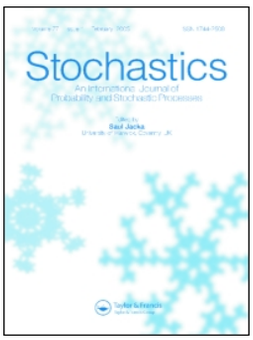

Taylor & Francis
Taylor & Francis Group

Stochastics

An International Journal of Probability and Stochastic Processes

ISSN: 1744-2508 (Print) 1744-2516 (Online) Journal homepage: https://www.tandfonline.com/loi/gssr20

On quantum versions of the classical Wasserstein distance

J. Agredo & F. Fagnola

To cite this article: J. Agredo & F. Fagnola (2017) On quantum versions of the classical Wasserstein distance, Stochastics, 89:6-7, 910-922, DOI: 10.1080/17442508.2016.1276914

To link to this article: https://doi.org/10.1080/17442508.2016.1276914

STOCHASTICS, 2017
VOL. 89, NOS. 6–7, 910-922
https://doi.org/10.1080/17442508.2016.1276914

Taylor & Francis
Taylor & Francis Group

On quantum versions of the classical Wasserstein distance

J. Agredo[a,b] and F. Fagnola[c]

[a]Facultad de Ciencias, Departamento de Matemáticas, Universidad Nacional de Colombia, Carrera, Colombia; [b]Departamento de Matemáticas, Ciudad Universitaria Bogotá, Escuela Colombiana de Ingeniería Julio Garavito, Carrera, Colombia; [c]Dipartimento di Matematica, Politecnico di Milano, Milano, Italy

ABSTRACT

We investigate a definition of quantum Wasserstein distance of two states based on their couplings on the product algebra as in the classical case. A detailed analysis of the two qubit model leads to a formal definition fulfilling some minimal requirements. It also shows that a clear-cut definition, by direct generalization of the classical one, is not achievable.

ARTICLE HISTORY
Received 20 April 2016
Accepted 22 December 2016

KEYWORDS
Wasserstein distance; quantum probability; qbit state

1. Introduction

Let (M, d) be a separable metric space. The reader can think of a subset of a Euclidean space with the Euclidean distance to fix the ideas. Denote by $\mathcal{P}_1(M)$ be the collection of all probability measures μ on M for which $\int_M d(x,z)d\mu(x) < \infty$ for some (and hence for all by Proposition 11.8.1 of [8]) $z \in M$.

The classical Wasserstein distance of two probability measures $\mu, \nu \in \mathcal{P}_1(E)$ is defined by

$$W(\mu,\nu) = \inf_{\vartheta \in \Xi(\mu,\nu)} \int_{M\times M} d(x,y)d\vartheta(x,y)$$

where $\Xi(\mu,\nu)$ is the set of all probability measures ϑ on $M \times M$ with marginals μ and ν, i.e. such that for all Borel subsets A, B of M

$$\vartheta(A \times M) = \mu(A), \quad \vartheta(M \times B) = \nu(B).$$

The Kantorovich–Rubinstein theorem (see [8] Section 11.8 for a proof) provides another representation for $W(\mu,\nu)$

$$W(\mu,\nu) = \sup\left\{ \int_M f d(\mu-\nu) \,\middle|\, |f(x)-f(y)| \le d(x,y) \quad \forall x,y,\in M \right\}$$

i.e. the supremum is over all f with Lipschitz constant not bigger than 1.

It is well-known (see [11]) that, for a finite set M, $W(\mu,\nu)$ is equivalent to the total variation distance $\|\mu-\nu\|_{\mathrm{TV}}$; more precisely

$$d_{\min}\|\mu-\nu\|_{\mathrm{TV}} \le W(\mu,\nu) \le \mathrm{diam}(M)\,\|\mu-\nu\|_{\mathrm{TV}} \tag{1}$$

CONTACT F. Fagnola franco.fagnola@polimi.it

where $d_{\min}(M) = \min\{d(x,y) \mid x \neq y\}$, $\operatorname{diam}(M) = \max\{d(x,y) \mid x, y \in M\}$.

Notions corresponding to Wasserstein distance appeared in several fields in mathematics (see [13] introduction, Chapters 1 and 8). This investigation aims at finding a good quantum analogue of the Wasserstein distance. The norm distance for normal states (as linear functionals on a von Neumann algebra) is the appropriate quantum generalization of the total variation distance of probability measures, therefore a good quantum analogue of the classical Wasserstein distance should be equivalent to the total variation distance for systems with finite degrees of freedom.

Our interest in the Wasserstein distance for quantum states is motivated by [3,4,6,7,14,15] where several different analogues of this notion have been proposed and none emerged as a good, universal, concept as in the classical case.

In Zyczkowski and Slomczynski [14,15] define a Monge distance between quantum states of some special classes (coherent and number states, for instance) as the Wasserstein distance between the corresponding Husimi distributions of quantum optics. In Biane and Voiculescu [4] investigate the free noncommutative analogue of the classical Wasserstein distance between two probability measures. In D'Andrea and Martinetti [6] discuss the relationship between the Wasserstein distance between probability distributions on a metric space, arising in the study of Monge-Kantorovich transport problem, and the spectral distance of noncommutative geometry. In [7] the author defines the cost-distance between two density matrices as the classical Wasserstein of the corresponding two probability distributions on a suitable quotient space with cost given by the geodesic distance on this space. Concepts arising from these investigations do not fully reflect the meaning of this distance in classical probability.

In this paper we investigate possible direct generalizations of the classical definition based on minimization of a functional on all possible couplings. Following the usual path for finding extensions of classical probabilistic notions in quantum probability, see e.g. the introduction of [2] for quantum Markov chains and the introduction of [10] for entropy production), we list below a few minimal requirements that ought to be fulfilled by any candidate:

(1) it should be a distance,
(2) it should be equivalent to the norm distance for states on a finite dimensional von Neumann algebra because the classical Wasserstein distance is equivalent to the norm distance on finite probability spaces by (1), and
(3) its restriction to states on a maximal abelian subalgebra should coincide with their classical Wasserstein distance because classical probability spaces arise in a quantum model by considering observables, i.e. random variables, belonging to an abelian subalgebra and restrictions of states to this subalgebra.

The case study of two q-bit states shows that it is possible to define such a distance. However, one has to give up the classical idea of looking for the minimum of a single functional. This seems due to the presence of many incompatible observables contrary to what happens in classical models. As a result, mimicking of the classical definition seems unfeasible.

The paper is organized as follows. In Section 2 we analyse the case of a two-point probability space $\{-1,+1\}$ in the quantum language. Additionally, we show that the Wasserstein distance of two probability measures in $\{-1,+1\}$ is the same distance if

the defining functional is minimized over all possible *quantum* couplings. Moreover, we give a formal definition of Wasserstein distance for quantum states, with the above three properties, however, minimizing three functionals. In Section 3 we show how some reasonable choices of a single functional for minimization do not lead to a good definition.

2. The classical case in the quantum language

We begin by considering the simplest case of two probability measures on a two-point space $\{-1,+1\}$, say, and two probability densities μ,ν

$$\frac{1}{2}\left(1-a,1+a\right), \qquad \frac{1}{2}\left(1-b,1+b\right)$$

with $-1\leq a,b\leq 1$. We analyse this simple case because, on one hand if a notion admits a reasonable non-commutative analogue the translation should be transparent at least in the simplest situation, on the other hand non-commutative analogues of finite probability spaces already reveal non-trivial features ([1,5] and also [12], Chapter 1).

In this case $W(\mu,\nu)$ is twice the total variation distance

$$\|\mu-\nu\|_{\mathrm{TV}}=\frac{1}{2}\sum_j|\mu_j-\nu_j|=\frac{1}{2}|a-b|$$

because, for any function $f:\{-1,+1\}\to\mathbb{R}$, since $\mu_1-\nu_1=-(\mu_{-1}-\nu_{-1})$,

$$\begin{aligned} f(-1)(\mu_{-1}-\nu_{-1})+f(1)(\mu_1-\nu_1) &= \big(f(1)-f(-1)\big)(\mu_1-\nu_1)\\ &= \big(f(-1)-f(1)\big)(\mu_{-1}-\nu_{-1})\end{aligned}$$

and $|f(1)-f(-1)|\leq 2=|1-(-1)|$ by the condition on the Lipschitz norm.

A little thought shows that a convenient parametrization of densities on $\{-1,+1\}\times\{-1,+1\}$ with the above marginals is

$$\begin{aligned}\vartheta(-1,-1)&=\frac{1}{4}(1-a-b+c), \quad \vartheta(1,-1)=\frac{1}{4}(1+a-b-c),\\ \vartheta(-1,1)&=\frac{1}{4}(1-a+b-c), \quad \vartheta(1,1)=\frac{1}{4}(1+a+b+c),\end{aligned}$$

where c is a real parameter such that the above four numbers, that sum up to 1, are non-negative, i.e. such that

$$1+c\geq|a+b|,\ 1-c\geq|a-b|,$$

namely

$$|a+b|-1\leq c\leq 1-|a-b|. \tag{2}$$

(Note that $|a+b|-1\leq 1-|a-b|$ because $|a+b|+|a-b|=2\max\{|a|,|b|\}\leq 2$). The above joint density can be viewed as a coupling of μ and ν depending on the constant c.

Clearly, in this case

$$\int_{\{-1,+1\}\times\{-1,+1\}}|x-y|d\theta(x,y)=2\left(\theta(-1,1)+\theta(1,-1)\right)=\left(1-c\right)$$

so that, taking into account (2), we find $W(\mu,\nu) = |a-b|$.

It is instructive to see how to tackle the problem in the quantum language. We refer to P.-A. Meyer's book [12], chapter 1, for definitions and notations on two-level quantum systems.

The above probability measures are represented by two states μ, ν on the algebra $M_2(\mathbb{C})$ of 2×2 complex matrices.

$$\mu = \frac{1}{2}(\mathbb{1} + a\sigma_3), \qquad \nu = \frac{1}{2}(\mathbb{1} + b\sigma_3) \tag{3}$$

where $\mathbb{1}$ is the 2×2 identity matrix and σ_j are the Pauli matrices

$$\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \qquad \sigma_2 = \begin{pmatrix} 0 & -\mathrm{i} \\ \mathrm{i} & 0 \end{pmatrix}, \qquad \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

States on $M_2(\mathbb{C}) \otimes M_2(\mathbb{C})$ with μ and ν as marginals are parametrized by ([9] formula (4))

$$\rho = \frac{1}{4}\left(\mathbb{1}_4 + a\sigma_3 \otimes \mathbb{1} + b\mathbb{1} \otimes \sigma_3 + \sum_{i,j=1}^{3} c_{ij}\sigma_i \otimes \sigma_j\right) \tag{4}$$

where $\mathbb{1}_4 = \mathbb{1} \otimes \mathbb{1}$ is the identity matrix on $\mathbb{C}^2 \otimes \mathbb{C}^2$ and $(c_{ij})_{1 \le i,j \le 3}$ are nine real parameters subject to certain constraints for positivity of ρ that we do not need in the present discussion (see [9] discussion on positivity of 4×4 matrices as (4) in Section 2 and [9] formula (32)).

Looking at the above formula it becomes clear that quantum coupling is much more complex. Moreover, constraints on numbers c_{ij} are very difficult to handle because they are expressed by nonlinear functions. Nevertheless, for a pair of classical states, we can still solve the minimization problem in a simple way discovering a new feature of the quantum framework: the minimum point is not unique as in the classical case even if the minimum value is the same.

If we consider the Euclidean distance in the classical functional for minimization

$$\int_{E\times E} d(x,y)\,d\vartheta(x,y) = \sum_{i,j=\pm 1} |i-j|\vartheta(i,j),$$

thinking of the function $(i,j) \to |i-j|$ as the multiplication operator

$$\sum_{i,j=\pm 1} |i-j|\, E_i^i \otimes E_j^j = \left|\sum_{i=\pm 1} iE_i^i \otimes \mathbb{1} - \sum_{j=\pm 1} \mathbb{1} \otimes jE_j^j\right| = |\sigma_3 \otimes \mathbb{1} - \mathbb{1} \otimes \sigma_3|\,,$$

where E_{-1}^{-1}, E_1^1 are one-dimensional projectors such that $E_{-1}^{-1} + E_1^1 = \mathbb{1}$, becomes

$$\mathrm{tr}\left(|\sigma_3 \otimes \mathbb{1} - \mathbb{1} \otimes \sigma_3|\rho\right).$$

The above discussion makes it clear that $\sigma_3 \otimes \mathbb{1}$ and $\mathbb{1} \otimes \sigma_3$ play the role of position observables of the first system and second system.

A little computation yields

$$\mathrm{tr}\left(|\sigma_3 \otimes \mathbb{1} - \mathbb{1} \otimes \sigma_3|\rho\right) = (1 - c_{33}),$$

showing that we are essentially looking for minima of the same function as in the classical case.

Choosing the basis e_-, e_+ of $\mathbb{C}^2$

$$e_- = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \quad e_+ = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

and the basis $e_+ \otimes e_+, e_+ \otimes e_-, e_- \otimes e_+, e_- \otimes e_-$ of $\mathbb{C}^2 \otimes \mathbb{C}^2$ a long but straightforward computation shows that ρ is represented by the density matrix $1/4$ times

$$\begin{pmatrix} 1+a+b+c_{33} & c_{31} - \mathrm{i}c_{32} & c_{13} - \mathrm{i}c_{23} & d - \mathrm{i}s' \\ c_{31} + \mathrm{i}c_{32} & 1+a-b-c_{33} & s + \mathrm{i}d' & -c_{13} + \mathrm{i}c_{23} \\ c_{13} + \mathrm{i}c_{23} & s - \mathrm{i}d' & 1-a+b-c_{33} & -c_{31} + \mathrm{i}c_{32} \\ d + \mathrm{i}s' & -c_{13} - \mathrm{i}c_{23} & -c_{31} - \mathrm{i}c_{32} & 1-a-b+c_{33} \end{pmatrix} \tag{5}$$

where

$$s = c_{11} + c_{22}, \quad s' = c_{12} + c_{21}, \quad d = c_{11} - c_{22}, \quad d' = c_{12} - c_{21}.$$

Here, from non-negativity of the diagonal elements, we recover the classical condition (2)

$$|a+b| - 1 \leq c_{33} \leq 1 - |a-b|. \tag{6}$$

Therefore, we fix $c_{33} = 1 - |a-b|$ and show that we can choose the other parameters c_{ij} in such a way that ρ is positive semi-definite.

Suppose, to fix the ideas, that $a \leq b$. The other case can be dealt with the same way. The diagonal element $1 + a - b - c_{33}$ vanishes, the positive semi-definiteness of ρ implies

$$c_{13} = c_{23} = 0, \quad c_{31} = c_{32} = 0, \quad c_{12} = c_{21}, \quad c_{11} + c_{22} = 0,$$

and the density becomes

$$\frac{1}{2}\begin{pmatrix} 1+a & 0 & 0 & c_{11} - \mathrm{i}c_{12} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & b-a & 0 \\ c_{11} + \mathrm{i}c_{12} & 0 & 0 & 1-b \end{pmatrix} \tag{7}$$

which is clearly positive semi-definite whenever

$$c_{11}^2 + c_{12}^2 \leq (1+a)(1-b).$$

We summarize this conclusion in the following

Proposition 1: *The minimum value of*

$$\mathrm{tr}\left(|\sigma_3 \otimes \mathbb{1} - \mathbb{1} \otimes \sigma_3|\rho\right)$$

on the convex set of all states ρ with given marginals μ, ν is $|a-b|$. In particular it coincides with the classical Wasserstein distance of the probability densities (3).

It is worth noticing that minimum points are not unique, even if both states are pure i.e. $a = \pm 1, b = \mp 1$, and are entangled states unless $c_{11} = c_{12} = c_{21} = c_{22} = 0$.

Remarks.

(1) In view of analysing arbitrary states, note that, if the quantity to minimize was $1 - (c_{33} + c_{11} + c_{22})$ we would find, in general, a strictly smaller minimum value.
(2) For all $p > 0$, a simple computation yields

$$|\sigma_3 \otimes \mathbb{1} - \mathbb{1} \otimes \sigma_3|^p = 2^{p-1} |\sigma_3 \otimes \mathbb{1} - \mathbb{1} \otimes \sigma_3|.$$

Therefore, considering a power p, the final result only changes by a multiplicative constant.

Minimizing $(1 - c_{33})$ we get $|a_3 - b_3|$. Thus, by a permutation of the axes, minimizing $(1 - c_{jj})$ we get $|a_j - b_j|$. A general pair of two qubit states can be represented by

$$\mu = \frac{1}{2}(\mathbb{1} + a_1\sigma_1 + a_2\sigma_2 + a_3\sigma_3), \qquad \nu = \frac{1}{2}(\mathbb{1} + b_1\sigma_1 + b_2\sigma_2 + b_3\sigma_3) \tag{8}$$

where $a = (a_1, a_2, a_3), b = (b_1, b_2, b_3)$ are real vectors in the unit sphere of $\mathbb{R}^3$, i.e.

$$a_1^2 + a_2^2 + a_3^2 \leq 1, \qquad b_1^2 + b_2^2 + b_3^2 \leq 1.$$

As a consequence, a possible formal definition of the 'quantum' $W(\mu, \nu)$ distance is

$$W(\mu, \nu) = \sum_{j=1}^{3} \left(\inf_{\rho_j} \operatorname{tr} \left(|\sigma_j \otimes \mathbb{1} - \mathbb{1} \otimes \sigma_j| \rho_j \right) \right) \tag{9}$$

where each infimum is over the set of all states ρ_j with marginals μ, ν. In this way you get

$$\sum_{j=1}^{3} |a_j - b_j|$$

and, by changing the definition to

$$W_p(\mu, \nu) = \left(\sum_{j=1}^{3} \left(\inf_{\rho_j} \operatorname{tr} \left(|\sigma_j \otimes \mathbb{1} - \mathbb{1} \otimes \sigma_j| \rho_j \right) \right)^p \right)^{1/p} \tag{10}$$

for $p = 2$ we get the trace norm because

$$|\mu - \nu|^2 = 2^{-2} \, |(a-b) \cdot \sigma|^2 = \|a-b\|^2 \, \mathbb{1}$$

where $\|a-b\| = (\sum_{j=1}^{3} (a_j - b_j)^2)^{1/2}$ is the Euclidean distance. It follows that $|\mu - \nu| = \|a-b\| \, \mathbb{1}/2$ and so

$$\operatorname{tr} \left(|\mu - \nu| \right) = \|a-b\| = W_2(\mu, \nu).$$

Remarks.

(1) (Heuristics) By the uncertainty principle components of a state along non-commutative variables (observables) can not be measured simultaneously. Formulae (9) and (10) seem reminiscent of this physical obstruction which as no classical counterpart. An experimenter trying to measure the difference between two states observes the values of a set of maximal observables and the result will depend on this choice. In the classical case there is a single maximal observable and so a single functional to minimize.

(2) Note that

$$\begin{aligned}\sigma_3 \otimes \mathbb{1} - \mathbb{1} \otimes \sigma_3 &= \left((\sigma_3 + \mathbb{1}) \otimes \mathbb{1} - \mathbb{1} \otimes (\sigma_3 + \mathbb{1})\right) \\ &= 2\left(|e_+\rangle\langle e_+| \otimes \mathbb{1} - \mathbb{1} \otimes |e_+\rangle\langle e_+|\right).\end{aligned}$$

Therefore, if the system space is multidimensional, the obvious generalization seems minimization of

$$2\mathrm{tr}\left(\left|\, |e_j\rangle\langle e_j| \otimes \mathbb{1} - \mathbb{1} \otimes |e_j\rangle\langle e_j| \,\right|^2 \rho_j\right)$$

for a suitable collection of vectors $(e_j)_j$ on states ρ_j with the given marginals.

3. Two qubit quantum states

Formulae (9) and (10) define a distance fulfilling the requirements 1, 2 and 3 stated in the introduction, however, adopting this definition, we have to give up the classical idea of minimizing a single functional. This section is aimed at convincing those classical probabilists who are left with a bad taste in their mouths that some reasonable choices of a single functional do not work.

In the attempt to find a single functional, getting rid of summation over j and minimization on different states in (10) one would be tempted to try to minimize over all states ρ with given marginals

$$\max_{\nu\in\mathbb{R}^3} \mathrm{tr}\left(|\nu\cdot\sigma \otimes \mathbb{1} - \mathbb{1} \otimes \nu\cdot\sigma|^2\,\rho\right) = \max_{\nu\in\mathbb{R}^3} \left\langle \nu, \left(2\mathbb{1} - (c + c^T)\right)\nu\right\rangle$$

namely the biggest eigenvalue of $\left(2\mathbb{1} - (c + c^T)\right)$ (where c^T denotes the transpose matrix).

This choice is motivated by the analogy with the classical case where one minimizes the probability of finding different values in the position observable in the first and second factor. Here, however, there are several non commuting observables, therefore one may consider one dimensional projections $(\mathbb{1} + \nu\cdot\sigma)/2$ with ν a vector in the unit sphere ($\nu \in \mathbb{R}^3$, $\|\nu\| = 1$ and compute the probability of finding different values in the first and second factor for observables in the abelian algebra of observables commuting with $\nu\cdot\sigma$. Since the orthogonal projection of $(\mathbb{1} + \nu\cdot\sigma)/2$ is $(\mathbb{1} - \nu\cdot\sigma)/2$ this is

$$\begin{aligned}&\frac{1}{4}\mathrm{tr}\left(\left(\mathbb{1} + \nu\cdot\sigma\right) \otimes (\mathbb{1} - \nu\cdot\sigma) + \left(\mathbb{1} - \nu\cdot\sigma\right) \otimes (\mathbb{1} + \nu\cdot\sigma)\rho\right) \\ = \;&\frac{1}{2}\mathrm{tr}\left(\left(\mathbb{1} \otimes \mathbb{1} - \nu\cdot\sigma \otimes \nu\cdot\sigma\right)\rho\right) = \left\langle \nu, \left(2\mathbb{1} - (c + c^T)\right)\nu\right\rangle.\end{aligned}$$

Therefore we try to minimize, for instance,

(1) the biggest probability of finding different values

$$\max_{v\in\mathbb{R}^3}\left\langle v,\left(2\mathbb{1}-(c+c^T)\right)v\right\rangle \tag{11}$$

namely the biggest eigenvalue of $\left(2\mathbb{1}-(c+c^T)\right)$,

(2) the 'average probability' of finding different values

$$\int_{S^3}\left\langle v,\left(2\mathbb{1}-(c+c^T)\right)v\right\rangle \mathrm{d}v$$

(here the integral is with respect to the normalized Haar measure on the sphere) namely, since

$$\int_{S^3} v_j^2\mathrm{d}v=\frac{1}{2},\qquad \int_{S^3} v_jv_k\mathrm{d}v=0\quad \text{for } j\neq k$$

we find

$$\int_{S^3}\left\langle v,\left(2\mathbb{1}-(c+c^T)\right)v\right\rangle \mathrm{d}v=\sum_{j=1}^{3}\left(1-c_{jj}\right), \tag{12}$$

(3) the sum of probabilities of finding different values in three orthogonal directions (the analogue of the choice of [4] Section 1.1 for the analogue in free probability) yields twice the same functional.

However, none of the above choices work even for commuting states. Indeed, for $a=b$, after minimization over all choices of c leaving ρ positive semi-definite, we would like to find the value 0. This does not happen because:

(1) in the former case this would imply $1-c_{jj}\leq 0$ for all j, thus, by positive semi-definiteness of the matrix (5), we immediately see that this implies $c_{33}=1$ and so $c_{12}=c_{21}, c_{11}+c_{22}=0$ which is incompatible with $c_{11}\geq 1, c_{22}\geq 1$.

(2) (and 3.) in the latter cases $3-(c_{11}+c_{22}+c_{33})=0$ implies $c_{11}+c_{22}=3-c_{33}$, but positivity of the inner 2×2 block of (5) also implies

$$(1-c_{33})^2\geq (c_{11}+c_{22})^2+(c_{12}-c_{21})^2\geq (c_{11}+c_{22})^2=(3-c_{33})^2$$

This inequality is equivalent to $c_{33}\geq 2$. However, positivity of diagonal elements of the matrix (5) implies $1-c_{33}\geq |a-b|\geq 0$, i.e. $c_{33}\leq 1$ which is in contradiction with $c_{33}\geq 2$.

Another possible attempt is minimization of

$$\mathrm{tr}\left(|v\cdot\sigma\otimes\mathbb{1}-\mathbb{1}\otimes v\cdot\sigma|\,\rho\right) \tag{13}$$

for a unit vector v parallel to $b-a$ (for $b\neq a$ and $v=e_3$ for $b=a$). This choice, surprisingly, works when components of a and b orthogonal to $b-a$ are 'small' with respect to $b-a$. In a more precise way, we have the following

Theorem 2: *Suppose* $0<\|b-a\|<2$ *and let* $v=(b-a)/\|b-a\|$. *Then the minimum of*

$$\mathrm{tr}\left(|v\cdot\sigma\otimes\mathbb{1}-\mathbb{1}\otimes v\cdot\sigma|\,\rho\right) \tag{14}$$

on all states ρ with marginals (8) is $\|b-a\|$ if and only if

$$\begin{aligned}\|a-\langle a,v\rangle\, v\|^2 &= \|b-\langle b,v\rangle\, v\|^2 \\ &\leq \|b-a\|\left((1-\langle b,v\rangle\vee\langle a,v\rangle\,)\wedge(1+\langle a,v\rangle\wedge\langle b,v\rangle\,)\right)\end{aligned} \tag{15}$$

where $r\vee s$ (resp. $r\wedge s$) denotes the maximum (resp. minimum) of the real numbers r,s.

***Proof*:** By a change of coordinates in $\mathbb{R}^3$, exchanging the roles of a and b if necessary, we can suppose that

$$\|b-a\| = (b_3-a_3), \qquad a_2=b_2=0 \tag{16}$$

and, as a consequence, $a_1 = b_1$. The coupling coefficients matrix $(c_{ij})_{1\leq i,j\leq 3}$ will also change the new coordinates, and a little computation, as in the previous section, shows that

$$\mathrm{tr}\left(|v\cdot\sigma\otimes\mathbb{1}-\mathbb{1}\otimes v\cdot\sigma|\,\rho\right) = (1-c_{33}) \tag{17}$$

The state ρ is represented by the density matrix $1/4$ times

$$\begin{pmatrix} 1+a_3+b_3+c_{33} & b_1+c_{31}-\mathrm{i}c_{32} & a_1+c_{13}-\mathrm{i}c_{23} & d-\mathrm{i}s' \\ b_1+c_{31}+\mathrm{i}c_{32} & 1+a_3-b_3-c_{33} & s+\mathrm{i}d' & a_1-c_{13}+\mathrm{i}c_{23} \\ a_1+c_{13}+\mathrm{i}c_{23} & s-\mathrm{i}d' & 1-a_3+b_3-c_{33} & b_1-c_{31}+\mathrm{i}c_{32} \\ d+\mathrm{i}s' & a_1-c_{13}-\mathrm{i}c_{23} & b_1-c_{31}-\mathrm{i}c_{32} & 1-a_3-b_3+c_{33} \end{pmatrix}$$

where $s=c_{11}+c_{22}, s'=c_{12}+c_{21}, d=c_{11}-c_{22}, d'=c_{12}-c_{21}$. A necessary condition for the above matrix to be positive semi-definite clearly is $1-a_3-b_3-c_{33}\geq 0$, namely $1-c_{33}\geq b_3-a_3=\|b-a\|$.

We now show that, if $c_{33}=1-(b_3-a_3)$, it is possible to choose the other coefficients c_{ij} so as to get a positive semi-definite matrix. First of all, since a diagonal matrix element is zero, all the matrix elements on the same row and column must be zero i.e., recalling that $b_1=a_1$,

$$c_{32}=c_{23}=0,\ c_{13}=a_1,\ c_{31}=-a_1,\ c_{11}+c_{22}=0,\ c_{12}=c_{21}.$$

Now, neglecting the zero row and the zero column, and putting $c_{11}=x$, we remain with the 3×3 matrix

$$\frac{1}{2}\begin{pmatrix} 1+a_3 & a_1 & x \\ a_1 & b_3-a_3 & a_1 \\ x & a_1 & 1-b_3 \end{pmatrix} \tag{18}$$

which is positive semi-definite if and only if all the principal minors, namely the diagonal elements, the determinants of the 2×2 matrices

$$\begin{pmatrix} 1+a_3 & a_1 \\ a_1 & b_3-a_3 \end{pmatrix},\quad \begin{pmatrix} b_3-a_3 & a_1 \\ a_1 & 1-b_3 \end{pmatrix},\quad \begin{pmatrix} 1+a_3 & x \\ x & 1-b_3 \end{pmatrix} \tag{19}$$

and the determinant of the matrix itself are non-negative. Now, recalling $-1\leq a_3\leq 1$ and $b_3-a_3\geq 0$, the matrix (18) turns out to be positive semi-definite if and only if

$$a_1^2\leq(b_3-a_3)(1+a_3),\ \ a_1^2\leq(b_3-a_3)(1-b_3),\ \ x^2\leq(1+a_3)(1-b_3) \tag{20}$$

and its determinant, which is 1/8 times,

$$-(b_3 - a_3)x^2 + 2a_1^2 x + (b_3 - a_3)(1 - b_3)(1 + a_3) - a_1^2(2 - b_3 + a_3) \tag{21}$$

is non-negative. The discriminant of this second order polynomial in x is 1/4 times

$$\begin{aligned} &a_1^2 - a_1^2(b_3 - a_3)(2 - b_3 + a_3) + (b_3 - a_3)(1 - b_3)(1 + a_3) \\ &= a_1^2\left(1 - b_3 + a_3\right)^2 + (b_3 - a_3)(1 - b_3)(1 + a_3) \end{aligned}$$

and so it is also non-negative. The determinant of the 3×3 matrix (18) is non-negative for all x in the interval $[x_1, x_2]$ determined by the roots x_1, x_2 of the second order polynomial (21). It turns out that we can find an x such that the 3×3 matrix (18) is positive semi-definite if and only if the intervals $[-((1 + a_3)(1 - b_3))^{1/2}, ((1 + a_3)(1 - b_3))^{1/2}]$ and $[x_1, x_2]$ have non-empty intersection.

Note that the second order polynomial (21) has at least a positive root since $-(b_3 - a_3) < 0$ and $2a_1^2 \geq 0$ so that $x_2 > 0$. Moreover, if we evaluate it at $\pm((1 + a_3)(1 - b_3))^{1/2}$ we find

$$\begin{aligned} &\pm 2a_1^2((1 + a_3)(1 - b_3))^{1/2} - a_1^2(2 - b_3 + a_3) \\ &= -a_1^2\left((1 + a_3)^{1/2} \mp (1 - b_3)^{1/2}\right)^2 \leq 0 \end{aligned} \tag{22}$$

Therefore both $\pm((1 + a_3)(1 - b_3))^{1/2}$ lie outside the open interval $]x_1, x_2[$ and we can find and x for which the matrix (18) is positive semi-definite, namely an x such that $|x| \leq ((1 + a_3)(1 - b_3))^{1/2}$ and (21) is non-negative, the in the following cases.

(1) $x_2 \leq ((1+a_3)(1-b_3))^{1/2}$. This happens if and only if the middle point $a_1^2/(b_3 - a_3)$ of the interval $[x_1, x_2]$ satisfies $a_1^2/(b_3 - a_3) \leq ((1 + a_3)(1 - b_3))^{1/2}$ which is (15).
(2) $x_1 = ((1 + a_3)(1 - b_3))^{1/2}$. In this case the determinant of the 3×3 matrix (18) vanishes irrespectively of a_1, moreover $1 + a_3 = 1 - b_3$ by (22) and positive semi-definiteness of the three 2×2 diagonal blocks holds if and only if $a_1^2 \leq (b_3 - a_3)(1 + a_3) = (b_3 - a_3)(1 - b_3) = (b_3 - a_3)(1 - b_3) = (b_3 - a_3)((1 + a_3)(1 - b_3))^{1/2}$ which is again (15).

The proof is completed noting that

$$(b_3 - a_3)\left((1 - b_3) \wedge (1 + a_3)\right) \leq (b_3 - a_3)((1 - b_3)(1 + a_3))^{1/2}$$

and so also the three 2×2 matrices (19) and the 3×3 matrix (18) are positive semi-definite, if and only the inequality (20) holds. □

The next negative result shows that, for $v = (b - a)/\|b - a\|$, by minimizing (13) it is impossible to obtain a norm which is *equivalent* to the total variation norm.

Theorem 3: *With the notation of Theorem* (2), $v = (b - a)/\|b - a\|$, *there exists no positive constant* κ, *independent of* a, b, *such that*

$$\|b - a\| \leq \min_{\rho} \operatorname{tr}\left(|v \cdot \sigma \otimes \mathbb{1} - \mathbb{1} \otimes v \cdot \sigma|\, \rho\right) \leq \kappa \|b - a\|$$

for all a, b *in the unit ball of* $\mathbb{R}^3$.

***Proof*:** By a change of coordinates in $\mathbb{R}^3$, as in the proof of Theorem 2, exchanging the roles of a and b if necessary, we can suppose that (16) holds.

If a constant κ satisfying the above inequality existed, it would be bigger than 1. Moreover, by non-negativity of the second diagonal element of ρ, from (17) we get the constraint $1 - c_{33} \leq \kappa(b_3 - a_3)$.

Theorem 2 suggests that something may go wrong when $a_1 = b_1$ is big and $b_3 - a_3$ is small. For this reason, we now consider two states depending on a parameter $\varepsilon > 0$ small enough with

$$a_1 = b_1 = 1 - \varepsilon, \qquad b_3 = \frac{\varepsilon}{2}, \qquad a_3 = -\frac{\varepsilon}{2}.$$

The reader can easily check that, for any $\varepsilon \in (0, 3 - \sqrt{7})$ we find states which do not satisfy (15).

The state ρ is represented by the density matrix $1/4$ times

$$\begin{pmatrix} 1 + c_{33} & 1 - \varepsilon + c_{31} - \mathrm{i}c_{32} & 1 - \varepsilon + c_{13} - \mathrm{i}c_{23} & d - \mathrm{i}s' \\ 1 - \varepsilon + c_{31} + \mathrm{i}c_{32} & 1 - c_{33} - \varepsilon & s + \mathrm{i}d' & 1 - \varepsilon - c_{13} + \mathrm{i}c_{23} \\ 1 - \varepsilon + c_{13} + \mathrm{i}c_{23} & s - \mathrm{i}d' & 1 - c_{33} + \varepsilon & 1 - \varepsilon - c_{31} + \mathrm{i}c_{32} \\ d + \mathrm{i}s' & 1 - \varepsilon - c_{13} - \mathrm{i}c_{23} & 1 - \varepsilon - c_{31} - \mathrm{i}c_{32} & 1 + c_{33} \end{pmatrix}$$

where $s = c_{11} + c_{22}, s' = c_{12} + c_{21}, d = c_{11} - c_{22}, d' = c_{12} - c_{21}$.

Now, since $1 - c_{33} \leq \kappa(b_3 - a_3) = \kappa\varepsilon$, the matrix

$$\begin{pmatrix} 1 + c_{33} & 1 - \varepsilon + c_{31} - \mathrm{i}c_{32} & 1 - \varepsilon + c_{13} - \mathrm{i}c_{23} & d - \mathrm{i}s' \\ 1 - \varepsilon + c_{31} + \mathrm{i}c_{32} & (\kappa - 1)\varepsilon & s + \mathrm{i}d' & 1 - \varepsilon - c_{13} + \mathrm{i}c_{23} \\ 1 - \varepsilon + c_{13} + \mathrm{i}c_{23} & s - \mathrm{i}d' & (\kappa + 1)\varepsilon & 1 - \varepsilon - c_{31} + \mathrm{i}c_{32} \\ d + \mathrm{i}s' & 1 - \varepsilon - c_{13} - \mathrm{i}c_{23} & 1 - \varepsilon - c_{31} - \mathrm{i}c_{32} & 1 + c_{33} \end{pmatrix}$$

has to be positive definite, for some choice of c_{ij} with $i \neq 3, j \neq 3$, for all $\varepsilon \in (0, 1)$. A lengthy computation shows that its characteristic polynomial is

$$\lambda^4 - 2(\kappa\varepsilon + 1 + c_{33})\lambda^3 + \Big(- (3 + c_{33})(1 - c_{33}) - 2 \sum_{(i,j) \neq (3,3)} c_{ij}^2 + 4\,(2 + \kappa(1 + c_{33}))\varepsilon + (\kappa^2 - 5)\varepsilon^2 \Big)\lambda^2 + \text{l.o.t.}$$

where l.o.t. are lower order terms in λ. Now, we have to consider only coefficients $c_{ij}(\varepsilon)$ (we emphasize here dependence on the given marginal states, i.e. on ε) in such a way that ρ is positive semi-definite. By positivity of diagonal elements of ρ we have $-1 \leq c_{33}(\varepsilon) \leq 1 - \varepsilon$.

Note that, the λ^3 coefficient is strictly negative for $\varepsilon > 0$. Since the four real eigenvalues of the above 4×4 Hermitian matrix must be nonnegative, by the well-known Descartes' rule of signs of roots of polynomials (the number of positive roots of a polynomial is either equal to the number of sign differences between consecutive nonzero coefficients, or is less than it by an even number), the coefficient of λ^2 must be non-negative i.e. our choice of $c_{ij}(\varepsilon)$ must satisfy

$$(3 + c_{33}(\varepsilon))(1 - c_{33}(\varepsilon)) + 2 \sum_{(i,j) \neq (3,3)} c_{ij}^2(\varepsilon) \leq 4(2 + \kappa(1 + c_{33}(\varepsilon)))\varepsilon + (\kappa^2 - 5)\varepsilon^2$$

for ε in a right neighbourhood of 0, in particular, since $c_{33}(\varepsilon) \leq 1 - \varepsilon$,

$$(3 + c_{33}(\varepsilon))(1 - c_{33}(\varepsilon)) + 2 \sum_{(i,j)\neq(3,3)} c_{ij}^2(\varepsilon) \leq \varepsilon \left(4(2 + \kappa(2 + \varepsilon)) + |\kappa^2 - 5|\varepsilon\right).$$

As a consequence

$$\lim_{\varepsilon\to 0^+} c_{33}(\varepsilon) = 1, \qquad \lim_{\varepsilon\to 0^+} c_{ij}(\varepsilon) = 0,$$

for all $(i, j) \neq (3, 3)$. Positive semi-definiteness of the 2×2 left-upper corner of ρ implies

$$\left(1 - \varepsilon + c_{31}(\varepsilon)\right)^2 + c_{32}(\varepsilon)^2 \leq (1 + c_{33}(\varepsilon))(1 - c_{33}(\varepsilon))$$

and letting $\varepsilon \to 0^+$ we find the contradiction $1 \leq 0$. □

Acknowledgements

Part of this work was done while the authors were visiting the ANESTOC Centre in September 2015. We wish to thank the director R. Rebolledo for valuable discussions and comments.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

JA and Universidad Nacional de Colombia want to highlight the financial support from 'Patrimonio autónomo fondo nacional de financiamiento para la ciencia, tecnología y la innovación, Francisco Jose de Caldas' and 'Universidad Nacional de Colombia (UN)'.

References

[1] L. Accardi, F. Fagnola, and S. Hachicha, *Generic q-Markov semigroups and speed of convergence of q-algorithms, Infin*, Dimens. Anal. Quantum Probab. Relat. Top. 9 (2006), pp. 567–594.
[2] L. Accardi, and F. Fidaleo, *Entangled Markov chains*, Ann. Mat. Pur. Appl. 184 (2005), pp. 327–346.
[3] J. Agredo, *A Wasserstein-type distance to measure deviation from equilibrium of quantum Markov semigroups*, Open Syst. Inf. Dyn. 20 (2013), pp. 1350009-1–1350009-20.
[4] Ph Biane, and D. Voiculescu, *A free probability analogue of the Wasserstein metric on the trace-state space*, Geom. Funct. Anal. 11 (2001), pp. 1125–1138.
[5] R. Carbone, E. Sasso, and V. Umanità, *Decoherence for quantum Markov semi-groups on matrix algebras*, Ann. Henri Poincaré 14 (2013), pp. 681–697.
[6] F. D'Andrea, and P. Martinetti, *A view on transport theory from noncommutative geometry*, SIGMA 6 (2010), pp. 057.
[7] F. D'Andrea, *Pythagoras Theorem in Noncommutative Geometry*, Proceedings of the Conference on Optimal Transport and Noncommutative Geometry, Besancon (2014). Available at arXiv:1507.08773.
[8] R. Dudley, *Real Analysis and Probability*, Cambridge in Advanced Mathematics Vol. 74 , Cambridge University Press, Cambridge, 2002.
[9] B.G. Englert, and M. Metwally, *Remarks on 2-q-bit states*, Appl. Phys. B 72 (2001), pp. 35–42.
[10] F. Fagnola, and R. Rebolledo, *Entropy production for quantum Markov semigroups*, Commun. Math. Phys. 335 (2015), pp. 547–570.

[11] Alison L. Gibbs, and Francis Edward Su, *On choosing and bounding probability metrics*, Int. Stat. Rev. 70(3) (2002), pp. 419–435. Available at arXiv:math/0209021.
[12] Paul-André Meyer, *Quantum Probability for Probabilists*, Lecture Notes in Mathematics, Vol. 1538., 2nd ed., Springer-Verlag, Berlin, 1995.
[13] C. Villani, *Topics in Optimal Transportation*, Graduate Studies in Mathematics Vol. 58, American Mathematical Society, Providence, 2003.
[14] K. Zyczkowski, and W. Slomczynski, *Monge distance between quantum states*, J. Phys. A 31 (1998), pp. 9095–9104.
[15] K. Zyczkowski, and W. Slomczynski, *Monge metric on the sphere and geometry of quantum states*, J. Phys. A 34 (2001), pp. 6689–6722.