ABSTRACT. We give a characterization of the contraction ratio of bounded linear maps in Banach space with respect to Hopf’s oscillation seminorm, which is the infinitesimal distance associated to Hilbert’s projective metric, in terms of the extreme points of a certain abstract “simplex”. The formula is then applied to abstract Markov operators defined on arbitrary cones, which extend the row stochastic matrices acting on the standard positive cone and the completely positive unital maps acting on the cone of positive semidefinite matrices. When applying our characterization to a stochastic matrix, we recover the formula of Dobrushin’s ergodicity coefficient. When applying our result to a completely positive unital map, we therefore obtain a noncommutative version of Dobrushin’s ergodicity coefficient, which gives the contraction ratio of the map (representing a quantum channel or a “noncommutative Markov chain”) with respect to the diameter of the spectrum. The contraction ratio of the dual operator (Kraus map) with respect to the total variation distance will be shown to be given by the same coefficient. We derive from the noncommutative Dobrushin’s ergodicity coefficient an algebraic characterization of the convergence of a noncommutative consensus system or equivalently the ergodicity of a noncommutative Markov chain.

1. INTRODUCTION

1.1. Motivation: from Birkhoff’s theorem to consensus dynamics. Hilbert’s projective metric $d_H$ on the interior of a (closed, convex, and pointed) cone $\mathcal{C}$ in a Banach space $\mathcal{X}$ can be defined by:

$$d_H(x, y) := \log \inf \{ \frac{\beta}{\alpha} : \alpha, \beta > 0, \alpha x \preceq y \preceq \beta x \},$$

where $\preceq$ is the partial order induced by $\mathcal{C}$. Birkhoff [Bir57] characterized the contraction ratio with respect to $d_H$ of a linear map $T$ preserving the interior $\mathcal{C}^0$ of the cone $\mathcal{C}$,

$$\sup_{x, y \in \mathcal{C}^0} \frac{d_H(Tx, Ty)}{d_H(x, y)} = \tanh \left( \frac{\text{diam} T(\mathcal{C}^0)}{4} \right), \quad \text{diam} T(\mathcal{C}^0) := \sup_{x, y \in \mathcal{C}^0} d_H(Tx, Ty).$$

This fundamental result, which implies that a linear map sending the cone $\mathcal{C}$ into its interior is a strict contraction in Hilbert’s metric, can be used to derive the Perron-Frobenius theorem from the Banach contraction mapping theorem, see [Bus73, KP82, EN95] for more information.

Hilbert’s projective metric is related to the following family of seminorms. To any point $e \in \mathcal{C}^0$ is associated the seminorm

$$x \mapsto \omega(x/e) := \inf \{ \beta - \alpha : \alpha e \preceq x \preceq \beta e \}$$

which is sometimes called Hopf’s oscillation [Hop63, Bus73] or Hilbert’s seminorm [GG04]. Nussbaum [Nus94] showed that $d_H$ is precisely the weak Finsler metric obtained when taking
\( \omega(\cdot / e) \) to be the infinitesimal distance at point \( e \). In other words,

\[
d_H(x, y) = \inf \gamma \int_0^1 \omega(\gamma(s)/\gamma(s))ds
\]

where the infimum is taken over piecewise \( C^1 \) paths \( \gamma : [0, 1] \rightarrow \mathcal{C}^0 \) such that \( \gamma(0) = x \) and \( \gamma(1) = y \). He deduced that the contraction ratio, with respect to Hilbert’s projective metric, of a nonlinear map \( f : \mathcal{C}^0 \rightarrow \mathcal{C}^0 \) that is positively homogeneous of degree 1 (i.e. \( f(\lambda x) = \lambda f(x) \) for all \( \lambda > 0 \)), can be expressed in terms of the Lipschitz constants of the linear maps \( Df(x) \) with respect to a family of Hopf’s oscillation seminorms:

\[
(1) \quad \sup_{x, y \in U} \frac{d_H(f(x), f(y))}{d_H(x, y)} = \sup_{x \in U} \frac{\omega(Df(x)z/f(x))}{\omega(z/x)}.
\]

Hence, to arrive at an explicit formula for the contraction rate of nonlinear maps in Hilbert’s projective metric, a basic issue is to determine the Lipschitz constant of a bounded linear map \( T : \mathcal{X} \rightarrow \mathcal{X}^* \) with respect to Hilbert’s oscillation seminorm, i.e.,

\[
(2) \quad \|T\|_H := \sup_{z \in \mathcal{X}, \omega(z/e) \neq 0} \frac{\omega(T(z)/T(e))}{\omega(z/e)}.
\]

The problem of computing the contraction rate \( (2) \) also arises in the study of consensus algorithms. A Markov operator is a linear map \( T \) which preserves the positive cone \( \mathcal{C} \) and fixes a unit element \( e \in \mathcal{C}^0 \):

\[
T(e) = e.
\]

A discrete time consensus system can be described by

\[
(3) \quad x_{k+1} = T_{k+1}(x_k), \quad k = 1, 2, \ldots,
\]

where \( T_1, T_2, \ldots \) is a sequence of Markov operators preserving the same unit element \( e \). The main concern of consensus theory is the convergence of the orbit \( x_k \) to a consensus state, which is represented by a scalar multiple of the unit element.

This model includes in particular the classical linear consensus system case when \( \mathcal{X} = \mathbb{R}^n \), \( \mathcal{C} = \mathbb{R}_+^n \), \( e = (1, \cdots, 1)^T \) and

\[
(4) \quad x_{k+1} = Ax_k, \quad k = 1, 2, \ldots,
\]

where \( A \) is a row stochastic matrix. This has been studied in the field of communication networks, control theory and parallel computation [Hir89, BT89, BGPS06, Mor05, VJAJ05, OT09, AB09]. A widely used Lyapunov function for the consensus dynamics, first considered by Tsi-}

\[
\Delta(x) = \max_{1 \leq i, j \leq n} (x_i - x_j),
\]

which is precisely Hopf’s oscillation seminorm \( \omega(x/e) \). It turns out that the latter seminorm can still be considered as a Lyapunov function for a Markov operator \( T \), with respect to an arbitrary cone. When \( \mathcal{C} = \mathbb{R}_+^n \), it is well known that if the contraction ratio of the stochastic matrix \( A \) with respect to the diameter is strictly less than one, then the orbits of the consensus dynamics \( (4) \) converge exponentially to a consensus state. We shall see here that the same remains true in general (Theorem 5.1).

For time-dependent consensus systems, a common approach is to bound the contraction ratio of every product of \( p \) consecutive operators \( T_{i+p} \circ \cdots \circ T_i \), \( i = 1, 2, \ldots \), for a fixed \( p \), see for example [Mor05]. Moreover, if \( \{T_k : k \geq 1\} \) is a stationary ergodic random process, then the almost sure convergence of the orbits of \( (3) \) to a consensus state can be deduced by showing that \( \mathbb{E}[\log \|T_{1+p} \cdots T_i\|_H] < 0 \) for some \( p > 0 \), see Bougerol [Bou93]. Hence, in consensus applications, a central issue is again to compute the contraction ratio \( (2) \).
1.2. Main results. Our first result characterizes the contraction ratio \(T\), in a slightly more general setting. We consider a bounded linear map \(T\) from a Banach space \(\mathcal{X}_1\) to a Banach space \(\mathcal{X}_2\). The latter are equipped with normal cones \(\mathcal{C}_i \subset \mathcal{X}_i\), and unit elements \(e_i \in \mathcal{C}_i^0\).

**Theorem 1.1** (Contraction rate in Hopf’s oscillation seminorm). Let \(T : \mathcal{X}_1 \to \mathcal{X}_2\) be a bounded linear map such that \(T(e_1) \in \mathbb{R}e_2\). Then

\[
\sup_{\omega(z/e_1) \neq 0} \frac{\omega(T(z)/e_2)}{\omega(z/e_1)} = \frac{1}{2} \sup_{\nu, \pi \in \mathcal{P}(\mathcal{E}(e_2))} \sup_{v, \pi \perp \nu} \|T^*(v) - T^*(\pi)\|_T = \sup_{v, \pi \in \mathcal{P}(\mathcal{E}(e_2))} \sup_{x \in [0, e_1]} \langle v - \pi, T(x) \rangle.
\]

The notations and notions used in this theorem are detailed in Section 5. In particular, we denote by the same symbol \(\preceq\) the order relations induced by the two cones \(\mathcal{C}_i, i = 1, 2; \mathcal{P}(\mathcal{E}_2) = \{\mu \in \mathcal{C}_2^\ast : \langle \mu, e_2 \rangle = 1\}\) denotes the abstract *simplices* of the dual Banach space \(\mathcal{X}_2^\ast\) of \(\mathcal{X}_2\), where \(\mathcal{C}_2^\ast\) is the dual cone of \(\mathcal{C}_2\); extr denotes the extreme points of a set; \(\perp\) denotes a certain *disjointness* relation, which will be seen to generalize the condition that two measures have disjoint supports; and \(T^*\) denotes the adjoint of \(T\). We shall make use of the following norm, which we call *Thompson’s norm*,

\[
\|z\|_T = \inf\{\alpha > 0 : -\alpha e_1 \preceq z \preceq \alpha e_1\}
\]
on the space \(\mathcal{X}_1\), and denote by \(\|\cdot\|_T^\ast\) the dual norm.

When \(\mathcal{C} = \mathbb{R}^n_+\), and \(T(z) = Az\) for some stochastic matrix \(A\), we shall see that the second supremum in Theorem 1.1 is simply

\[
\frac{1}{2} \max_{i < j} \sum_{1 \leq k \leq n} |A_{ik} - A_{jk}| = \frac{1}{2} \max_{i < j} \|A_i - A_j\|_{\ell_1} ,
\]

where \(A_i\) denotes the \(i\)th row of the matrix \(A\). This quantity is called *Dobrinsk contraction coefficient* in the theory of Markov chains; it is known to determine the contraction rate of the adjoint \(T^\ast\) with respect to the \(\ell_1\) (or total variation) metric, see [LPW09]. Moreover, the last supremum in Theorem 1.1 can be rewritten more explicitly as

\[
1 - \min_{i < j} \sum_{i=1}^n \min(A_{is}, A_{js}) ,
\]
a term which is known as *Dobrushin’s ergodicity coefficient* [Dob56]. Note that in general, the norm \(\|\cdot\|_T^\ast\) can be thought of as an abstract version of the \(\ell_1\) or total variation norm.

When specializing to a unital completely positive map \(T\) on the cone of positive semidefinite matrices, representing a quantum channel [SSR10, RKW11], we shall see that the last supremum in Theorem 1.1 coincides with the following expression, which provides a noncommutative analogue of Dobrushin’s ergodicity coefficient (see Corollary 8.1):

\[
1 - \min_{x=(x_1, \ldots, x_n)} \min_{u,v} \sum_{i=1}^n \min\{u^T(x_ix_i^\ast)u, v^T(x_ix_i^\ast)v\} .
\]

We use the above formula to show that the convergence of a noncommutative consensus system or equivalently the ergodicity of a noncommutative Markov chain can be characterized by the existence of a rank one matrix in certain subspace of matrices (Theorem 8.5 and 8.6).

2. Thompson’s norm and Hilbert’s seminorm

We start by some preliminary results. Throughout the paper, \((\mathcal{X}, \|\cdot\|)\) is a real Banach space. Denote by \(\mathcal{X}^\ast\) the dual space of \(\mathcal{X}\). For any \(x \in \mathcal{X}\) and \(q \in \mathcal{X}^\ast\), denote by \(\langle q, x \rangle\) the value of
Let $C \subset \mathcal{X}$ be a closed pointed convex cone, i.e., $\alpha C \subset C$ for $\alpha \in \mathbb{R}^+$, $C + C \subset C$ and $C \cap (-C) = 0$. The dual cone of $C$ is defined by

$$C^* = \{z \in \mathcal{X}^* : \langle z, x \rangle \geq 0 \ \forall x \in C\}.$$  

We denote by $C_0$ the interior of $C$. We define the partial order $\preceq$ induced by $C$ on $\mathcal{X}$ by

$$x \preceq y \iff y - x \in C,$$

so that

$$x \preceq y \Rightarrow \langle z, x \rangle \leq \langle z, y \rangle, \ \forall z \in C^*.$$  

We also define the relation $\prec$ by

$$x \prec y \iff y - x \in C_0.$$  

For $x \preceq y$ we define the order intervals:

$$[x, y] := \{z \in \mathcal{X} : x \preceq z \preceq y\}, \quad (x, y) := \{z \in \mathcal{X} : x \prec z \prec y\}.$$  

For $x \in \mathcal{X}$ and $y \in C_0$, following [Nus88], we define

$$M(x/y) := \inf\{t \in \mathbb{R} : x \preceq ty\}, \quad m(x/y) := \sup\{t \in \mathbb{R} : x \preceq ty\},$$

(5)

Observe that since $y \in C_0$, and since $C$ is closed and pointed, the two sets in (5) are non-empty, closed, and bounded from below and from above, respectively. In particular, $m$ and $M$ take finite values. For $x \in \mathcal{X}$ and $y \in C_0$, we call oscillation [Bus73] the difference between $M(x/y)$ and $m(x/y)$:

$$\omega(x/y) := M(x/y) - m(x/y).$$

Let $e$ denote a distinguished element in $C_0$, which we shall call a unit. For $x \in \mathcal{X}$, define

$$\|x\|_T := \max(M(x/e), -m(x/e))$$

which we call Thompson’s norm, with respect to the element $e$, and

$$\|x\|_H := \omega(x/e)$$

which we call Hilbert’s seminorm with respect to the element $e$.

Remark 2.1. These terminologies are motivated by the fact that Thompson’s part metric and Hilbert’s projective metric are Finsler metrics (see [Nus94]) for which the infinitesimal distances at the point $e \in C_0$ are respectively given by $\| \cdot \|_T$ and $\| \cdot \|_H$. The seminorm $\| \cdot \|_H$ is also called Hopf’s oscillation seminorm [Bus73].

We assume that the cone is normal, that is, there is a constant $K > 0$ such that

$$0 \preceq x \preceq y \Rightarrow \|x\| \leq K\|y\|.$$  

It is known that under this assumption the two norms $\| \cdot \|$ and $\| \cdot \|_T$ are equivalent, see [Nus94]. Therefore the space $\mathcal{X}$ equipped with the norm $\| \cdot \|_T$ is a Banach space. Since Thompson’s norm $\| \cdot \|_T$ is defined with respect to a particular element $e$, we write $(\mathcal{X}, e, \| \cdot \|_T)$ instead of $(\mathcal{X}, \| \cdot \|_T)$. By the definition and (5), Thompson’s norm can be rewritten by:

$$\|x\|_T = \sup_{z \in C^*} \frac{\langle z, x \rangle}{\langle z, e \rangle}.$$  

(6)
Remark 3.2. The dual space \( \mathcal{X}^{*} = \mathbb{R}^{n} \), the standard orthant cone \( C = \mathbb{R}^{n}_{+} \) and the unit vector \( e = 1 := (1, \ldots, 1)^{T} \). It can be checked that Thompson’s norm with respect to \( e \) is nothing but the sup norm
\[
\|x\|_{T} = \max_{i} |x_{i}| = \|x\|_{\infty},
\]
whereas Hilbert’s seminorm with respect to \( e \) is the so called diameter:
\[
\|x\|_{H} = \max_{1 \leq i, j \leq n} (x_{i} - x_{j}) = \Delta(x).
\]

Example 2.3. Let \( \mathcal{X} = S_{n} \), the space of Hermitian matrices of dimension \( n \) and \( C = S_{n}^{+} \), the cone of positive semidefinite matrices. Let the identity matrix \( I_{n} \) be the unit element: \( e = I_{n} \). Then Thompson’s norm with respect to \( I_{n} \) is nothing but the sup norm of the spectrum of \( X \), i.e.,
\[
\|X\|_{T} = \max_{1 \leq i, j \leq n} \lambda_{i}(X) = \|\lambda(X)\|_{\infty},
\]
where \( \lambda(X) := (\lambda_{1}(X), \ldots, \lambda_{n}(X)) \), is the vector of ordered eigenvalues of \( X \), counted with multiplicities, whereas Hilbert’s seminorm with respect to \( I_{n} \) is the diameter of the spectrum:
\[
\|X\|_{H} = \max_{1 \leq i, j \leq n} (\lambda_{i}(X) - \lambda_{j}(X)) = \Delta(\lambda(X)).
\]

3. Abstract Simplex in the Dual Space and Dual Unit Ball

We denote by \( (\mathcal{X}^{*}, e, \| \cdot \|_{T}) \) the dual space of \( (\mathcal{X}, e, \| \cdot \|_{T}) \) where the dual norm \( \| \cdot \|_{T} \) of a continuous linear functional \( z \in \mathcal{X}^{*} \) is defined by:
\[
\|z\|_{T} := \sup_{\|x\|_{T} = 1} \langle z, x \rangle.
\]

We define the abstract simplex in the dual space by:
\[
(7) \quad \mathcal{P}(e) := \{ \mu \in C^{*} \mid \langle \mu, e \rangle = 1 \}.
\]

Remark 3.1. For the standard orthant cone (Example 2.2, \( \mathcal{X} = \mathbb{R}^{n}, C = \mathbb{R}^{n}_{+} \) and \( e = 1 \)), the dual space \( \mathcal{X}^{*} \) is \( \mathcal{X} = \mathbb{R}^{n} \) itself and the dual norm \( \| \cdot \|_{T} \) is the \( \ell_{1} \) norm:
\[
\|x\|_{T} = \sum_{i} |x_{i}| = \|x\|_{1}.
\]

The abstract simplex \( \mathcal{P}(1) \) is the standard simplex in \( \mathbb{R}^{n} \):
\[
\mathcal{P}(1) = \{ v \in \mathbb{R}^{n}_{+} : \sum_{i} v_{i} = 1 \},
\]
i.e., the set of probability measures on the discrete space \( \{1, \ldots, n\} \).

Remark 3.2. For the cone of semidefinite matrices (Example 2.3, \( \mathcal{X} = S_{n}, C = S_{n}^{+} \) and \( e = I_{n} \)), the dual space \( \mathcal{X}^{*} \) is \( \mathcal{X} = S_{n} \) itself and the dual norm \( \| \cdot \|_{T} \) is the trace norm:
\[
\|X\|_{T} = \sum_{1 \leq i \leq n} |\lambda_{i}(X)| = \|X\|_{1}, \quad X \in S_{n}
\]

The simplex \( \mathcal{P}(I_{n}) \) is the set of positive semidefinite matrices with trace 1:
\[
\mathcal{P}(I_{n}) = \{ \rho \in S_{n}^{+} : \text{trace}(\rho) = 1 \}.
\]

The elements of this set are called density matrices in quantum physics. They are thought of as noncommutative analogues of probability measures.
We denote by $B^*_T(e)$ the dual unit ball:

$$B^*_T(e) = \{ x \in \mathcal{X}^* \mid \|x\|^*_T \leq 1 \} .$$

We denote by $\text{conv}(S)$ the convex hull of a set $S$. The next lemma relates the abstract simplex $\mathcal{P}(e)$ to the dual unit ball $B^*_T(e)$.

**Lemma 3.1.** The dual unit ball $B^*_T(e)$ of the space $(\mathcal{X}^*, e, \| \cdot \|_T)$, satisfies

$$B^*_T(e) = \text{conv}(\mathcal{P}(e) \cup -\mathcal{P}(e)).$$

**Proof.** For simplicity we write $\mathcal{P}$ instead of $\mathcal{P}(e)$ and $B^*_T$ instead of $B^*_T(e)$ in the proof. It follows from (6) that

$$\|x\|_T = \sup_{\mu \in \mathcal{P}} |\langle \mu, x \rangle| = \sup_{\mu \in \mathcal{P} \cup -\mathcal{P}} \langle \mu, x \rangle .$$

Hence $\|z\|_T^* \leq 1$ if and only if,

$$\langle z, x \rangle \leq \|x\|_T = \sup_{\mu \in \mathcal{P} \cup -\mathcal{P}} \langle \mu, x \rangle, \quad \forall x \in \mathcal{X} .$$

By the strong separation theorem [FHH’01, Thm 3.18], if $z$ did not belong to the closed convex hull $\text{conv}(\mathcal{P} \cup -\mathcal{P})$, the closure being understood in the weak star topology of $\mathcal{X}^*$, there would exist a vector $x \in \mathcal{X}$ and a scalar $\gamma$ such that

$$\langle z, x \rangle > \gamma \geq \langle \mu, x \rangle, \quad \forall \mu \in \mathcal{P} \cup -\mathcal{P} ,$$

contradicting (10). Hence,

$$B^*_T = \text{conv}(\mathcal{P} \cup -\mathcal{P}) .$$

We claim that the latter closure operation can be dispensed with. Indeed, by the Banach Alaoglu theorem, $B^*_T$ is weak-star compact. Hence, its subset $\mathcal{P}$, which is weak-star closed, is also weak-star compact. If $\mu \in B^*_T$, by the characterization of $B^*_T$ above, $\mu$ is a limit, in the weak star topology, of a net $\{\mu_a = s_a \nu_a - t_a \pi_a : a \in \mathcal{A}\}$ with $s_a + t_a = 1$, $s_a, t_a \geq 0$ and $\nu_a, \pi_a \in \mathcal{P}$ for all $a \in \mathcal{A}$. By passing to a subnet we can assume that $\{s_a, t_a : a \in \mathcal{A}\}$ converge respectively to $s, t \in [0, 1]$ such that $s + t = 1$ and $\{\nu_a, \pi_a : a \in \mathcal{A}\}$ converge respectively to $\nu, \pi \in \mathcal{P}$. It follows that $\mu = s\nu - t\pi \in \text{conv}(\mathcal{P} \cup -\mathcal{P})$.

**Remark 3.3.** We make a comparison with [RKW11]. In a finite dimensional setting, Reeb, Kastoryano, and Wolf defined a base $\mathcal{B}$ of a proper cone $\mathcal{K}$ in a vector space $\mathcal{V}$ to be a cross section of this cone, i.e., $\mathcal{B}$ is the intersection of the cone $\mathcal{K}$ with a hyperplane given by a linear functional in the interior of the dual cone $\mathcal{K}^*$. Their vector space $\mathcal{V}$ corresponds to our dual space $\mathcal{X}^*$, and, since $\mathcal{V}$ is of finite dimension, their dual space $\mathcal{V}^*$ corresponds to our primal space $\mathcal{X}$. Our cone $\mathcal{C} \subset \mathcal{K}$ corresponds to their dual cone $\mathcal{K}^*$. Modulo this identification, the base $\mathcal{B}$ can be written precisely as

$$\mathcal{B} = \{ \mu \in \mathcal{K} \mid \langle \mu, e \rangle = 1 \} ,$$

for some $e$ in the interior of $\mathcal{K}^*$, so that the base $\mathcal{B}$ coincides with our abstract simplex $\mathcal{P}(e)$. They defined the base norm of $\mu \in \mathcal{V}$ with respect to $\mathcal{B}$ by:

$$\|\mu\|_{\mathcal{B}} = \inf \{ \lambda > 0 : \mu \in \lambda \text{conv}(\mathcal{B} \cup -\mathcal{B}) \} .$$

They also defined the distinguishability norm of $\mu \in \mathcal{V}$ by:

$$\|\mu\|_{\tilde{M}} = \sup_{0 < x < e} \langle \mu, 2x - e \rangle .$$

And Theorem 14 in their paper [RKW11] states that the distinguishability norm is equal to the base norm:

$$\|\mu\|_{\tilde{M}} = \|\mu\|_{\mathcal{B}} .$$
In a finite dimensional setting, Lemma 3.1 is equivalent to the duality result (12) of Reeb et al. and the two approaches are dual to each other.

4. Characterization of Extreme Points of the Dual Unit Ball

The next lemma states that Hilbert’s seminorm coincides with the quotient norm on the quotient Banach space $\mathcal{X}/\mathbb{R}e$.

**Lemma 4.1.** For all $x \in \mathcal{X}$, we have:

$$\|x\|_H = 2 \inf_{\lambda \in \mathbb{R}} \|x + \lambda e\|_T$$

**Proof.** The expression

$$\|x + \lambda e\|_T = \max(\Gamma(x/e, \lambda), -\gamma(x/e) - \lambda)$$

is minimal when $\Gamma(x/e, \lambda) = -\gamma(x/e) - \lambda$. Substituting the value of $\lambda$ obtained in this way in $\|x + \lambda e\|_T$, we arrive at the announced formula. □

A standard result $[\text{Con90, P.88}]$ of functional analysis shows that if $W$ is a closed subspace of a Banach space $(\mathcal{X}, \| \cdot \|)$, then the quotient space $\mathcal{X}/W$ is complete. Besides, the dual of the quotient space $\mathcal{X}/W$ can be identified isometrically to the space of continuous linear forms on $\mathcal{X}$ that vanish on $W$, equipped with the dual norm $\| \cdot \|_H$. Specializing this result to $W = \mathbb{R}e$, we get:

**Lemma 4.2.** The quotient normed space $(\mathcal{X}/\mathbb{R}e, \| \cdot \|_H)$ is a Banach space. Its dual is $(M(e), \| \cdot \|_H^*)$ where

$$M(e) := \{ \mu \in \mathcal{X}^* | (\mu, e) = 0 \},$$

and

$$\|\mu\|_{H}^* := \frac{1}{2} \|\mu\|_T^*, \quad \forall \mu \in M(e).$$

The above lemma implies that the unit ball of the space $(M(e), \| \cdot \|_H^*)$, denoted by $B^*_H(e)$, satisfies:

$$B^*_H(e) = 2B^*_T(e) \cap M(e).$$

**Remark 4.1.** In the case of standard orthant cone $(\mathcal{X} = \mathbb{R}^n, C = \mathbb{R}^n_+ \text{ and } e = 1)$, Lemma 4.2 implies that for any two probability measures $\mu, \nu \in \mathcal{P}(1)$, the dual norm $\|\mu - \nu\|_H^*$ is the total variation distance between $\mu$ and $\nu$:

$$\|\mu - \nu\|_{H}^* = \frac{1}{2} \|\mu - \nu\|_1 = \|\mu - \nu\|_{TV}$$

Before giving a representation of the extreme points of $B^*_H(e)$, we define a *disjointness* relation $\perp$ on $\mathcal{P}(e)$.

**Definition 4.1.** For all $\nu, \pi \in \mathcal{P}(e)$, we say that $\nu$ and $\pi$ are *disjoint*, denoted by $\nu \perp \pi$, if

$$\mu = \frac{\nu + \pi}{2}$$

for all $\mu \in \mathcal{P}(e)$ such that $\mu \succ \frac{\nu}{2}$ and $\mu \succ \frac{\pi}{2}$.

We have the following characterization of the disjointness property.

**Lemma 4.3.** Let $\nu, \pi \in \mathcal{P}(e)$. The following assertions are equivalent:

(a) $\nu \perp \pi$. 

(b) The only elements \( \rho, \sigma \in \mathcal{P}(e) \) satisfying

\[
v - \pi = \rho - \sigma
\]

are \( \rho = \nu \) and \( \sigma = \pi \).

Proof. (a)⇒(b): Let any \( \rho, \sigma \in \mathcal{P}(e) \) such that

\[
v - \pi = \rho - \sigma.
\]

Then it is immediate that

\[
v + \sigma = \pi + \rho.
\]

Let \( \mu = \frac{\pi + \rho}{2} \). Then \( \mu \in \mathcal{P}(e) \), \( \mu \geq \frac{\nu}{2} \) and \( \mu \geq \frac{\pi}{2} \). Since \( \nu \perp \pi \), we obtain that \( \mu = \frac{\nu + \pi}{2} \).

(b)⇒(a): Let any \( \mu \in \mathcal{P}(e) \) such that \( \mu \geq \frac{\nu}{2} \) and \( \mu \geq \frac{\pi}{2} \). Then

\[
v - \pi = (2\mu - \pi) - (2\mu - \nu).
\]

From (b) we know that \( 2\mu - \pi = \nu \).

\[ \square \]

We denote by \( \text{extr}(\cdot) \) the set of extreme points of a convex set.

**Proposition 4.3.1.** The set of extreme points of \( B^*_H(e) \), denoted by \( \text{extr}B^*_H(e) \), is characterized by:

\[
\text{extr}B^*_H(e) = \{ \nu - \pi \mid \nu, \pi \in \text{extr}\mathcal{P}(e), \nu \perp \pi \}.
\]

Proof. It follows from (3) that every point \( \mu \in B^*_H(e) \) can be written as

\[
\mu = s\nu - t\pi
\]

with \( s + t = 1, s, t \geq 0, \nu, \pi \in \mathcal{P}(e) \). Moreover, if \( \mu \in \mathcal{M}(e) \), then

\[
0 = \langle \mu, e \rangle = s\langle \nu, e \rangle - t\langle \pi, e \rangle = s - t,
\]

thus \( s = t = \frac{1}{2} \). Therefore every \( \mu \in B^*_H(e) \cap \mathcal{M}(e) \) can be written as

\[
\mu = \frac{\nu - \pi}{2}, \nu, \pi \in \mathcal{P}(e).
\]

Therefore by (14) we proved that

\[
B^*_H(e) = \{ \nu - \pi : \nu, \pi \in \mathcal{P}(e) \}.
\]

Now let \( \nu, \pi \in \text{extr}\mathcal{P}(e) \) and \( \nu \perp \pi \). We are going to prove that \( \nu - \pi \in \text{extr}B^*_H(e) \). Let \( \nu_1, \pi_1, \nu_2, \pi_2 \in \mathcal{P}(e) \) such that

\[
\nu - \pi = \frac{\nu_1 - \pi_1}{2} + \frac{\nu_2 - \pi_2}{2}.
\]

Then

\[
\nu - \pi = \frac{\nu_1 + \nu_2}{2} - \frac{\pi_1 + \pi_2}{2}.
\]

By Lemma 4.3, the only possibility is \( 2\nu = \nu_1 + \nu_2 \) and \( 2\pi = \pi_1 + \pi_2 \). Since \( \nu, \pi \in \text{extr}\mathcal{P}(e) \) we obtain that \( \nu_1 = \nu_2 = \nu \) and \( \pi_1 = \pi_2 = \pi \). Therefore \( \nu - \pi \in \text{extr}B^*_H(e) \).

Now let \( \nu, \pi \in \mathcal{P}(e) \) such that \( \nu - \pi \in \text{extr}B^*_H(e) \). Assume by contradiction that \( \nu \) is not extreme in \( \mathcal{P}(e) \) (the case in which \( \pi \) is not extreme can be dealt with similarly). Then, we can find \( \nu_1, \nu_2 \in \mathcal{P}(e), \nu_1 \neq \nu_2 \), such that \( \nu = \frac{\nu_1 + \nu_2}{2} \). It follows that

\[
\mu = \frac{\nu_1 - \pi}{2} + \frac{\nu_2 - \pi}{2},
\]

where \( \nu_1 - \pi, \nu_2 - \pi \) are distinct elements of \( B^*_H(e) \), which is a contradiction. Next we show that \( \nu \perp \pi \). To this end, let any \( \rho, \sigma \in \mathcal{P}(e) \) such that

\[
v - \pi = \rho - \sigma.
\]
Then
\[ \nu - \pi = \frac{\nu - \pi + \rho - \sigma}{2} = \frac{\nu - \sigma}{2} + \frac{\rho - \pi}{2}. \]

If \( \sigma \neq \pi \), then \( \nu - \sigma \neq \nu - \pi \) and this contradicts the fact that \( \nu - \pi \) is extremal. Therefore \( \sigma = \pi \) and \( \rho = \nu \). From Lemma 4.3 we deduce that \( \nu \perp \pi \).

**Remark 4.2.** In the case of standard orthant cone \( (\mathcal{X} = \mathbb{R}^n, \mathcal{C} = \mathbb{R}^n_+ \) and \( e = 1 ) \), the set of extreme points of \( \mathcal{P}(1) \) is the set of standard basis vectors \( \{e_i\}_{i=1,\ldots,n} \). The extreme points are pairwise disjoint.

**Remark 4.3.** In the case of cone of semidefinite matrices \( (\mathcal{X} = \mathcal{S}_n, \mathcal{C} = \mathcal{S}_n^+ \) and \( e = I_n ) \), the set of extreme points of \( \mathcal{P}(I_n) \) is
\[ \text{extr} \mathcal{P}(I_n) = \{ xx^* \mid x \in \mathbb{C}^n, x^* x = 1 \}, \]
which are called pure states in quantum information terminology. Two extreme points \( xx^* \) and \( yy^* \) are disjoint if and only if \( x^* y = 0 \). To see this, note that if \( x^* y = 0 \) then any Hermitian matrix \( X \) such that \( X \succeq xx^* \) and \( X \succeq yy^* \) should satisfy \( X \succeq xx^* + yy^* \). Hence by definition \( xx^* \) and \( yy^* \) are disjoint. Inversely, suppose that \( xx^* \) and \( yy^* \) are disjoint and consider the spectral decomposition of the matrix \( xx^* - yy^* \), i.e., there is \( \lambda \leq 1 \) and two orthonormal vectors \( u, v \) such that \( xx^* - yy^* = \lambda (uu^* - vv^*) \). It follows that \( xx^* - yy^* = uu^* - ((1 - \lambda)uu^* + \lambda vv^*) \). By Lemma 4.3 the only possibility is \( yy^* = (1 - \lambda)uu^* + \lambda vv^* \) and \( xx^* = uu^* \) thus \( \lambda = 1, u = x \) and \( v = y \). Therefore \( x^* y = 0 \).

5. THE OPERATOR NORM INDUCED BY HOPF’S OSCILLATION SEMINORM

Consider two real Banach spaces \( \mathcal{X}_1 \) and \( \mathcal{X}_2 \). Let \( \mathcal{E}_1 \subset \mathcal{X}_1 \) and \( \mathcal{E}_2 \subset \mathcal{X}_2 \) be respectively two closed pointed convex normal cones with non empty interiors \( \mathcal{E}_1^0 \) and \( \mathcal{E}_2^0 \). Let \( e_1 \in \mathcal{E}_1^0 \) and \( e_2 \in \mathcal{E}_2^0 \). Then, we know from Section 4 that the two quotient spaces \( (\mathcal{X}_1 / \mathbb{R} e_1, \| \cdot \|_H) \) and \( (\mathcal{X}_2 / \mathbb{R} e_2, \| \cdot \|_H) \) are Banach spaces. The dual spaces of \( (\mathcal{X}_1 / \mathbb{R} e_1, \| \cdot \|_H) \) and \( (\mathcal{X}_2 / \mathbb{R} e_2, \| \cdot \|_H) \) are respectively \( (\mathcal{M}(e_1), \| \cdot \|_H^*) \) and \( (\mathcal{M}(e_2), \| \cdot \|_H^*) \) (see Lemma 4.2).

Let \( T \) denote a continuous linear map from \( (\mathcal{X}_1 / \mathbb{R} e_1, \| \cdot \|_H) \) to \( (\mathcal{X}_2 / \mathbb{R} e_2, \| \cdot \|_H) \). The operator norm of \( T \), denoted by \( \| T \|_H \), is given by:
\[ \| T \|_H := \sup_{\| x \|_H=1} \| T(x) \|_H = \sup \frac{\omega(T(x)/e_2)}{\omega(x/e_1)}. \]

By definition, the adjoint operator \( T^* : (\mathcal{M}(e_2), \| \cdot \|_H^*) \to (\mathcal{M}(e_1), \| \cdot \|_H^*) \) of \( T \) is:
\[ \langle T^*(\mu), x \rangle = \langle \mu, T(x) \rangle, \quad \forall \mu \in \mathcal{M}(e_2), x \in \mathcal{X}_1 / \mathbb{R} e_1. \]

The operator norm of \( T^* \), denoted by \( \| T^* \|_{H^*} \), is then:
\[ \| T^* \|_{H^*}^* := \sup_{\mu \in B_{\mathcal{M}(e_2)}} \| T^*(\mu) \|_{H^*}. \]

A classical duality result (see [AB99], § 6.8)) shows that an operator and its adjoint have the same operator norm. In particular,
\[ \| T \|_H = \| T^* \|_{H^*}. \]

**Theorem 5.1.** Let \( T : \mathcal{X}_1 \to \mathcal{X}_2 \) be a bounded linear map such that \( T(e_1) \subset \mathbb{R} e_2 \). Then,
\[ \| T \|_H = \frac{1}{2} \sup_{v, \pi \in \mathcal{P}(e_2)} \| T^*(v) - T^*(\pi) \|_{H^*}^* = \sup_{v, \pi \in \mathcal{P}(e_2)} \sup_{x \in [0, e_1]} \langle v - \pi, T(x) \rangle. \]
Moreover, the supremum can be restricted to the set of extreme points:

\[(16) \quad \|T\|_H = \frac{1}{2} \sup_{v, \pi \in \text{extr} \mathcal{P}(e_2)} \|T^*(v) - T^*(\pi)\|_T^* = \sup_{v, \pi \in \text{extr} \mathcal{P}(e_2)} \sup_{x \in \{0, e_1\}} \langle v - \pi, T(x) \rangle.\]

**Proof.** We already noted that \(\|T\|_H = \|T^*\|_T^*\). Moreover,

\(\|T^*\|_T^* = \sup_{\mu \in B_H^*(e_2)} \|T^*(\mu)\|_H^*\).

By the characterization of \(B_H^*(e_2)\) in (15) and the characterization of the norm \(\cdot \|_T^*\) in Lemma 4.2, we get

\[
\sup_{\mu \in B_H^*(e_2)} \|T^*(\mu)\|_H^* = \sup_{v, \pi \in \mathcal{P}(e_2)} \|T^*(v) - T^*(\pi)\|_H^* = \frac{1}{2} \sup_{v, \pi \in \mathcal{P}(e_2)} \|T^*(v) - T^*(\pi)\|_T^*. 
\]

For the second equality, note that

\[
\|T^*(v) - T^*(\pi)\|_T^* = \sup_{x \in [0, e_1]} \langle T^*(v) - T^*(\pi), 2x - e_1 \rangle = 2 \sup_{x \in [0, e_1]} \langle T^*(v) - T^*(\pi), x \rangle.
\]

We next show that the supremum can be restricted to the set of extreme points. By the Banach-Alaoglu theorem, \(B_H^*(e_2)\) is weak-star compact, and it is obviously convex. The dual space \(\mathcal{M}(e_2)\) endowed with the weak-star topology is a locally convex topological space. Thus by the Krein-Milman theorem, the unit ball \(B_H^*(e_2)\), which is a compact convex set in \(\mathcal{M}(e_2)\) with respect to the weak-star topology, is the closed convex hull of its extreme points. So every element \(\rho\) of \(B_H^*(e_2)\) is the limit of a net \((\rho_\alpha)_\alpha\) of elements in \(\text{conv} \{\text{extr} B_H^*(e_2)\}\). Observe now that the function

\[\varphi : \mu \mapsto \|T^*(\mu)\|_H^* = \sup_{x \in B_H(e_1)} \langle T^*(\mu), x \rangle = \sup_{x \in B_H(e_1)} \langle \mu, T(x) \rangle\]

which is a sup of weak-star continuous maps is convex and weak-star lower semi-continuous. This implies that

\[\varphi(\rho) \leq \liminf_{\alpha} \varphi(\rho_\alpha) \leq \sup \{\varphi(\mu) : \mu \in \text{conv} \{\text{extr} B_H^*(e_2)\}\} = \sup \{\varphi(\mu) : \mu \in \text{extr} B_H^*(e_2)\}.
\]

Using the characterization of the extreme points in Proposition 4.3.1, we get:

\[\sup_{\mu \in B_H^*(e_2)} \|T^*(\mu)\|_H^* = \sup_{\mu \in \text{extr} B_H^*(e_2)} \|T^*(\mu)\|_H^* = \sup_{v, \pi \in \text{extr} \mathcal{P}(e_2)} \sup_{x \in \{0, e_1\}} \langle v - \pi, T(x) \rangle. \]

**Remark 5.1.** When \(\mathcal{D}_1\) is of finite dimension, the set \([0, e_1]\) is the convex hull of the set of its extreme points, hence, the supremum over the variable \(x \in [0, e_1]\) in (16) is attained at an extreme point. Similarly, if \(\mathcal{D}_2\) is of finite dimension, the suprema over \((v, \pi)\) in the same equation are also attained, because the map \(\varphi\) in the proof of the previous theorem, which is a supremum of an equi-Lipschitz family of maps, is continuous (in fact, Lipschitz).

**Remark 5.2.** Theorem 5.1 should be compared with Proposition 12 of [RKW11] which can be stated as follows.
**Proposition 5.1.1** (Proposition 12 in [RKW11]). Let \( \mathcal{V}, \mathcal{V}' \) be two finite dimensional vector spaces and \( L : \mathcal{V} \to \mathcal{V}' \) be a linear map and let \( \mathcal{B} \subset \mathcal{V} \) and \( \mathcal{B}' \subset \mathcal{V}' \) be bases. Then

\[
\sup_{v_1 \neq v_2 \in \mathcal{B}} \frac{\|L(v_1) - L(v_2)\|_{\mathcal{B}'}}{\|v_1 - v_2\|_{\mathcal{B}}} = \frac{1}{2} \sup_{v_1, v_2 \in \text{extr } \mathcal{B}} \frac{\|L(v_1) - L(v_2)\|_{\mathcal{B}'}}{\|v_1 - v_2\|_{\mathcal{B}}} \tag{17}
\]

The first term in (17) is called the **contraction ratio** of the linear map \( L \), with respect to the base norms (see Remark 5.3). One important application of this proposition concerns the **base preserving** maps \( L \) such that \( L(\mathcal{B}) \subset \mathcal{B}' \). Let us translate this proposition in the present setting.

Consider a linear map \( T : \mathcal{X}_1 / \mathbb{R} e_1 \to \mathcal{X}_2 / \mathbb{R} e_2 \). Then \( T^* : \mathcal{X}_2^* \to \mathcal{X}_1^* \) is a base preserving linear map \( (T^*(\mathcal{P}(e_2)) \subset \mathcal{P}(e_1)) \) and so, Proposition 12 of [RKW11] shows that:

\[
\sup_{v, \pi \in \mathcal{P}(e_2)} \frac{\|T^*(v - \pi)\|_T}{\|v - \pi\|_T} = \frac{1}{2} \sup_{v, \pi \in \text{extr } \mathcal{P}(e_2)} \frac{\|T^*(v) - T^*(\pi)\|_T}{\|T^*(v)\|_T} \tag{18}
\]

Hence, by comparison with [RKW11], the additional information here is the equality between the contraction ratio in Hilbert’s seminorm of a unit preserving linear map, and the contraction ratio with respect to the base norms of the dual base preserving map. The latter is the primary object of interest in quantum information theory whereas the former is of interest in the control/consensus literature. We also proved that the supremum in (18) can be restricted to pairs of **disjoint** extreme points \( v, \pi \). Finally, the expression of the contraction rate as the last supremum in Theorem 5.1.1 leads here to an abstract version of Dobrushin’s ergodic coefficient, see Eqn (23) and Corollary 8.1 below.

Let us recall the definition of Hilbert’s projective metric.

**Definition 5.1** ([Bir57]). Hilbert’s projective metric between two elements \( x \) and \( y \) of \( \mathcal{C}_0 \) is

\[
d_H(x, y) = \log(M(x/y)/m(x/y)). \tag{19}
\]

Consider a linear operator \( T : \mathcal{X}_1 \to \mathcal{X}_2 \) such that \( T(\mathcal{C}_0) \subset \mathcal{C}_0 \). Following [Bir57], [Bus73], the **projective diameter** of \( T \) is defined as below:

\[
diam T = \sup\{d_H(T(x), T(y)) : x, y \in \mathcal{C}_0 \}. \tag{18}
\]

Birkhoff’s contraction formula [Bir57], [Bus73] states that the oscillation ratio equals to the contraction ratio of \( T \) and they are related to its projective diameter.

**Theorem 5.2** ([Bir57], [Bus73]).

\[
\sup_{x, y \in \mathcal{C}_0} \frac{\omega(T(x)/T(y))}{\omega(x/y)} = \sup_{x, y \in \mathcal{C}_0} \frac{d_H(T(x), T(y))}{d_H(x, y)} = \tanh\left(\frac{\text{diam } T}{4}\right). \tag{19}
\]

The projective diameter of \( T^* \) is defined by:

\[
diam T^* = \sup\{d_H(T^*(u), T^*(v)) : u, v \in \mathcal{C}_2^* \setminus \{0\}\}. \tag{19}
\]

Note that \( \text{diam } T = \text{diam } T^* \). This is because

\[
\sup_{x, y \in \mathcal{C}_1} \frac{M(T(x)/T(y))}{m(T(x)/T(y))} = \sup_{x, y \in \mathcal{C}_1} \frac{\langle u, T(x) \rangle \langle v, T(y) \rangle}{\langle u, T(y) \rangle \langle v, T(x) \rangle} \tag{19}
\]

\[
= \sup_{u, v \in \mathcal{C}_2^* \setminus \{0\}} \frac{M(T(u)/T(v))}{m(T(u)/T(v))}. \tag{19}
\]

**Corollary 5.3** (Compare with [RKW11]). Let \( T : \mathcal{X}_1 \to \mathcal{X}_2 \) be a bounded linear map such that \( T(e_1) \in \mathbb{R} e_2 \) and \( T(e_1^0) \subset \mathcal{C}_2^0 \), then:

\[
\|T^*\|_H = \|T\|_H \leq \tanh\left(\frac{\text{diam } T}{4}\right) = \tanh\left(\frac{\text{diam } T^*}{4}\right). \tag{19}
\]
Proof. It is sufficient to prove the inequality. For this, note that
\[ \|T\|_H = \sup_{x \in \mathcal{X}} \frac{\omega(T(x)/e_2)}{\omega(x/e_1)} = \sup_{x \in \mathcal{X}} \frac{\omega(T(x)/e_2)}{\omega(x/e_1)}. \]
Then we apply Birkhoff’s contraction formula. \(\square\)

Remark 5.3. Reeb et al. [RKW11] showed in a different way that
\[ \|T^*\|_H^* \leq \tanh\left(\frac{\text{diam} T^*}{4}\right), \]
in a finite dimensional setting. Corollary 5.3 shows that as soon as the duality formula \(\|T^*\|_H^* = \|T\|_H\) is established, the latter inequality follows from Birkhoff’s contraction formula.

6. Application to discrete Markov operators on cones

A classical result, which goes back to Doeblin and Dobrushin, characterizes the Lipschitz constant of a Markov matrix acting on the space of measures (i.e., a row stochastic matrix acting on the left), with respect to the total variation norm (see the discussion in Section 7 below). The same constant characterizes the contraction ratio with respect to the “diameter” (Hilbert’s seminorm) of the consensus system driven by this Markov matrix (i.e., a row stochastic matrix acting on the right). Markov operators on cones extend Markov matrices. In this section, we extend to these abstract operators a number of known properties of Markov matrices.

A bounded linear map \(T : \mathcal{X} \to \mathcal{X}\) is a Markov operator with respect to a unit vector \(e\) in the interior \(\mathcal{C}^0\) of a closed convex pointed cone \(\mathcal{C} \subset \mathcal{X}\) if it satisfies the two following properties:

(i) \(T\) is positive, i.e., \(T(\mathcal{C}) \subset \mathcal{C}\).

(ii) \(T\) preserves the unit element \(e\), i.e., \(T(e) = e\).

The case when \(\|T\|_H < 1\) or equivalently \(\|T^*\|_H^* < 1\) is of special interest; the following theorem shows that the iterates of \(T\) converge to a rank one projector with a rate bounded by \(\|T\|_H\).

Theorem 6.1 (Geometric convergence to consensus/invariant measure). Let \(T : \mathcal{X} \to \mathcal{X}\) be a Markov operator with respect to the unit element \(e\). If \(\|T\|_H < 1\) or equivalently \(\|T^*\|_H^* < 1\), then there is \(\pi \in \mathcal{P}(e)\) such that for all \(x \in \mathcal{X}\)
\[ \|T^n(x) - (\pi,x)e\|_T \leq (\|T\|_H)^n \|x\|_H, \]
and for all \(\mu \in \mathcal{P}(e)\)
\[ \|(T^*)^n(\mu) - \pi\|^*_H \leq (\|T\|_H)^n. \]

Proof. The intersection
\[ \cap_n [m(T^n(x)/e), M(T^n(x)/e)] \subset \mathbb{R} \]
is nonempty (as a non-increasing intersection of nonempty compact sets), and since \(\|T\|_H < 1\) and
\[ \omega(T^n(x)/e) \leq (\|T\|_H)^n \omega(x/e), \]
this intersection must be reduced to a real number \(\{c(x)\} \subset \mathbb{R}\) depending on \(x\), i.e.,
\[ c(x) = \cap_n [m(T^n(x)/e), M(T^n(x)/e)]. \]
Thus for all \(n \in \mathbb{N}\),
\[ -\omega(T^n(x)/e)e \leq T^n(x) - c(x)e \leq \omega(T^n(x)/e)e. \]
Therefore by definition:
\[ \|T^n(x) - c(x)e\|_T \leq \omega(T^n(x)/e)e \leq (\|T\|_H)^n \|x\|_H. \]
It is immediate that:
\[ c(x)e = \lim_{n \to \infty} T^n(x) \]
from which we deduce that \( c : \mathcal{X} \to \mathbb{R} \) is a continuous linear functional. Thus there is \( \pi \in \mathcal{X}^* \) such that \( c(x) = \langle \pi, x \rangle \). Besides it is immediate that \( \langle \pi, e \rangle = 1 \) and \( \pi \in \mathcal{C}^* \) because \( x \in C \Rightarrow c(x) e \in C \Rightarrow c(x) \geq 0 \Rightarrow \langle \pi, x \rangle \geq 0 \).

Therefore \( \pi \in \mathcal{P}(e) \). Finally for all \( \mu \in \mathcal{P}(e) \) and all \( x \in \mathcal{X} \) we have
\[ \langle (T^*)^n(\mu) - \pi, x \rangle = \langle \mu, T^n(x) - \langle \pi, x \rangle e \rangle \leq \|\mu\| T^n(x) - \langle \pi, x \rangle e \|_T \leq (\|T\|_H)^n \|x\|_H. \]

Hence
\[ \| (T^*)^n(\mu) - \pi \|_H \leq (\|T\|_H)^n. \]

A time invariant discrete time consensus system can be described by
\[ x_{k+1} = T(x_k), \quad k = 1, 2, \ldots \quad (20) \]
The main concern of consensus theory is the convergence of the orbit \( x_k \) to a consensus state, which is represented by a scalar multiple of the unit element \( e \). The dual system of (20) represents a homogeneous discrete time Markov system:
\[ \pi_{k+1} = T^*(\pi_k), \quad k = 1, 2, \ldots \quad (21) \]

One of the central issues in Markov chain study is the strong ergodicity, i.e., the convergence of the distribution \( \pi_k \) to an invariant measure, given by a fixed point of \( T^* \). Theorem 6.1 shows that if \( \|T\|_H < 1 \) or equivalently \( \|T^*\|_H < 1 \), then the consensus system (20) is globally convergent and the homogeneous Markov chain (21) is strongly ergodic.

A time-dependent consensus system is described by
\[ x_{k+1} = T_{k+1}(x_k), \quad k = 1, 2, \ldots \quad (22) \]
where \( \{T_k : k \geq 1\} \) is a sequence of Markov operators sharing a common unit element \( e \in \mathcal{C}^0 \). Then if there is an integer \( p > 0 \) and a constant \( \alpha < 1 \) such that for all \( i \in \mathbb{N} \)
\[ \|T_{i+p} \cdots T_{i+1}\|_H \leq \alpha, \]
then the same lines of proof of Theorem 6.1 imply the existence of \( \pi \in \mathcal{P}(e) \) such that for all \( \{x_k\} \) satisfying (22),
\[ \|x_k - \langle \pi, x_0 \rangle e\|_T \leq \alpha^{\frac{k}{p}} \|x_0\|_H, \quad n \in \mathbb{N}. \]
Moreover, if \( \{T_k : k \geq 1\} \) is a stationary ergodic random process, then the almost sure convergence of the orbits of (22) to a consensus state can be deduced by showing that
\[ \mathbb{E}[\log \|T_{i+p} \cdots T_i\|_H] < 0 \]
for some \( p > 0 \), see Bougerol [Bou93]. The ergodicity of a inhomogeneous Markov chain can be studied in a dual approach. Hence, in Markov chain and consensus applications, a central issue is to compute the operator norm \( \|T\|_H \) of a Markov operator \( T \).

A direct application of Theorem 5.1 leads to following characterization of the operator norm.

**Theorem 6.2** (Abstract Dobrushin’s ergodicity coefficient). Let \( T : \mathcal{X} \to \mathcal{X} \) be a Markov operator with respect to \( e \). Then,
\[ \|T\|_H = \|T^*\|_H^* = 1 - \inf_{v, \pi \in \mathcal{P}(e)} \inf_{x \in [0, e]} \langle \pi, T(x) \rangle + \langle v, T(e - x) \rangle. \]
Proof. Since $T(e) = e$, we have:

$$\sup_{\nu, \pi \in \text{extr } P(e)} \sup_{x \in [0, e]} \langle v - \pi, T(x) \rangle = \sup_{\nu, \pi \in \text{extr } P(e)} \sup_{x \in [0, e]} 1 - \langle \pi, T(x) \rangle - \langle v, T(e - x) \rangle.$$ 

\[\square\]

7. APPLICATIONS TO STOCHASTIC MATRICES

In this section, we specialize the previous general results to the case of the standard orthant cone ($\mathcal{X} = \mathbb{R}^n$, $\mathcal{C} = \mathbb{R}^n_+$ and $e = 1$, Example 2.2). We recover the classical Dobrushin’s ergodicity coefficient and some known convergence results of the consensus system.

A linear map $T : \mathbb{R}^n \to \mathbb{R}^n$ given by $T(x) = Ax$, $x \in \mathbb{R}^n$ is a Markov operator if and only if $A$ is a row stochastic matrix. The operator norm corresponds to the contraction ratio of the matrix $A$ with respect to the diameter $\Delta$:

$$\|T\|_H = \tau(A) := \sup_{\Delta(x) \neq 0} \frac{\Delta(Ax)}{\Delta(x)},$$

and the dual operator norm corresponds to the Lipschitz constant of $A'$ with respect to the total variation distance on the space of probability measures:

$$\|T\|_H = \delta(A) := \sup_{\mu \neq v \in \mathcal{P}(1)} \frac{\|A'\mu - A'v\|_1}{\|\mu - v\|_1}.$$  

The value $\delta(A)$ is known as Dobrushin’s ergodicity coefficient of the Markov chain with transition probability matrix $A'$, see [LPW09]. Specializing Theorem 6.2 to this case, we get

$$\tau(A) = \delta(A) = 1 - \min_{i \neq j} \min_{I \subset \{1, \ldots, n\}} \left( \sum_{k \in I} A_{ik} + \sum_{k \in I^c} A_{jk} \right).$$  

The latter formula yields directly the following explicit form of Dobrushin’s ergodicity coefficient [Dob56]:

$$\tau(A) = \delta(A) = 1 - \min_{i \neq j} \sum_{s=1}^{n} \min(A_{is}, A_{js}).$$  

The above equality is a known result in the study of Markov chain. It is known that if $\tau(A) < 1$, then the Markov chain associated to $A$ is strongly ergodic [Sen91].

A (time-invariant) consensus system associated to the matrix $A$ is described by:

$$x_{k+1} = Ax_k, \ k = 1, 2, \ldots$$

By Theorem 6.1 if $\tau(A) < 1$, then the consensus system (24) converges to a multiple of $\mathbf{1}$ with an exponential rate $\tau(A)$.

Remark 7.1. A simple classical situation in which $\tau(A) < 1$ is when there is a Doeblin state, i.e., an element $j \in \{1, \ldots, n\}$ such that $A_{ij} > 0$ holds for all $i \in \{1, \ldots, n\}$. Besides, a Doeblin state is represented by a node connected to all other nodes in the graph associated to a matrix $A$. Based on this observation, some graph connectivity conditions [VJA05] [OT09] [Mor05] [AB09] characterizing the exponential convergence of consensus systems can be derived directly from Dobrushin’s ergodicity coefficient (23). As far as we know, such connection between Dobrushin’s ergodicity coefficient and the convergence of consensus system has been firstly observed in [MDA05].
Remark 7.2. For example, consider a time-variant linear consensus system:
\begin{equation}
  x_{k+1} = A_k x_k, \quad k = 1, 2, \ldots,
\end{equation}
where \( \{A_k\} \) is a sequence of stochastic matrices. Moreau [Mor05] showed that if all the non-zero entries of the matrices \( \{A_k\} \) are bounded from below by a positive constant \( \alpha > 0 \) and if there is \( p \in \mathbb{N} \) such that for all \( i \in \mathbb{N} \) there is a node connected to all other nodes in the graph associated to the matrix \( A_{i+p} \cdots A_{i+1} \), then the system (25) is globally uniformly convergent. These two conditions imply exactly that there is a Doeblin state associated to the matrix \( A_{i+p} \cdots A_{i+1} \). The uniform bound \( \alpha \) is to have an upper bound on the contraction rate, more precisely,
\[ \tau(A_{i+p} \cdots A_{i+1}) \leq 1 - \alpha, \quad \forall i = 1, 2, \ldots \]

8. Applications to noncommutative Markov operators

In this section, we specialize the previous general results to a finite dimensional noncommutative space (\( \mathcal{X} = S_n \), \( \mathcal{Y} = S_n^+ \) and \( \mathcal{E} = I_n \), Example [2.3]).

A completely positive unital linear map \( \Phi : S_n \to S_n \) is characterized by a set of matrices \( \{V_1, \ldots, V_m\} \) satisfying
\begin{equation}
  \sum_{i=1}^m V_i^* V_i = I_n
\end{equation}
such that the map \( \Phi \) is given by:
\begin{equation}
  \Phi(X) = \sum_{i=1}^m V_i^* X V_i, \quad \forall X \in S_n.
\end{equation}
The matrices \( \{V_1, \ldots, V_m\} \) are called Kraus operators. It is clear that \( \Phi : S_n \to S_n \) defines a Markov operator. The dual operator of \( \Phi \) is given by:
\begin{equation}
  \Psi(X) = \sum_{i=1}^m V_i X V_i^*, \quad X \in S_n.
\end{equation}
It is a completely positive and trace-preserving map, called Kraus map. The map \( \Phi \) and \( \Psi \) represent a purely quantum channel [SSR10, RKW11]. The map \( \Phi \) acts between spaces of measures while the adjoint map \( \Psi \) is trace-preserving and acts between spaces of states (density matrices). The operator norm of \( \Phi : S_n \to S_n \) is characterized by a set of matrices \( \{V_1, \ldots, V_m\} \) such that for all \( i \in \mathbb{N} \) there is a node connected to all other nodes in the graph associated to the matrix \( A_{i+p} \cdots A_{i+1} \). The uniform bound \( \alpha \) is to have an upper bound on the contraction rate, more precisely,
\[ \tau(A_{i+p} \cdots A_{i+1}) \leq 1 - \alpha, \quad \forall i = 1, 2, \ldots \]

Specializing Theorem 6.2 to Kraus maps, we obtain the noncommutative version of Dobrushin’s ergodicity coefficient.

Corollary 8.1 (Noncommutative Dobrushin’s ergodicity coefficient). Let \( \Phi \) be a completely positive unital linear map defined in (27). Then,
\begin{equation}
  \|\Phi\|_H = \|\Psi\|_H = 1 - \min_{u, v: u^* v = 0} \min_{X \in \{x_1, \ldots, x_n\}} \sum_{i=1}^n \min\{u^* \Phi(x_i) u, v^* \Phi(x_i) v\}
\end{equation}
Proof. It can be easily checked that
\[ \text{extr}[0,I_n] = \{ P \in S_n : P^2 = P \}. \]

Hence, Theorem 6.2 and Remark 4.3 yield:
\[
\| \Phi \|_H = \| \Psi \|_H^* = 1 - \min_{u,v} \min_{u^*v=0} (I_n-P)u + v^* \Phi(P)v = 1 - \min_{u,v} \min_{u^*v=0} \sum_{i \in J} u^* \Phi(x_{ij})u + \sum_{i \notin J} v^* \Phi(x_{ij})v
\]
from which (28) follows.

Remark 8.1. For the noncommutative case, it is not evident whether more effective characteri-
zation of the contraction rate exists. Note that the dual operator norm was studied in quantum
information theory, see [RKW11] and references therein. They provided a Birkhoff type upper
bound (Corollary 9 in [RKW11]):
\[
\| \Psi \|_H^* \leq \tanh(\text{diam} \Psi / 4).
\]
The value \( \text{diam} \Psi \) is not directly computable. This upper bound is equal to 1 if and only if
\( \text{diam} \Psi = \infty \), which is satisfied if and only if there exist a pair of nonzero vectors \( u, v \in \mathbb{C}^n \) such
that:
\[
\text{span}\{ V_i u : 1 \leq i \leq m \} \neq \text{span}\{ V_i v : 1 \leq i \leq m \}.
\]
We next provide a much tighter, in fact necessary and sufficient, condition for the operator norm
to be 1.

Corollary 8.2. The following conditions are equivalent:
1. \( \| \Phi \|_H = \| \Psi \|_H^* = 1 \).
2. There are nonzero vectors \( u, v \in \mathbb{C}^n \) such that
\[
\langle V_i u, V_j v \rangle = 0, \ \forall i, j \in \{ 1, \ldots, m \}.
\]
3. There is a rank one matrix \( Y \subset \mathbb{C}^{m \times n} \) such that
\[
\text{trace}(V_i^* V_j Y) = 0, \ \forall i, j \in \{ 1, \ldots, m \}.
\]

Proof. From Corollary 8.1 we know that \( \| \Phi \|_H = 1 \) if and only if there exist an orthonormal
basis \( \{ x_1, \ldots, x_n \} \) and two vectors \( u, v \in \mathbb{C}^n \) of norm 1 such that
\[
\sum_{i=1}^n \min_{j=1}^m u^* V_j^* x_i x_i^* V_j u, \ \sum_{j=1}^m v^* V_j^* x_i x_i^* V_j v = 0.
\]
This is equivalent to that for each \( i \in \{ 1, \ldots, n \} \), either
\[
x_i^* V_j u = 0, \ \forall j = 1, \ldots, m
\]
is true, or
\[
x_i^* V_j v = 0, \ \forall j = 1, \ldots, m
\]
is true. This is equivalent to
\[
\langle V_i u, V_j v \rangle = 0, \ \forall i, j \in \{ 1, \ldots, m \}.
\]
The equivalence between the second and the third condition is trivial by taking \( Y = vu^* \).
We consider a time-invariant noncommutative consensus system:

\[
X_{k+1} = \Phi(X_k), \quad k = 1, 2, \ldots
\]

where \(\Phi\) is a completely positive unital map. To study the convergence of such system, Sepulchre, Sarlette and Rouchon [SSR10] proposed to study the contraction ratio

\[
\alpha := \sup_{X>0} \frac{d_H(\Phi(X), I_n)}{d_H(X, I_n)}.
\]

They applied Birkhoff’s contraction formula (Theorem 5.2) to give an upper bound on the contraction ratio \(\alpha\):

\[
\alpha \leq \tanh\left(\frac{\text{diam} \Phi}{4}\right).
\]

The following theorem is a direct corollary of Nussbaum [Nus94].

**Theorem 8.3.** (Corollary of [Nus94, Thm2.3])

\[
\|\Phi\|_H = \lim_{\varepsilon \to 0^+} \left( \sup_{0 < d_H(X, I_n) \leq \varepsilon} \frac{d_H(\Phi(X), I_n)}{d_H(X, I_n)} \right).
\]

By this theorem, it is clear that the contraction ratio used in [SSR10] is an upper bound of the operator norm \(\|\Phi\|_H\):

\[
\|\Phi\|_H \leq \alpha.
\]

We next provide an algebraic characterization of the global convergence of system (29), based on the result established in Corollary 8.2. Let us consider a sequence of matrix subspaces defined as follows:

\[
H_0 = \text{span}\{I_n\},
\]

\[
H_{k+1} = \text{span}\{V_i^*XV_j : X \in H_k, i, j = 1, \ldots, m\}, \quad k = 0, 1, \ldots.
\]

**Lemma 8.4.** There is \(k_0 \leq n^2 - 1\) such that

\[
H_{k_0+s} = H_{k_0}, \quad \forall s \in \mathbb{N}.
\]

**Proof.** It follows from (26) that \(H_{k+1} \supseteq H_k\) for all \(k \in \mathbb{N}\). Besides, if for some \(k_0 \in \mathbb{N}\) such that

\[
H_{k_0+1} = H_{k_0},
\]

then

\[
H_{k_0+s} = H_{k_0}, \quad \forall s \in \mathbb{N}.
\]

This property also implies that if for some \(k_0 \in \mathbb{N}\)

\[
H_{k_0+1} \neq H_{k_0},
\]

then

\[
H_{k_0-s+1} \neq H_{k_0-s}, \quad \forall 1 \leq s \leq k_0.
\]

Since the dimension of \(H_k\) can not exceed \(n^2\), the case

\[
H_{k_0+1} \neq H_{k_0},
\]

can not happen more than \(n^2\) times. \( \square \)

For all \(k \in \mathbb{N}\), let \(G_k\) be the orthogonal complement of \(H_k\). Then there is \(k_0 \leq n^2 - 1\) such that

\[
G_k \supseteq G_{k+1}, \quad \forall k \in \mathbb{N}; \quad G_{k_0} = G_{k_0+s}, \quad \forall s \in \mathbb{N}
\]

**Theorem 8.5.** The following conditions are equivalent:

1. There exists \(k\) such that \(\|\Phi^k\|_H < 1\).
2. Every orbit of the system (29) converges to an equilibrium co-linear to \(I_n\).
3. The subspace \(\cap_k G_k\) does not contain a rank one matrix.
(4) There exists $k_0 \leq n^2 - 1$ such that $\|\Phi^{k_0}\|_H < 1$.

**Proof.** (1) $\Rightarrow$ (2): We apply Theorem 6.1 to the application $\Phi^k$.

(2) $\Rightarrow$ (1): Hilbert’s seminorm defines a norm in the orthogonal space to the identity matrix $I_n$. It follows from Gelfand’s formula that

$$\lim_{k \to +\infty} \|\Phi^k\|_1/k = \max\{\lambda : \Phi(X) = \lambda X, X \in S_n, X \perp I_n, X \neq 0\}$$

Thus if (1) is not true, then there is $X \in S_n$ and $\lambda \in \mathbb{C}$ such that $|\lambda| = 1$, $X \perp I$ and $\Phi(X) = \lambda X$. The system is therefore not globally convergent to an equilibrium co-linear to $I_n$.

(3) $\Leftrightarrow$ (1): Note that for all $k \in \mathbb{N}$,

$$\Phi^k(X) = \sum_{i_1, \ldots, i_k} V_{i_1}^* \cdots V_{i_k}^* X V_{i_1} \cdots V_{i_k}.$$ 

By Corollary 8.2 we know that $\|\Phi^k\|_H = 1$ if and only if the subspace $G_k$ contains a a rank one matrix. Therefore, $\|\Phi^k\|_H = 1$ for all $k \in \mathbb{N}$ if and only if the subspace $\cap_k G_k$ contains a rank one matrix.

(3) $\Rightarrow$ (4): By (30), there is $k_0 \leq n^2 - 1$ such that $G_{k_0} = \cap_k G_k$. It follows that if (3) is true then there is $k_0 \leq n^2 - 1$ such that $G_{k_0}$ does not contain a rank one matrix. Then by Corollary 8.2 we deduce that $\|\Phi^{k_0}\|_H < 1$ if (3) is true.

In a dual way, the above analysis also applies to the ergodicity study of noncommutative Markov chain given by:

$$(31) \quad \Pi_{k+1} = \Psi(\Pi_k), \quad k = 1, 2, \ldots$$

Below is a dual version of Theorem 8.5.

**Theorem 8.6.** The following conditions are equivalent:

1. There exists $k$ such that $\|\Psi^k\|_{I_H} < 1$.
2. The Markov chain (31) converges to a unique invariant measure regardless of initial distribution.
3. The subspace $\cap_k G_k$ does not contain a rank one matrix.
4. There exists $k_0 \leq n^2 - 1$ such that $\|\Psi^{k_0}\|_{I_H} < 1$.

**Remark 8.2.** A sufficient condition for the global convergence of the noncommutative consensus system (29) or equivalently, the ergodicity of the noncommutative Markov chain (31) would be that there is $k_0 \leq n^2 - 1$ such that

$$H_{k_0} = \mathbb{C}^{n \times n}.$$ 

Such condition can be checked in polynomial time.

**REFERENCES**

[AB99] C. D. Aliprantis and K. C. Border. *Infinite Dimensional Analysis: A Hitchiker’s Guide*. Springer, 1999.

[AB09] David Angeli and Pierre-Alexandre Bliman. Convergence speed of unsteady distributed consensus: decay estimate along the settling spanning-trees. *SIAM J. Control Optim.*, 48(1):1–32, 2009.

[BGPS06] Stephen Boyd, Arpita Ghosh, Balaji Prabhakar, and Devavrat Shah. Convergence speed of unsteady distributed consensus: decay estimate along the settling spanning-trees. *SIAM J. Control Optim.*, 48(1):1–32, 2009.

[Bir57] Garrett Birkhoff. Extensions of Jentzsch’s theorem. *Trans. Amer. Math. Soc.*, 85:219–227, 1957.

[Bou93] Philippe Bougerol. Kalman filtering with random coefficients and contractions. *SIAM J. Control Optim.*, 31(4):942–959, 1993.

[BT89] Dimitri P. Bertsekas and John N. Tsitsiklis. *Parallel and distributed computation: numerical methods*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 1989.

[Bus73] P. J. Bushell. Hilbert’s metric and positive contraction mappings in a Banach space. *Arch. Rational Mech. Anal.*, 52:330–338, 1973.
[Con90] John B. Conway. *A course in functional analysis*, volume 96 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, second edition, 1990.

[Dob56] R. Dobrushin. Central limit theorem for non-stationary Markov chains. I. *Teor. Veroyatnost. i Primenen.*, 1:72–89, 1956.

[EN95] Simon P. Eveson and Roger D. Nussbaum. An elementary proof of the Birkhoff-Hopf theorem. *Math. Proc. Cambridge Philos. Soc.*, 117(1):51–55, 1995.

[FHH+01] Marián Fabian, Petr Habala, Petr Hájek, Vicente Montesinos Santalucia, Jan Pelant, and Václav Zizler. *Functional analysis and infinite-dimensional geometry*. CMS Books in Mathematics/Ouvrages de Mathématiques de la SMC, 8. Springer-Verlag, New York, 2001.

[GG04] S. Gaubert and J. Gunawardena. The Perron-Frobenius theorem for homogeneous, monotone functions. *Trans. of AMS*, 356(12):4931–4950, 2004.

[Hir89] Morris W. Hirsch. Convergent activation dynamics in continuous time networks. *Neural Networks*, 2(5):331–349, 1989.

[Hop63] E. Hopf. An inequality for positive linear integral operators. *Journal of Mathematics and Mechanics*, 12(5):683–692, 1963.

[KP82] Elon Kohlberg and John W. Pratt. The contraction mapping approach to the Perron-Frobenius theory: why Hilbert’s metric? *Math. Oper. Res.*, 7(2):198–210, 1982.

[LPW09] David A. Levin, Yuval Peres, and Elizabeth L. Wilmer. *Markov chains and mixing times*. American Mathematical Society, Providence, RI, 2009. With a chapter by James G. Propp and David B. Wilson.

[MDA05] Cao M., Spielman D., A., and Morse A., S. A lower bound on convergence of a distributed network consensus algorithm. In *Proc. of the joint 44th IEEE Conference on Decision and Control and European Control Conference*, pages 2356–2361. IEEE, 2005.

[Mor05] Luc Moreau. Stability of multiagent systems with time-dependent communication links. *IEEE Trans. Automat. Control*, 50(2):169–182, 2005.

[Nus88] R. D. Nussbaum. Hilbert’s projective metric and iterated nonlinear maps. *Mem. Amer. Math. Soc.*, 75(391):iv+137, 1988.

[Nus94] Roger D. Nussbaum. Finsler structures for the part metric and Hilbert’s projective metric and applications to ordinary differential equations. *Differential Integral Equations*, 7(5-6):1649–1707, 1994.

[OT09] Alex Olshevsky and John N. Tsitsiklis. Convergence speed in distributed consensus and averaging. *SIAM J. Control Optim.*, 48(1):33–55, 2009.

[RKW11] David Reeb, Michael J. Kastoryano, and Michael M. Wolf. Hilbert’s projective metric in quantum information theory. *J. Math. Phys.*, 52(8):082201, 33, 2011.

[Sen91] E. Seneta. Applications of ergodicity coefficients to homogeneous Markov chains. In *50 Years after Doeblin: Developments in the Theory of Markov Chains, Markov Processes and Sums of Random Variables*, Blaubeuren, Germany, 2-7 November 1991.

[SSR10] Rodolphe Sepulchre, Alain Sarlette, and Pierre Rouchon. Consensus in noncommutative spaces. In *Proceedings of the 49th IEEE Conference on Decision and Control*, pages 6596–6601, Atlanta, USA, Dec 2010.

[TBA86] John N. Tsitsiklis, Dimitri P. Bertsekas, and Michael Athans. Distributed asynchronous deterministic and stochastic gradient optimization algorithms. *IEEE Trans. Automat. Control*, 31(9):803–812, 1986.

[VJA05] D. Blondel Vincent, Hendrickx Julien, M., Olshevsky Alex, and Tsitsiklis John, N. Convergence in multiagent coordination, consensus, and flocking. In *Proceedings of the joint 44th IEEE Conference on Decision and European Control Conference*, pages 2996–3000. IEEE, 2005.