Ergodicity Coefficients are Induced Matrix Seminorms

Giulia De Pasquale, Francesco Bullo and M. Elena Valcher

Abstract—Ergodicity coefficients are a useful algebraic tool in the study of the convergence properties of inhomogeneous Markov chains and averaging algorithms. Their study has a rich history, going back all the way to the original works by Markov.

In this work we show how ergodicity coefficients are equal to certain induced matrix seminorms and the induced norm of optimally-deflated matrices. This equivalence clarifies their use as contraction factors for semicontractive dynamical systems. In particular, the Dobrushin ergodicity coefficient \( \tau_1 \) is shown to be equal to two different induced matrix seminorms. In the context of Markov chains, we provide an expression for the mixing time in terms of \( \ell_\infty \) ergodicity coefficient. Finally, we show that, for primitive matrices, induced matrix seminorms minimize the efficiency gap in the estimation of the convergence factor of semicontractive dynamical systems, thus proving to be a more accurate tool compared to ergodicity coefficients.

I. INTRODUCTION

Problem description: Ergodicity coefficients are a useful algebraic tool for the study of long term behaviour of inhomogeneous Markov chains [18]. In this context, ergodicity coefficients determine the convergence properties of products of stochastic matrices with an increasing number of factors [6]. Determining the inclusion region of the subdominant eigenvalues, they provide an upper bound on the convergence factor of algorithms to compute the stationary distributions for stochastic matrices. In particular, such coefficients evaluate how fast the matrix product of row stochastic matrices converges to a rank one matrix, thus determining if a Markov chain is ergodic.

Interestingly, in the past two decades, as Markov chains with large state spaces have attracted more interest, a different asymptotic analysis has arisen. In this large scale setting, a target distance from the stationary distribution is often fixed and one wants to estimate how the number of steps required to reach the target grows as the size of the state space increases [10]. The number of steps required to reach the target distance is known in the literature as mixing time of the chain.

Problem motivation: The concept of ergodicity is developed starting from the more general concept of contractivity [17]. Contractivity of a dynamical system guarantees all solutions to converge exponentially fast towards each other [13].

However, there are numerous applications where the contractivity conditions are not satisfied and where, however, the distance between trajectories diminishes in a weaker sense, i.e., with respect to a seminorm; this is the setup motivating semicontraction theory.

The semicontraction theory was first developed by Wang and Slotine [20] who introduced the concept of partial contraction and exploited the theory in the study of dynamic behaviours of nonlinear coupled oscillators. A recent re-elaboration and extension of the theory is due to Saber et al. [7] who provided a comprehensive unifying framework for semicontraction, partial contraction and horizontal contracton [5] theories. Specifically, [7] provides a reinterpretation of the theory based on the concept of induced matrix seminorm; arbitrary semi-measures are allowed in the study of semicontractive systems. The results are then applied to affine averaging and flow systems, networks of diffusively coupled dynamical systems, and other examples.

Building upon these references, this work aims to develop semicontraction theory in the context of discrete time dynamical systems. We are interested in understanding ergodicity coefficients as contraction factors and, specifically, in defining them as induced matrix seminorms.

Contraction factors, in fact, play a key role in the context of contraction theory, in determining the exponential convergence and synchronization properties of dynamical systems. They also come into play in the context of distributed information and distributed averaging consensus algorithms, e.g., gossiping algorithms [11] and algorithms over time-varying networks. Typically, in fact, a gossiping process can be modeled as a discrete time linear system in which the state matrix is time-varying, doubly stochastic with a special structure. In this context, it becomes of particular interest the rate at which a sequence of agent gossip variables converges to a common value. From the study in [11], it turns out that, the worst case convergence rate of a periodic gossip sequence can be expressed in terms of suitably defined seminorms. One wants to find a suitable seminorm with respect to which the (stochastic) matrix derived by the subsequence of gossip over a determined period is a semicontraction: the analogy with ergodicity coefficients arises spontaneously.

Future works aim to apply our study also in the context of coupled discrete time oscillators.

Literature review: The study of ergodicity coefficients is traced back to the pioneering work of Markov [14], in 1906, in which a first expression of ergodicity coefficients was provided in the context of the Weak Law of Large Numbers. In 1931 a formalization of the concept of weak ergodicity was then...
introduced for the first time by Kolmogorov [8], who stated that a sequence of stochastic matrices is weakly ergodic if the rows of the matrix product tend to become identical as the number of factors increases. Subsequent works from Doeblin [4] and Dobrushin [3] provide conditions for weak ergodicity. The key results in this research area were extended and then reviewed by Seneta in the 80’s, e.g., see [18]. A comprehensive survey of ergodicity coefficients is given by Ipsen and Selee [6] and a recent treatment on their connection with spectral graph theory is given by Marsli and Hall [15].

A characterization of “Convergability” [12], namely the convergence of a product of an infinite number of stochastic matrices, is studied by Liu et. al in [12], where a different approach, based on optimally deflated matrices, is proposed.

**Contribution:** The scientific literature on ergodicity coefficients and matrix norms is quite vast. However, notably, it becomes rather sparse when it comes to definitions and properties of induced matrix seminorms [9]. Different seminorms are suitable for different problems, therefore when one takes the induced seminorm approach in determining the semi-contractivity property of a dynamical system, a natural question arises in terms of what are correct seminorms corresponding to the ergodicity coefficients $\tau_p$, for $p \in \{1, 2, \infty\}$. Both the existence and uniqueness in the correspondence between ergodicity coefficients and induced (weighted) matrix seminorm are open questions.

Our first and main contribution is to verify that (i) the traditional definition of $\tau_p$, for all $p$, in Ipsen and Selee [6], (ii) the optimally deflated matrix defined as Liu et al. [12], and (iii) a third definition based on a novel induced matrix seminorm, are identical. Specifically, our novel definition involves the orthogonal projection matrix $P_v$, associated with the vector $v$, and the seminorm $\|x\|_{p,R}$, for $p \in \{1, 2, \infty\}$. We then particularize the result to row stochastic matrices, in which case an oblique projection can also be considered as weight for the induced matrix seminorm. As a second contribution, we show that, for the “Dobrushin coefficient” $\tau_1$ a second possible seminorm, weighted by the incidence matrix of a complete directed graph, is possible. As a third contribution we provide an equivalent expression for the mixing time of a Markov chain in terms of $\ell_\infty$ ergodicity coefficient.

Finally, a formalization of the essential spectral radius of primitive matrices, as the result of a constrained optimization problem over all possible weighted induced matrix seminorms, is provided, thus showing how induced matrix seminorms determine a tighter bound, with respect to ergodicity coefficients, for the convergence rate of semicontractive dynamical systems.

**Paper organization:** In Section II some basic and preliminary concepts together with the main definitions are given. Section III provides our main contributions. Finally, Section IV concludes the work.

**Notation:** Let $I_n \in \mathbb{R}^{n \times n}$ denote the identity matrix of size $n$. Let $\mathbb{N}$ and $\mathbb{Q}_n$ denote the $n$ dimensional column vectors all whose entries equal 1 and 0, respectively. Let $e_i$ denote the $i$-th vector of the canonical basis in $\mathbb{R}^n$. We let $\mathbb{N}_\infty$ denote the set of extended natural numbers, $\mathbb{N}_\infty = \mathbb{N} \cup \{\infty\}$. For $v \in \mathbb{R}^n$, define $\text{ave}(v) = \frac{1}{n} \sum_{i=1}^n v_i$.

For a matrix $A \in \mathbb{R}^{n \times n}$, let $\text{spec}(A)$ denote its spectrum, $A^T$ its transpose, $[A]_{i,j}$ its $(i, j)$th entry, and $[A]_{i,**}$ its $i$-th row.

The matrix $A$ is nonnegative if all its entries are nonnegative, it is row stochastic if it is nonnegative and $A1_n = 1_n$, and it is primitive if it is nonnegative and there exists $k \in \mathbb{N}$ such that $A_k$ is entry-wise positive. For $A \in \mathbb{R}^{n \times n}$ the Moore-Penrose inverse of $A$ is the unique matrix $A^* \in \mathbb{R}^{n \times n}$ such that $AA^*A = A = A^*A$.

Given $A \in \mathbb{R}^{n \times n}$, a vector subspace $W \subseteq \mathbb{R}^n$ is $A$-invariant if $AW \subseteq W$. Given $v \in \mathbb{R}^n$, let $P_v$ denote the orthogonal projector onto $\langle v \rangle^\perp$, where the symbol $\langle v \rangle^\perp$ denotes the orthogonal complement of $\langle v \rangle$. Note that, if $v = 1_n$, then $P_v = I_n - \frac{1}{n} 1_n 1_n^\top/n =: \Pi_n$. Define the $n$-simplex as $\Delta_n = \{v \in \mathbb{R}^n | \sum_{i=1}^n v_i = 1, v_i \geq 0 \}$ and the sign function, sign : $\mathbb{R} \to \{-1, 0, 1\}$ as sign$(x) = \frac{x}{|x|}$ if $x \neq 0$, and sign$(0) = 0$. A square matrix $Q$, is an oblique projector if $Q^2 = Q$, $Q^{1/2}$ denotes a matrix such that $Q^{1/2}Q^{1/2} = Q$.

An unweighted undirected graph is a couple $G = (V, \mathcal{E})$, where $V = \{1, \ldots, n\}$ is the set of vertices and $\mathcal{E} \subseteq V \times V$ is the set of edges. If $(i, j)$, $i \neq j$, is an edge in $G$, $i$ represents the head, while $j$ its tail. An undirected graph is complete if, for every pair of distinct vertices, there exists an edge connecting them. The complete graph with $n$ nodes (and no self-loops) has $m = n \cdot (n - 1)$ edges. Let $C_n \in \{-1, 0, 1\}^{n \times n}$ denote the oriented incidence matrix of the complete undirected graph with $n$ nodes. For every vertex $i \in V$ and every edge $e \in \{1, \ldots, m\}$, we have $[C_n]_{i,e} = 1$ if $i$ is the head of edge $e$, $[C_n]_{i,e} = -1$ if $i$ is the tail of edge $e$, $[C_n]_{i,e} = 0$ otherwise.

**II. Technical Background**

**A. Basic concepts**

For $x \in \mathbb{R}^n$ and $p \in \mathbb{N}$, the $\ell_p$-norm of $x$ is

$$
\|x\|_p := \left( \sum_{i=1}^n |x_i|^p \right)^{1/p}.
$$

For the $\ell_\infty$-norm,

$$
\|x\|_\infty = \lim_{p \to \infty} \left( \sum_{i=1}^n |x_i|^p \right)^{1/p} = \max_i |x_i|.
$$

The essential spectral radius of a row stochastic matrix $A \in \mathbb{R}^{n \times n}$ is

$$
\rho_{\text{ess}}(A) := \begin{cases} 0, & \text{if } \text{spec}(A) = \{1, \ldots, 1\}; \\
\max\{||\lambda|| \mid \lambda \in \text{spec}(A) \setminus \{1\}\}, & \text{otherwise}
\end{cases}
$$

**Definition 1.** (Seminorms). A function $\|\cdot\| : \mathbb{R}^n \to \mathbb{R}_{\geq 0}$ is a seminorm on $\mathbb{R}^n$ if it satisfies the following properties:

- (homogeneity): $\|ax\| = |a| \|x\|$ for every $x \in \mathbb{R}^n$ and $a \in \mathbb{R}$;
- (subadditivity): $\|x + y\| \leq \|x\| + \|y\|$, for every $x, y \in \mathbb{R}^n$.

The kernel $W \subseteq \mathbb{R}^n$ of a seminorm $\|\cdot\| : \mathbb{R}^n \to \mathbb{R}_{\geq 0}$ is the subspace of vectors $w \in \mathbb{R}^n$ satisfying $\|w\| = 0$.

**Definition 2.** (equivalent norms) Let $\|\cdot\|_p : \mathbb{R}^k \to \mathbb{R}_{\geq 0}$ be the $\ell_p$-norm on $\mathbb{R}^k$ and $R \in \mathbb{R}^{k \times n}$. The $R$-weighted seminorm on $\mathbb{R}^n$ associated with the $\ell_p$-norm on $\mathbb{R}^k$ is

$$
\|x\|_{p,R} := \|Rx\|_p, \quad \forall x \in \mathbb{R}^n.
$$
Clearly, in this case the kernel $W$ of the seminorm coincides with $\ker R$.

In the following we show examples of two distinct weighted seminorms.

**Example 3** (Two weighted seminorms). Given a vector $x \in \mathbb{R}^n$ and $C_n \in \mathbb{R}^{n \times n}$, the $C_n^T$-weighted seminorm of $x$ associated with the $\ell_p$-norm is

$$\|x\|_{p,C_n^T} = \|C_n^T x\|_p = \left( \sum_{i,j} |x_i - x_j|^p \right)^{\frac{1}{p}}$$

and, for $p = \infty$,

$$\|x\|_{\infty,C_n^T} = \max_{i,j} |x_i - x_j|.$$

Similarly, the $\Pi_n$ weighted seminorm of $x$ associated with the $\ell_p$-norm is

$$\|x\|_{p,\Pi_n} = \|\Pi_n x\|_p = \left( \sum_i |x_i - \text{ave}(x)|^p \right)^{\frac{1}{p}}$$

and, for $p = \infty$,

$$\|x\|_{\infty,\Pi_n} = \max_i |x_i - \text{ave}(x)| = \max \{ \max_i x_i - \text{ave}(x), \text{ave}(x) - \min_i x_i \}.$$

Note that these two seminorms have the same kernel $W = (\Pi_n)$ and they are generally different.

Also recall that, for $p \in \mathbb{N}_e$, the induced $\ell_p$-norm on $\mathbb{R}^{m \times n}$ is defined as

$$\|A\|_p := \max_{\|x\|_p \leq 1} \|Ax\|_p.$$  \hspace{1cm} (1)

**Definition 4** (Induced matrix seminorm [7]). Given a seminorm $\|\| : \mathbb{R}^n \to \mathbb{R}_+ \geq 0$ with kernel $W \subseteq \mathbb{R}^n$, the induced seminorm on $\mathbb{R}^{n \times n}$ is defined as

$$\|A\| := \max_{\|x\| \leq 1} \|Ax\|.$$  \hspace{1cm} (2)

Induced matrix seminorms enjoy the conditional submultiplicativity property relative to the orthogonal complement to their kernel, i.e., for every $x \perp W$,

$$\|Ax\| \leq \|A\| \|x\|.$$  \hspace{1cm} (3)

Inequality (3) is generalized by the following lemma.

**Lemma 5** (Lemma 1 in [9]). Let $\|\| : \mathbb{R}^n \to \mathbb{R}_+ \geq 0$ be a seminorm on $\mathbb{R}^n$. Then, for every $A, B \in \mathbb{R}^{n \times n}$,

$$\|AB\| \leq \|A\| \|B\|.$$  \hspace{1cm} (4)

In the context of Markov chains it is of interest to compute the time required for the distance of the Markov chain to the stationary distribution to become sufficiently small [10].

**Definition 6** (Distance from the stationary distribution [10]). Given a primitive row-stochastic matrix $A \in \mathbb{R}^{n \times n}$ with dominant left eigenvector $\pi \in \Delta_n$, the distance of $A$ from the stationary distribution $\pi$ at the time instant $k \in \mathbb{N}$ is

$$d(A,k) := \frac{1}{2} \max_i \sum_{j=1}^n |(A^k)_{ij} - \pi_j|.$$  \hspace{1cm} (5)

Given two probability distributions $\xi, \nu \in \Delta_n$, the total variation distance between $\xi$ and $\nu$ [10] is

$$\|\xi - \nu\|_{\text{tv}} = \frac{1}{2} \sum_{j=1}^n |\xi_j - \nu_j|.$$  \hspace{1cm} (6)

With this notation, equation (5) reads:

$$d(A,k) = \max_{i} \|\{A^k\}_{ij} - \pi\|_{\text{tv}}.$$  \hspace{1cm} (7)

**Definition 7** (Mixing time). For $0 < \epsilon \ll 1$, the $\epsilon$-mixing time of a Markov chain is the time required for the distance to the stationary distribution to be less than or equal to $\epsilon$, namely

$$t_{\max}(A,\epsilon) := \min \{k \in \mathbb{N} | d(A,k) \leq \epsilon \}.$$  \hspace{1cm} (8)

**B. The three main concepts**

**Definition 8.** (Weighted $\ell_p$ induced seminorm [7]) Given the $\ell_p$-norm on $\mathbb{R}^k$, $\|\|_p : \mathbb{R}^k \to \mathbb{R}_+ \geq 0$ and $R \in \mathbb{R}^{k \times k}$,

Given $p \in \mathbb{N}_e$ and a matrix $R \in \mathbb{R}^{k \times k}$, the $R$-weighted induced seminorm on $\mathbb{R}^{n \times n}$, associated with the $\ell_p$-norm on $\mathbb{R}^k$, $\|\|_{p,R} : \mathbb{R}^n \to \mathbb{R}_+ \geq 0$, is

$$\|A\|_{p,R} := \max_{\|x\|_p \leq 1} \|Ax\|_{p,R}.$$  \hspace{1cm} (9)

Note that, generalizing the treatment in [7], we allow $k$ to be greater than or smaller than $n$. The $R$-weighted induced seminorm (6) is the operator norm [16] of the operator defined on the real (normed) linear space $W^\perp$ by $x \to R Ax$.

**Definition 9** ($\ell_p$ ergodicity coefficient [19]). For $p \in \mathbb{N}_e$, the $\ell_p$ ergodicity coefficient $\tau_p : \mathbb{R}^m \times \mathbb{R}^{m \times n} \to \mathbb{R}_+ \geq 0$ is defined as

$$\tau_p(v,A) := \max_{\|x\|_p \leq 1} \|A^\top x\|_p.$$  \hspace{1cm} (10)

The ergodicity coefficient (7) is the operator norm of the operator defined on the real (normed) linear space $\langle v \rangle^\perp$ by $x \to A^\top x$ [16].

For the special class of row-stochastic matrices we steadily assume $v = \mathbb{1}_n$ and hence the first argument in (7) is omitted. The $\ell_1$-norm ergodicity coefficient of a row stochastic matrix is also known in the literature as “Dobrushin coefficient” [6].

**Definition 10** ($\ell_p$ deflated induced norm [12]). For $p \in \mathbb{N}_e$, the $\ell_p$ deflated induced norm $\Psi_p : \mathbb{R}^m \times \mathbb{R}^{m \times n} \to \mathbb{R}_+ \geq 0$ is defined as

$$\Psi_p(v,A) := \min_{c \in \mathbb{R}^n} \|A - vc^\top\|_p.$$  \hspace{1cm} (11)

**C. Relevant prior results**

**Lemma 11** (Theorem 6.3 in [6]). Given a matrix $A \in \mathbb{R}^{m \times n}$ and a vector $v \in \mathbb{R}^m$, $v \neq 0$, then for all vectors $c \in \mathbb{R}^n$

$$\tau_p(v,A) \leq \|\langle A - vc^\top \rangle^\perp\|_p.$$  \hspace{1cm} (12)

By restricting our attention now to row stochastic matrices we introduce the following lemma.
Lemma 12 (Theorem 3.7 in [6]). If $A \in \mathbb{R}^{n \times n}$ is a row stochastic matrix, then

$$
\tau_1(A) = \frac{1}{2} \max_{i,j} \sum_{k=1}^{n} |[A]_{i,k} - [A]_{j,k}|
$$

$$= 1 - \min_{i,j} \sum_{k=1}^{n} \min\{[A]_{i,k}, [A]_{j,k}\}
$$

Lemma 13 (Proposition 7 in [2]). For a row stochastic matrix $A \in \mathbb{R}^{n \times n}$ it holds

$$\|A\|_{\infty,C_n^2} \leq 1 - \min_{i,j} \sum_{k=1}^{n} \min\{[A]_{i,k}, [A]_{j,k}\}
$$

III. RESULTS

A. Preliminary lemmas

As a preliminary result we provide a simplified expression for induced seminorms of a special class of matrices.

Lemma 14 (Simplified induced seminorm). Given a seminorm $\| \cdot \| : \mathbb{R}^n \to \mathbb{R}_{\geq 0}$ with kernel $\mathcal{W} \subseteq \mathbb{R}^n$, any matrix $A \in \mathbb{R}^{n \times n}$ such that $AW \subseteq \mathcal{W}$ satisfies

$$\|A\| = \max_{\|x\| \leq 1} \|Ax\|.
$$

Proof. Clearly,

$$\max_{\|x\| \leq 1} \|Ax\| \geq \max_{x \in \mathcal{W}} \|Ax\| = \|A\|.
$$

On the other hand, every $x \in \mathbb{R}^n$ can be expressed as $x = x_1 + x_2$, for some $x_1 \in \mathcal{W}$, $x_2 \in \mathcal{W}^\perp$, and condition $\|x\| \leq 1$ implies $\|x_2\| \leq 1$. Therefore

$$\|Ax\| = \|A(x_1 + x_2)\| \leq \|A\|_1 + \|A\|_2 = \|A\|_2
$$

where we used the fact that $Ax_1 \in \mathcal{W}$ and hence $\|Ax_1\| \leq 0$. Therefore (9) holds.

The following results are more or less known in the literature; we report a proof here for completeness sake.

Lemma 15 (Properties of oblique projectors). Given a row stochastic matrix $A \in \mathbb{R}^{n \times n}$, and any vector $w \in \mathbb{R}^n$, such that $w^\top A = w^\top$, $w^\top 1_n = 1$, and

(i) $A^k - 1_n w^\top = (A - 1_n w^\top)^k$ for all $k \in \mathbb{N}$,

(ii) $Q_w := I_n - 1_n w^\top$ is an oblique projector onto $\langle w \rangle^\perp$,

(iii) $w^\top Q_w = 0^\top_1$,

(iv) $AQ_w = Q_w A$,

(v) $A \|p,Q_w\| = \|Q_w A\|_p = \|Q_w A Q_w\|_p$.

Proof. To prove statement (i) one can proceed by induction and notice that, for $k = 1$, the equality trivially holds. So, by supposing the equality to be true for $k - 1$ one can show that it also holds true for $k$. In fact,

$$\left(A - 1_n w^\top\right)^k = (A - 1_n w^\top)(A - 1_n w^\top)^{k-1}
$$

$$= (A - 1_n w^\top)\left(A^{k-1} - 1_n w^\top\right)
$$

$$= A^k - 21_n w^\top + 1_n w^\top = A^k - 1_n w^\top.
$$

To prove statement (ii) we show that

$$Q_w^2 = (I_n - 1_n w^\top)(I_n - 1_n w^\top)
$$

$$= (I_n - 1_n w^\top) - 1_n w^\top + 1_n w^\top = Q_w,
$$

and this proves that $Q_w$ is an oblique projector. In order to prove that $Q_w$ is an oblique projector onto $\langle w \rangle^\perp$ we show that

$$x \perp w \iff x \in \text{Im} Q_w.
$$

In fact, if $x \perp w$, that is $w^\top x = 0$, then $x = (I_n - 1_n w^\top)x = Q_w x$, hence $x \in \text{Im} Q_w$.

On the other hand, if $x \in \text{Im} Q_w$, then there exists $z \in \mathbb{R}^n$ such that $x = Q_w z = (I_n - 1_n w^\top)z$. Then $w^\top x = w^\top(z - 1_n w^\top z) = w^\top z - w^\top z = 0$.

Statement (iii) follows from

$$w^\top Q_w = w^\top (I_n - 1_n w^\top) = w^\top - w^\top = 0^\top.
$$

Statement (iv) follows from

$$AQ_w = A(I_n - 1_n w^\top) = A - 1_n w^\top A = (I_n - 1_n w^\top)A = Q_w A.
$$

To prove statement (v), we notice that since the kernel $\mathcal{W} = \ker Q_w$, of the original seminorm is $A$-invariant, Lemma 14 applies, and therefore

$$\|A\|_{p,Q_w} = \max_{\|x\| \leq 1} \|Ax\|_{p,Q_w} = \max_{\|x\|_p,Q_w \leq 1} \|Q_w Ax\|_p
$$

$$= \max_{\|x\|_p,Q_w \leq 1} \|Q_w Ax\|_p
$$

and this concludes the proof.

The following lemma extends the result from Theorem 6 in [7], on the equivalence between induced weighted matrix seminorms and induced matrix norms, to square weight matrices that are rank-deficient and admit a special factorization.

Lemma 16 (Equivalence for weighted induced matrix seminorms). Let $v$ be a vector in $\mathbb{R}^n$, let $P_v$ be the orthogonal projector onto $\langle v \rangle^\perp$ and assume as weight matrix $R = SP_v$, with $S \in \mathbb{R}^{n \times n}$ non singular. Then for all $p \in \mathbb{N}$, and for all $A$ such that $AW \subseteq \mathcal{W}$, where $W = \langle v \rangle$ denotes the kernel of the seminorm $\| \cdot \|_{p,R}$, one has

$$\|A\|_{p,R} = \|RAR^\top\|_p.
$$

Proof. The proof follows from the fact that

$$\|RAR^\top\|_p = \max_{\|z\|_p \leq 1} \|SP_v AP_v S^{-1}z\|_p
$$

$$= \max_{\|x\|_p, z \leq 1} \|SP_v AP_v x\|_p
$$

where it has been exploited the fact the matrix $S$ is nonsingular and that $P_v$ is the orthogonal projection onto the orthogonal complement to $\ker R$ and hence the constraint $x \perp \ker R$ is equivalent to require $x$ to satisfy $x = P_v x$. The last equality follows from Lemma 14.
B. Main equivalence results

We are now ready to state our main result on the equivalence among ergodicity coefficients, deflated induced norms and weighted induced matrix seminorms.

**Theorem 17.** (Main equivalence) Given a matrix \( A \in \mathbb{R}^{n \times n} \), a right eigenvector of \( A \), \( v \in \mathbb{R}^n \), and integers \( p, q \in \mathbb{N} \) such that \( \frac{1}{p} + \frac{2}{q} = 1 \),

(i) the following main equivalence holds:

\[ \tau_p(v, A) = \Psi_q(v, A) = \|A\|_{q,p}, \]

(ii) \( c^* = \arg \min_{c \in \mathbb{R}^n} \|A - vc^T\|_q \), and

(iii) \( \Psi_q(v, A) = \|P_vA\|_q \).

**Proof.** We prove the equalities in statement (i) in two steps:

Step 1: \( \tau_p(v, A) = \Psi_q(v, A) \)

Step 2: \( \Psi_q(v, A) = \|P_vA\|_q \)

(Step 1): In order to prove that \( \tau_p(v, A) = \Psi_q(v, A) \), we first notice that, since \( \frac{1}{p} + \frac{2}{q} = 1 \), the duality property of the norm operator implies \( \|A\|_q = \|A^T\|_p \) and hence, \( \Psi_q(v, A) = \min_{c \in \mathbb{R}^n} \|\langle A - vc^T \rangle\|_p \).

So, to prove that \( \tau_p(v, A) = \Psi_q(v, A) \), we will show that \( \tau_p(v, A) = \min_{c \in \mathbb{R}^n} \|\langle A - vc^T \rangle\|_p \).

To do so, we exploit Lemma 11 and we show that there always exists a vector \( c^* \in \mathbb{R}^n \) for which the inequality (8) holds as an equality.

By Definition 9, \( \tau_p(v, A) := \max_{\|x\|_p \leq 1} \langle A^T x \rangle_p = \max_{\|x\|_p \leq 1} \langle (A - vc^T)^T x \rangle_p \)

for any \( c \in \mathbb{R}^n \). From Lemma 11 and from the definition of induced matrix norm (1), one gets

\[ \tau_p(v, A) = \max_{\|x\|_p \leq 1} \langle (A - vc^T)^T x \rangle_p \]

for any \( c \in \mathbb{R}^n \), and hence

\[ \frac{\|\langle A - vc^T \rangle\|_p}{\|x\|_p \leq 1}, \]

(10)

We will show that there always exist an optimal vector \( c^* \in \mathbb{R}^n \) for which the lower bound in (10) is reached.

One can always write the vector \( x \) in (10) as \( x = x_\perp + x_\parallel v \), with \( v^T x_\perp = 0 \) and \( x_\parallel \in \mathbb{R}^n \), therefore

\[ \|\langle A - vc^T \rangle\|_p = \|A^T x_\perp + A^T x_\parallel v - c \|v\|_2^2 \|x_\parallel \|_p \].

Let \( c^* \) be such that \( c^* \cdot \|v\|_2^2 \cdot x_\parallel = A^T x_\parallel v \), namely \( c^* = A^T \|v\|_2^{-2} \).

Then

\[ \min_{c \in \mathbb{R}^n} \|\langle A - vc^T \rangle\|_p = \|\langle A - vc^T \rangle\|_p |_{c = c^*} = \max_{\|x\|_p \leq 1} \langle (A - vc^T)^T x \rangle_p = \max_{\|x\|_p \leq 1} \|A^T x\|_p = \tau_p(v, A), \]

from which the equality in Step 1, together with statement (ii), follow.

We are now in the position to prove the second equality in statement (i).

(Step 2): From Step 1 it follows that \( c^* = \frac{A^T v}{\|v\|_2^2} \) is the vector that minimizes the \( q \)-norm of the matrix \( A - vc^T \), namely

\[ c^* = \frac{A^T v}{\|v\|_2^2} = \arg \min_{c \in \mathbb{R}^n} \|A - vc^T\|_q, \]

and that \( \Psi_q(v, A) = \|A - vc^T\|_q = \tau_p(v, A) \). Therefore, to prove that \( \Psi_q(v, A) = \|A\|_{q,p} \), we can show that

\[ \|A - vc^T\|_q = \|A\|_{q,p} \]

(12)

By plugging the expression (11) of \( c^* \), into (12), one gets

\[ \|A - \frac{v v^T}{\|v\|_2^2} A\|_q = \|I - \frac{v v^T}{\|v\|_2^2} A\|_q = \|P_v A\|_q \]

where \( P_v = I_n - \frac{v v^T}{\|v\|_2^2} \) is the orthogonal projector onto \( \langle v \rangle \).

Therefore, statement (iii) follows.

Finally, we show that \( \|P_v A\|_q = \|A\|_{q,p} \). From the definition of induced matrix norm (1), one gets

\[ \|P_v A\|_q = \max_{\|x\|_q \leq 1} \|P_v A x\|_q \]

Without loss of generality one can write the vector \( x \) in (13) as \( x = x_\perp + x_\parallel v \), with \( v^T x_\perp = 0 \) and \( x_\parallel \in \mathbb{R}^n \). From the fact that \( v \) is an eigenvector of \( A \), and hence \( A v = \lambda v \), for some \( \lambda \in \mathbb{R} \), also \( x_\perp = P_v x \) and \( P_v x = 0 \), it holds

\[ \max_{\|x\|_q \leq 1} \|P_v A x\|_q = \max_{\|y\|_q \leq 1} \|P_v A y\|_q = \max_{\|y\|_q \leq 1} \|P_v A y\|_q = \|A\|_{q,p} \]

where the last equality follows from a direct extension of Lemma 14, since the kernel of the seminorm is \( \text{ker} \ P_v \) and it is \( A \)-invariant. This concludes the proof of Step 2.

Both the first equality in statement (i), namely the fact that \( \tau_p(v, A) = \Psi_q(v, A) \), and statement (ii) from Theorem 17 hold true for generic pairs \((v, A) \in \mathbb{R}^n \times \mathbb{R}^{n \times n} \). Theorem 17 becomes of special interest in the context of averaging systems and Markov chains, which the next two corollaries refer to.

**Corollary 18 (Averaging Systems).** Given a sequence of \( n \times n \) row stochastic matrices \( \{A(k)\}_{k=0}^{\infty} \), \( x(0) \in \mathbb{R}^n \), \( p, q \in \mathbb{N} \) with \( \frac{1}{p} + \frac{1}{q} = 1 \), if \( s := \sup_k \tau_p(A(k)) < 1 \), then the system

\[ x(k + 1) = A(k) x(k) \]

is semi-contracting with rate \( s \) in the \( \Pi_p \) weighted \( \ell_q \) seminorm, namely \( \|x(k)\|_{p,\Pi_p} \leq s^k \|x(0)\|_{p,\Pi_p} \rightarrow 0 \) for \( k \rightarrow \infty \).

**Proof.** The proof follows from Theorem 17, for \( v = 1_n \) and hence \( P_v = \Pi_p \), and from the submultiplicative property of induced matrix seminorms (4):

\[ \|x(k + 1)\|_{p,\Pi_p} = \|A(k) x(k)\|_{p,\Pi_p} = \tau_q(A(k)) \|x(k)\|_{p,\Pi_p}. \]

**Corollary 19 (Markov chains).** Given a row-stochastic matrix \( A \in \mathbb{R}^{n \times n} \), with a (dominant) left eigenvector \( w \in \mathbb{R}^n \),
\[ \pi(0) \in \Delta_n, \ p, q \in \mathbb{N}_e \text{ with } \frac{1}{p} + \frac{1}{q} = 1, \text{ if } \tau_q(w, A^T) < 1, \text{ then the system} \]
\[ \pi(k + 1) = A^T \pi(k) \]
is semi-contracting in the \( P_w \) weighted \( \ell_p \) seminorm, namely \( \| \pi(k) \|_{p,P_w} \to 0 \) for \( k \to \infty \).

**Proof.** The result directly follows from Theorem 17, from which
\[
\| \pi(k + 1) \|_{p,P_w} \leq \| A^T \|_{p,P_w} \| \pi(k) \|_{p,P_w} = \tau_q(w, A^T) \| \pi(k) \|_{p,P_w}.
\]

For the special class of row stochastic matrices a second class of equivalences is also provided.

**Theorem 20** (Ergodicity coefficients and oblique projection). For a row stochastic matrix \( A \), with a dominant left eigenvector \( w, w^\top 1_n = 1 \), and \( p \in \mathbb{N}_e \), it holds that
\[
\tau_p(w, A^T) = \| A - 1_n w^\top \|_p = \| A \|_{p,Q_w},
\]
where \( Q_w := I - 1_n w^\top \) is an oblique projection matrix.

**Proof.** The result is proved in two steps

Step 1: \( \tau_p(w, A^T) = \| A - 1_n w^\top \|_p \)

Step 2: \( \tau_p(w, A^T) = \| A \|_{p,Q_w} \)

(Step 1): By Definition 9 it follows that
\[
\tau_p(w, A^T) = \max_{\| x \|_p \leq 1} \| A x \|_p
\]
and since each vector \( x \in \mathbb{R}^n \), \( x \perp w \), satisfies \( x = Q_w x \), one can write
\[
\tau_p(w, A^T) = \max_{\| x \|_p \leq 1} \| A x \|_p = \max_{\| x \|_p \leq 1} \| A Q_w x \|_p = \max_{\| x \|_p \leq 1} \| A Q_w x \|_p
\]
and, by exploiting statement (iv) from Lemma 15, one gets
\[
\tau_p(w, A^T) = \max_{\| x \|_p \leq 1} \| A Q_w x \|_p = \max_{\| x \|_p \leq 1} \| (A - 1_n w^\top) x \|_p
\]
from which the equality \( \tau_p(w, A^T) = \| A - 1_n w^\top \|_p \) directly follows.

(Step 2): Follows from identity (15) and from statement (iv) and (v) of Lemma 15 from which
\[
\tau_p(w, A^T) = \max_{\| x \|_p \leq 1} \| A Q_w x \|_p = \max_{\| x \|_p \leq 1} \| A Q_w x \|_p = \| A \|_{p,Q_w}.
\]

**C. Mixing time results**

**Lemma 21.** Given a Markov chain with primitive transition probability matrix \( A \in \mathbb{R}^{n \times n} \), its distance from the stationary distribution \( \pi \in \Delta_n \), is given by
\[
d(A, k) = \frac{1}{2} \tau_\infty(\pi, (A^k)^\top).
\]

**Proof.** From Definition 6 one can notice that the distance \( d(A, k) \) can be equivalently rewritten as
\[
d(A, k) = \frac{1}{2} \| A^k - 1_n \pi^\top \|_\infty
\]
and the identity \( \| A^k - 1_n \pi^\top \|_\infty = \tau_\infty(\pi, (A^k)^\top) \) follows directly from Theorem 20.

**Remark 22.** A direct consequence of Lemma 21 is that the \( \epsilon \)-mixing time of a Markov chain with transition probability matrix \( A \in \mathbb{R}^{n \times n} \) and stationary distribution \( \pi \in \Delta_n \), can be also expressed as
\[
t_{\text{max}}(A, \epsilon) = \min\{ k \in \mathbb{N} \mid \tau_\infty(\pi, (A^k)^\top) \leq 2\epsilon \}.
\]

**D. Efficient approximation of contraction rates**

We have seen how, when the system dynamics can be described by stochastic matrices, ergodicity coefficients and induced matrix seminorms provide upper bounds on the contraction factor of the system. In this context one defines the efficiency bound as the gap between the contraction factor \( s \) and its estimate \( \hat{s} \), namely \( |\hat{s} - s| \). The estimate \( \hat{s} \) is said to be \( \epsilon \) efficient if it is such that \( |\hat{s} - s| \leq \epsilon \) and optimal if \( \hat{s} = s \).

While, in general, ergodicity coefficients provide a fixed efficiency bound, we show how, in certain specific settings, by optimizing over the weights, induced (weighted) matrix seminorms provide optimal/efficient estimate for the contraction factor.

**Lemma 23** (Optimal induced seminorms). For a primitive matrix \( A \), with dominant right eigenvector \( v \),

(i) \( \rho_{\text{ess}}(A) = \inf\{ \| A \|_{p,R} \mid \ker R = \langle v \rangle, p \in \mathbb{N}_e \} \),

(ii) if \( A \) is also doubly stochastic with positive diagonal entries, then \( \tau_2(A) < 1 \).

Additionally, if \( A \) is diagonalizable, then \( \rho_{\text{ess}}(A) = \min\{ \| A \|_{p,R} \mid \ker R = \langle v \rangle, p \in \mathbb{N}_e \} \).

**Proof.** The proof of statement (i) is obtained in two steps.

Step 1: \( \rho_{\text{ess}}(A) \leq \| A \|_{p,R}, \forall p, \forall R \) such that \( \ker R = \langle v \rangle \)

Step 2: \( \forall \epsilon > 0 \exists p, R \) with \( \ker R = \langle v \rangle \) such that
\[
\| A \|_{p,R} \leq \rho_{\text{ess}}(A) + \epsilon
\]
(Step 1): We first recall that the kernel of the seminorm \( \| \cdot \|_{p,R} \) is \( \ker R \) and we assume that the eigenvalues of \( A \) are labelled in a way such that \( \lambda_1 > |\lambda_2| \geq \cdots \geq |\lambda_n| \). So we show that
\[
\| A^T \|_{p,R} = \max_{\| x \|_p \leq 1} \| A^T x \|_{p,R} = \max_{\| x \|_p \leq 1} \| A^T x \|_{p,R}
\]
\[
\geq |\lambda_2| \| w_2 \| = \rho_{\text{ess}}(A^T)
\]
with \( w_2 \) a second dominant left eigenvector of \( A \) (note: \( w_2 \perp v \)). From (17), Step 1 immediately follows.

(Step 2): We define \( D_\varepsilon := \text{diag}(1, 1/\varepsilon, \ldots, 1/\varepsilon^{n-1}) \), and let \( T \) be a matrix such that \( T(P_v A P_v)T^{-1} \) is in Jordan form and hence can be expressed as \( \Lambda + N \), where \( A \) is a diagonal matrix whose diagonal entries are the eigenvalues of \( P_v A P_v \) and \( N \) is an upper triangular matrix whose nonzero entries are the off-diagonal entries of \( T(P_v A P_v)T^{-1} \). Set \( R = D_\varepsilon TP_\varepsilon \) and note that \( \ker R = \langle v \rangle \). We also notice that \( P_v A = P_v A P_v \) and that, for a primitive, \( A, \rho_{\text{ess}}(A) = \rho(P_v A) \) and \( \text{spec}(P_v A) = \text{spec}(A) \setminus \{ \lambda_1 \} \cup \{0\} \).

By applying Lemma 16 with \( S = TD_\varepsilon \), for \( p = \infty \), one gets

\[
\|A\|_{\infty,R} = \|RAR^\top\|_\infty = \|D_\varepsilon TP_\varepsilon A P_\varepsilon T^{-1}D_\varepsilon^{-1}\|_\infty = (18)
\]

Finally, equation (18), together with the monotonicity property of the \( \ell_\infty \) norm (with respect to the absolute value of the argument), prove Step 2.

Moreover, if \( A \) is diagonalizable, the matrix \( N \) is the zero matrix and the inequality in (18) holds as an equality, hence \( \rho_{\text{ess}}(A) = \min\{\|A\|_p,R \mid \ker R = \langle v \rangle\} \).

Statement (ii) follows from Theorem 8 in [12].

E. Computational approaches

This section refers to special cases of interest for which induced matrix seminorms admit an explicit expression.

In the following we present an equivalence theorem with respect to the \( \ell_\infty \) weighted seminorm. The result holds only for \( p = \infty \), however numerical simulations suggest that the first equality in (19) also holds for \( p = 1,2 \).

**Theorem 24** (Equivalence for \( C_n^* \)-weighted, \( \ell_\infty \) induced seminorm). For a row stochastic matrix \( A \in \mathbb{R}^{n \times n} \),

\[
\|A\|_{\infty,C_n^*} = \|A\|_{\infty,\Pi_n} = \tau_1(A) = 1 - \min_{i,j} \min_{k=1}^n \{ |A|_{i,j,k} | A|_{j,k} \}.
\]

**Proof.** The equality \( \|A\|_{\infty,\Pi_n} = \tau_1(A) \) directly follows from Lemma 12 and from statement (i) in Theorem 17, for \( v = 1_n \). Therefore, in order to complete the proof we can show that \( \tau_1(A) = \|A\|_{\infty,C_n^*} \).

To do so we exploit the \( A \)-invariance of \( \ker C_n^* \) together with Lemma 12 and the inequality from Lemma 13, from which:

\[
\|A\|_{\infty,C_n^*} = \max_{\|x\|_{\infty,C_n^*}} \|Ax\|_{\infty,C_n^*} \leq \tau_1(A)
\]

and we will show that there always exists a vector \( x^* \) for which the inequality in (20) holds as an equality. In particular, by exploiting Lemma 12, we will show that

\[
\max_{\|x\|_{\infty,C_n^*}} \|Ax\|_{\infty,C_n^*} = \frac{1}{2} \sum_{k=1}^n |A|_{i,k} - |A|_{j,k}.
\]

In fact, by Definition 2 together with Example 3, it follows that

\[
\max_{\|x\|_{\infty,C_n^*}} \|Ax\|_{\infty,C_n^*} \leq \max_{\|x\|_{\infty,C_n^*}} \|e_i^\top - e_j^\top\|Ax\| \leq \max_{\|x\|_{\infty,C_n^*}} \|\sum_{k=1}^n (|A|_{i,k} - |A|_{j,k})x_k\| \leq \frac{1}{2} \max_{i,j} \sum_{k=1}^n |A|_{i,k} - |A|_{j,k},
\]

and the inequality in (21) holds as an equality for \( x^* \) such that \( x_k^* = \frac{1}{2} \text{sign}(|A|_{i,k} - |A|_{j,k}), k = \{1, \ldots, N\}, \) with \( (i,j) = \arg \max_{i,j} \sum_k |A|_{i,k} - |A|_{j,k} \) and this concludes the proof. □

**Conjecture 25.** For a row stochastic matrix \( A \in \mathbb{R}^{n \times n} \), and \( p = 1,2 \):

\[
\|A\|_{p,\Pi_n} = \|A\|_{p,C_n^*}.
\]

The following lemma provides a Linear Matrix Inequality (LMI) formulation for the \( P \)-weighted induced seminorm associated with the \( \ell_2 \) norm of a matrix \( A \), for the case in which the kernel of the seminorm is \( A \)-invariant. It can be interpreted as a generalization of the discrete time contractivity LMI [1].

**Lemma 26** (LMI formulation for weighted \( \ell_2 \) seminorms). Given a matrix \( A \in \mathbb{R}^{n \times n} \), having a real eigenvalue with largest absolute value and associated right eigenvector \( v \) and consider any \( P = P^\top \geq 0 \), such that \( \ker(P) = \langle v \rangle \), then

\[
\|A\|_{2,P,\frac{1}{2}} = \min \{ b \mid A^\top PA \leq bP \}.
\]

**Proof.** The proof directly follows from the fact that

\[
\|A\|_{2,P,\frac{1}{2}} = \max_{\|x\|_{2,P,\frac{1}{2}} \leq 1} \|Ax\|_{2,P,\frac{1}{2}} = \max_{x \in \ker(P)} \frac{x^\top A^\top PAx}{x^\top Px}.
\]

Therefore,

\[
\|A\|_{2,P,\frac{1}{2}} \leq b \quad \text{if and only if} \quad A^\top PA \leq bP,
\]

from which (23) directly follows. □

**Lemma 27.** For a primitive, symmetric matrix \( A \), with dominant right eigenvector \( v \),

\[
\|A\|_{2,P,\frac{1}{2}} = \rho_{\text{ess}}(A).
\]

**Proof.** Assume the real eigenvalues of \( A \) are numbered in such a way that \( |\lambda_1(A)| > \cdots \geq |\lambda_n(A)| \). From the definition of \( \|A\|_{2,P,\frac{1}{2}} \), from the Spectral theorem and from statements (i) and (ii) of Theorem 17, it holds that

\[
\|A\|_{2,P,\frac{1}{2}} = \|P_v A\|_{2} = \max_{\|x\|_{2} \leq 1} (x^\top A^\top P_v Ax)^{\frac{1}{2}} = \max_{x \neq 0} \left( \frac{x^\top A^\top P_v Ax}{x^\top x} \right)^{\frac{1}{2}} (25)
\]

where the last equality exploits the fact that \( P_{\frac{1}{2}} \). Moreover, since the matrix \( P_v \) is the orthogonal projection
onto the space orthogonal to the dominant eigenvector of $A$, $v$, it immediately follows that the the vector $x^*$ that maximizes (25) is the second dominant eigenvector of $A$, say $v_2$, in correspondence of which one gets

$$
\left( \frac{v_2^T A^T P_2 A v_2}{\|v_2\|^2} \right)^{\frac{1}{2}} = |\lambda_2| \tag{26}
$$

where the equality exploits the fact that $P_2 v_2 = v_2$. \hfill \Box

IV. CONCLUSIONS

The scientific literature related to ergodicity coefficients and induced norms is quite vast. However, the same can not be said about induced matrix seminorms. Induced matrix seminorms play an important role in the study of the stability of a system and its convergence behaviour. In this paper we attempt to fill a gap in the scientific literature by formalizing a connection among these algebraic tools. We show how induced matrix seminorms minimize the efficiency gap in the convergence rate of semi contractive dynamical systems, thus proving to be a more powerful tool, in certain specific settings, with respect to ergoficity coefficients. We have also provided an equivalent between two different weighted induced matrix seminorms and shown how seminorms and ergodicity coefficients come into play in the study of convergence to stationary distributions of Markov chains. Future works aim to apply our study in the context of coupled discrete time oscillators.

REFERENCES

[1] S. Boyd, L. El Ghaoui, E. Feron, and V. Balakrishnan. *Linear Matrix Inequalities in System and Control Theory*. SIAM, 1994, ISBN 089871334X.

[2] B. Charron-Bost. Orientation and connectivity based criteria for asymptotic consensus, March 2013. Available at https://arxiv.org/abs/1303.2043.

[3] R. L. Dobrushin. Central limit theorem for nonstationary Markov chains. I. *Theory of Probability & Its Applications*, 1(1):65–80, 1956. doi:10.1137/1101006.

[4] W. Doeblin. Le cas discontinu des probabilités en chaıne. *Publ. Faculty of Science University Masaryk (Brno)*, pages 3–13, 1937.

[5] F. Forni and R. Sepulchre. A differential Lyapunov framework for contraction analysis. *IEEE Transactions on Automatic Control*, 59(3):614–628, 2014. doi:10.1109/TAC.2013.2285771.

[6] I. C. F. Ipsen and T. M. Selee. Ergodicity coefficients defined by vector norms. *SIAM Journal on Matrix Analysis and Applications*, 32(1):153–200, 2011. doi:10.1137/090752948.

[7] S. Jafarpour, P. Cisneros-Velarde, and F. Bullo. Weak and semi-contraction for network systems and diffusively-coupled oscillators. *IEEE Transactions on Automatic Control*, 2021. To appear. Extended report with proofs at http://arxiv.org/abs/2005.09774. doi:10.1109/TAC.2021.3073096.

[8] A. N. Kolmogorov. Über die analytischen Methoden in der Wahrscheinlichkeitsrechnung. *Mathematische Annalen*, 104:415–158, 1931. doi:10.1007/BF01457949.

[9] V. V. Kolpakov. Matrix seminorms and related inequalities. *Journal of Soviet Mathematics*, 23:2094–2106, 1983. doi:10.1007/BF01093289.

[10] D. A. Levin and Y. Peres. *Markov Chains and Mixing Times*. American Mathematical Society, 2 edition, 2017, ISBN 978147042962. With contributions by E. L. Wilmer.

[11] J. Liu, A. S. Morse, B. D. O. Anderson, and C. Yu. Contraction for consensus processes. In *IEEE Conf. on Decision and Control and European Control Conference*, pages 1974–1979, 2011. doi:10.1109/CDC.2011.6160989.

[12] J. Liu, S. Mou, A. S. Morse, B. D. O. Anderson, and C. Yu. Deterministic gossiping. *Proceedings of the IEEE*, 99(9):1505–1524, 2011. doi:10.1109/JPROC.2011.2159689.

[13] W. Lohmiller and J.-J. E. Slotine. On contraction analysis for non-linear systems. *Automatica*, 34(6):683–696, 1998. doi:10.1016/S0005-1098(98)00019-3.

[14] A. A. Markov. Extensions of the law of large numbers to dependent quantities. *Izvestiia Fiz. - Matem. Obsch. Kazan Univ.*, 15, 1906. (in Russian).

[15] R. Marsli and F. J. Hall. Some properties of ergodicity coefficients with applications in spectral graph theory. *Linear and Multilinear Algebra*, 0(0):1–21, 2020. doi:10.1080/03081087.2020.1777251.

[16] U. G. Rothblum. Algebraic eigenspaces of nonnegative matrices. *Linear Algebra and its Applications*, 12(3):282–292, 1975. doi:10.1016/0024-3795(75)90050-6.

[17] E. Seneta. Coefficients of ergodicity: Structure and applications. *Advances in Applied Probability*, 11(3):576–590, 1979. doi:10.2307/1426955.

[18] E. Seneta. *Non-negative Matrices and Markov Chains*. Springer, 2 edition, 1981, ISBN 0387297650.

[19] E. Seneta. Explicit forms for ergodicity coefficients and spectrum localization. *Linear Algebra and its Applications*, 60:187–197, 1984. doi:10.1016/0024-3795(84)90079-X.

[20] W. Wang and J.-J. E. Slotine. On partial contraction analysis for coupled nonlinear oscillators. *Biological Cybernetics*, 92(1):38–53, 2005. doi:10.1007/s00422-004-0527-x.