

Dobrushin's ergodicity coefficient for Markov operators on cones

Stéphane Gaubert and Zheng Qu

Abstract. Doeblin and Dobrushin characterized the contraction rate of Markov operators with respect to the total variation norm. We generalize their results by giving an explicit formula for the contraction rate of a Markov operator over a cone in terms of pairs of extreme points with disjoint support in a set of abstract probability measures. By duality, we derive a characterization of the contraction rate of consensus dynamics over a cone with respect to Hopf's oscillation seminorm (the infinitesimal seminorm associated with Hilbert's projective metric). We apply these results to Kraus maps (noncommutative Markov chains, representing quantum channels), and characterize the ultimate contraction of the map in terms of the existence of a rank one matrix in a certain subspace.

Keywords. Markov operator, Dobrushin's ergodicity coefficient, ordered linear space, invariant measure, contraction ratio, consensus, noncommutative Markov chain, quantum channel, zero error capacity, rank one matrix.

1. Introduction

A basic result in the theory of Markov chains, due to Doeblin and Dobrushin, is the characterization of the contraction rate of a Markov operator acting on a space of measures equipped with the total variation norm.

The authors were partially supported by the PGMO Programme of FMJH and EDF, and by the programme "Ingénierie Numérique & Sécurité" of the French National Agency of Research, project "MALTHY", number ANR-13-INSE-0003.

An announcement of some of the present results has appeared in the Proceedings of the ECC'13 (European Control Conference), July 17-18 2013, Zurich.

Consider in particular a finite Markov chain with transition (row stochastic) matrix $A \in \mathbb{R}^{n \times n}$. The associated Markov operator is the map $v \mapsto vA$ from \mathbb{R}^n to \mathbb{R}^n , where the elements of \mathbb{R}^n are thought of as row vectors. The set of probability measures can be identified to the standard simplex $\mathcal{P} := \{v \in \mathbb{R}^n \mid v_i \geq 0 \text{ for } 1 \leq i \leq n, \sum_{1 \leq i \leq n} v_i = 1\}$, and the total variation norm is nothing but one half of the ℓ_1 norm $\|\cdot\|_1$ on \mathbb{R}^n . We are interested in the Lipschitz constant of the map $v \mapsto vA$, $\mathcal{P} \rightarrow \mathcal{P}$ with respect to the total variation norm, or equivalently, in the operator norm of the map $v \mapsto vA$ on the subspace of vectors of \mathbb{R}^n with zero sum, equipped with the same norm,

$$\delta(A) := \max_{v, \pi \in \mathcal{P}, v \neq \pi} \frac{\|vA - \pi A\|_1}{\|v - \pi\|_1} = \max_{\substack{v \in \mathbb{R}^n, \|v\|_1=1 \\ \sum_{1 \leq i \leq n} v_i=0}} \|vA\|_1 .$$

The Doeblin-Dobrushin characterization reads

$$\delta(A) = \frac{1}{2} \max_{i < j} \sum_{1 \leq s \leq n} |A_{is} - A_{js}| , \quad (1)$$

$$= 1 - \min_{i < j} \sum_{s=1}^n \min(A_{is}, A_{js}) . \quad (2)$$

The expression of $\delta(A)$ given by (1) is known as *Doeblin contraction coefficient*, see [LPW09], whereas the second expression, in (2), is known as *Dobrushin ergodicity coefficient* [Dob56]. The latter is often used to show that $\delta(A) < 1$. This holds in particular if there is a *Doeblin state*, i.e., a distinguished state t such that $A_{it} \geq \varepsilon > 0$ for all $1 \leq i \leq n$. Then, $\delta(A) \leq 1 - \varepsilon$.

A dual characterization of $\delta(A)$ has been used in linear consensus theory. The latter is motivated by communication networks, control theory and parallel computation [Hir89, BT89, BGPS06, Mor05, BHOT05, OT09, AB09]. It studies dynamics of the form

$$x_{k+1} = A_k x_k, \quad k = 1, 2, \dots, \quad x_k \in \mathbb{R}^n \quad (3)$$

where A_k are row stochastic matrices, acting on column vectors. One looks for conditions which guarantee the convergence of x_k to a *consensus state*, i.e., to a scalar multiple of the unit vector e of \mathbb{R}^n . To this end, one considers the following seminorm, sometimes called *diameter* or *Tsitsiklis' Lyapunov function* [TBA86]

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

for all $x, y \in \mathbb{R}^n$. It is known [CSM05] that

$$\delta(A) = \max_{x \notin \mathbb{R}e} \frac{\Delta(Ax)}{\Delta(x)} \quad (4)$$

so that the Doeblin-Dobrushin ergodicity coefficients coincides with the one-step contraction rate of the consensus dynamics with respect to the diameter seminorm. We note that the same seminorm Δ is a fundamental tool in Perron-Frobenius theory, where it is called *Hopf's oscillation* [Hop63, Bus73] or *Hilbert's seminorm* [GG04].

In this paper, we extend the Doeblin-Dobrushin theorem, as well as the dual characterization (4), to Markov operators over cones. We consider a bounded linear self-map T of a Banach space \mathcal{X} , equipped with a normal cone $\mathcal{C} \subset \mathcal{X}$, and a *unit* element \mathbf{e} belonging to the interior of \mathcal{C} . We say that T is an *abstract Markov operator* if $T(\mathcal{C}) \subset \mathcal{C}$ and $T(\mathbf{e}) = \mathbf{e}$. The *Hopf oscillation* in the space \mathcal{X} is the seminorm defined by

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

where \preceq denotes the partial order induced by \mathcal{C} . Our main result reads:

Theorem 1.1 (Contraction rate in Hopf's oscillation seminorm). *Let $T : \mathcal{X} \rightarrow \mathcal{X}$ be an abstract Markov operator. Then*

$$\begin{aligned} \sup_{\substack{z \in \mathcal{X} \\ \omega(z/\mathbf{e}) \neq 0}} \frac{\omega(T(z)/\mathbf{e})}{\omega(z/\mathbf{e})} &= \sup_{\substack{v, \pi \in \mathcal{P}(\mathbf{e}) \\ v \neq \pi}} \frac{\|T^*(v - \pi)\|_T^*}{\|v - \pi\|_T^*} \\ &= \frac{1}{2} \sup_{\substack{v, \pi \in \text{extr } \mathcal{P}(\mathbf{e}) \\ v \perp \pi}} \|T^*(v) - T^*(\pi)\|_T^* \\ &= 1 - \inf_{\substack{v, \pi \in \text{extr } \mathcal{P}(\mathbf{e}) \\ v \perp \pi}} \inf_{x \in [0, \mathbf{e}]} \langle \pi, T(x) \rangle + \langle v, T(\mathbf{e} - x) \rangle. \end{aligned}$$

This theorem follows from Theorems 5.1 and 6.2 below. The notations and notions used here are detailed in Section 5. In particular, T^* denotes the adjoint of T and we make use of the following norm, which we call *Thompson's norm*,

$$\|z\|_T = \inf\{\alpha > 0 : -\alpha\mathbf{e} \preceq z \preceq \alpha\mathbf{e}\}$$

on the space \mathcal{X} , and denote by $\|\cdot\|_T^*$ the dual norm. The notation $\mathcal{P}(\mathbf{e}) = \{\mu \in \mathcal{C}^* : \langle \mu, \mathbf{e} \rangle = 1\}$ refers to the abstract *simplex* of the dual Banach space \mathcal{X}^* of \mathcal{X} , where \mathcal{C}^* is the dual cone of \mathcal{C} ; *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.

Taking $\mathcal{X} = \mathbb{R}^n$, \mathcal{C} the standard positive cone \mathbb{R}_+^n , and \mathbf{e} the standard unit vector $(1, \dots, 1)^\top$, we recover from Theorem 1.1 the Doeblin-Dobrushin characterization (1),(2), as well as its dual form in linear consensus theory (4).

Results related to Theorem 1.1 have previously appeared. In a finite dimensional setting, Reeb, Kastoryano, and Wolf [RKW11] gave a characterization analogous to the second equality of the above theorem without

the disjointness condition. We refer to Remark 5.3 for a comparison. Also, Mukhamedov gave in [Muk13], in the setting of von Neumann algebras, a characterization similar to the same equality, still without the disjointness condition. He established some other properties of the ergodicity coefficient, and derived ergodic type theorems for nonhomogeneous Markov chains.

Several motivations lead to consider Markov operators over cones which differ from the standard positive cone of \mathbb{R}^n .

First, Sepulchre, Sarlette, and Rouchon [SSR10] and independently, Reeb, Kastoryano and Wolf [RKW11], have shown that tools from Perron-Frobenius theory (specially contraction results in different metrics over cones) provide a unifying general approach to address issues from quantum information and control. Here, quantum channels are represented by self-maps T of the cone of positive semidefinite matrices, preserving the Loewner order, and the identity matrix. Relations with classical “consensus” theory were also addressed in [SSR10]. We derive further results, showing that Theorem 1.1 leads to a noncommutative analogue of Dobrushin’s ergodicity coefficient (see Corollary 7.1):

$$1 - \min_{\substack{X=(x_1, \dots, x_n) \\ XX^*=I_n}} \min_{\substack{u, v: u^*v=0 \\ u^*u=v^*v=1}} \sum_{i=1}^n \min\{u^*T(x_i x_i^*)u, v^*T(x_i x_i^*)v\} .$$

Then, 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 7.7 and 7.8). Also, it follows from these results that an operator T representing a quantum channel has a contraction rate of 1 (absence of contraction) with respect to Hopf’s oscillation if and only if there exists two distinguishable pure states, i.e., a quantum clique of cardinality 2 [BS08], or equivalently if the quantum channel has a positive zero-error capacity [MA05].

We also derive as a direct illustration a convergence result (geometric convergence of the iterates of the operator to a rank one operator, or geometric convergence to a “consensus state”) in Theorem 6.1. Actually, the present contraction results are useful more generally when considering iterates of random contractions. Then, almost sure convergence to a consensus state can be obtained by adapting ideas of Bougerol [Bou93], see the discussion in §6 below. We limited our convergence treatment here to simple illustrations of our results: we note that the question of proving “weak ergodicity results” in their best generality has been thoroughly studied, we refer the reader to the work of Mukhamedov [Muk13], and to the references therein.

Our second and original motivation arises from *non-linear*, rather than linear, Perron-Frobenius theory, i.e., from the study of non-linear maps over cones. In this setting, the interior of a cone \mathcal{C} is equipped with *Hilbert's projective metric*, defined by:

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

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} ,

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

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

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.

The generalization of Birkhoff's theorem to non-linear maps, and in particular, the computation of the Lipschitz constant of nonlinear maps with respect to Hilbert's projective metric, has important applications (including population dynamics), and it has motivated several works, specially the one of Nussbaum [Nus94], who observed that d_H is the weak Finsler metric obtained when taking $\omega(\cdot/\mathbf{e})$ to be the infinitesimal distance at point \mathbf{e} . In other words,

$$d_H(x, y) = \inf_{\gamma} \int_0^1 \omega(\dot{\gamma}(s)/\gamma(s)) ds$$

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 $\lambda > 0$), restricted to a geodesically convex subset $U \subset \mathcal{C}^0$, 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:

$$\sup_{x, y \in U} \frac{d_H(f(x), f(y))}{d_H(x, y)} = \sup_{x \in U} \kappa(x) \tag{5}$$

where

$$\kappa(x) := \sup_{\substack{z \in \mathcal{C} \\ \omega(z/x) \neq 0}} \frac{\omega(Df(x)z/f(x))}{\omega(z/x)} . \tag{6}$$

We recognize in $\kappa(x)$ a variant of the Doeblin-Dobrushin coefficient $\delta(A)$, in which the domain and range of A are equipped with different unit elements, namely x and $f(x)$. Our characterization carries over to this case. In particular, Theorem 5.1 below gives an explicit formula for $\kappa(x)$, which, in combination with Nussbaum's characterization (5) allows one to compute the contraction rate of a non-linear map with respect to Hilbert's projective metric.

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}^* the dual space of \mathcal{X} . For any $x \in \mathcal{X}$ and $q \in \mathcal{X}^*$, denote by $\langle q, x \rangle$ the value of $q(x)$. Let $\mathcal{C} \subset \mathcal{X}$ be a closed pointed convex cone with nonempty interior \mathcal{C}_0 , in particular, $\alpha\mathcal{C} \subset \mathcal{C}$ for $\alpha \in \mathbb{R}^+$, $\mathcal{C} + \mathcal{C} \subset \mathcal{C}$ and $\mathcal{C} \cap (-\mathcal{C}) = 0$. The partial order \preceq induced by \mathcal{C} on \mathcal{X} is defined as follows:

$$x \preceq y \Leftrightarrow y - x \in \mathcal{C}.$$

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

$$[x, y] := \{z \in \mathcal{X} \mid x \preceq z \preceq y\}.$$

For $x \in \mathcal{X}$ and $y \in \mathcal{C}_0$, following [Nus88], we define

$$\begin{aligned} M(x/y) &:= \inf\{t \in \mathbb{R} : x \preceq ty\} \\ m(x/y) &:= \sup\{t \in \mathbb{R} : x \succeq ty\} \end{aligned} \tag{7}$$

Observe that since $y \in \mathcal{C}_0$, and since \mathcal{C} is closed and pointed, the two sets in (7) 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 $y \in \mathcal{C}_0$, we call *oscillation* [Bus73] the difference between $M(x/y)$ and $m(x/y)$:

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

Let \mathbf{e} denote a distinguished element in \mathcal{C}_0 , which we shall call a *unit*. For $x \in \mathcal{X}$, define

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

which we call *Thompson's norm*, with respect to the element \mathbf{e} , and

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

which we call *Hilbert's seminorm* with respect to the element \mathbf{e} .

Remark 2.1. These terminologies are motivated by the fact that Thompson's part metric and Hilbert's projective metric are Finsler metrics for which the infinitesimal distances at the point $\mathbf{e} \in \mathcal{C}_0$ are respectively given by $\|\cdot\|_T$ and $\|\cdot\|_H$, see [Nus94]. The seminorm $\|\cdot\|_H$ is also called Hopf's

oscillation seminorm [Bus73]. Besides, it is clear that the unit \mathbf{e} is an *order unit* and Thompson's norm $\|\cdot\|_T$ is the corresponding *order unit norm*, see [Ell64, Alf71, Nag74].

We assume that the cone \mathcal{X} 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 an order unit Banach space. Since Thompson's norm $\|\cdot\|_T$ is defined with respect to a particular element \mathbf{e} , we write $(\mathcal{X}, \mathbf{e}, \|\cdot\|_T)$ instead of $(\mathcal{X}, \|\cdot\|_T)$.

Example 2.2. We consider the finite dimensional vector space $\mathcal{X} = \mathbb{R}^n$, the standard positive cone $\mathcal{C} = \mathbb{R}_+^n$ and the unit vector $\mathbf{e} = \mathbf{1} := (1, \dots, 1)^T$. It can be checked that Thompson's norm with respect to $\mathbf{1}$ is nothing but the sup norm

$$\|x\|_T = \max_i |x_i| = \|x\|_\infty,$$

whereas Hilbert's seminorm with respect to $\mathbf{1}$ is the so called *diameter*:

$$\|x\|_H = \max_{1 \leq i, j \leq n} (x_i - x_j) = \Delta(x).$$

Example 2.3. Let $\mathcal{X} = \mathbb{S}_n$, the space of Hermitian matrices of dimension n and $\mathcal{C} = \mathbb{S}_n^+$, the cone of positive semidefinite matrices. Let the identity matrix I_n be the unit element: $\mathbf{e} = I_n$. Then Thompson's norm with respect to I_n is nothing but the sup norm of the spectrum of X , i.e.,

$$\|X\|_T = \max_{1 \leq i \leq n} \lambda_i(X) = \|\lambda(X)\|_\infty,$$

where $\lambda(X) := (\lambda_1(X), \dots, \lambda_n(X))$, 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 $(\mathcal{X}^*, \mathbf{e}, \|\cdot\|_T^*)$ the dual space of $(\mathcal{X}, \mathbf{e}, \|\cdot\|_T)$ where the dual norm $\|\cdot\|_T^*$ of a continuous linear functional $z \in \mathcal{X}^*$ is defined by:

$$\|z\|_T^* := \sup_{\|x\|_T=1} \langle z, x \rangle.$$

The *abstract simplex* in the dual space is defined by:

$$\mathcal{P}(\mathbf{e}) := \{\mu \in \mathcal{C}^* \mid \langle \mu, \mathbf{e} \rangle = 1\}, \quad (8)$$

where \mathcal{C}^* is the *dual cone* of \mathcal{C} :

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

Remark 3.1. For the standard positive cone (Example 2.2, $\mathcal{X} = \mathbb{R}^n$, $\mathcal{C} = \mathbb{R}_+^n$ and $\mathbf{e} = \mathbf{1}$), the dual space \mathcal{X}^* is $\mathcal{X} = \mathbb{R}^n$ itself and the dual norm $\|\cdot\|_T^*$ is the ℓ_1 norm:

$$\|x\|_T^* = \sum_i |x_i| = \|x\|_1 .$$

The abstract simplex $\mathcal{P}(\mathbf{1})$ is the standard simplex in \mathbb{R}^n :

$$\mathcal{P}(\mathbf{1}) = \{v \in \mathbb{R}_+^n : \sum_i v_i = 1\} ,$$

i.e., the set of probability measures on the discrete space $\{1, \dots, n\}$.

Remark 3.2. For the cone of semidefinite matrices (Example 2.3, $\mathcal{X} = S_n$, $\mathcal{C} = S_n^+$ and $\mathbf{e} = I_n$), 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.

By the duality between order unit and base normed spaces [Ell64], the space $(\mathcal{X}^*, \mathbf{e}, \|\cdot\|_T^*)$ is a base normed space. The abstract simplex $\mathcal{P}(\mathbf{e})$ coincides with the *base* and the dual norm $\|\cdot\|_T^*$ with the *base norm*. We denote by $B_T^*(\mathbf{e})$ the dual unit ball:

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

We denote by $\text{conv}(S)$ the convex hull of a set S . The next lemma relates the abstract simplex $\mathcal{P}(\mathbf{e})$ to the dual unit ball $B_T^*(\mathbf{e})$. The proof can be found in [Ell64, Alf71].

Lemma 3.3 ([Ell64]). *The dual unit ball $B_T^*(\mathbf{e})$ of the space $(\mathcal{X}^*, \mathbf{e}, \|\cdot\|_T^*)$, satisfies*

$$B_T^*(\mathbf{e}) = \text{conv}(\mathcal{P}(\mathbf{e}) \cup -\mathcal{P}(\mathbf{e})) . \quad (9)$$

Remark 3.4. Reeb, Kastoryano, and Wolf [RKW11] defined a *base* \mathcal{B} of a proper cone \mathcal{K} in a finite dimensional vector space \mathcal{V} , which coincides with the definition of our “abstract simplex”. They defined the *base norm* of $\mu \in \mathcal{V}$ with respect to \mathcal{B} by:

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

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

$$\|\mu\|_{\tilde{M}} = \sup_{0 \preceq x \preceq \mathbf{e}} \langle \mu, 2x - \mathbf{e} \rangle. \quad (10)$$

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

$$\|\mu\|_{\tilde{M}} = \|\mu\|_{\mathcal{B}}. \quad (11)$$

Lemma 3.3 is equivalent to the duality result (11) of Reeb et al., in a finite dimensional setting, and their approach can be seen as a dual one to ours.

4. Characterization of extreme points of the dual unit ball

A standard result of functional analysis shows that if \mathcal{W} is a closed subspace of a Banach space $(\mathcal{X}, \|\cdot\|)$, then the quotient space \mathcal{X}/\mathcal{W} is a Banach space, canonically equipped with the *quotient norm*

$$x \mapsto \inf_{w \in \mathcal{W}} \|x + w\|,$$

see [Con90, Chap. III, § 4]. The next lemma shows that when \mathcal{X} is equipped with Thompson's norm, Hilbert's seminorm coincides with the quotient norm of $\mathcal{X}/\mathbb{R}\mathbf{e}$, up to a factor 2.

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

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

Proof. The expression

$$\|x + \lambda \mathbf{e}\|_T = \max(M(x/\mathbf{e}) + \lambda, -m(x/\mathbf{e}) - \lambda)$$

is minimal when $M(x/\mathbf{e}) + \lambda = -m(x/\mathbf{e}) - \lambda$. Substituting the value of λ obtained in this way in $\|x + \lambda \mathbf{e}\|_T$, we arrive at the announced formula. \square

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

$$\mathcal{M}(\mathbf{e}) := \{\mu \in \mathcal{X}^* \mid \langle \mu, \mathbf{e} \rangle = 0\},$$

and

$$\|\mu\|_H^* := \frac{1}{2} \|\mu\|_T^*, \quad \forall \mu \in \mathcal{M}(\mathbf{e}). \quad (12)$$

Proof. It is shown in [Con90, Chap. III, Theorem 10.2] that if \mathcal{W} is a closed subspace of a Banach space $(\mathcal{X}, \|\cdot\|)$, the dual of the quotient space \mathcal{X}/\mathcal{W} can be identified isometrically to the space of continuous linear forms on \mathcal{X} that vanish on \mathcal{W} , equipped with the dual norm $\|\cdot\|^*$ of \mathcal{X}^* . Specializing this result to the case in which \mathcal{X} is equipped with twice the Thompson norm and $\mathcal{W} = \mathbb{R}\mathbf{e}$, and noting that multiplying the norm on the

space \mathcal{X} by a given positive factor divides the corresponding dual norm by the same factor, we obtain (12). \square

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

$$B_H^*(\mathbf{e}) = 2B_T^*(\mathbf{e}) \cap \mathcal{M}(\mathbf{e}). \quad (13)$$

Remark 4.3. In the case of the standard positive cone ($\mathcal{X} = \mathbb{R}^n$, $\mathcal{C} = \mathbb{R}_+^n$ and $\mathbf{e} = \mathbf{1}$), Lemma 4.2 implies that for any two probability measures $\mu, \nu \in \mathcal{P}(\mathbf{1})$, the dual norm $\|\mu - \nu\|_H^*$ is the total variation distance between μ and ν :

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

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

Definition 4.4. For all $\nu, \pi \in \mathcal{P}(\mathbf{e})$, we say that ν and π are *disjoint*, denoted by $\nu \perp \pi$, if

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

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

Example 4.5. In the case of the standard positive cone ($\mathcal{X} = \mathbb{R}^n$, $\mathcal{C} = \mathbb{R}_+^n$ and $\mathbf{e} = \mathbf{1}$), two points ν, π in $\mathcal{P}(\mathbf{1})$ are disjoint if and only if for all $1 \leq i \leq n$, $\nu_i = 0$ or $\pi_i = 0$ holds, meaning that ν and π , thought of as discrete probability measures, have disjoint supports.

We have the following characterization of the disjointness property.

Lemma 4.6. Let $\nu, \pi \in \mathcal{P}(\mathbf{e})$. The following assertions are equivalent:

- (a) $\nu \perp \pi$.
- (b) The only elements $\rho, \sigma \in \mathcal{P}(\mathbf{e})$ satisfying

$$\nu - \pi = \rho - \sigma$$

are $\rho = \nu$ and $\sigma = \pi$.

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

$$\nu - \pi = \rho - \sigma.$$

Then it is immediate that

$$\nu + \sigma = \pi + \rho.$$

Let $\mu = \frac{\nu + \sigma}{2} = \frac{\pi + \rho}{2}$. Then $\mu \in \mathcal{P}(\mathbf{e})$, $\mu \succ \frac{\nu}{2}$ and $\mu \succ \frac{\pi}{2}$. Since $\nu \perp \pi$, we obtain that $\mu = \frac{\nu + \pi}{2}$. It follows that $\rho = \nu$ and $\sigma = \pi$.

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

$$\nu - \pi = (2\mu - \pi) - (2\mu - \nu).$$

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 4.7. *The set of extreme points of $B_H^*(\mathbf{e})$, denoted by $\text{extr} B_H^*(\mathbf{e})$, is characterized by:*

$$\text{extr} B_H^*(\mathbf{e}) = \{v - \pi \mid v, \pi \in \text{extr } \mathcal{P}(\mathbf{e}), v \perp \pi\}.$$

Proof. It follows from (9) that every point $\mu \in B_T^*(\mathbf{e})$ can be written as

$$\mu = sv - t\pi$$

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

$$0 = \langle \mu, \mathbf{e} \rangle = s\langle v, \mathbf{e} \rangle - t\langle \pi, \mathbf{e} \rangle = s - t,$$

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

$$\mu = \frac{v - \pi}{2}, \quad v, \pi \in \mathcal{P}(\mathbf{e}).$$

Therefore by (13) we proved that

$$B_H^*(\mathbf{e}) = \{v - \pi : v, \pi \in \mathcal{P}(\mathbf{e})\}. \quad (14)$$

Now let $v, \pi \in \text{extr } \mathcal{P}(\mathbf{e})$ and $v \perp \pi$. We are going to prove that $v - \pi \in \text{extr} B_H^*(\mathbf{e})$. Let $v_1, \pi_1, v_2, \pi_2 \in \mathcal{P}(\mathbf{e})$ such that

$$v - \pi = \frac{v_1 - \pi_1}{2} + \frac{v_2 - \pi_2}{2}.$$

Then

$$v - \pi = \frac{v_1 + v_2}{2} - \frac{\pi_1 + \pi_2}{2}.$$

By Lemma 4.6, the only possibility is $2v = v_1 + v_2$ and $2\pi = \pi_1 + \pi_2$. Since $v, \pi \in \text{extr } \mathcal{P}(\mathbf{e})$ we obtain that $v_1 = v_2 = v$ and $\pi_1 = \pi_2 = \pi$. Therefore $v - \pi \in \text{extr} B_H^*(\mathbf{e})$.

Now let $v, \pi \in \mathcal{P}(\mathbf{e})$ such that $v - \pi \in \text{extr} B_H^*(\mathbf{e})$. Assume by contradiction that v is not extreme in $\mathcal{P}(\mathbf{e})$ (the case in which π is not extreme can be dealt with similarly). Then, we can find $v_1, v_2 \in \mathcal{P}(\mathbf{e}), v_1 \neq v_2$, such that $v = \frac{v_1 + v_2}{2}$. It follows that

$$\mu = \frac{v_1 - \pi}{2} + \frac{v_2 - \pi}{2},$$

where $v_1 - \pi, v_2 - \pi$ are distinct elements of $B_H^*(\mathbf{e})$, which is a contradiction. Next we show that $v \perp \pi$. To this end, let any $\rho, \sigma \in \mathcal{P}(\mathbf{e})$ such that

$$v - \pi = \rho - \sigma.$$

Then

$$v - \pi = \frac{v - \pi + \rho - \sigma}{2} = \frac{v - \sigma}{2} + \frac{\rho - \pi}{2}.$$

If $\sigma \neq \pi$, then $v - \sigma \neq v - \pi$ and this contradicts the fact that $v - \pi$ is extremal. Therefore $\sigma = \pi$ and $\rho = v$. From Lemma 4.6, we deduce that $v \perp \pi$.

□

Remark 4.8. In the case of standard positive cone ($\mathcal{X} = \mathbb{R}^n$, $\mathcal{C} = \mathbb{R}_+^n$ and $\mathbf{e} = \mathbf{1}$), the set of extreme points of $\mathcal{P}(\mathbf{1})$ is the set of standard basis vectors $\{e_i\}_{i=1, \dots, n}$. The extreme points are pairwise disjoint.

Remark 4.9. In the case of cone of semidefinite matrices ($\mathcal{X} = S_n$, $\mathcal{C} = S_n^+$ and $\mathbf{e} = I_n$), 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 $X \succcurlyeq xx^*$ and $X \succcurlyeq yy^*$ should satisfy $X \succcurlyeq xx^* + yy^*$. 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 4.6, 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 $\mathcal{C}_1 \subset \mathcal{X}_1$ and $\mathcal{C}_2 \subset \mathcal{X}_2$ be respectively two closed pointed convex normal cones with non empty interiors \mathcal{C}_1^0 and \mathcal{C}_2^0 . Let $\mathbf{e}_1 \in \mathcal{C}_1^0$ and $\mathbf{e}_2 \in \mathcal{C}_2^0$. Then, we know from Section 4 that the two quotient spaces $(\mathcal{X}_1/\mathbb{R}\mathbf{e}_1, \|\cdot\|_H)$ and $(\mathcal{X}_2/\mathbb{R}\mathbf{e}_2, \|\cdot\|_H)$ are Banach spaces. The dual spaces of $(\mathcal{X}_1/\mathbb{R}\mathbf{e}_1, \|\cdot\|_H)$ and $(\mathcal{X}_2/\mathbb{R}\mathbf{e}_2, \|\cdot\|_H)$ are respectively $(\mathcal{M}(\mathbf{e}_1), \|\cdot\|_H^*)$ and $(\mathcal{M}(\mathbf{e}_2), \|\cdot\|_H^*)$ (see Lemma 4.2).

Let T be a continuous linear map from the space $(\mathcal{X}_1/\mathbb{R}\mathbf{e}_1, \|\cdot\|_H)$ to $(\mathcal{X}_2/\mathbb{R}\mathbf{e}_2, \|\cdot\|_H)$. The operator norm of T , denoted by $\|T\|_H$, is given by:

$$\|T\|_H := \sup_{\|x\|_H=1} \|T(x)\|_H = \sup \frac{\omega(T(x)/\mathbf{e}_2)}{\omega(x/\mathbf{e}_1)}. \quad (15)$$

By definition, the *adjoint operator* $T^* : (\mathcal{M}(\mathbf{e}_2), \|\cdot\|_H^*) \rightarrow (\mathcal{M}(\mathbf{e}_1), \|\cdot\|_H^*)$ of T is:

$$\langle T^*(\mu), x \rangle = \langle \mu, T(x) \rangle, \quad \forall \mu \in \mathcal{M}(\mathbf{e}_2), x \in \mathcal{X}_1/\mathbb{R}\mathbf{e}_1.$$

The operator norm of T^* , denoted by $\|T^*\|_H^*$, is then:

$$\|T^*\|_H^* := \sup_{\mu \in B_H^*(\mathbf{e}_2)} \|T^*(\mu)\|_H^*.$$

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 $T : \mathcal{X}_1 \rightarrow \mathcal{X}_2$ be a bounded linear map such that $T(\mathbf{e}_1) \in \mathbb{R}\mathbf{e}_2$. Then,*

$$\begin{aligned} \|T\|_H = \|T^*\|_H^* &= \frac{1}{2} \sup_{\mathbf{v}, \boldsymbol{\pi} \in \mathcal{P}(\mathbf{e}_2)} \|T^*(\mathbf{v}) - T^*(\boldsymbol{\pi})\|_T^* \\ &= \sup_{\mathbf{v}, \boldsymbol{\pi} \in \mathcal{P}(\mathbf{e}_2)} \sup_{x \in [0, \mathbf{e}_1]} \langle \mathbf{v} - \boldsymbol{\pi}, T(x) \rangle. \end{aligned}$$

Moreover, the supremum can be restricted to the set of mutually disjoint extreme points:

$$\begin{aligned} \|T\|_H = \|T^*\|_H^* &= \frac{1}{2} \sup_{\substack{\mathbf{v}, \boldsymbol{\pi} \in \text{extr } \mathcal{P}(\mathbf{e}_2) \\ \mathbf{v} \perp \boldsymbol{\pi}}} \|T^*(\mathbf{v}) - T^*(\boldsymbol{\pi})\|_T^* \\ &= \sup_{\substack{\mathbf{v}, \boldsymbol{\pi} \in \text{extr } \mathcal{P}(\mathbf{e}_2) \\ \mathbf{v} \perp \boldsymbol{\pi}}} \sup_{x \in [0, \mathbf{e}_1]} \langle \mathbf{v} - \boldsymbol{\pi}, T(x) \rangle. \end{aligned} \quad (16)$$

Proof. We already noted that $\|T\|_H = \|T^*\|_H^*$. Moreover,

$$\|T^*\|_H^* = \sup_{\boldsymbol{\mu} \in B_H^*(\mathbf{e}_2)} \|T^*(\boldsymbol{\mu})\|_H^*.$$

By the characterization of $B_H^*(\mathbf{e}_2)$ in (14) and the characterization of the norm $\|\cdot\|_H^*$ in Lemma 4.2, we get

$$\begin{aligned} \sup_{\boldsymbol{\mu} \in B_H^*(\mathbf{e}_2)} \|T^*(\boldsymbol{\mu})\|_H^* &= \sup_{\mathbf{v}, \boldsymbol{\pi} \in \mathcal{P}(\mathbf{e}_2)} \|T^*(\mathbf{v}) - T^*(\boldsymbol{\pi})\|_T^* \\ &= \frac{1}{2} \sup_{\mathbf{v}, \boldsymbol{\pi} \in \mathcal{P}(\mathbf{e}_2)} \|T^*(\mathbf{v}) - T^*(\boldsymbol{\pi})\|_T^* . \end{aligned}$$

For the second equality, note that

$$\begin{aligned} \|T^*(\mathbf{v}) - T^*(\boldsymbol{\pi})\|_T^* &= \sup_{x \in [0, \mathbf{e}_1]} \langle T^*(\mathbf{v}) - T^*(\boldsymbol{\pi}), 2x - \mathbf{e}_1 \rangle \\ &= 2 \sup_{x \in [0, \mathbf{e}_1]} \langle T^*(\mathbf{v}) - T^*(\boldsymbol{\pi}), x \rangle . \end{aligned}$$

We next show that the supremum can be restricted to the set of extreme points. By the Banach-Alaoglu theorem, $B_H^*(\mathbf{e}_2)$ is weak-star compact, and it is obviously convex. The dual space $\mathcal{M}(\mathbf{e}_2)$ endowed with the weak-star topology is a locally convex topological space. Thus by the Krein-Milman theorem, the unit ball $B_H^*(\mathbf{e}_2)$, which is a compact convex set in $\mathcal{M}(\mathbf{e}_2)$ with respect to the weak-star topology, is the closed convex hull of

its extreme points. So every element ρ of $B_H^*(\mathbf{e}_2)$ is the limit of a net $(\rho_\alpha)_\alpha$ of elements in $\text{conv}(\text{extr} B_H^*(\mathbf{e}_2))$. Observe now that the function

$$\varphi : \mu \mapsto \|T^*(\mu)\|_H^* = \sup_{x \in B_H(\mathbf{e}_1)} \langle T^*(\mu), x \rangle = \sup_{x \in B_H(\mathbf{e}_1)} \langle \mu, T(x) \rangle$$

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

$$\begin{aligned} \varphi(\rho) &\leq \liminf_{\alpha} \varphi(\rho_\alpha) \\ &\leq \sup\{\varphi(\mu) : \mu \in \text{conv}(\text{extr} B_H^*(\mathbf{e}_2))\} \\ &= \sup\{\varphi(\mu) : \mu \in \text{extr} B_H^*(\mathbf{e}_2)\} . \end{aligned}$$

Using the characterization of the extreme points in Proposition 4.7, we get:

$$\begin{aligned} \sup_{\mu \in B_H^*(\mathbf{e}_2)} \|T^*(\mu)\|_H^* &= \sup_{\mu \in \text{extr} B_H^*(\mathbf{e}_2)} \|T^*(\mu)\|_H^* \\ &= \sup_{\substack{v, \pi \in \text{extr} \mathcal{P}(\mathbf{e}_2) \\ v \perp \pi}} \|T^*(v) - T^*(\pi)\|_H^* . \quad \square \end{aligned}$$

Remark 5.2. When \mathcal{X}_1 is of finite dimension, the set $[0, \mathbf{e}_1]$ is the convex hull of the set of its extreme points, hence, the supremum over the variable $x \in [0, \mathbf{e}_1]$ in (16) is attained at an extreme point. Similarly, if \mathcal{X}_2 is of finite dimension, the suprema over (v, π) in the same equation are also attained, because the map φ in the proof of the previous theorem, which is a supremum of an equi-Lipschitz family of maps, is continuous (in fact, Lipschitz).

Remark 5.3. Theorem 5.1 should be compared with a result of [RKW11] which can be stated as follows.

Proposition 5.4 (Proposition 12 in [RKW11]). *Let $\mathcal{V}, \mathcal{V}'$ be two finite dimensional vector spaces and $L : \mathcal{V} \rightarrow \mathcal{V}'$ be a linear map and let $\mathcal{B} \subset \mathcal{V}$ and $\mathcal{B}' \subset \mathcal{V}'$ be bases. Then*

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

The first term in (17) is called the *contraction ratio* of the linear map L , with respect to base norms. 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 $T : \mathcal{X}_1/\mathbb{R}\mathbf{e}_1 \rightarrow \mathcal{X}_2/\mathbb{R}\mathbf{e}_2$. Then $T^* : \mathcal{X}_2^* \rightarrow \mathcal{X}_1^*$ is a base preserving linear map ($T^*(\mathcal{P}(\mathbf{e}_2)) \subset \mathcal{P}(\mathbf{e}_1)$) and so, Proposition 12 of [RKW11] shows

that:

$$\|T^*\|_H^* = \sup_{\substack{\nu, \pi \in \mathcal{P}(\mathbf{e}_2) \\ \nu \neq \pi}} \frac{\|T^*(\nu - \pi)\|_T^*}{\|\nu - \pi\|_T^*} = \frac{1}{2} \sup_{\nu, \pi \in \text{extr } \mathcal{P}(\mathbf{e}_2)} \|T^*(\nu) - T^*(\pi)\|_T^* \quad (18)$$

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 $\|T\|_H$, and the contraction ratio with respect to the base norms of the dual base preserving map $\|T^*\|_H^*$. The latter is the primary object of interest in quantum information theory whereas the former is of interest in the control/consensus literature [TBA86, Mor05]. We also proved that the supremum in (18) can be restricted to pairs of *disjoint* extreme points ν, π . Finally, the expression of the contraction rate as the last supremum in Theorem 5.1 leads here to an abstract version of Dobrushin's ergodic coefficient, see Eqn (2) and Corollary 7.1 below.

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

Definition 5.5 ([Bir57]). *Hilbert's projective metric* between two elements x and y of \mathcal{C}_0 is

$$d_H(x, y) = \log(M(x/y)/m(x/y)) \quad . \quad (19)$$

(The notation $M(\cdot/\cdot)$ was defined in (7).)

Consider a linear operator $T : \mathcal{X}_1 \rightarrow \mathcal{X}_2$ such that $T(\mathcal{C}_1^0) \subset \mathcal{C}_2^0$. Following [Bir57, Bus73], the *projective diameter* of T is defined as below:

$$\text{diam } T = \sup\{d_H(T(x), T(y)) : x, y \in \mathcal{C}_1^0\}.$$

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 5.6 ([Bir57, Bus73]).

$$\sup_{x, y \in \mathcal{C}_1^0} \frac{\omega(T(x)/T(y))}{\omega(x/y)} = \sup_{x, y \in \mathcal{C}_1^0} \frac{d_H(T(x), T(y))}{d_H(x, y)} = \tanh\left(\frac{\text{diam } T}{4}\right).$$

Following [RKW11], we define the projective diameter of T^* :

$$\text{diam } T^* = \sup\{d_H(T^*(u), T^*(v)) : u, v \in \mathcal{C}_2^* \setminus \{0\}\}.$$

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

$$\begin{aligned} \sup_{x, y \in \mathcal{C}_1^0} \frac{M(T(x)/T(y))}{m(T(x)/T(y))} &= \sup_{x, y \in \mathcal{C}_1^0} \sup_{u, v \in \mathcal{C}_2^* \setminus \{0\}} \frac{\langle u, T(x) \rangle \langle v, T(y) \rangle}{\langle u, T(y) \rangle \langle v, T(x) \rangle} \\ &= \sup_{u, v \in \mathcal{C}_2^* \setminus \{0\}} \frac{M(T^*(u)/T^*(v))}{m(T^*(u)/T^*(v))} \end{aligned}$$

Corollary 5.7 (Compare with [RKW11]). *Let $T : \mathcal{X}_1 \rightarrow \mathcal{X}_2$ be a bounded linear map such that $T(\mathbf{e}_1) \in \mathbb{R}\mathbf{e}_2$ and $T(\mathcal{C}_1^0) \subset \mathcal{C}_2^0$, then:*

$$\|T^*\|_H^* = \|T\|_H \leq \tanh\left(\frac{\text{diam } T}{4}\right) = \tanh\left(\frac{\text{diam } T^*}{4}\right)$$

Proof. It is sufficient to prove the inequality. For this, note that

$$\|T\|_H = \sup_{\substack{x \in \mathcal{X}_1 \\ \omega(x/\mathbf{e}_1) \neq 0}} \omega(T(x)/\mathbf{e}_2) / \omega(x/\mathbf{e}_1) = \sup_{\substack{x \in \mathcal{C}_1^0 \\ \omega(x/\mathbf{e}_1) \neq 0}} \omega(T(x)/\mathbf{e}_2) / \omega(x/\mathbf{e}_1).$$

Then we apply Birkhoff's contraction formula. \square

Remark 5.8. 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 5.7 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 Markov operators on cones and discrete time consensus dynamics

A bounded linear map $T : \mathcal{X} \rightarrow \mathcal{X}$ is a *Markov operator* with respect to a unit vector \mathbf{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 \mathbf{e} , i.e., $T(\mathbf{e}) = \mathbf{e}$.

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 $T : \mathcal{X} \rightarrow \mathcal{X}$ be a Markov operator with respect to the unit element \mathbf{e} . If $\|T\|_H < 1$ or equivalently $\|T^\|_H^* < 1$, then there is $\boldsymbol{\pi} \in \mathcal{P}(\mathbf{e})$ such that for all $x \in \mathcal{X}$*

$$\|T^n(x) - \langle \boldsymbol{\pi}, x \rangle \mathbf{e}\|_T \leq (\|T\|_H)^n \|x\|_H,$$

and for all $\boldsymbol{\mu} \in \mathcal{P}(\mathbf{e})$

$$\|(T^*)^n(\boldsymbol{\mu}) - \boldsymbol{\pi}\|_H^* \leq (\|T\|_H)^n.$$

Proof. The intersection

$$\cap_n [m(T^n(x)/\mathbf{e}), M(T^n(x)/\mathbf{e})] \subset \mathbb{R}$$

is nonempty (as a non-increasing intersection of nonempty compact sets), and since $\|T\|_H < 1$ and

$$\omega(T^n(x)/\mathbf{e}) \leq (\|T\|_H)^n \omega(x/\mathbf{e}),$$

this intersection must be reduced to a real number $\{c(x)\} \subset \mathbb{R}$ depending on x , i.e.,

$$c(x) = \bigcap_n [m(T^n(x)/\mathbf{e}), M(T^n(x)/\mathbf{e})] .$$

Thus for all $n \in \mathbb{N}$,

$$-\omega(T^n(x)/\mathbf{e})\mathbf{e} \leq T^n(x) - c(x)\mathbf{e} \leq \omega(T^n(x)/\mathbf{e})\mathbf{e}.$$

Therefore by definition:

$$\|T^n(x) - c(x)\mathbf{e}\|_T \leq \omega(T^n(x)/\mathbf{e}) \leq (\|T\|_H)^n \|x\|_H.$$

It is immediate that:

$$c(x)\mathbf{e} = \lim_{n \rightarrow \infty} T^n(x)$$

from which we deduce that $c : \mathcal{X} \rightarrow \mathbb{R}$ 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 $\langle \pi, \mathbf{e} \rangle = 1$ and $\pi \in \mathcal{C}^*$ because

$$x \in \mathcal{C} \Rightarrow c(x)\mathbf{e} \in \mathcal{C} \Rightarrow c(x) \geq 0 \Rightarrow \langle \pi, x \rangle \geq 0.$$

Therefore $\pi \in \mathcal{P}(\mathbf{e})$. Finally for all $\mu \in \mathcal{P}(\mathbf{e})$ and all $x \in \mathcal{X}$ we have

$$\begin{aligned} \langle (T^*)^n(\mu) - \pi, x \rangle &= \langle \mu, T^n(x) - \langle \pi, x \rangle \mathbf{e} \rangle \\ &\leq \|\mu\|_T^* \|T^n(x) - \langle \pi, x \rangle \mathbf{e}\|_T \\ &\leq (\|T\|_H)^n \|x\|_H. \end{aligned}$$

Hence

$$\|(T^*)^n(\mu) - \pi\|_H^* \leq (\|T\|_H)^n.$$

□

A time invariant discrete time consensus system can be described by

$$x_{k+1} = T(x_k), \quad k = 1, 2, \dots \quad (20)$$

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 \mathbf{e} . The dual system of (20) represents a homogeneous discrete time Markov system:

$$\pi_{k+1} = T^*(\pi_k), \quad k = 1, 2, \dots \quad (21)$$

One of the central issues in Markov chain study is the ergodic property, i.e., the convergence of the distribution π_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 ergodic.

A time-dependent consensus system is described by

$$x_{k+1} = T_{k+1}(x_k), \quad k = 1, 2, \dots \quad (22)$$

where $\{T_k : k \geq 1\}$ is a sequence of Markov operators sharing a common unit element $\mathbf{e} \in \mathcal{C}^0$. Then if there is an integer $p > 0$ and a constant $\alpha < 1$ such that for all $i \in \mathbb{N}$

$$\|T_{i+p} \dots T_{i+1}\|_H \leq \alpha,$$

then the same lines of proof of Theorem 6.1 imply the existence of $\pi \in \mathcal{P}(\mathbf{e})$ such that for all $\{x_k\}$ satisfying (22),

$$\|x_k - \langle \pi, x_0 \rangle \mathbf{e}\|_T \leq \alpha^{\lfloor \frac{k}{p} \rfloor} \|x_0\|_H, \quad n \in \mathbb{N}.$$

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

$$\mathbb{E}[\log \|T_{1+p} \dots T_1\|_H] < 0$$

for some $p > 0$, see Bougerol [Bou93]. The ergodicity of an 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 $T : \mathcal{X} \rightarrow \mathcal{X}$ be a Markov operator with respect to \mathbf{e} . Then,*

$$\|T\|_H = \|T^*\|_H^* = 1 - \inf_{\substack{\mathbf{v}, \pi \in \text{extr } \mathcal{P}(\mathbf{e}) \\ \mathbf{v} \perp \pi}} \inf_{x \in [0, \mathbf{e}]} \langle \pi, T(x) \rangle + \langle \mathbf{v}, T(\mathbf{e} - x) \rangle.$$

Proof. Since $T(\mathbf{e}) = \mathbf{e}$, we have:

$$\begin{aligned} & \sup_{\substack{\mathbf{v}, \pi \in \text{extr } \mathcal{P}(\mathbf{e}) \\ \mathbf{v} \perp \pi}} \sup_{x \in [0, \mathbf{e}]} \langle \mathbf{v} - \pi, T(x) \rangle \\ &= \sup_{\substack{\mathbf{v}, \pi \in \text{extr } \mathcal{P}(\mathbf{e}) \\ \mathbf{v} \perp \pi}} \sup_{x \in [0, \mathbf{e}]} 1 - \langle \pi, T(x) \rangle - \langle \mathbf{v}, T(\mathbf{e} - x) \rangle. \end{aligned}$$

□

Example 6.3. Let us specialize Theorem 6.2 to the case of the standard positive cone of \mathbb{R}^n (Example 2.2). Then, a Markov operator $\mathbb{R}^n \rightarrow \mathbb{R}^n$ is of the form $T(x) = Ax$, where A is a row-stochastic matrix. We get

$$\delta(A) = 1 - \min_{i < j} \min_{I \subset \{1, \dots, n\}} \left(\sum_{k \in I} A_{ik} + \sum_{k \notin I} A_{jk} \right).$$

This formula yields directly the explicit form of Dobrushin's ergodicity coefficient recalled in the introduction,

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

Remark 6.4. Several results of linear consensus theory can be interpreted, or proved, in terms of ergodicity coefficient. Consider the time-variant linear consensus system:

$$x_{k+1} = A_k x_k, \quad k = 1, 2, \dots, \quad (23)$$

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 $A_{i+p} \dots A_{i+1}$, then the system (23) is globally uniformly convergent. These two conditions imply exactly that there is a Doeblin state associated to the matrix $A_{i+p} \dots A_{i+1}$. The uniform bound α is to have an upper bound on the contraction rate, more precisely,

$$\delta(A_{i+p} \dots A_{i+1}) \leq 1 - \alpha, \quad \forall i = 1, 2, \dots$$

.

7. Applications to noncommutative Markov operators

In this section, we specialize the previous general results to a finite dimensional noncommutative space ($\mathcal{X} = \mathbb{S}_n$, $\mathcal{C} = \mathbb{S}_n^+$ and $\mathbf{e} = I_n$, Example 2.3).

A completely positive unital linear map $\Phi : \mathbb{S}_n \rightarrow \mathbb{S}_n$ is characterized by a set of matrices $\{V_1, \dots, V_m\}$ satisfying

$$\sum_{i=1}^m V_i^* V_i = I_n \quad (24)$$

such that the map Φ is given by:

$$\Phi(X) = \sum_{i=1}^m V_i^* X V_i, \quad \forall X \in \mathbb{S}_n. \quad (25)$$

The matrices $\{V_1, \dots, V_m\}$ are called *Kraus operators*. It is clear that $\Phi : \mathbb{S}_n \rightarrow \mathbb{S}_n$ defines a Markov operator. The dual operator of Φ is given by:

$$\Psi(X) = \sum_{i=1}^m V_i X V_i^*, \quad X \in \mathbb{S}_n.$$

It is a completely positive and trace-preserving map, called Kraus map. The map Φ and Ψ represent a purely quantum channel [SSR10, RKW11].

In particular, the adjoint map Ψ is trace-preserving and acts on density matrices. The operator norm of $\Phi : S_n/\mathbb{R}I_n \rightarrow S_n/\mathbb{R}I_n$ is the contraction ratio with respect to the diameter of the spectrum:

$$\|\Phi\|_H = \sup_{X \in S_n} \frac{\lambda_{\max}(\Phi(X)) - \lambda_{\min}(\Phi(X))}{\lambda_{\max}(X) - \lambda_{\min}(X)}.$$

The operator norm of the adjoint map $\Psi : \mathcal{P}(I_n) \rightarrow \mathcal{P}(I_n)$ is the contraction ratio with respect to the trace norm (the total variation distance):

$$\|\Psi\|_H^* = \sup_{\rho_1, \rho_2 \in \mathcal{P}(I_n)} \frac{\|\Psi(\rho_1) - \Psi(\rho_2)\|_1}{\|\rho_1 - \rho_2\|_1}.$$

The values $\|\Phi\|_H$ and $\|\Psi\|_H^*$ are the noncommutative counterparts of $\delta(\cdot)$.

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

Corollary 7.1 (Noncommutative Dobrushin's ergodicity coefficient).

Let Φ be a completely positive unital linear map defined in (25). Then,

$$\|\Phi\|_H = \|\Psi\|_H^* = 1 - \min_{\substack{u, v: u^*v=0 \\ u^*u=v^*v=1}} \min_{\substack{X=(x_1, \dots, x_n) \\ XX^*=I_n}} \sum_{i=1}^n \min\{u^*\Phi(x_i x_i^*)u, v^*\Phi(x_i x_i^*)v\} \quad (26)$$

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 4.9 yield:

$$\begin{aligned} \|\Phi\|_H = \|\Psi\|_H^* &= 1 - \min_{\substack{u, v: u^*v=0 \\ u^*u=v^*v=1}} \min_{P^2=P} u^*\Phi(I_n - P)u + v^*\Phi(P)v \\ &= 1 - \min_{\substack{u, v: u^*v=0 \\ u^*u=v^*v=1}} \min_{\substack{X=(x_1, \dots, x_n) \\ XX^*=I_n}} \min_{J \subset \{1, \dots, n\}} \sum_{i \in J} u^*\Phi(x_i x_i^*)u + \sum_{i \notin J} v^*\Phi(x_i x_i^*)v \end{aligned}$$

from which (26) follows. \square

Remark 7.2. For the noncommutative case, it is not evident whether more effective characterization 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 tighter, in fact necessary and sufficient, condition for the operator norm to be 1.

Corollary 7.3. *The following conditions are equivalent:*

1. $\|\Phi\|_H = \|\Psi\|_H^* = 1$.
2. *There are nonzero vectors $u, v \in \mathbb{C}^n$ such that*

$$\langle V_i u, V_j v \rangle = 0, \quad \forall i, j \in \{1, \dots, m\}.$$
3. *There is a rank one matrix $Y \subset \mathbb{C}^{n \times n}$ such that*

$$\text{trace}(V_i^* V_j Y) = 0, \quad \forall i, j \in \{1, \dots, m\}.$$

Proof. From Corollary 7.1 we know that $\|\Phi\|_H = 1$ if and only if there exist an orthonormal basis $\{x_1, \dots, x_n\}$ and two vectors $u, v \in \mathbb{C}^n$ of norm 1 such that

$$\sum_{i=1}^n \min\left\{ \sum_{j=1}^m u^* V_j^* x_i x_i^* V_j u, \sum_{j=1}^m v^* V_j^* x_i x_i^* V_j v \right\} = 0.$$

This is equivalent to that for each $i \in \{1, \dots, n\}$, either

$$x_i^* V_j u = 0, \quad \forall j = 1, \dots, m$$

is true, or

$$x_i^* V_j v = 0, \quad \forall j = 1, \dots, m$$

is true. This is equivalent to

$$\langle V_i u, V_j v \rangle = 0, \quad \forall i, j \in \{1, \dots, m\}.$$

The equivalence between the second and the third condition is trivial by taking $Y = v u^*$. \square

Remark 7.4. The condition appearing in item 2 of Corollary 7.3 is equivalent to the positivity of the zero-error capacity [MA05], or to the existence of a quantum clique of cardinality 2 [BS08]. The latter problem is known to be QMA₁ complete (proof of Theorem 3.2 in [BS08]). Thus, Corollary 7.3 relates the absence of contraction to a known hard problem in quantum computing.

We consider a time-invariant noncommutative consensus system:

$$X_{k+1} = \Phi(X_k), \quad k = 1, 2, \dots \tag{27}$$

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

$$\alpha := \sup_{X > 0} d_H(\Phi(X), I_n) / d_H(X, I_n).$$

They applied Birkhoff's contraction formula (Theorem 5.6) to give an upper bound on the contraction ratio α :

$$\alpha \leq \tanh(\text{diam } \Phi/4) .$$

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

Theorem 7.5. (Corollary of [Nus94, Thm2.3])

$$\|\Phi\|_H = \lim_{\varepsilon \rightarrow 0^+} \left(\sup \left\{ \frac{d_H(\Phi(X), I_n)}{d_H(X, I_n)} : 0 < d_H(X, I_n) \leq \varepsilon \right\} \right),$$

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 (27), based on the result established in Corollary 7.3. Let us consider a sequence of matrix subspaces defined as follows:

$$\mathcal{H}_0 = \text{span}\{I_n\} ,$$

$$\mathcal{H}_{k+1} = \text{span}\{V_i^* X V_j : X \in \mathcal{H}_k, i, j = 1, \dots, m\} , \quad k = 0, 1, \dots,$$

Lemma 7.6. *There is $k_0 \leq n^2 - 1$ such that*

$$\mathcal{H}_{k_0+s} = \mathcal{H}_{k_0}, \quad \forall s \in \mathbb{N}.$$

Proof. It follows from (24) that $\mathcal{H}_{k+1} \supseteq \mathcal{H}_k$ for all $k \in \mathbb{N}$. Besides, if for some $k_0 \in \mathbb{N}$ such that

$$\mathcal{H}_{k_0+1} = \mathcal{H}_{k_0} ,$$

then

$$\mathcal{H}_{k_0+s} = \mathcal{H}_{k_0}, \quad \forall s \in \mathbb{N}.$$

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

$$\mathcal{H}_{k_0+1} \neq \mathcal{H}_{k_0} ,$$

then

$$\mathcal{H}_{k_0-s+1} \neq \mathcal{H}_{k_0-s}, \quad \forall 1 \leq s \leq k_0 .$$

Since the dimension of \mathcal{H}_k can not exceed n^2 , the case

$$\mathcal{H}_{k_0+1} \neq \mathcal{H}_{k_0} ,$$

can not happen more than n^2 times. □

For all $k \in \mathbb{N}$, let \mathcal{G}_k be the orthogonal complement of \mathcal{H}_k . Then there is $k_0 \leq n^2 - 1$ such that

$$\mathcal{G}_k \supseteq \mathcal{G}_{k+1}, \quad \forall k \in \mathbb{N}; \quad \mathcal{G}_{k_0} = \mathcal{G}_{k_0+s}, \quad \forall s \in \mathbb{N} \quad (28)$$

Theorem 7.7. *The following conditions are equivalent:*

- (1) *There exists k such that $\|\Phi^k\|_H < 1$.*

- (2) Every orbit of the system (27) converges to an equilibrium co-linear to I_n .
- (3) The subspace $\cap_k \mathcal{G}_k$ 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 Φ^k .

(2) \Rightarrow (1): Consider the quotient real linear space $\mathcal{W} := S_n/\mathbb{R}I_n$. Since $\Phi(I_n) = I_n$, Φ yields a quotient linear map $\mathcal{W} \mapsto \mathcal{W}$. We already observed in (15) that $\|\Phi\|_H$ is the operator norm induced by the norm $\|\cdot\|_H$ on \mathcal{W} . It follows that $\|\Phi_1\Phi_2\|_H \leq \|\Phi_1\|_H\|\Phi_2\|_H$ holds for all linear maps $\Phi_1, \Phi_2: \mathcal{W} \rightarrow \mathcal{W}$, and so, by Fekete's subadditive lemma,

$$\inf_{k \geq 1} \|\Phi^k\|_H^{1/k} = \lim_{k \rightarrow +\infty} \|\Phi^k\|_H^{1/k}$$

Observe also that $\|\Phi\|_H \leq 1$, so that $\|\Phi^k\|_H \leq 1$ holds for all $k \geq 1$. Then, if (1) is not true, we deduce that

$$\lim_{k \rightarrow +\infty} \|\Phi^k\|_H^{1/k} = 1 . \quad (29)$$

Now, for any real normed vector space $(\mathcal{V}, \|\cdot\|_{\mathcal{V}})$, let $\mathcal{V}_{\mathbb{C}} = \mathcal{V} + i\mathcal{V}$ denote the complexification of \mathcal{V} , and for any \mathbb{R} -linear self-map T of \mathcal{V} , let $T_{\mathbb{C}}$ denote the complexification of T , so that $T_{\mathbb{C}}(x + iy) = T(x) + iT(y)$ for all $x, y \in \mathcal{V}$. Recall that $\mathcal{V}_{\mathbb{C}}$ can be equipped with the norm

$$\|x + iy\|_{\mathcal{V}_{\mathbb{C}}} = \sup_{0 \leq \theta < 2\pi} \|x \cos \theta - y \sin \theta\|_{\mathcal{V}}$$

and that the operator norm of T induced by the norm $\|\cdot\|_{\mathcal{V}}$ on \mathcal{V} , denoted by $\|T\|_{\mathcal{V}}$, as well as the operator norm of $T_{\mathbb{C}}$ induced by $\|\cdot\|_{\mathcal{V}_{\mathbb{C}}}$ on $\mathcal{V}_{\mathbb{C}}$, denoted by $\|T_{\mathbb{C}}\|_{\mathcal{V}_{\mathbb{C}}}$, coincide,

$$\|T_{\mathbb{C}}\|_{\mathcal{V}_{\mathbb{C}}} = \|T\|_{\mathcal{V}} . \quad (30)$$

Consider in particular $(S_n)_{\mathbb{C}} = S_n + iS_n \simeq \mathbb{C}^{n \times n}$, observe that $(S_n/\mathbb{R}I_n)_{\mathbb{C}} \simeq \mathbb{C}^{n \times n}/\mathbb{C}I_n$. It follows from (30) that

$$\|\Phi_{\mathbb{C}}^k\|_{\mathcal{W}_{\mathbb{C}}} = \|\Phi^k\|_{\mathcal{W}} = \|\Phi^k\|_H$$

holds for all k , and so, by (29),

$$\lim_{k \rightarrow \infty} \|\Phi_{\mathbb{C}}^k\|_{\mathcal{W}_{\mathbb{C}}}^{1/k} = 1 . \quad (31)$$

By Gelfand's formula, the left-hand side of (31) is the spectral radius of the \mathbb{C} -linear map $\Phi_{\mathbb{C}}: \mathcal{W}_{\mathbb{C}} \rightarrow \mathcal{W}_{\mathbb{C}}$. Hence, $\Phi_{\mathbb{C}}$ has an eigenvalue on the unit circle, meaning that there exists $\theta \in [0, 2\pi)$, $X, Y \in S_n$, with $X + iY \notin \mathbb{C}I_n$, such that

$$\Phi(X + iY) - e^{i\theta}(X + iY) \in \mathbb{C}I_n ,$$

and so

$$\Phi^k(X + iY) - e^{ik\theta}(X + iY) \in \mathbb{C}I_n ,$$

for all $k \geq 1$. Identifying the real and imaginary parts, we get $\Phi^k(X) = \cos(k\theta)X - \sin(k\theta)Y + \alpha_k I_n$ and $\Phi^k(Y) = \sin(k\theta)X + \cos(k\theta)Y + \beta_k I_n$, for some $\alpha_k, \beta_k \in \mathbb{R}$. Observe that since $X + iY \notin \mathbb{C}I_n$, we have $X, Y \notin \mathbb{R}I_n$. It follows that the orbit $(\Phi^k(X))_{k \geq 1}$ does not converge to a scalar multiple of I_n , contradicting (2).

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

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

By Corollary 7.3, we know that $\|\Phi^k\|_H = 1$ if and only if the subspace \mathcal{G}_k contains a rank one matrix. Therefore, $\|\Phi^k\|_H = 1$ for all $k \in \mathbb{N}$ if and only if the subspace $\cap_k \mathcal{G}_k$ contains a rank one matrix.

(3) \Rightarrow (4): By (28), there is $k_0 \leq n^2 - 1$ such that $\mathcal{G}_{k_0} = \cap_k \mathcal{G}_k$. It follows that if (3) is true then there is $k_0 \leq n^2 - 1$ such that \mathcal{G}_{k_0} does not contain a rank one matrix. Then by Corollary 7.3 we deduce that $\|\Phi^{k_0}\| < 1$ if (3) is true. \square

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

$$\Pi_{k+1} = \Psi(\Pi_k), \quad k = 1, 2, \dots \quad (32)$$

Below is a dual version of Theorem 7.7.

Theorem 7.8. *The following conditions are equivalent:*

- (1) *There exists k such that $\|\Psi^k\|_H^* < 1$.*
- (2) *The Markov chain (32) converges to a unique invariant measure regardless of initial distribution.*
- (3) *The subspace $\cap_k \mathcal{G}_k$ does not contain a rank one matrix.*
- (4) *There exists $k_0 \leq n^2 - 1$ such that $\|\Psi^{k_0}\|_H^* < 1$.*

Remark 7.9. A sufficient condition for the global convergence of the noncommutative consensus system (27) or equivalently, the ergodicity of the noncommutative Markov chain (32) would be that there is $k_0 \leq n^2 - 1$ such that

$$\mathcal{H}_{k_0} = \mathbb{C}^{n \times n}.$$

Thus, checking the global convergence appears to be more tractable than checking the one step contraction (compare this characterization with the one of Corollary 7.3).

References

- [AB99] C. D. Aliprantis and K. C. Border. *Infinite Dimensional Analysis. A Hirschiker's Guide*. Springer, 1999.

- [AB09] D. Angeli and P.-A. Bliman. Convergence speed of unsteady distributed consensus: decay estimate along the settling spanning-trees. *SIAM J. Control Optim.*, 48(1):1–32, 2009.
- [Alf71] E. M. Alfsen. *Compact convex sets and boundary integrals*. Springer-Verlag, New York, 1971. *Ergebnisse der Mathematik und ihrer Grenzgebiete, Band 57*.
- [BGPS06] S. Boyd, A. Ghosh, B. Prabhakar, and D. Shah. Randomized gossip algorithms. *IEEE Trans. Inform. Theory*, 52(6):2508–2530, 2006.
- [BHOT05] V. D. Blondel, J. M. Hendrickx, A. Olshevsky, and J. N. Tsitsiklis. Convergence in multiagent coordination, consensus, and flocking. In *Proceedings of the joint 44th IEEE Conference on Decision and Control and European Control Conference*, pages 2996–3000. IEEE, 2005.
- [Bir57] G. Birkhoff. Extensions of Jentzsch’s theorem. *Trans. Amer. Math. Soc.*, 85:219–227, 1957.
- [Bou93] Ph. Bougerol. Kalman filtering with random coefficients and contractions. *SIAM J. Control Optim.*, 31(4):942–959, 1993.
- [BS08] S. Beigi and P. W. Shor. On the complexity of computing zero-error and holevo capacity of quantum channels. arxiv:0709.2090v3, 2008.
- [BT89] D. P. Bertsekas and J. 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] J. B. Conway. *A course in functional analysis*, volume 96 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, second edition, 1990.
- [CSM05] M. Cao, D. A. Spielman, and A. S. Morse. 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.
- [Dob56] R. Dobrushin. Central limit theorem for non-stationary Markov chains. *I. Teor. Veroyatnost. i Primenen.*, 1:72–89, 1956.
- [Ell64] A. J. Ellis. The duality of partially ordered normed linear spaces. *J. London Math. Soc.*, 39:730–744, 1964.
- [EN95] S. P. Eveson and R. D. Nussbaum. An elementary proof of the Birkhoff-Hopf theorem. *Math. Proc. Cambridge Philos. Soc.*, 117(1):31–55, 1995.
- [GG04] S. Gaubert and J. Gunawardena. The Perron-Frobenius theorem for homogeneous, monotone functions. *Trans. of AMS*, 356(12):4931–4950, 2004.
- [Hir89] M. 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] E. Kohlberg and J. W. Pratt. The contraction mapping approach to the Perron-Frobenius theory: why Hilbert’s metric? *Math. Oper. Res.*, 7(2):198–210, 1982.
- [LPW09] D. A. Levin, Y. Peres, and E. L. Wilmer. *Markov chains and mixing times*. American Mathematical Society, Providence, RI, 2009. With a chapter by James G. Propp and David B. Wilson.
- [MA05] R. A. C. Medeiros and F. M. De Assis. Quantum zero-error capacity. *Int. J. Quantum Inform.*, 03:135, 2005.
- [Mor05] L. Moreau. Stability of multiagent systems with time-dependent communication links. *IEEE Trans. Automat. Control*, 50(2):169–182, 2005.
- [Muk13] F. Mukhamedov. The Dobrushin ergodicity coefficient and the ergodicity of noncommutative Markov chains. *J. Math. Anal. Appl.*, 408(1):364–373, 2013.
- [Nag74] R. J. Nagel. Order unit and base norm spaces. In *Foundations of quantum mechanics and ordered linear spaces (Advanced Study Inst., Marburg, 1973)*, pages 23–29. Lecture Notes in Phys., Vol. 29. Springer, Berlin, 1974.
- [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] A. Olshevsky and J. N. Tsitsiklis. Convergence speed in distributed consensus and averaging. *SIAM J. Control Optim.*, 48(1):33–55, 2009.
- [RKW11] D. Reeb, M. J. Kastoryano, and M. M. Wolf. Hilbert’s projective metric in quantum information theory. *J. Math. Phys.*, 52(8):082201, 33, 2011.
- [SSR10] R. Sepulchre, A. Sarlette, and P. Rouchon. Consensus in noncommutative spaces. In *Proceedings of the 49th IEEE Conference on Decision and Control*, pages 6596–6601, Atlanta, USA, Dec 2010.
- [TBA86] J. N. Tsitsiklis, D. P. Bertsekas, and M. Athans. Distributed asynchronous deterministic and stochastic gradient optimization algorithms. *IEEE Trans. Automat. Control*, 31(9):803–812, 1986.

Stéphane Gaubert
 INRIA and CMAP UMR 7641 CNRS
 École Polytechnique
 91128 Palaiseau Cédex, France
 e-mail: Stephane.Gaubert@inria.fr

Zheng Qu
 School of Mathematics, University of Edinburgh, Edinburgh, EH9 3JZ, UK
 e-mail: zheng.qu@ed.ac.uk