$$

[3]One representation with linearly independent elements is the *regular representation*: take $\mathcal{X} = \mathbb{R}^{|\mathcal{G}|}$, index the vectors of the canonical basis of $\mathcal{X}$ by $\{\, v(g) \in \mathcal{X} : v(g)_h = \delta_{h,g} \;\; \forall\, g,h \in \mathcal{G}\,\}$ where $\delta_{g,h}$ is the Kronecker delta and define the linear action of $\mathcal{G}$ on $\mathcal{X}$ by $a(h, v(g)) = v(hg)$ for all $g,h \in \mathcal{G}$. To see that the actions associated to different $h \in \mathcal{G}$ are all linearly independent, it suffices to notice that $a(h, v(\mathrm{e}_{\mathcal{G}})) = v(h)$. This is essentially the representation used in (15).

for all $g \in \mathcal{G}$. If the group actions are linearly dependent, then several weights $\mathsf{s}(t)$ or $\mathsf{p}(t)$ can be associated to any map of the form $\mathcal{F}$ and clearly (14) is not the unique dynamics corresponding to (7). However, if we want to study (7) by focusing on the group properties, and prove convergence in a way that is valid for *all possible actions associated to the group*, then (14) is the unique lift of (7) that achieves this goal. In the current section we hence study the behavior of (14).

Again, let us choose an ordering of $\mathcal{G}$ and consider $\mathsf{p}(t)$, $\mathsf{s}(t)$ as column vectors in $\mathbb{R}^{|\mathcal{G}|}$, i.e. indices $g \in \mathcal{G}$ are identified with rows in the column vector. Then (14) becomes:

$$\mathsf{p}(t+1) \;=\; \left(\sum_{h\in\mathcal{G}} \mathsf{s}_h(t)\Pi_h\right)\mathsf{p}(t) \;=\; \tilde{M}(t)\mathsf{p}(t) \;=\; \left(\prod_{i=0}^{t} \tilde{M}(i)\right)\mathsf{p}(0)\,, \tag{15}$$

where we define $\tilde{M}(t) = \sum_{h\in\mathcal{G}} \mathsf{s}_h(t)\Pi_h$, and $\Pi_h \in \mathbb{R}^{|\mathcal{G}|\times|\mathcal{G}|}$ denotes the unique permutation matrix such that, for any $\mathsf{p} \in \mathbb{R}^{|\mathcal{G}|}$ and $\mathsf{q} = \Pi_h\mathsf{p}$, we have $\mathsf{p}_g = \mathsf{q}_{(hg)}$. For each given sequence $\mathsf{s}(0), \mathsf{s}(1), ...$, equation (15) looks like the transition dynamics of a (time-inhomogeneous) *Markov chain* on the distribution $\mathsf{p}(t)$ over $\mathcal{G}$, in the sense that the corresponding $\tilde{M}(t)$ are a sequence of doubly stochastic matrices. In fact, since $(\Pi_h\,\mathsf{p})_g = \mathsf{p}_{h^{-1}g}$, $\tilde{M}(t)$ implements the (group) convolution of $p(t)$ with $s(t)$.

Definition 3.1 is satisfied independently of the particular actions associated to $\mathcal{G}$ if we can ensure convergence to a vector p such that:

$$\mathsf{p}_g = \mathsf{p}_{h^{-1}g} \quad \forall\, g, h \in \mathcal{G}\,. \tag{16}$$

Since for g fixed $\{h^{-1}g : h \in \mathcal{G}\} = \mathcal{G}$, this is consistently equivalent to

$$\mathsf{p}_g = 1/|\mathcal{G}| =: \hat{\mathsf{p}}_g \;\; \forall\, g \in \mathcal{G}\,, \tag{17}$$

in accordance with Proposition 3.1. To attain symmetrization, we thus require that the dynamics of p converges to the unique value $\mathsf{p} = \hat{\mathsf{p}}$ given by (17).

The targeted convergence to a uniform distribution $\hat{\mathsf{p}}$ under switched dynamics (15) with doubly stochastic transition matrix $\tilde{M}$, is reminiscent of the standard average consensus problem between $|\mathcal{G}|$ agents in $\mathbb{R}$. There are however at least two major differences between these frameworks.

1. The state $\mathsf{p}(t)$ models $\mathcal{E}_{t,0}$ from the original problem. In particular, $\mathsf{p}(0)$ models $\mathcal{E}_{0,0}$ which is the identity. Hence, in principle, we would only need to study the evolution from this *known initial state.*

2. The transition matrix has a different structure inherited from its constituents. For average consensus the transition matrix is essentially the identity plus a sum of symmetric edge-interaction-matrices, with 4 nonzero entries of equal magnitude per edge of the graph. For p, it is a sum of permutation matrices, each of them with $|\mathcal{G}|$ nonzero entries.

The second point actually alleviates the first one: by group translation, convergence to $\hat{\mathsf{p}}$ from the particular initial condition $\mathsf{p}(0)$ corresponding to identity $\mathcal{E}_{0,0}$, implies convergence to $\hat{\mathsf{p}}$ from *any* initial convex weights vector $\mathsf{p}(0)$. The following section investigates when the system defined by (15) converges to symmetrization. The resemblance with classical consensus will guide us to derive convergence conditions, although they will have to be translated to match the $\mathsf{p}(t)$ and $\mathsf{s}(t)$ structure (see second point).

4.1 Associated continuous-time dynamics

A standard procedure to obtain continuous-time dynamics corresponding to the abstract symmetrization framework is to take infinitesimal steps of (15):

$$\mathsf{p}(t+dt) = (1-\beta dt)\mathsf{p}(t) + \beta dt \tilde{M}(t)\mathsf{p}(t)$$

and the limit for dt going to zero gives

$$\tfrac{d}{dt}\mathsf{p}(t) = -\beta \tilde{L}\mathsf{p}(t) \quad \text{with} \quad \tilde{L} = I_{|\mathcal{G}|} - \tilde{M}\ , \tag{18}$$

where $\beta > 0$ is just a scalar gain (i.e. it governs the continuous-time speed). The matrix $\tilde{L}$ in (18) is a Laplacian matrix for a balanced graph, as is standard in conventional average consensus, with all off-diagonal elements ≤ 0, all diagonal elements ≥ 0, and satisfying $\tilde{L}\hat{\mathsf{p}} = \tilde{L}^T\hat{\mathsf{p}} = 0$ i.e. symmetrization is a stationary solution.

The present paper shall focus on the discrete-time iteration (15). Similar convergence results for the continuous-time dynamics and discussions for a particular application can be found in our paper [31].

4.2 Example: $\mathsf{p}(t)$ for gossip consensus

Let us quickly formulate the gossip algorithm in the action-independent form. In Section 3.4, we illustrated how $x(t+1) = A(t)x(t)$, with

$$A(t) \quad = \quad (1-\alpha)\, I_{mn} + \alpha(t)\, P_{(j,k)}$$

when edge (j,k) is selected at time t. The doubly-stochastic $\tilde{M}(t) = (1-\alpha)I + \alpha\Pi_{(j,k)}$ describing the $\mathsf{p}(t)$ dynamics has dimensions $m! \times m!$ (independently of n), with two nonzero entries *on each row and column*: $\tilde{M}_{g,g} = (1-\alpha)$ and $\tilde{M}_{g,\pi_{(j,k)}g} = \alpha$ for all $g \in \mathcal{G}$. The corresponding continuous-time dynamics would have as nonzero entries $\tilde{L}_{g,g} = \alpha$ and $\tilde{L}_{g,\pi_{(j,k)}g} = -\alpha$ for all $g \in \mathcal{G}$, when the link (j,k) is active.

Convergence of the p-dynamics is not necessary for convergence of the linear gossip algorithm. Indeed, a dimension counting argument suffices to show that the corresponding actions of $\mathfrak{P}$ are not linearly independent for $m \geq 4$: the space of possible actions has dimension m^2 (consider $A(t) = I_n \otimes A_m(t)$ and count the number of entries in matrix $A_m(t)$), while there are $m!$ permutations and $m! > m^2$ for $m \geq 4$. This means that ensuring convergence of the switched $\tilde{M}$ dynamics for p is in principle more demanding than for the switched A for x. However, as we prove in the next section, convergence on p follows from the typical assumptions of consensus, and allows us to draw conclusions that are valid for all possible $\mathcal{X}$ and actions of $\mathfrak{P}$.

5 Convergence analysis

We now examine the convergence properties of (15) with a switching signal $\mathsf{s}(t)$. This reduces to analyzing an infinite product of doubly stochastic matrices $\tilde{M}(t)$. This problem has been investigated in much detail in other contexts, including standard linear consensus [32, 12, 27, 22]. Among others, [27] proposes a common quadratic Lyapunov function for all possible switchings, which shows that instability is not possible. The question is then, under which conditions is $\hat{\mathsf{p}}$ *asymptotically* stable. We first give convergence results for deterministic $\mathsf{s}(t)$. Their adaptation to a randomly selected $\mathsf{s}(t)$ is explained at the end of the section.

5.1 Formal conditions and convergence proof

In the context of consensus on graphs, a sufficient condition for convergence is given in terms of a requirement that the union of all edges that appear during a uniformly bounded time interval, must form a connected graph at all times (see e.g. [22]). This result could be applied to (15), if we view each group element as a node of a *Cayley graph* and draw the directed edges that correspond to the group translations Π_g with $\mathsf{s}_g(t) \geq \underline{\alpha} > 0$ at time t. The problem at hand however has more structure: an arbitrary adjacency matrix for a graph on N nodes has order N^2 parameters, while (15) shows that $\tilde{M}(t)$ is defined by $m! = N$ elements only — namely the vector $\mathsf{s}(t)$. In fact we can define a vector of convex weights $\mathsf{q}_g(t,T)$ such that the evolution from time t to time $t+T$ writes

$$\prod_{i=0}^{T-1} \tilde{M}(t+i) = \sum_{g\in\mathcal{G}} \mathsf{q}_g(t,T)\,\Pi_g\,. \tag{19}$$

This again involves only $m! = N$ elements $\mathsf{q}_g(t,T)$. We therefore give independent convergence proofs, in the hope to highlight the role of the assumptions in a way that is more natural in the group-theoretic framework. We next formulate a condition that essentially translates the connected-graph requirement (in fact rather its essential consequence, i.e. that the transition matrix from t to $t+T$ is primitive) into our framework.

Assumption 5.1. *Assume the sequence* $\mathsf{s}(t)$ *to be such that there exist some finite* $T, \delta > 0$*, such that for each time* t*:*

$$\mathsf{q}_g(t,T) > \delta \quad \forall g \in \mathcal{G}\,. \tag{20}$$

This assumption can be translated into properties of the transition matrices in (15). If $M(t) = M$ for each t, then the assumption is equivalent to M being primitive. In the general case, we request that each $\prod_{i=0}^{T-1} \tilde{M}(t+i)$ is primitive, with all entries at least δ.

Notice how Assumption 5.1 does not require that $\{g \in \mathcal{G} : \mathsf{s}_g(i) > \delta \text{ for some } i \in [t, t+T]\} = \mathcal{G}$. Thus a priori, the (combination of) available actions for all t may be restricted to a subset $\mathcal{S}$ of $\mathcal{G}$; a necessary condition for Assumption 5.1 to hold is then that $\mathcal{S}$ generates $\mathcal{G}$. This is similar to requiring that the union of edges appearing during a time interval T in the corresponding Cayley graph form a connected graph, but not necessarily the complete graph. We will further examine Assumption 5.1 in Section 5.2.

Now let us formally establish that Assumption 5.1 is a sufficient condition to ensure convergence to $\hat{\mathsf{p}}$.

Theorem 5.1. *For any switching sequence* $\mathsf{s}(t)$ *satisfying Assumption 5.1, the algorithm* (15) *makes any initial condition* $\mathsf{p}(0)$ *converge to the uniform vector* $\hat{\mathsf{p}}$*, elementwise with exponential convergence factor* $(1-|\mathcal{G}|\delta)^{1/T}$*. Furthermore, the Euclidean norm* $\|\mathsf{p}-\hat{\mathsf{p}}\|^2$ *is a Lyapunov function.*

Proof. We can uniformly bound the evolution of the entries of $\mathsf{p}(N\cdot T)$ for integers N and show that they converge to $1/|\mathcal{G}|$ at the announced rate.

Consider the sequences of numbers $y(k)$ and $x(k)$ given by:

$$y(k+1) = (1-|\mathcal{G}|\delta)y(k)+\delta \quad \text{with} \quad y(0)=0, \tag{21}$$
$$x(k+1) = (1-|\mathcal{G}|\delta)x(k)+\delta \quad \text{with} \quad x(0)=1, \tag{22}$$

or equivalently since $0 < \delta \leq 1/|\mathcal{G}|$ (the minimal entry of p cannot be larger than for the uniform distribution),

$$y(k) = \tfrac{1}{|\mathcal{G}|} - \tfrac{1}{|\mathcal{G}|}(1-|\mathcal{G}|\delta)^k\,, \quad x(k) = \tfrac{1}{|\mathcal{G}|} + \tfrac{|\mathcal{G}|-1}{|\mathcal{G}|}(1-|\mathcal{G}|\delta)^k\,.$$

These two sequences respectively increase / decrease monotonously and exponentially towards $1/|\mathcal{G}|$. Hence we conclude the first part of the proof by showing that

$$x(k) \geq \mathsf{p}_g(k \cdot T) \geq y(k) \tag{23}$$

for every integer k and every g. We do this by induction on k.

For $k = 0$ we have of course $x(0) = 1 \geq \mathsf{p}_g(0) \geq y(0) = 0$. Now assuming that the inequality holds for k, let us prove that it then holds for $k + 1$. For each $g \in \mathcal{G}$, we have

$$\begin{aligned} \mathsf{p}_g((k+1)T) &= \sum_h \mathsf{q}_h(t,T)\mathsf{p}_{h^{-1}g}(kT) = \delta \sum_h \mathsf{p}_{h^{-1}g}(kT) + \sum_h (\mathsf{q}_h(t,T) - \delta)\,\mathsf{p}_{h^{-1}g}(kT) \\ &= \delta + \sum_h (\mathsf{q}_h(t,T) - \delta)\,\mathsf{p}_{h^{-1}g}(kT) \end{aligned}$$

since $\sum_h \mathsf{p}_{h^{-1}g}(kT) = 1$ for each g. From the assumptions $\mathsf{p}_{h^{-1}g}(kT) \geq y(k)$ and $\mathsf{q}_h(t,T) > \delta$, and using $\sum_h \mathsf{q}_h(t,t') = 1$ for all t, t', we then get:

$$\mathsf{p}_g((k+1)T) \geq \delta + \sum_h (\mathsf{q}_h(t,T) - \delta)\, y(kT) \geq \delta + (1 - |\mathcal{G}|\delta) y(k) = y(k+1)\,.$$

An analog reasoning shows that $\mathsf{p}_g((k+1)T) \leq x(k+1)$.

The exponential convergence of the Euclidean norm for t being a multiple of T is a direct consequence of the exponential elementwise convergence. The fact that for *any* admissible switching sequence this Lyapunov function never increases between *any t and $t+1$*, is shown as follows. Denoting $\dagger$ the transpose of a vector or matrix and I an identity matrix of appropriate dimension, we have

$$\begin{aligned} \|\mathsf{p}(t+1) - \hat{\mathsf{p}}\|^2 &= (\mathsf{p}(t+1) - \hat{\mathsf{p}})^\dagger (\mathsf{p}(t+1) - \hat{\mathsf{p}}) \\ &= (\tilde{M}(t)\mathsf{p}(t) - \hat{\mathsf{p}})^\dagger (\tilde{M}(t)\mathsf{p}(t) - \hat{\mathsf{p}}) \\ &= \|\mathsf{p}(t) - \hat{\mathsf{p}}\|^2 + \mathsf{p}(t)^\dagger (\tilde{M}(t)^\dagger \tilde{M}(t) - I)\mathsf{p}(t)\,. \end{aligned}$$

by using $\tilde{M}(t)\hat{\mathsf{p}} = \hat{\mathsf{p}}$.
Since $\tilde{M}(t)^\dagger \tilde{M}(t)$ is doubly stochastic and symmetric, $(\tilde{M}(t)^\dagger \tilde{M}(t) - I)$ is negative semidefinite for any t. □

We observe (see appendix) that the relative entropy, or Kullback-Leibler pseudo-distance [7] between $\mathsf{p}(t)$ and $\hat{\mathsf{p}}$ can also be used as a Lyapunov function to show asymptotic convergence, although in that case it is not as direct to show that convergence is exponential.

As an immediate corollary, we have symmetrization on $\mathcal{X}$ with the associated actions, for *any $\mathcal{X}$, any linear group action and any $\mathsf{s}(t)$ satisfying Assumption 5.1*.

Corollary 5.1. *Any algorithm of the form (7) on a vector space $\mathcal{X}$ with $\mathsf{s}(t)$ satisfying Assumption 5.1, asymptotically converges to $\lim_{t\to+\infty} x(t) = \bar{\mathcal{F}}(x(0))$. The convergence is exponential and at least as fast as $(1 - |\mathcal{G}|\delta)^{t/T}$.*

If the actions associated to group elements are linearly dependent, as is the case for consensus, a faster convergence speed can be expected, since convergence at the group level, for the lifted p dynamics, is not necessary for convergence of the state.

5.2 Examining switching signals

Let us now provide some typical examples of switching signals $\mathsf{s}(t)$ and check if they satisfy Assumption 5.1. It is actually instructive to start by listing some cases that lead to a violation of the assumption.

- If (possibly after some initial transient) the vector $\mathsf{s}(t)$ contains a single nonzero entry at any time, then $\mathsf{q}(t, T)$ will also contain a single element.
- Consider that (after some initial transient) $\mathsf{s}_g(t)$ can be nonzero at any time only for $g \in \mathcal{S}$, a *subgroup* of $\mathcal{G}$. Then each $\tilde{M}(t)$ is a weighted sum of Π_g with $g \in \mathcal{S}$, and by subgroup properties the propagator $\prod_{i=0}^{t-1} \tilde{M}(i)$ is also a weighted sum of Π_g with g restricted to $\mathcal{S}$, such that we can have $q_g(t, T) \neq 0$ for at most all $g \in \mathcal{S}$.
- More generally, if $\mathsf{s}_g(t)$ can be nonzero at any time only for $g \in \mathcal{S}$, now being some subset of $\mathcal{G}$, and the elements of $\mathcal{S}$ do not generate the whole group, then Assumption 5.1 cannot hold.

Conversely, sufficient conditions for Assumption 5.1 to hold include the following.

- If there exists a set $\mathcal{J} \subset \mathcal{G}$ that generates $\mathcal{G}$ and such that for each t, there exists $i \in [t, t + T]$ such that $\mathcal{S}_i = \{g \in \mathcal{G} : \mathsf{s}_g(i) > \delta\}$ contains $\mathcal{J} \cup \{\mathrm{e}_{\mathcal{G}}\}$, then Assumption 5.1 is satisfied. We leave this simple proof to the reader.
- If $\mathcal{G}$ is Abelian, then the order in which the group elements are selected has no importance, but it is still relevant to know which ones are selected at the same time or not. Then we can use a reduced Cayley graph to investigate Assumption 5.1 as follows. For each time t, take the set $\mathcal{S}_t = \{g \in \mathcal{G} : \mathsf{s}_g(t) > \delta\}$, choose one $\bar{g}_t \in \mathcal{S}_t$ and let $\bar{\mathcal{S}}(t) = \{\bar{g}_t^{-1} g : g \in \mathcal{S} \setminus \{\bar{g}_t\}\}$. Then consider a starting time t_0 and recursively construct a graph as follows. Start with a single node $\mathrm{e}_{\mathcal{G}}$. At each step $i = 1, 2, ..., T$, add edges (and potentially vertices) to connect every vertex $h \in \mathcal{G}$ that is already present in the graph at step $i - 1$, with the set of nodes $\{s\, h : s \in \bar{\mathcal{S}}_{t_0+i}\}$. If for all t we have $\mathsf{s}_{\mathrm{e}_{\mathcal{G}}}(t) > \delta$, and for all t_0 the graph obtained at $i = T$ contains all the $g \in \mathcal{G}$, then Assumption 5.1 is satisfied.

5.3 Randomized Convergence

So far we have always formulated convergence properties for a given switching signal $\mathsf{s}(t)$. We now briefly indicate how they can be adapted when $\mathsf{s}(t)$ is selected at random. We thus consider that at each time t, $\mathsf{s}(t)$ is selected from a set $\mathfrak{S}$ according to some given probability distribution, independently of the $\mathsf{s}(i)$ for $i \neq t$. In other words, the $\mathsf{s}(t)$ are independent, not necessarily identically distributed, random variables over a set of vectors of convex weights. Then we get the following convergence result.

Theorem 5.2. *Assume that there exist some fixed values of T,δ, and $\varepsilon > 0$ for which the statement of Assumption 5.1 holds with probability at least ε at each time t. Then for any $\gamma > 0$, the probability of having an Euclidean distance $\|\mathsf{p}(t) - \hat{\mathsf{p}}\| < \gamma$ converges to 1 as t converges to $+\infty$.*

Proof. Assume that Assumption 5.1 holds for all times between t_0 and $t_0 + N_\gamma T$ for some $N_\gamma > 0$. Then we can apply Theorem 5.1 between t_0 and $t_0 + N_\gamma T$, and the resulting exponential convergence is guaranteed to reach $\|\mathsf{p}(t_0 + N_\gamma T) - \hat{\mathsf{p}}\| < \gamma$ for N_γ sufficiently large. (Note that the exponential convergence proof of Theorem 5.1, in particular the bounding by sequences ,, holds for any $\mathsf{p}(t_0)$.) Moreover, as proved at the end of Theorem 5.1, the Lyapunov function $\|\mathsf{p}(t) - \hat{\mathsf{p}}\|$ cannot increase

between t and $t+1$ under (15), for any vector of convex weights $\mathsf{s}(t)$. Hence we would also have $\|\mathsf{p}(t) - \hat{\mathsf{p}}\| < \gamma$ for any $t > t_0 + N_\gamma T$.

The proof is concluded by noting that, under the specified random choice of the signal $\mathsf{s}(t)$, the probability that a sequence of $B \cdot N_\gamma \cdot T$ elements contains no subsequence of $N_\gamma T$ consecutive elements satisfying Assumption 5.1, is at most $(1-\varepsilon^N)^B$. The latter converges to 0 as B goes to ∞, thus as t goes to ∞ for fixed γ,δ,T.

□

Let us briefly discuss some examples of randomized evolutions.

- If at each time, we randomly select a single element $h(t)$ from $\mathcal{G}$ with probability of $h(t) = g$ being greater than zero for all g, and take

$$\mathsf{s}_{h(t)}(t) = \alpha\,, \quad \mathsf{s}_{\mathrm{e}_\mathcal{G}}(t) = (1-\alpha)\,, \quad \mathsf{s}_g(t) = 0 \text{ for } g \notin \{h(t),\, \mathrm{e}_\mathcal{G}\}\,, \tag{24}$$

  then the requirements of Theorem 5.2 are clearly satisfied. Of course this situation directly generalizes to cases where more than one $h(t) \in \mathcal{G}$ is applied at each time.

- Like in the deterministic case, a similar result is obtained if in (24) we randomly select $h(t)$ from some subset $\mathcal{S}$ of $\mathcal{G}$, and this subset generates the whole group. The subset may also vary (e.g. cyclically) with time, as long as it allows with nonzero probability to construct one sequence satisfying Assumption 5.1. The linear gossip algorithm fits in this category, as the connected graph condition in Proposition 2.1 ensures that swaps of adjacent agents can be selected in a way that generates the whole group of permutations.

A few remarks are in order.

Remark 1 (Time-varying possibilities)**.** Theorem 5.2 only requires some uniform upper bound T on a time interval that guarantees that all group elements are associated with weights of at least $\delta > 0$. It thus allows for dynamics where $\mathsf{p}(t)$ does not evolve towards $\hat{\mathsf{p}}$ for shorter time intervals, as long as there is a nonzero probability to reduce the distance from $\hat{\mathsf{p}}$ in finite time. Therefore, we can ensure convergence if, for example, one strictly contractive evolution is applied only every T_0 steps, while we do not know how s_g is selected in between.

Remark 2 (Explicit robustness to α)**.** A major contribution of Theorem 5.2 is to establish the *robustness* of consensus-like algorithms with respect to uncertainties in the values of $\mathsf{s}_g(t)$ for a wide variety of applications (see Section 6). Indeed, if we consider that the $h \in \mathcal{S}$ for which $\mathsf{s}_h \neq 0$ are chosen deterministically, but the values $\mathsf{s}_h(t)$ are randomly chosen in some compact set strictly inside $[0,1]$ for all t, then Assumption 5.1 holds with given T either for all such sequences or for none; in the former case, compactness ensures that δ is bounded from below, and Theorem 5.2 holds. This shows that it is not important to control the exact proportions in which the chosen actions are applied. Typically in a gossip algorithm [5], one uses the maximally mixing value $\alpha = 1/2$. Nonetheless, convergence holds provided that $\alpha(t) \in [\underline{\alpha}, \overline{\alpha}] \subset (0,1)$ for all t. Of course, the choice of $\mathsf{s}(t)$ can severely affect convergence *speed*, but this discussion goes beyond the scope of the present paper.

Remark 3. In relation with Assumption 5.1, it is useful to work with sequences satisfying (with a given non-zero probability) $\mathsf{s}_{\mathrm{e}_\mathcal{G}}(t) \geq \beta$ at any t for some constant $\beta > 0$. Indeed, this ensures that once $q_g(t, t+t_1) \geq \delta' > 0$ for some $t_1 \leq T$, we have $q_g(t, t+T) \geq \delta = \delta' \beta^{T-t_1}$. Most results in linear consensus [32, 26, 22] explicitly make this assumption. Not assuming $\mathsf{s}_{\mathrm{e}_\mathcal{G}}(t) \geq \beta > 0$ for all t generally makes it necessary to perform a detailed analysis of the successions in $\mathsf{s}(t)$ in order to ensure Assumption 5.1.

6 Examples

We next illustrate the potential of our results by illustrating a variety of tasks covered by our framework. For these tasks, the gossip-inspired dynamics we have studied recover some relevant, existing class of algorithms or variations of these. We naturally start with consensus-type problems, including in Example 6.3 a quantum consensus algorithm which we have proposed and analyzed with a rather technical, ad-hoc approach in [20]. With the new *lifted* convergence results at hand, the solution is immediate. We then turn to more general symmetrization problems which do not include a network structure or a consensus-type task. These include random state generation protocols and quantum dynamical decoupling, two key tasks in quantum information theory and applications. In order to further illustrate the variety of the potential applications, we also include an academic example, showing how even the seemingly unrelated discrete Fourier transform can be seen as a symmetrization problem. The analysis of these protocols from a unified symmetrization viewpoint, and hence explicit proof of their robustness and randomization properties, are, to the best of our knowledge, new results. The list of examples is by no means assumed to be exhaustive, and we are confident that more areas of application will be identified.

6.1 Linear consensus

The gossip algorithm of Section 2 is one basic application of our framework. The group-theoretic language also encompasses other basic linear algorithms for average consensus of m subsystems in $\mathbb{R}^n$.

The most standard consensus algorithm implements, at each time, a motion of each subsystem towards the average of its neighbors in an *undirected graph* $G(t)$. Thus the edges of $G(t)$ model a set of interactions that are all simultaneously active. This corresponds to setting $\mathtt{s}_g(t) \neq 0$ for $g = e$ and for all $g \in \mathfrak{P}$ that model a pairwise permutation of two agents linked by an edge in $G(t)$, up to possibly having to use negative $\mathtt{s}_g(t)$. We recall that, since the actions associated to $\mathfrak{P}$ in standard consensus are not linearly independent, this is not the only way to lift the consensus dynamics to the permutation group; in particular, there is a way to do this without ever necessitating negative $\mathtt{s}_g(t)$, see next paragraph. Gossip, with a single edge active at a time and hence only two nonzero elements in $\mathtt{s}_g(t)$, is just a particular case.

In the group-theoretic formulation, there seems no reason to limit our algorithmic building blocks to pairwise permutations. Including more general permutations allows one to cover situations with explicit multipartite interactions, *e.g.* where subsystem 1 forwards its value to 2, who simultaneously transmits its value to 3, and so on. Selecting $\mathtt{s}_g \neq 0$ specifically for g corresponding to such situations, allows to model *synchronous* linear consensus iterations with symmetric or non-symmetric state transition matrix $A(t)$. The resulting $A(t)$ however will still be doubly-stochastic for any $\mathtt{s}$. As proved by Birkhoff [4], any doubly stochastic matrix can be decomposed as a convex sum of permutations. The corresponding network structure is called a *balanced directed graph* [27], and one could argue that the interpretation as a sum of general permutations gives a sensible rationale as why a graph might be ensured to be balanced in the consensus context. In this sense, any consensus algorithm on a balanced directed graph can be seen as a generalization of a gossip-type algorithm. Convergence, independently of the particular application, is guaranteed if Assumption 5.1 is satisfied.

Let us consider a concrete example of a consensus application: three vehicles need to establish agreement about the position of the center of a circle, on which they will move as a formation [30]. Let $x_k \in \mathbb{R}^2$ denote the center estimate for vehicle k, with $k = 1, 2, 3$. We assume that vehicles 2 and

3 cannot communicate. This corresponds to a consensus problem for a graph on 3 nodes $\{1, 2, 3\}$ and with edges $(1, 2)$, $(1, 3)$. A compatible consensus algorithm is:

$$x(t+1) = Ax(t) \quad : \quad \begin{array}{rcl} x_1(t+1) & = & (1-2\alpha)x_1(t) + \alpha x_2(t) + \alpha x_3(t) \\ x_2(t+1) & = & (1-\alpha)x_2(t) + \alpha x_1(t) \\ x_3(t+1) & = & (1-\alpha)x_3(t) + \alpha x_1(t) \end{array} \tag{25}$$

with $\alpha \leq 0.5$ to maintain double stochasticity.

From the symmetrization viewpoint, this problem considers all possible permutations of the initial estimates of the circle centers associated to the 3 vehicles:

$$\begin{array}{lcccccc} \text{permutation} & x_1(0) & x_1(0) & x_2(0) & x_3(0) & x_2(0) & x_3(0) \\ & x_2(0) & x_3(0) & x_1(0) & x_2(0) & x_3(0) & x_1(0) \\ & x_3(0) & x_2(0) & x_3(0) & x_1(0) & x_1(0) & x_2(0) \\ \text{weight} & \mathsf{p}_e & \mathsf{p}_{[1,3,2]} & \mathsf{p}_{[2,1,3]} & \mathsf{p}_{[3,2,1]} & \mathsf{p}_{[2,3,1]} & \mathsf{p}_{[3,1,2]} \end{array} \tag{26}$$

The vector $\mathsf{p}(t)$ represents the weight distribution over these 6 situations, labeling each permutation π of $[1, 2, 3]$ with the vector $[\pi(1), \pi(2), \pi(3)]$. According to (8), at any time $x_1(t)$ is the sum of the first element of each of the 6 columns, weighted by the corresponding entry of $\mathsf{p}(t)$. One can similarly compute $x_2(t)$ and $x_3(t)$. We start with all the weight concentrated on the trivial permutation, corresponding to $\mathsf{p}_e(0) = 1$. The consensus dynamics redistributes the weight such that finally all six situations have the same weight i.e. $\mathsf{p} = \hat{\mathsf{p}}$, the vector with all elements equal to $1/6$. When $\mathsf{p} = \hat{\mathsf{p}}$, the average positions of x_1, x_2 and x_3 are all the same and located at the barycenter of $x_1(0)$, $x_2(0)$ and $x_3(0)$, as expected from average consensus.

Following (15), the lifted dynamics associated to (25) would be modeled by:

$$\mathsf{s}_e = 1 - 2\alpha \; ; \quad \mathsf{s}_{[2,1,3]} = \alpha \; ; \quad \mathsf{s}_{[3,2,1]} = \alpha \; ; \quad \mathsf{s}_g = 0 \text{ for all other } g \, . \tag{27}$$

For example, the action associated to $[2, 1, 3]$, corresponding to active communication along the link $(1, 2)$, can be viewed as exchanging the first and second row of (26). Equivalently, leaving the first three rows of (26) in place, the action associated to $[2, 1, 3]$ "exchanges weight" between p_e and $\mathsf{p}_{[2,1,3]}$, between $\mathsf{p}_{[3,2,1]}$ and $\mathsf{p}_{[2,3,1]}$, and between $\mathsf{p}_{[3,1,2]}$ and $\mathsf{p}_{[1,3,2]}$.

We have mentioned that convergence in the permutation group is not necessary for convergence of the corresponding consensus algorithm. Related to this point, convergence *speed* may differ for p and for x. This can be illustrated already on the above simple example. The eigenvalues of the $\tilde{M}$ matrix corresponding to (27) indeed differ from those of the A matrix associated to consensus in (25). For $\alpha > 0.4$ we get $\sigma(\tilde{M}) > \sigma(A)$, where $\sigma(X)$ denotes the dominating singular value of X i.e. the largest modulus among all eigenvalues of X that differ from 1. Thus for $0.5 \geq \alpha > 0.4$, the eigenvalues of $\tilde{M}$ which govern convergence on the permutation group, underestimate the actual convergence speed of (25) on $\mathbb{R}^6$. For instance $\alpha = 0.45$ gives a geometric convergence rate with factor $\sigma(A) = 0.55$ for consensus, but only with $\sigma(\tilde{M}) = 0.8$ on the permutation group. Intuitively this can be understood by noting that the circle centers on the above schematic representation would all be located at the same central position already if e.g. $\mathsf{p}_e = \mathsf{p}_{[3,1,2]} = \mathsf{p}_{[2,3,1]} = 1/3$. Hence converging to $\mathsf{p} = \hat{\mathsf{p}}$, while it is actually attained by the algorithm (25), is not necessary for reaching consensus towards controlling the circular formation. Therefore the effective convergence speed can be faster for the original, "un-lifted" dynamics.

6.2 Gossip symmetrizing probability distributions

Consider a collection of m subsystems, each one possessing a random variable y_j on the same outcome set Y, for $j = 1, 2, \ldots, m$. We denote $\mathbb{P}$ the joint probability distribution of the y_j. In order to maintain a compact notation we will consider Y countable, but the uncountable case does not present additional technical difficulties. We are interested in symmetrizing the joint probability distribution, i.e. attaining a distribution $\hat{\mathbb{P}}$ such that

$$\begin{aligned} &\hat{\mathbb{P}}[y_1 = a_1, ..., y_j = a_j, ..., y_k = a_k, ..., y_m = a_m] \\ &= \hat{\mathbb{P}}[y_1 = a_1, ..., y_j = a_k, ..., y_k = a_j, ..., y_m = a_m] \end{aligned} \tag{28}$$

for all choices of j, k and of the considered outcomes $\{a_i\}$. The invariance then also holds for general permutations in $\mathfrak{P}$. We want to achieve this in a distributed way, where at each time t a reduced set $E(t)$ of pairwise interactions are available.

Our framework suggests the following randomized way to perform this task. At each time t a pair (j, k) is selected from $E(t)$, the random variables at these locations are swapped with probability α, and remain in place with probability $1 - \alpha$. This random action still leaves $y_j(t + 1), y_k(t + 1)$ two random variables on Y, but their probability distributions have changed: e.g. the new random variable $y_j(t + 1)$ at location j follows the marginal distribution of $y_j(t)$ with probability $1 - \alpha$, or it follows the marginal distribution of $y_k(t)$, with probability α. Overall, *not knowing whether the random variables have been exchanged or not*, the resulting probability distribution for the $y_i(t+1)$, $i = 1, 2, ..., m$ writes:

$$\begin{aligned} \mathbb{P}_{t+1}[y_1 = a_1, ..., y_j = a_j, ..., y_k = a_k, ..., y_m = a_m] \; = \\ (1 - \alpha)\, \mathbb{P}_t[y_1 = a_1, ..., y_j = a_j, ..., y_k = a_k, ..., y_m = a_m] \\ +\alpha\, \mathbb{P}_t[y_1 = a_1, ..., y_j = a_k, ..., y_k = a_j, ..., y_m = a_m] \end{aligned} \tag{29}$$

In the group symmetrization picture, this framework (goal (28) and dynamics (29)) corresponds to the exact same setting as standard gossip consensus, with $\mathcal{G} = \mathfrak{P}$ the group of permutations on m objects. Only the action is different, now implementing a swap on probability distributions (*including all correlations with other random variables* than the ones involved in the swap), instead of a swap of real numbers.

6.3 Gossip symmetrizing quantum subsystems

A classical random variable can be viewed as a special, commutative case in the framework of quantum, non-commutative probability theory. Following this analogy, the previous example can be extended to quantum observables – that is, self-adjoint linear operators on some Hilbert space $\mathcal{H}$. This is done in [20] with an ad-hoc approach, independently of the present general framework.

Consider a multipartite quantum system, composed of m isomorphic subsystems with individual Hilbert space $\mathcal{H}_1 = \mathcal{H}_2 = ... = \mathcal{H}_m$. The state of the overall system, which has the role of a probability distribution, is described by a density operator ρ on the tensor product of the individual Hilbert spaces, $\mathcal{H} = \mathcal{H}_1 \otimes \mathcal{H}_2 \otimes ... \otimes \mathcal{H}_m$. Let $\mathcal{X}$ be the set of self-adjoint operators on $\mathcal{H}$, associated to observable physical quantities. With $\mathcal{G}$ still being the permutation group of m objects, represented on the integers $1, 2, ..., m$ by elements π, we define the action $a_q(\pi, X)$ on $\mathcal{X}$ by

$$a_q(\pi, X) = X_{\pi(1)} \otimes X_{\pi(2)} \otimes ... \otimes X_{\pi(m)}$$

for operators of the form $X = X_1 \otimes X_2 \otimes ...X_m$ on $\mathcal{H}$, and extend it to the whole set $\mathcal{X}$ of self-adjoint operators on $\mathcal{H}$ by linearity. To each such action, we can associate a unitary operator U_π on $\mathcal{H}$ such

that

$$a_q(\pi, X) = U_\pi^\dagger \, X \, U_\pi \quad \text{for all } X \in \mathcal{X} \,,$$

where $U^\dagger$ denotes the adjoint of U (i.e. the complex conjugate transpose in matrix notation).

For this quantum system, the group dynamics corresponding to linear gossip would apply at each step a convex combination of the identity and the permutation of two physical subsystems j, k. Explicitly, the dynamics of X is given by:

$$X(t+1) = (1-\alpha)X(t) + \alpha U_{(j,k)}^\dagger X(t) U_{(j,k)}, \quad \alpha \in [0,1] \,.$$

This is a completely-positive, trace-preserving and unital map on $\mathcal{X}$. The latter two properties mirror double stochasticity of $\tilde{M}(t)$.

The convergence of the action-independent dynamics to $\hat{\rho}$ directly implies that both the cyclic and randomized versions of this quantum gossip algorithm will drive any initial $X \in \mathcal{X}$ to

$$\hat{X} = \frac{1}{m!} \sum_{\pi \in \mathfrak{P}} U_\pi^\dagger \, X \, U_\pi \,.$$

Physically, this implies that the measurement of any joint property on a subset of $n < m$ quantum systems will give the same statistics irrespective of the particular n subsytems that are selected.

Equivalently, we could consider as $\mathcal{X}$ the set of all density operators on $\mathcal{H}$, with the action $a'_q(g, \cdot) := a_q(g^{-1}, \cdot)$. These two equivalent viewpoints on quantum mechanics are well-known as the "Heisenberg picture" and the "Schrödinger picture". Example 6.2 is retrieved when all considered operators are diagonal in a fixed basis, and the diagonal of the density operator is then equivalent to a classical probability density. In the language of [20], this dynamics attains *symmetric state consensus*.

6.4 Randomized discrete Fourier transform

The above applications all involve permutations as the underlying group. The permutation group and the set of generators that can be activated encodes the network structure for the distributed computation task. We next show, starting with an academic example, how *the same class of algorithms can be used to tackle different problems that do not involve any network or consensus-reaching task. Specifically, a choosing a different group structure can lead to a* randomized *algorithm computing the discrete Fourier transform.*

The discrete Fourier transform of a (column) vector $x = (x_0, x_1, ..., x_{N-1}) \in \mathbb{C}^N$ is the (column) vector $\chi = (\chi_0, \chi_1, ..., \chi_{N-1})$ with

$$\chi_k = \frac{1}{N} \sum_{n=0}^{N-1} e^{-i \frac{kn 2\pi}{N}} \, x_n \quad \text{for } k = 0, 1, ..., N-1 \,, \tag{30}$$

up to normalization[4]. The complex numbers $\{e^{i\,k\,2\pi/N} : k = 0, 1, ..., N-1\}$ characterizing the Fourier transform form a faithful representation of the cyclic group of order N, that is the Abelian group generated by a single element $\bar{g}$,

$$\mathcal{G}_{c,N} = \{ \mathrm{e} = \bar{g}^0 = \bar{g}^N, \bar{g}, \bar{g}^2, \bar{g}^3, ..., \bar{g}^{N-1} \} \,.$$

We next show how the computation of (30) can be obtained as a byproduct of a symmetrization task with respect to an action of $\mathcal{G}_{c,N}$.

[4]Our developments can be extended to functions on finite Abelian groups, with the Fourier transform defined on characters.

It is convenient to consider the vector space $\mathbb{R}^{N\times N}$ and associate to the (column) vector $x \in \mathbb{R}^N$ the square matrix $X = x\,\mathbf{1}^T$, where $\mathbf{1}^T$ is the row vector of ones. To $\bar{g} \simeq e^{i\,2\pi/N}$ we associate the group action $a(\bar{g}, \cdot) = Q(\cdot)$ defined by:

$$\begin{aligned} X \;\mapsto \quad & Q(X) = \sigma\, X\, D^{-1} \\ \text{with} \quad & D = \mathrm{diag}(1,\, e^{i\,2\pi/N},\, e^{i\,4\pi/N},\, ...,\, e^{i\,(N-1)2\pi/N}) \\ & \sigma = \begin{pmatrix} 0 & 1 & 0 & 0 & ... & 0 \\ 0 & 0 & 1 & 0 & ... & 0 \\ 0 & 0 & 0 & 1 & ... & 0 \\ \vdots & & & & & \\ 0 & 0 & 0 & 0 & ... & 1 \\ 1 & 0 & 0 & 0 & ... & 0 \end{pmatrix}. \end{aligned} \tag{31}$$

The action corresponding to a general group element is obtained by composition. Direct computation shows that the m, n element of $\hat{X} = \frac{1}{N}\sum_{k=0}^{N-1} Q^k(X)$, resulting from the symmetrization of X under the action Q, equals

$$\hat{X}_{[m,n]} \;=\; \frac{1}{N}\sum_{k=0}^{N} x_{(m+k \bmod (N-1))}\; e^{-i\frac{2\pi k}{N}n}\,.$$

Hence symmetrization under this action of $\mathcal{G}_{c,N}$ gives the Fourier transform of x as:

$$\chi^T = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \end{bmatrix} \hat{X}\,.$$

The robust convergence of algorithm (15) thus indicates that the Fourier transform does not necessarily have to be computed in an orderly fashion, but can asymptotically result from rather arbitrary convex combinations of the actions Q^k with different k, as long as the $\mathsf{s}(t)$ ensure sufficient mixing. Note that the actions $\{Q^0, Q^1, ..., Q^{N-1}\}$ are all linearly independent, so the map from dynamics on group actions to dynamics on p is one-to-one.

6.5 Random state generation

A variety of applications require to generate random numbers, codewords or, more generally, *states* with a target probability distribution. This includes among others the Markov chain Montecarlo methods [3] as well as classical and quantum cryptography protocols [25]. A typical, and fundamental, target probability distribution is the uniform or Haar measure on compact sets. Random sample generators must hence be able to transform some *generic* source of randomness – i.e. not necessarily uniform nor in fact exactly known – into a (almost) *uniform* probability distribution. There are various ways of doing this, and our framework points to a particular class of so-called random circuits [14, 13]. Indeed, group symmetrization provides a robust way to obtain a uniform distribution on a finite set of states $\mathcal{Y}$ that are linked by a group of transformations $\mathcal{G}$, if we can pick elements of $\mathcal{G}$ with some generic probability distribution.

More precisely, consider a finite group $\mathcal{G}$, and its linear action $a(g, \cdot)$ on a vector space $\mathcal{X}$. For some fixed $y_e \in \mathcal{X}$, consider its *orbit*, i.e. the set $\mathrm{Orb}_{\mathcal{G}}(y_e) = \{y_g = a(g, y_e),\, g \in \mathcal{G}\}$. We want to generate a state $y(T)$ that is uniformly (pseudo-)randomly distributed over $\mathrm{Orb}_{\mathcal{G}}(y_e)$, by passing a deterministic $y(0) \in \mathrm{Orb}_{\mathcal{G}}(y_e)$ through a sequence of (pseudo-)random operations, labeled for convenience by time $t = 0, 1, ..., T-1$. Each operation is associated to a $g(t) \in \mathcal{G}$, drawn according to some possibly unknown probability distributions $\mathsf{s}_g(t)$, mutually independent at each time. We make the technical assumption that $g \neq h \Rightarrow a(g, y(0)) \neq a(h, y(0))$ i.e. $|\mathrm{Orb}_{\mathcal{G}}(y_e)| = |\mathcal{G}|$.

As y propagates through the sequence according to $y(t+1) = a(g(t), y(t))$, the probability $\mathsf{p}_h(t)$ to have $y(t+1) = a(h, y(0))$ follows dynamics (15). Hence according to Theorem 5.1, it is sufficient that $\mathsf{s}(t)$ allows to satisfy Assumption 5.1 to ensure that the distribution of $y(T)$ converges to the *uniform* distribution over $\mathrm{Orb}_{\mathcal{G}}(y_e)$ as $T \to \infty$. Note that for a fixed circuit distribution $\mathsf{s}_g(t)$, we indeed apply Theorem 5.1 as we are modeling the *deterministic* evolution (as t increases) of a probability distribution.

Remark 4. *In addition to finite groups, the case in which $\mathcal{G}$ becomes a continuous Lie group is of great interest for practical applications, including quantum information and more specifically random quantum circuit theory [14, 13]. In that framework, the space of interest is associated to a register of N quantum bits, so that $\mathcal{X} \cong \mathbb{C}^{2^N}$; the group of physically relevant unitary evolutions for the register, or* gates*, is $\mathcal{G} = SU(2^N)$. The finite group setting can effectively approximate such continuous distribution by considering a sufficiently dense subset of the Lie group. It is well known [25] that there exist finite* universal sets *of gates which generate a mathematically dense subset of $SU(2^N)$; ensuring $\mathsf{s}_g(t) > 0$ on such a universal set, is sufficient to satisfy Assumption 5.1 for any finite subset of a dense subset of $SU(2^N)$.*

6.6 Dynamical decoupling

Quantum Dynamical Decoupling (DD) is a set of open-loop control techniques that are primarily used to reduce the effect of unknown Hamiltonian drifts, or couplings to the environment, on a target quantum system [33]. The main idea is to apply a sequence of "switching" unitary rotations to the system, such that effects of the undesired dynamics over the sequence of unitary rotations compensate each other and the net effect is negligible. This task can be translated into a symmetrization task [37], and we show here how our results suggest a robust DD scheme. For the sake of simplicity, we restrict ourselves to the suppression of the drift Hamiltonian in finite dimensional systems. The extension to decoupling from the environment is straightforward.

The quantum evolution of an isolated finite-dimensional system is driven by its Hamiltonian H, a Hermitian matrix whose spectrum is associated to the energy levels of the system. The propagator for the system is then the unitary operator

$$U_t = e^{-iHt}$$

when H is constant. When H is time-varying, the propagator must be computed as an ordered product of exponentials over infinitesimal intervals. The resulting unitary operator can be associated to an effective Hamiltonian H_{eff} such that

$$U_T = e^{-iH_{\mathrm{eff}}\,T}\ .$$

A DD strategy consists in a time-dependent control Hamitlonian $H_c(t)$ such that, for any constant H_d in a class of expected perturbations, the effective Hamiltonian associated to $H_d + H_c(t)$ is "close" to a scalar matrix after a predefined time T: $H_{\mathrm{eff}} \approx \lambda I$ with $\lambda \in \mathbb{R}$. Indeed, this would suppress any physical effect of H_d at time T since global phases of the form $U_t = e^{i\lambda t}$ are irrelevant for predictions in quantum mechanics [29]. DD in its simplest form entails a sequence of fast, impulsive control operations that induce a group of "instantaneous" unitary transformations on the system, and achieves first-order suppression of H_d. The relevant time interval $[0, T)$ is subdivided into N subintervals of length $dt = T/N$ and instantaneous controls are applied at the end of each sub-interval so that the effective Hamiltonian for subinterval $[(k-1)\,dt,\ k\,dt)$ is $g_k H_d g_k^\dagger$ with $g_k \in \mathcal{G}$. Then, the Magnus expansion [19] allows to approximate the exact evolution from time 0 to T to first order as:

$$e^{-i\,dt\,g_1 H_d g_1^\dagger}\ e^{-i\,dt\,g_2 H_d g_2^\dagger}\ \ldots\ e^{-i\,dt\,g_N H_d g_N^\dagger} \quad\approx\quad e^{-i\,dt\,\sum_{k=1}^N g_k H_d g_k^\dagger} \quad =: e^{-i\,T\,\bar{H}}\ , \tag{32}$$

where † denotes matrix conjugate transpose. Accuracy improves as the product of H_d with dt gets smaller. Hence, given a class $\mathfrak{H}_0$ of drift Hamiltonians on some finite-dimensional Hilbert space $\mathcal{H} \cong \mathbb{C}^n$, first-order DD follows from identifying a finite subgroup $\mathcal{G}$ of unitaries such that

$$\tfrac{1}{|\mathcal{G}|} \sum_{g\in\mathcal{G}} g\, H_d\, g^\dagger = \lambda I \tag{33}$$

for all $H_d \in \mathfrak{H}_0$. In the language of our paper, DD achieves symmetrization with respect to a group $\mathcal{G}$, and the latter is selected such that the action $a(g, H) = g\, H\, g^\dagger$ on the space $\mathcal{X}$ of all Hamiltonians H satisfies $\bar{\mathcal{F}}(\mathfrak{H}_0) \subseteq \{\lambda I, \lambda \in \mathbb{R}\}$.

Achieving symmetrization in (32) means choosing each $g \in \mathcal{G}$ an equal number of times over the N subintervals. An obvious choice is just to take $N = m|\mathcal{G}|$ and iterate m times a predefined path through the elements of $\mathcal{G}$. However, when H_d is not really constant for a duration $|\mathcal{G}|\, dt$ or when considering higher-order Magnus terms, the potential advantage of randomized [34, 18] or concatenated [15] sequences of g_k has been recognized. Our general dynamics (15) allows to retrieve and combine these two variants of DD and, in particular, to highlight their robustness.

Consider an iterative construction of the sequence of unitaries g_k, where at the n-th iteration the time interval $[0, T)$ is subdivided into $N = 2^n$ subintervals. Denote $\mathcal{S} \subseteq \mathcal{G}$ the set of available control actions. We start at $n = 0$ from the situation with no control pulses, so $g_1 = \mathrm{e} \cong I_{\mathcal{H}}$ over $[0, T)$ and $\bar{H} = H_d$. Increasing n, we then choose one element $h(n) \in \mathcal{S}$, we divide each subinterval $\left[(m-1)\frac{T}{2^n}, m\frac{T}{2^n}\right)$ into two equal time intervals $\left[(2m-2)\frac{T}{2^{n+1}}, (2m-1)\frac{T}{2^{n+1}}\right)$ and $\left[(2m-1)\frac{T}{2^{n+1}}, 2m\frac{T}{2^{n+1}}\right)$, and we update the sequence as follows for $m = 1, ..., 2^n$:

$$\text{At } n: \quad g_m = \bar{g} \qquad \Rightarrow \qquad \text{At } n+1: \quad g_{2m-1} = \bar{g}\,, \quad g_{2m} = h(n)\bar{g}\,. \tag{34}$$

Denoting by $\mathsf{p}_g(n)$ the fraction of time $[0, T)$ during which $g_k = g \in \mathcal{G}$, the procedure (34) correponds to (15) with t replaced by n, and the switching signal:

$$\mathsf{s}_g(n) = 1/2 \;\text{ for } g \in \{\mathrm{e}_{\mathcal{G}}, h(n)\}\,, \quad \mathsf{s}_g(n) = 0 \;\text{ for all other } g \in \mathcal{G}\,. \tag{35}$$

In action form, the average Hamiltonian at the n-th iteration is

$$\bar{H}_n = \textstyle\sum_{g\in\mathcal{G}} \mathsf{p}_g(n)\, a(g, H_d) \;=\; \sum_{g\in\mathcal{G}} \mathsf{s}_g(n-1)\, a(g, \bar{H}_{n-1})\,.$$

Our theorems ensure the convergence of $\bar{H}_n$ towards the $\mathcal{G}$-symmetrized form (33) of H_d as n is increased, if Assumption 5.1 holds. This is valid both for deterministic or random choices of the $h(n)$. Furthermore, our results indicate a remarkable generality and robustness of the procedure: (i) the control actions $h(n)$ don't have to be chosen uniformly in $\mathcal{G}$, actually any deterministic choice or probabilistic distribution over enough elements will work; (ii) the set $\mathcal{S}$ of control actions does not have to be all $\mathcal{G}$, e.g. a set of generators would be sufficient; and (iii) the subdivision can be more general than a "perfect average": any $\mathsf{s}_{h(n)}(n) = 1 - \mathsf{s}_{\mathrm{e}}(n) = \alpha$ with $\alpha \in (0, 1)$ would asymptotically work, not just (35) where $\alpha = 1/2$.

7 Conclusion

The present paper shows how the simple dynamics of linear gossip consensus can inspire robust iterative procedures for tasks that can be formulated as *symmetrization with respect to a finite group.* We prove convergence for a general symmetrization process with either deterministic or randomized

choices of the individual iterations. We have shown how a variety of existing algorithms, some unrelated to any network structure, are covered by the framework. We expect that in many other applications the *robustness* of the consensus formulation can be advantageously carried over to symmetrizations tasks, e.g. including actions on infinite-dimensional spaces. Natural directions for expanding our results in the short term include the development of (approximate) symmetrization procedures for infinite and continuous groups, as well as an in-depth study of convergence *speed* for specific protocols. Regarding the latter, our bound in Theorem 5.1 can be unnecessarily pessimistic especially when the concerned group actions are not linearly independent, as is the case e.g. for consensus. The possibility to lift, to the abstract symmetrization framework, several speed-up strategies for faster mixing is also being investigated. Replacing the linear action on a vector field by abstract algebraic structures could also offer a rewarding way to unify more algorithmic procedures, hopefully including e.g. alternating directions optimization or dominant eigenvector computations, under the symmetrization viewpoint.

8 Acknowledgments

The authors would like to thank Lorenza Viola for suggesting the application of these techniques to random quantum circuits, for pointing out some key references and for numerous, fruitful and pleasant discussions on these topics. This work has been partially supported by the QUINTET and QFUTURE strategic projects of the Dept. of Information Engineering and University of Padua, and by the Belgian Inter-University Attraction Poles network DYSCO.

References

[1] C. Altafini and F. Ticozzi. Modeling and control of quantum systems: An introduction. *IEEE Trans. Aut. Cont,*, 57(8):1898 –1917, 2012.

[2] Baruch Awerbuch. Optimal distributed algorithms for minimum weight spanning tree, counting, leader election, and related problems. In *Proceedings of the nineteenth annual ACM symposium on Theory of computing*, pages 230–240. ACM, 1987.

[3] B.A. Berg, D.P. Landau, W.S. Kendall, R. Chen, and E.A. Thompson. *Markov Chain Monte Carlo: innovations and applications*. World Scientific Publishing, 2005.

[4] G. Birkhoff. Three observations on linear algebra. *Univ. Nac. Tucuan. Revista A*, 5:147–151, 1946.

[5] S. Boyd, A. Ghosh, B. Prabhakar, and D. Shah. Randomized gossip algorithms. *IEEE Trans. Information Theory (Special issue)*, 52(6):2508–2530, 2006.

[6] Ruggero Carli, Edoardo D'Elia, and Sandro Zampieri. A pi controller based on asymmetric gossip communications for clocks synchronization in wireless sensors networks. In *Decision and Control and European Control Conference (CDC-ECC), 2011 50th IEEE Conference on*, pages 7512–7517. IEEE, 2011.

[7] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. Wiley, 1991.

[8] Alexandros G Dimakis, Soummya Kar, José MF Moura, Michael G Rabbat, and Anna Scaglione. Gossip algorithms for distributed signal processing. *Proceedings of the IEEE*, 98(11):1847–1864, 2010.

[9] D.L. Donoho. Compressed sensing. *IEEE Trans.Information Theory*, 52(4):1289–1306, 2006.

[10] C. D. Godsil and Gordon. Royle. *Algebraic graph theory*. Springer New York, 2001.

[11] Hideaki Ishii and Roberto Tempo. Distributed randomized algorithms for the pagerank computation. *Automatic Control, IEEE Transactions on*, 55(9):1987–2002, 2010.

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

[13] J.Emerson, E.Livine, and S.Lloyd. Convergence conditions for random quantum circuits. *Physical Review A*, 72(6):060302, 2005.

[14] J.Emerson, Y. Weinstein, M. Saraceno, S. Lloyd, and D. Cory. Pseudo-random unitary operators for quantum information processing. *Science*, 302:2098–2100, 2003.

[15] K.Khodjasteh and D.A.Lidar. Fault-tolerant quantum dynamical decoupling. *Phys. Rev. Lett.*, 95(18):180501, 2005.

[16] E. Knill, R. Laflamme, and L. Viola. Theory of quantum error correction for general noise. *Phys. Rev. Lett.*, 84(11):2525–2528, 2000.

[17] Naomi Ehrich Leonard, Derek A Paley, Francois Lekien, Rodolphe Sepulchre, David M Fratantoni, and Russ E Davis. Collective motion, sensor networks, and ocean sampling. *Proceedings of the IEEE*, 95(1):48–74, 2007.

[18] L.F.Santos and L.Viola. Enhanced convergence and robust performance of randomized dynamical decoupling. *Phys. Rev. Lett.*, 97(15):150501, 2006.

[19] W. Magnus. On the exponential solution of differential equations for a linear operator. *Commun. Pure and Appl. Math.*, 7:649–673, 1954.

[20] L. Mazzarella, A. Sarlette, and F. Ticozzi. Consensus for quantum networks: from symmetry to gossip iterations. *IEEE Trans. Automatic Control*, 60(1):158–172, 2015.

[21] Ciamac Cyrus Moallemi and Benjamin Van Roy. Consensus propagation. *Information Theory, IEEE Transactions on*, 52(11):4753–4766, 2006.

[22] L. Moreau. Stability of multi-agent systems with time-dependent communication links. *IEEE Trans. Automatic Control*, 50(2):169–182, 2005.

[23] R. Motwani and P. Raghavan. *Randomized algorithms*. Cambridge University Press, 1995.

[24] A. Nedic and A. Ozdaglar. Distributed subgradient methods for multiagent optimization. *Automatic Control, IEEE Transactions on*, 54(1):48–61, 2009.

[25] M. A. Nielsen and I. L. Chuang. *Quantum Computation and Information*. Cambridge University Press, Cambridge, 2002.

[26] R. Olfati-Saber, J.A. Fax, and R.M. Murray. Consensus and cooperation in networked multiagent systems. *Proc. IEEE*, 95(1):215–233, 2007.

[27] R. Olfati-Saber and R.M. Murray. Consensus problems in networks of agents with switching topology and time delays. *IEEE Trans. Automatic Control*, 49(9):1520–1533, 2004.

[28] Judea Pearl. Fusion, propagation, and structuring in belief networks. *Artificial intelligence*, 29(3):241–288, 1986.

[29] J. J. Sakurai. *Modern Quantum Mechanics*. Addison-Wesley, New York, 1994.

[30] Rodolphe Sepulchre, Derek A Paley, and Naomi Ehrich Leonard. Stabilization of planar collective motion: All-to-all communication. *IEEE Trans.Aut.Control*, 52(5):811–824, 2007.

[31] F. Ticozzi, L. Mazzarella, and A. Sarlette. Symmetrization for quantum networks: a continuous-time approach. *Proc. Conf. on Math. Theory of Networks and Systems (MTNS)*, 2014.

[32] J.N. Tsitsiklis and M. Athans (advisor). Problems in decentralized decision making and computation. *PhD Thesis, MIT*, 1984.

[33] L. Viola, E. Knill, and S. Lloyd. Dynamical decoupling of open quantum system. *Phys. Rev. Lett.*, 82(12):2417–2421, 1999.

[34] Lorenza Viola and Emanuel Knill. Random decoupling schemes for quantum dynamical control and error suppression. *Phys. Rev. Lett.*, 94:060502, 2005.

[35] Peng-Jun Wan, Khaled M Alzoubi, and Ophir Frieder. Distributed construction of connected dominating set in wireless ad hoc networks. In *INFOCOM 2002. Twenty-First annual joint conference of the IEEE computer and communications societies. Proceedings. IEEE*, volume 3, pages 1597–1604. IEEE, 2002.

[36] Lin Xiao, Stephen Boyd, and Sanjay Lall. A scheme for robust distributed sensor fusion based on average consensus. In *Information Processing in Sensor Networks, 2005. IPSN 2005. Fourth International Symposium on*, pages 63–70. IEEE, 2005.

[37] P. Zanardi. Symmetrizing evolutions. *Phys. Lett. A*, 258:77, 1999.

A Convergence in relative entropy

We here show that the relative entropy, or Kullback-Leibler pseudo-distance, is also a Lyapunov function for the convergence of $\mathsf{p}(t)$ to $\hat{\mathsf{p}}$ under our symmetrizing dynamics. Before giving the proof, let us recall some basic facts about relative entropy and the log sum inequality.

The relative entropy, or Kullback-Leibler pseudo-distance [7] of a vector of convex weights $\{\mathsf{q}_g\}_{g\in\mathcal{G}}$ with respect to another one $\{\mathsf{p}_g\}_{g\in\mathcal{G}}$ is given by:

$$K(\mathsf{p}\|\mathsf{q}) = \sum_{g\in\mathcal{G}} \mathsf{p}_g \left(\log \mathsf{p}_g - \log \mathsf{q}_g\right). \tag{36}$$

This expression is not symmetric in p, q, but $K(\mathsf{p}\|\mathsf{q}) \geqslant 0$ and the equality holds if and only if $\mathsf{p} = \mathsf{q}$. We shall also use the following [7].

Proposition A.1 (Log Sum Inequality)**.** *Let $\{a_i\}_{i=1}^n$ and $\{b_i\}_{i=1}^n$ be nonnegative numbers. Then it holds:*

$$\sum_{i=1}^n a_i \log \frac{a_i}{b_i} \geqslant \left(\sum_{i=1}^n a_i\right) \log \frac{\sum_i a_i}{\sum_i b_i}. \tag{37}$$

Furthermore, excluding the singular cases where $\sum_i a_i = 0$ or $\sum_i b_i = 0$, the equality holds if and only if $\frac{a_i}{b_i} = \alpha$ is constant over $i = 1, \ldots, n$.

We can now turn to the convergence proof using $K(\mathsf{p}(t)\|\hat{\mathsf{p}})$ as Lyapunov function. The corresponding statement would be equivalent to Theorem 5.1 except that we do not prove the exponential character of the convergence. $K(\mathsf{p}(t)\|\hat{\mathsf{p}})$ is nonnegative and it equals zero if and only if $\mathsf{p}(t) = \hat{\mathsf{p}}$. To use it as a strict Lyapunov function, it remains to prove that, under Assumption 5.1, this relative entropy of $\mathsf{p}(t)$ with respect to $\hat{\mathsf{p}}$ strictly decreases after (any) T steps. For every t we have that:

$$\begin{aligned} K(\mathsf{p}(t+T)\|\hat{\mathsf{p}}) &= \sum_{g\in\mathcal{G}} \mathsf{p}_g(t+T) \log \frac{\mathsf{p}_g(t+T)}{\hat{\mathsf{p}}_g} \\ &= \sum_{g\in\mathcal{G}} \left(\sum_{h\in\mathcal{G}} \mathsf{q}_h(t,T)\mathsf{p}_{h^{-1}g}(t) \right) \log \frac{\sum_h \mathsf{q}_h(t,T)\mathsf{p}_{h^{-1}g}(t)}{\sum_h \mathsf{q}_h(t,T)\hat{\mathsf{p}}_g}. \end{aligned}$$

Now by applying the log sum inequality over h for each fixed g we get:

$$\begin{aligned} &\left(\sum_{h\in\mathcal{G}} \mathsf{q}_h(t,T)\mathsf{p}_{h^{-1}g}(t) \right) \log \frac{\sum_h \mathsf{q}_h(t,T)\mathsf{p}_{h^{-1}g}(t)}{\sum_h \mathsf{q}_h(t,T)\hat{\mathsf{p}}_g} \\ &\qquad \leq \sum_{h\in\mathcal{G}} \left(\mathsf{q}_h(t,T)\mathsf{p}_{h^{-1}g}(t) \log \frac{\mathsf{q}_h(t,T)\mathsf{p}_{h^{-1}g}(t)}{\mathsf{q}_h(t,T)\hat{\mathsf{p}}_{h^{-1}g}} \right). \end{aligned} \tag{38}$$

Furthermore, Assumption 5.1 allows us: (i) to divide by $\mathsf{q}_h(t,T)$; and (ii) in conjunction with the fact that $\sum_g \mathsf{p}_g(t) = 1$ for all t, to exclude the singular cases in Proposition A.1. Therefore the equality in (38) holds if and only if

$$\frac{\mathsf{q}_h(t,T)\mathsf{p}_{h^{-1}g}(t)}{\mathsf{q}_h(t,T)\hat{\mathsf{p}}_{h^{-1}g}} = \frac{\mathsf{p}_{h^{-1}g}(t)}{\hat{\mathsf{p}}_{h^{-1}g}}$$

is constant over all $g' = h^{-1}g \in \mathcal{G}$. Since $\sum_{g'\in\mathcal{G}} \mathsf{p}_{g'\in\mathcal{G}}(t) = \sum_{g'} \hat{\mathsf{p}}_{g'} = 1$ for every t, the equality holds if and only if $\mathsf{p}(t) = \hat{\mathsf{p}}$. Returning to the sum over g, we thus get

$$0 \leqslant K(\mathsf{p}(t+T)\|\hat{\mathsf{p}}) \leqslant K(\mathsf{p}(t)\|\hat{\mathsf{p}}) \tag{39}$$

and each equality holds if and only if $\mathsf{p}(t) = \hat{\mathsf{p}}$. Henceforth the Lyapunov function $K(\mathsf{p}(t)\|\hat{\mathsf{p}})$ strictly decreases after any T steps, as the requirement $q_h(t,T) > \delta$ ensures that for any given $\mathsf{p}(t) \neq \hat{\mathsf{p}}$, we get in (38) a strict contraction factor independent of $\mathsf{s}(t)$. This ensures, by Lyapunov arguments, that the system asymptotically converges to $\mathsf{p} = \hat{\mathsf{p}}$.

The fact that exponential convergence is not as direct, would also require another approach for the randomized case, that is Theorem 5.2.