A spectral gap for POVMs

Victoria Kaminker, Leonid Polterovich and Dor Shmoish\textsuperscript{a}

June 30, 2017

Abstract

For a class of positive operator valued measures, we introduce the spectral gap, an invariant which shows up in a number of contexts: the quantum noise operator responsible for the unsharpness of quantum measurements, the Markov chain describing the state reduction for repeated quantum measurements, and the Berezin transform on compact Kähler manifolds. The spectral gap admits a transparent description in terms of geometry of certain metric measure spaces, is related to the diffusion distance, and exhibits a robust behaviour under perturbations in the Wasserstein metric.

Contents

1 Introduction 2
2 Preliminaries on POVMs 3
3 Introducing spectral gap 3
4 Minimal noise of quantum measurement 4
5 Repeated quantum measurement 6
6 Spectral gap for quantization 9
7 A geometric interpretation 10

\textsuperscript{a}The authors are partially supported by the European Research Council Advanced grant 338809.
1 Introduction

Positive operator valued measures (POVMs) appear in quantum mechanics on at least two different occasions. First, they model quantum measurements [3, 4]. Second, they provide a natural language for the Berezin-Toeplitz quantization of classical phase spaces [13, 6]. In the present note we deal with a special class of POVMs which, roughly speaking, admit an operator-valued density with respect to a probability measure. With such a POVM, one naturally associates a Markov operator and a quantum channel which possess the same positive eigenvalues. The difference between the two maximal eigenvalues of either of these operators is called the spectral gap of a POVM and lies in the focus of the present note. We refer to Sections 2 and 3 for preliminaries and precise definitions.

Our first group of results provides an interpretation of the spectral gap in a number of seemingly remote contexts: the quantum noise operator responsible for the unsharpness of quantum measurements (Section 4), the Markov chain describing the state reduction for repeated quantum measurements (Section 5), and the Berezin transform on compact Kähler manifolds (Section 6).

Furthermore, following Oreshkov and Calsamiglia [19, VII.C] we adopt a geometric viewpoint at POVMs of the above class encoding them as probability measures in the space of quantum states $S$ equipped with the Hilbert-Schmidt metric. It turns out that the spectral gap admits a transparent description in terms of the geometry of such metric measure spaces and exhibits a robust behaviour under perturbations of POVMs in the Wasserstein metric (Section 7). In a similar spirit, one can consider a POVM as a data cloud in $S$, which leads us to a link between the spectral gap and the diffusion distance, a notion coming from geometric data analysis (Section 8).
2 Preliminaries on POVMs

The mathematical model of quantum mechanics starts with a complex Hilbert space $\mathcal{H}$. In what follows we consider finite-dimensional Hilbert spaces only. Observables are represented by Hermitian operators whose space is denoted by $\mathcal{L}(\mathcal{H})$. Quantum states are provided by density operators, i.e., positive trace-one operators $\rho \in \mathcal{L}(\mathcal{H})$. They form a subset $\mathcal{S}(\mathcal{H}) \subset \mathcal{L}(\mathcal{H})$.

**Notation:** We write $((A, B))$ for the scalar product $\text{trace}(AB)$ on $\mathcal{L}(\mathcal{H})$.

Let $\Omega$ be a set equipped with a $\sigma$-algebra $\mathcal{C}$ of its subsets. By default, we assume that $\Omega$ is a Polish topological space (i.e., it is homeomorphic to a complete metric space possessing a countable dense subset) and $\mathcal{C}$ is the Borel $\sigma$-algebra. An $\mathcal{L}(\mathcal{H})$-valued positive operator valued measure (POVM) $W$ on $(\Omega, \mathcal{C})$ is a countably additive map $W : \mathcal{C} \to \mathcal{L}(\mathcal{H})$ which takes each subset $X \in \mathcal{C}$ to a positive operator $W(X) \in \mathcal{L}(\mathcal{H})$ and which is normalized by $W(\Omega) = 1$. In the present paper we focus on POVMs possessing a bounded nowhere vanishing density with respect to some probability measure $\alpha$ on $(\Omega, \mathcal{C})$, that is having the form

$$dW(s) = n F(s) d\alpha(s), \tag{1}$$

where $n = \dim_{\mathbb{C}} \mathcal{H}$ and $F : \Omega \to \mathcal{S}(\mathcal{H})$ is a measurable function. For instance any POVM $W = \{W_1, \ldots, W_N\}$ on a finite set $\Omega = \{1, \ldots, N\}$ is of the form (1) with $\alpha(i) = n^{-1} \text{trace}(W_i)$ and $F(i) = W_i/\text{trace}(W_i))$. Such POVMs model quantum measurements with a finite number of device readings, see Section 4 below for further discussion. Another example of POVMs satisfying condition (1) is provided by the Berezin-Toeplitz quantization of closed Kähler manifolds, see Section 6.

3 Introducing spectral gap

Let us introduce the main character of our story, the spectral gap of a POVM of the form (1). Denote by $L_2(\Omega, \alpha)$ the $L_2$-space of real valued functions on $\Omega$. The scalar product is denoted by $\langle \phi, \psi \rangle := \int_{\Omega} \phi \psi d\alpha$. Define a map $T : L_2(\Omega, \alpha) \to \mathcal{L}(\mathcal{H})$ by

$$T(\phi) = \int_{\Omega} \phi dW = n \int_{\Omega} \phi(s) F(s) d\alpha(s).$$

The dual map $T^* : \mathcal{L}(\mathcal{H}) \to L_2(\Omega, \alpha)$ is given by $T^*(A)(s) = n((F(s), A))$.  


Put
\[ \mathcal{E} = \frac{1}{n} T T^* : \mathcal{L}(\mathcal{H}) \to \mathcal{L}(\mathcal{H}), \]
\[ \mathcal{E}(A) = n \int_{\Omega} ((F(s), A)) F(s) d\alpha(s). \quad (2) \]
Observe that \( \mathcal{E} \) is a unital trace-preserving completely positive map. In the terminology of [12, Example 5.4], this is an example of an entanglement-breaking quantum channel.

Furthermore, set
\[ \mathcal{B} = \frac{1}{n} T^* T : L_2(\Omega, \alpha) \to L_2(\Omega, \alpha), \]
\[ \mathcal{B}(\phi)(t) = n \int_{\Omega} \phi(s)((F(s), F(t))) d\alpha(s). \quad (3) \]
Observe that \( \mathcal{B} \) is a Markov operator [2], i.e., it is bounded in \( L_1(\Omega, \alpha) \), normalized by \( \mathcal{B}(1) = 1 \), and preserves positivity: \( \mathcal{B}(\phi) \geq 0 \) for \( \phi \geq 0 \). The Markov kernel of \( \mathcal{B} \) considered as a map sending points of \( \Omega \) to Borel probability measures on \( \Omega \), is given by
\[ t \mapsto n((F(s), F(t))) d\alpha(s). \quad (4) \]
Furthermore, the image of \( \mathcal{B} \) is finite-dimensional as \( \mathcal{B} \) factors through \( \mathcal{L}(\mathcal{H}) \) and its spectrum belongs to \([0, 1]\). Note that 1 belongs to the spectrum.

Note now that positive eigenvalues of \( \mathcal{E} \) and \( \mathcal{B} \) coincide. Indeed, \( T^* \) maps isomorphically an eigenspace corresponding to a positive eigenvalue of \( \mathcal{E} \) to the eigenspace of \( \mathcal{B} \) corresponding to the same eigenvalue. Denote \( 1 = \gamma_1 \geq \gamma_2 \geq \ldots \) the positive part of the spectrum of \( \mathcal{E} \) and \( \mathcal{B} \). The number \( \gamma(W) := 1 - \gamma_2 \) is called the spectral gap of the POVM \( W \).

In the next sections we discuss several interpretations and some properties of the spectral gap.

### 4 Minimal noise of quantum measurement

Given an observable \( A \in \mathcal{L}(\mathcal{H}) \), write \( A = \sum \lambda_i P_i \) for its spectral decomposition, where \( P_i \)'s are pair-wise distinct orthogonal projectors. According to the statistical postulate of quantum mechanics, in a state \( \rho \) the
observable $A$ attains value $\lambda_i$ with probability $(P_i, \rho)$. It follows that the expectation of $A$ in $\rho$ equals $\mathbb{E}(A, \rho) = ((A, \rho))$ and the variance is given by $\text{Var}(A, \rho) = ((A^2, \rho)) - \mathbb{E}(A, \rho)^2$.

In quantum measurement theory [3, 4], a POVM $W$ represents a measuring device coupled with the system, while $\Omega$ is interpreted as the space of device readings. When the system is in a state $\rho \in \mathcal{S}(\mathcal{H})$, the probability of finding the device in a subset $X \in \mathcal{C}$ equals

$$\mu_{\rho}(X) := ((W(X), \rho)) .$$

An experimentalist performs a measurement whose outcome, at every state $\rho$, is distributed in $\Omega$ according to the measure $\mu_{\rho}$. Given a function $\phi \in L_2(\Omega, \alpha)$ (experimentalist’s choice), this procedure yields an unbiased approximate measurement of the quantum observable $A := T(\phi)$. The expectation of $A$ in every state $\rho$ equals $((A, \rho))$ and thus coincides with the one of the measurement procedure given by $\int_{\Omega} \phi d\mu_{\rho}$ (hence unbiased), in spite of the fact that actual probability distributions determined by the observable $A$ (see above) and the random variable $(\phi, \mu_{\rho})$ could be quite different (hence approximate). In particular, in general, the variance increases under an unbiased approximate measurement:

$$\text{Var}(\phi, \mu_{\rho}) = \text{Var}(A, \rho) + ((\Delta_W(\phi), \rho)) ,$$

where

$$\Delta_W(\phi) := T(\phi^2) - T(\phi)^2$$

is the noise operator. This operator, which is known to be positive, measures the increment of the variance. We wish to explore the relative magnitude of this increment for the “maximally mixed” state $\theta_0 = \frac{1}{n} \mathbb{I}$. To this end introduce the minimal noise of the POVM $W$ as

$$\mathcal{N}_{\text{min}}(W) := \inf_{\phi} \frac{((\Delta_W(\phi), \theta_0))}{\text{Var}(\phi, \mu_{\theta_0})} ,$$

where the infimum is taken over all non-constant functions $\phi \in L_2(\Omega, \alpha)$. It turns out that the minimal noise coincides with the spectral gap.

**Theorem 4.1.** $\mathcal{N}_{\text{min}}(W) = \gamma(W)$. 
Proof. Since \( \text{trace}(T(\phi^2)) = n(\phi, \phi) \), we readily get that
\[
((\Delta_W(\phi), \theta_0)) = ((\mathbb{I} - \mathcal{B})\phi, \phi) ,
\]
where \( \mathcal{B} = n^{-1}T^*T \) is the Markov operator given by (3), while
\[
\text{Var}(\phi, \mu_{\theta_0}) = (\phi, \phi) - (\phi, 1) ^2 .
\]
Therefore, by the variational principle,
\[
\mathcal{N}_{\text{min}}(W) = \gamma(W) .
\]

5 Repeated quantum measurement

Repeated measurement with the POVM \( dW(s) = nF(s)d\alpha(s) \) is modeled by two sequences of random variables defined on a probability space \((\Theta, \mathbb{P})\): the sequence of device readings \( s_i : \Theta \to \Omega, i = 1, 2, \ldots \) and the sequence of reduced quantum states \( \rho_i : \Theta \to \mathcal{S}(\mathcal{H}), i = 0, 1, \ldots \). Here \( \rho_0 \) is the initial quantum state and \( i \) stands for the discrete time. At each time moment \( i \in \mathbb{N} \), the experimentalist performs a measurement with the outcome \( s_{i+1} \), and at each step the system jumps from the state \( \rho_i \) to the state \( \rho_{i+1} \) due to the state reduction (a.k.a. wave function collapse). The above sequences are related by the deterministic recursive Lüders rule, (3),
\[
\rho_{i+1} = \frac{F(s_{i+1})^{1/2} \rho_i F(s_{i+1})^{1/2}}{((F(s_{i+1}), \rho_i))} , \quad i \geq 0 , \quad (6)
\]
as well as by the following axiom for the conditional probability \( \mathbb{P}(s_{i+1} | \rho_i) \): for a subset \( X \subset \Omega \) and a state \( \rho \in \mathcal{S}(\mathcal{H}) \)
\[
\mathbb{P}(s_{i+1} \in X | \rho_i = \rho) = \mu_{\rho}(X) := n \int _X ((F(s), \rho))d\alpha(s) . \quad (7)
\]
Recall that, by definition of the conditional probability, this means that for every measurable subset \( Y \subset \mathcal{S}(\mathcal{H}) \) we have
\[
\mathbb{P}(s_{i+1} \in X, \rho_i \in Y) = \int _Y \mu_{\rho}(X)d\rho_{i+1}\mathbb{P}(\rho) ,
\]

\[\square\]

\[\square\]
and
\[ d(s_{i+1} \times \rho_i) \mathbb{P}(s, \rho) = n((F(s), \rho))d\alpha(s)d\rho_i \mathbb{P}(\rho). \tag{8} \]

We shall discuss repeated quantum measurements for the following class of POVMs. A POVM \( W \) given by formula (1) is called pure if

(i) for every \( s \in \Omega \) the state \( F(s) \) is pure, i.e. a rank one projector;

(ii) the map \( F : \Omega \to \mathcal{S}(\mathcal{H}) \) is one to one.

Pure POVMs, under various names, arise in several areas of mathematics including the Berezin-Toeplitz quantization (see the next section), convex geometry (see \[11\] for the notion of an isotropic measure and \[11\] for the resolution of identity associated to John and Löwner ellipsoids), signal processing (see \[8\] for a link between tight frames and quantum measurements) and Hamiltonian group actions \[10\]. When \( \Omega \) is a finite set, a pure POVM with a given measure \( \alpha \) exists if and only if the measure \( \alpha(\{s\}) \) of each point \( s \in \Omega \) is \( \leq 1/n \), see \[10\] for a detailed account on the structure of the moduli spaces of pure POVMs on finite sets up to unitary conjugations.

Denote by \( \mathcal{P} \subset \mathcal{S}(\mathcal{H}) \) the space of all pure states. In view of (ii) we, without loss of generality, identify the set of device readings \( \Omega \) with a subset of \( \mathcal{P} \) so that \( F \) becomes the natural inclusion:

\[ F(s) = s \quad \forall s \in \Omega \subset \mathcal{P}. \tag{9} \]

Furthermore, recall that for any rank one projector \( P \) and any \( \rho \in \mathcal{S}(\mathcal{H}) \)

\[ P^{1/2} \rho P^{1/2} = P \rho P = ((P, \rho)) P. \]

Thus for pure POVMs the Lüders rule (5) reads \( \rho_{i+1} = s_{i+1} \) for all \( i \geq 1 \), and the conditional probability \( \mathbb{P}(\rho_{i+1} | \rho_i = \rho) \) is given by the measure \( \mu_\rho \). It follows that the sequence \( \rho_i \) describes a Markov chain with the state space \( \Omega \subset \mathcal{P} \) with the Markov kernel (4). Thus the corresponding Markov operator equals \( \mathcal{B} \), and the spectral gap of the Markov chain coincides with the spectral gap \( \gamma(W) \) of the POVM \( W \) defined above.

Next, we claim that the expectations of \( \rho_{i+1} \) and \( \rho_i \) are related by the quantum channel \( \mathcal{E} \) defined in (2) above, where for pure POVMs (given our convention (9))

\[ \mathcal{E}(r) = n \int_{\Omega} ((t, r)) t d\alpha(t). \]
Theorem 5.1. For all $i$, $E(\rho_{i+1}) = E(E(\rho_i))$.

Proof. Indeed, putting without loss of generality $i = 0$, we get that $\rho_0$ and $\rho_1$ are $\Omega$-valued random variables defined on a probability space $(\Theta, \mathbb{P})$ and

$$E(\rho_1) = \int_{\Omega \times \Omega} r_1 d(\rho_1 \times \rho_0) \mathbb{P}(r_1, r_0)$$

$$= n \int_{\Omega \times \Omega} r_1((r_1, r_0)) d\alpha(r_1) d\rho_0 \mathbb{P}(r_0)$$

$$= n \int_{\Omega} r_1((r_1, \int_{\Omega} r_0 d\rho_0 \mathbb{P}(r_0))) d\alpha(r_1) = E(E(\rho_0)).$$

Assume now that the spectral gap $\gamma(W)$ of $W$ is strictly positive. It follows that for every initial state with the expectation $\rho$ its images $E_k(\rho)$, $k \to \infty$ converge to the maximally mixed quantum state $\frac{1}{n} \mathbb{I}$ at the exponential rate $\sim (1 - \gamma(W))^k$. In other words, for pure POVMs the spectral gap controls the convergence rate to the maximally mixed state under repeated quantum measurements.

The Markov chain interpretation of pure POVMs of the form $dW(s) = nF(s) d\alpha(s)$ gives rise to the following estimate for the spectral gap. For an operator $B \in \mathcal{L}(\mathcal{H})$, $0 \leq B \leq \mathbb{I}$, put

$$\kappa(B) = \frac{\text{trace}(B^2)}{\text{trace}(B)}. \quad (10)$$

This quantity measures a deviation of $B$ from being an orthogonal projector. Indeed, $\kappa(B)$ is always $\leq 1$, and $\kappa(B) = 1$ if and only if $B$ is an orthogonal projector. Introduce the quantity

$$\Upsilon(W) = 1 - \sup_{Y \subset \Omega, \alpha(Y) \leq 1/2} \kappa(W(Y)). \quad (11)$$

With this notation, one has the following inequality:

Theorem 5.2. $\Upsilon(W)^2/2 \leq \gamma(W) \leq 2\Upsilon(W)$. 

8
Proof. Recall [17] that the bottleneck ratio for the Markov operator with the kernel \( t \mapsto n((F(s), F(t)))d\alpha(s) \) (see (4) above) is given by

\[
\inf_{Y \subset \Omega, \alpha(Y) \leq 1/2} \frac{\int_Y \int_{\Omega \setminus Y} n((F(s), F(t)))d\alpha(s)d\alpha(t)}{\alpha(Y)}.
\]

The expression we are minimizing can be rewritten as

\[
\frac{n^{-1}((W(Y), 1 - W(Y)))}{n^{-1}\text{trace}(W(Y))} = 1 - \kappa(W(Y)),
\]

which yields that the bottleneck ratio equals \( \Upsilon(W) \). Thus the proposition is an immediate consequence of the bottleneck ratio estimate for the spectral gap of Markov chains, see [17, Theorem 13.14] and [16]. \( \square \)

6 Spectral gap for quantization

Pure POVMs naturally appear in the context of the Berezin-Toeplitz quantization of closed Kähler manifolds which are quantizable in the following sense: the cohomology class \([\omega]/(2\pi)\) is integral, where \( \omega \) is the Kähler symplectic form on \( M \). Let us briefly recall the construction of this quantization (see [20, 14] for preliminaries). Pick a holomorphic Hermitian line bundle \( L \) over \( M \) whose Chern connection has curvature \( i\omega \). Define the Planck constant \( \hbar \) by \( 1/k \), where \( k \in \mathbb{N} \) is large enough. Write \( L^k \) for the \( k \)-th tensor power of \( L \). Consider the space \( \mathcal{H}_\hbar \) of all global holomorphic sections of \( L^k \). In this setting, one defines a family of pure \( \mathcal{L}(\mathcal{H}_\hbar) \)-valued POVMs \( dW_\hbar = n_\hbar F_\hbar(s)d\alpha_\hbar(s) \), where \( n_\hbar = \dim_{\mathbb{C}} \mathcal{H}_\hbar \), the map \( F_\hbar : M \to \mathcal{P} \) sends a point \( s \in M \) to the coherent state projector at \( z \) (from the viewpoint of algebraic geometry, the map \( F_\hbar \) comes from the Kodaira embedding theorem), and the measure \( \alpha_\hbar \) is given by \((R_\hbar/n_\hbar)d\text{Vol}\), where the density \( R_\hbar : M \to \mathbb{R} \) is the Rawnsley function.

Note that in the context of the Berezin-Toeplitz quantization, the operator \( \mathcal{B}_\hbar := \frac{1}{n_\hbar} T_\hbar T_\hbar \) given by formula (3) above is known as the Berezin transform acting on functions on \( M \). Our next result contains an estimate for the spectral gap of the Berezin-Toeplitz POVM \( W_\hbar \), or, equivalently, for the first eigenvalue of the operator \( 1 - \mathcal{B}_\hbar \). Denote by \( \lambda_1(M) \) the first eigenvalue of the Laplace-Beltrami operator \( \Delta \) of the Kähler Riemannian metric on \( M \). (Convention: We define \( \Delta f \) as \(-\text{div}\nabla f\), mind the minus sign, so \( \Delta \) is positive.)
Theorem 6.1. There exists $c > 0$ depending only on the Kähler structure on $M$ such that

$$ch + O(h^2) \leq \gamma(W_h) \leq \lambda_1(M)h + O(h^2).$$

(12)

The upper bound immediately follows from the asymptotic expansion of the Berezin transform (see [20])

$$B_h(f) = f - h\Delta f + O(h^2)$$

for every smooth function $f$ on $M$. (Warning: here the remainder $O(h^2)$ depends on $f$.) Thus choosing $f$ to be the $L^2$-normalized first eigenfunction of $\Delta$, we see that

$$\gamma(W_h) \leq (1 - B_h)f, f)_{L^2} \leq h\lambda_1(M) + O(h^2).$$

The proof of the lower bound involves a result from a paper [15] by Lebeau and Michel. It is given in Section 9 below.

Let us mention that for the standard quantization of the complex projective space $\mathbb{C}P^n$ the spectral gap can be calculated by means of representation theory [21]. It coincides with $\lambda_1(\mathbb{C}P^n)h + O(h^2)$, where $\lambda_1(\mathbb{C}P^n) = n + 1$. It would be interesting to understand whether the spectral gap of the Berezin transform equals $\hbar\lambda_1(M) + O(h^2)$ for any quantizable closed Kähler manifold.

7 A geometric interpretation

Let $\mathcal{V}$ be a finite-dimensional affine real vector space whose associated linear space is equipped with a scalar product. Write dist for the corresponding distance on $\mathcal{V}$. Given a compactly supported probability measure $\sigma$ on $\mathcal{V}$, introduce the following objects:

- the center of mass $C(\sigma) = \int_{\mathcal{V}} v d\sigma(v)$;
- the mean squared distance from the origin,

$$I(\sigma) = \int_{\mathcal{V}} \text{dist}(C, v)^2 d\sigma(v);$$
• the mean squared distance to the best fitting line

\[ J(\sigma) = \inf_\ell \int_\mathcal{V} \text{dist}(v, \ell)^2 d\sigma(v), \]

where the infimum is taken over all affine lines \( \ell \subset \mathcal{V} \).

The infimum in the definition of \( J \) is attained at the (not necessarily unique) best fitting line which is known to pass through the center of mass \( C \) (Pearson, 1901; see [9, p.188] for a historical account).

Assume now that we are given an \( \mathcal{L}(\mathcal{H}) \)-valued POVM on \( \Omega \) satisfying equation (1), i.e., of the form \( dW = n F \, d\alpha \) for some \( F : \Omega \to \mathcal{S}(\mathcal{H}) \). We write \( \mathcal{V} \subset \mathcal{L}(\mathcal{H}) \) for the affine subspace consisting of all trace 1 operators. Recall that \( \mathcal{L}(\mathcal{H}) \) is equipped with the scalar product \( \langle (A, B) \rangle = \text{trace}(AB) \) and \( \mathcal{S}(\mathcal{H}) \subset \mathcal{V} \). Consider a measure \( \sigma_W := F_* \alpha \) on \( \mathcal{S}(\mathcal{H}) \subset \mathcal{V} \). Observe that the center of mass \( C(\sigma_W) \) coincides with the maximally mixed state \( \frac{I}{n} \).

**Theorem 7.1.** The spectral gap \( \gamma(W) \) depends only on the push-forward measure \( \sigma_W \) on \( \mathcal{S}(\mathcal{H}) \):

\[ \gamma(W) = 1 - n(I(\sigma_W) - J(\sigma_W)). \]

**Proof.** Let \( \ell \subset \mathcal{V} \) be any line passing through the center of mass \( \frac{1}{n} I \) generated by a trace zero unit vector \( A \in \mathcal{L}(\mathcal{H}) \). For a point \( B \in \mathcal{V} \) we have

\[ \text{dist}(B, \ell)^2 = \langle (B - \frac{1}{n} I, B - \frac{1}{n} I) \rangle - \langle (B - \frac{1}{n} I, A) \rangle^2. \]

Integrating over \( \sigma_W \) and taking infimum over \( \ell \) we get that

\[ J(\sigma_W) = I(\sigma_W) - K, \tag{13} \]

with

\[ K = \sup_{\text{trace}(A) = 0, \text{trace}(A^2) = 1} \int_{\mathcal{V}} \langle (B, A) \rangle^2 dF_* \alpha(B). \]

The latter integral can be rewritten as

\[ \int_{\Omega} \langle (F(s), A) \rangle^2 d\alpha(s) = n^{-1}(\langle E(A), A \rangle), \]

so by definition \( K = n^{-1}(1 - \gamma(W)) \). Substituting this into (13), we deduce the theorem. \( \square \)

\(^a\)The problem of finding \( J \) and the corresponding minimizer \( \ell \) appears in the literature under several different names including “total least squares” and “orthogonal regression.”
Furthermore, the gap $\gamma(W)$, as a function of the measure $\sigma_W$, is Lipschitz with respect to the $L_2$-Wasserstein distance $\delta$ on the space of Borel probability measures on $S(\mathcal{H})$. This distance is defined as follows: for compactly supported Borel probability measures $\sigma_1, \sigma_2$ on a metric space $(X,d)$

$$\delta(\sigma_1, \sigma_2) := \inf_{\nu} \left( \int_{X \times X} \dist(x_1, x_2)^2 \, d\nu(x_1, x_2) \right)^{1/2},$$

where the infimum is taken over all Borel probability measures $\nu$ on $X \times X$ with marginals $\sigma_1$ and $\sigma_2$.

**Theorem 7.2.** Let $\sigma_V$ and $\sigma_W$ be measures on $S(\mathcal{H})$ associated to POVMs $V$ and $W$ respectively. Then

$$|\gamma(V) - \gamma(W)| \leq 12n\delta(\sigma_V, \sigma_W).$$

Note that this result enables us to compare spectral gaps of POVMs defined on different sets (but having values in the same Hilbert space).

Theorem 7.2 immediately follows from the fact that $C(\sigma), I(\sigma)$ and $J(\sigma)$ are Lipschitz in $\sigma$ with respect to the Wasserstein distance. For readers’ convenience, we present a proof in Section 10 below.

### 8 A link to the diffusion distance

Here we adopt the geometric viewpoint developed in the previous section. We start with a POVM $dW = nF\,d\alpha$ on $\Omega$ and view it as a measure $F_*\alpha$ on the space of states $S(\mathcal{H})$. Let $A \in \mathcal{L}(\mathcal{H})$ be the trace zero unit vector generating the best fitting line corresponding to $W$. In view of Theorem 7.1

$$\gamma_1 = 1 - \gamma(W) = n(I - J),$$

with $I = I(\sigma_W)$ and $J = J(\sigma_W)$.

**Theorem 8.1.** The function

$$\psi_1 : \Omega \to \mathbb{R}, \ s \mapsto \frac{((F(s), A))}{\sqrt{I - J}}$$

is an eigenfunction of the operator $\mathcal{B}$ with the eigenvalue $\gamma_1$. Furthermore, $||\psi_1||_{L_2(\Omega, \alpha)} = 1$.  

12
In other words, up to a multiplicative constant, the first eigenfunction is the projection to the best fitting line.

**Proof.** As we have seen in the proof of Theorem 7.1 above, the operator \( A \) generating the best fitting line is an eigenvector of the quantum channel \( \mathcal{E} \): 
\[
\mathcal{E}A = \gamma_1 A.
\]
Since \( \mathcal{E} = n^{-1}TT^* \) and \( B = n^{-1}T^*T \), we have \( B(T^*A) = \gamma_1 T^*A \) and \( (T^*A, T^*A) = n\gamma_1 \). Furthermore, \( T^*A(s) = n((F(s), A)) \) and \( n\gamma_1 = n^2(I - J) \). Choosing \( \psi_1 = T^*A/||T^*A||_{L^2(\Omega,\alpha)} \), we get (14). \( \square \)

The information on \( \gamma_1 = 1 - \gamma(W) \) and \( \psi_1 \) sheds light on the diffusion distance on \( \Omega \) associated to the Markov operator \( B \) (see [7]). This metric, which plays an important role in geometric analysis of data sets, depends on a positive parameter \( \tau \) (the time). It is defined through a collection of \( L^2 \)-normalized eigenfunctions \( \{\psi_k\} \) corresponding to the positive eigenvalues \( \gamma_1 \geq \gamma_2 \geq \ldots \) of \( B \) as follows:
\[
D_\tau(s,t) = \left( \sum_{k \geq 1} \gamma_k^{2\tau}(\psi_k(s) - \psi_k(t))^2 \right)^{1/2} \quad \forall s,t \in \Omega. 
\]
Assume now that \( \gamma_2 < \gamma_1 \). In this case the asymptotic behavior of \( D_\tau(s,t) \) as \( \tau \to \infty \) is given by
\[
D_\tau(s,t) = \gamma_1 \frac{|((F(s) - F(t), A))|}{(I - J)^{1/2}} \cdot (1 + o(1)) \quad \text{if} \quad ((F(s), A)) \neq ((F(t), A)),
\]
and \( D_\tau(s,t) = O(\gamma_2^{\tau}) \) otherwise. The difference in these asymptotic formulas highlights the fact that the metric space \( (\Omega, D_\tau) \) behaves differently on scales above and below \( \sim \gamma_1^{\tau} \).

**9 Estimating the gap for the Berezin transform**

In this section we prove the lower bound in Theorem 6.1. We fix a closed manifold \( M \) equipped with a smooth measure \( d\mu \) and focus on Markov chains on \( M \) given by its symmetric non-negative kernel \( K(x,y) = K(y,x) \) and the stationary measure \( r d\mu \), where \( r \) is a positive density. In particular, we have
\[
\int_M r d\mu = 1, \quad \int_M K(x,y) r(y) d\mu(y) = 1.
\]
The spectral gap $\gamma(K, r)$ admits a variational characterization

$$
\gamma(K, r) = \inf_{f} \frac{\mathcal{A}(f)}{\text{Var}(f, r d\mu)},
$$

where

$$
\mathcal{A}(f) := \frac{1}{2} \int_{M \times M} K(x, y)(f(x) - f(y))^2 r(x) r(y) d\mu(x) d\mu(y).
$$

We shall use the following comparison result for the spectral gaps of Markov chains, which is an immediate consequence of [17, Lemma 13.22]: Given two Markov chains $(K, r)$ and $(Q, \rho)$ on $(M, \mu)$ such that for some positive constants $c_1, c_2, c_3 > 0$

$$
c_2 \rho \leq r \leq c_1 \rho
$$

and

$$
Q \leq c_3 K,
$$

we have

$$
\gamma(K, r) \geq c_2^2 c_1^{-1} c_3^{-1} \gamma(Q, \rho).
$$

The Markov chain associated to the Berezin-Toeplitz quantization is given by $K_h(x, y) = n_h((F_h(x), F_h(y)))$, where $F_h(x)$ is the coherent state projector at $x \in M$, and by

$$
r_h(x) = n_h^{-1} \text{Vol}(M) R_h(x),
$$

where $R_h$ is the Rawnsley function. Here the measure $\mu$ is the normalized Riemannian volume on $M$,

$$
\mu = \frac{\text{Vol}}{\text{Vol}(M)}.
$$

In the notation of Theorem 6.1, $\gamma(W_h) = \gamma(K_h, r_h)$.

Denote by $d$ the Riemannian distance on $M$. The following properties of $K_h$ and $r_h$ are crucial for our purposes: first (see formula (50) in [6]), there exists positive constants $a, b, c > 0$ such that

$$
K(x, y) \geq n_h a \exp\{-d(x, y)^2/(bh)\} \quad \forall x, y \in M \quad \text{with} \quad d(x, y) \leq c \sqrt{h},
$$

and second (see e.g. [20, 6])

$$
r_h = 1 + O(h).
$$
In order to get the desired lower bound for the spectral gap $\gamma(K_\hbar, r_\hbar)$, we shall compare it with the gap of another Markov chain on $(M, \mu)$, called the semiclassical random walk. The latter gap was calculated in [15] for any closed Riemannian manifold. Write $v(x, t)$ for the volume of the Riemannian ball of radius $t$ centered at $x \in M$, and put

$$v_t = \int_M v(x, t)d\mu(x).$$

The semiclassical random walk depends on a parameter $t > 0$ and is given by the kernel

$$Q_t(x, y) = v_t \cdot \frac{\chi_{[0,t]}(d(x, y))}{v(x, t)v(y, t)},$$

where $\chi$ stands for the indicator function, and the stationary measure $\rho_t d\mu$ with

$$\rho_t(x) = \frac{v(x, t)}{v_t}.$$

Put $t = c\sqrt{\hbar}$. We shall write $c_1, c_2, \ldots$ for positive constants independent of $\hbar$ (and hence of $t$). By the inequality

$$a \exp\{-s^2/(b\hbar)\} \geq a \exp\{-c^2/b\} \chi_{[0,t]}(s) \quad \forall s \geq 0,$$

we get from (18) that $Q_t \leq c_3 K_\hbar$. In view of (19), $c_3 \rho_t \leq r_\hbar \leq c_1 \rho_\hbar$. Thus, the comparison theorem (17) yields $\gamma(K_\hbar, r_\hbar) \geq c_4 \gamma(Q_t, \rho_t)$. By the central result of Lebeau-Michel [15], we have $\gamma(Q_t, \rho_t) \geq c_5 t^2 = c_6 \hbar$. It follows that

$$\gamma(K_\hbar, r_\hbar) \geq c_7 \hbar. \quad (20)$$

This proves the desired lower bound on the spectral gap and completes the proof of Theorem 6.1.

10 Proving robustness

We use the notation of Section 7. Let $Q \subset V$ be a convex compact subset of diameter $D$, and $\nu$ be a measure on $Q \times Q$ with the marginals $\alpha$ and $\beta$. Put

$$\delta_0 := \left(\int_{Q \times Q} \text{dist}(v, w)^2d\nu(v, w)\right)^{1/2}.$$

We shall prove the following inequalities:
(i) $\text{dist}(C(\alpha), C(\beta)) \leq \delta_0$;

(ii) $|I(\alpha) - I(\beta)| \leq 4D\delta_0$;

(iii) $|J(\alpha) - J(\beta)| \leq 2D\delta_0$;

Together with Theorem 7.1 and the fact that the diameter of $\mathcal{S}(\mathcal{H})$ is $\leq 2$, this readily yields Theorem 7.2.

Observe that

$$|C(\alpha) - C(\beta)| = \left| \int_{Q \times Q} (v - w) d\nu(v, w) \right| \leq \delta_0 ,$$

where the last inequality follows from the Cauchy-Schwarz. This yields (i).

By (i), for all $v, w \in Q$

$$|\text{dist}(C(\alpha), v)^2 - \text{dist}(C(\beta), w)^2| \leq 2D(\text{dist}(C(\alpha), C(\beta)) + \text{dist}(v, w))$$

$$\leq 2D(\delta_0 + \text{dist}(v, w)) .$$

Integrating against $\nu$ and applying Cauchy-Schwarz, we get (ii).

Finally, let $\ell_{\beta}$ be the best fitting line for $\beta$ (in case it is not unique, choose any such line). Then

$$J(\alpha) \leq \int_{Q \times Q} \text{dist}(v, \ell_{\beta})^2 d\alpha(v) = \int_{Q \times Q} \text{dist}(v, \ell_{\beta})^2 d\nu(v, w) . \quad (21)$$

Note that since $Q$ is convex, the center of mass $C(\beta)$ lies in $Q$. Since $\ell_{\beta}$

passes through $C(\beta)$ we get that $\text{dist}(v, \ell_{\beta}) \leq D$ provided $v \in Q$. Thus for all $v, w \in Q$

$$\text{dist}(v, \ell_{\beta})^2 \leq \text{dist}(w, \ell_{\beta})^2 + 2D \cdot \text{dist}(v, w) .$$

Substituting this into the right hand side of (21) and applying Cauchy-

Schwarz we get that $J(\alpha) \leq J(\beta) + 2D\delta_0$. Similarly, $J(\beta) \leq J(\alpha) + 2D\delta_0$, which yields (iii). The proof of (i)-(iii), and hence of Theorem 7.1, is complete. \qed
11 Concluding remarks

In quantum measurement theory, there are two concepts of quantum noise: the increment of variance for unbiased approximate measurements as formalized by the noise operator, see Section 4, and a non-unitary evolution of a quantum system described by a quantum channel (a.k.a. a quantum operation, see, e.g. [18 Chapter 8]). Such a non-unitary evolution can be caused, for instance, by the quantum state reduction in the process of repeated quantum measurements, see Section 5. Interestingly enough, for pure POVMs of the form \( dW(s) = nF(s)\,d\alpha(s) \), i.e. when the density \( F : \Omega \rightarrow \mathcal{P} \subset \mathcal{S}(\mathcal{H}) \) is a one to one map taking values in rank one projectors, the spectral gap \( \gamma(W) \) brings together these two seemingly remote concepts: it measures the minimal magnitude of noise production in the context of the noise operator, and it equals the spectral gap of the Markov chain modeling repeated quantum measurements.

The Wasserstein distance can be used for estimating the quality of quantum measurements, see [5] and references therein. Here the Wasserstein distance is defined with respect to a metric on the space of device readings \( \Omega \). We deal, in the context of POVMs of the form \( dW(s) = nF(s)\,d\alpha(s) \), with another version of the Wasserstein distance which goes back to [19] and does not require a choice of a metric on \( \Omega \). Rather one looks at the push-forward measures \( F,\alpha \) on the space of quantum states \( \mathcal{S}(\mathcal{H}) \) and considers the Wasserstein distance on such measures with respect to the canonical Hilbert-Schmidt metric on \( \mathcal{S}(\mathcal{H}) \). It would be interesting to explore applicability of this approach to uncertainty relations appearing in the theory of quantum measurements (cf. [5]).

A logical extension of the idea of viewing POVMs as measures on the space of quantum states \( \mathcal{S} \) is to treat them as data clouds in \( \mathcal{S} \). This opens up a prospect of using various tools of geometric data analysis for studying POVMs. Our results on the diffusion distance associated to a POVM can be considered as a step in this direction.

Acknowledgment. We are grateful to D. Aharonov, S. Artstein-Avidan, P. Busch, L. Charles, G. Kalai, G. Kindler, M. Krivelevich, Y. Ostrover, A. Veselov and S. Weinberger for useful discussions on various aspects of this work. We thank O. Oreshkov for bringing our attention to the paper [19].

\[ \text{In [19] the authors consider the } L_1\text{-version of this distance, and call it the Kantorovich distance.} \]
and illuminating comments, as well I. Polterovich for very helpful remarks on the manuscript.

References

[1] Aubrun, G., and Szarek, S.J., *Alice and Bob meet Banach*, Book available at [http://math.univ-lyon1.fr/ aubrun/ABMB/index.html](http://math.univ-lyon1.fr/ aubrun/ABMB/index.html).

[2] Bakry, D., Gentil, I. and Ledoux, M., *Analysis and geometry of Markov diffusion operators*. Springer, 2013.

[3] Busch, P., Grabowski, M., and Lahti, P.J., *Operational Quantum Physics*. Lecture Notes in Physics, New Series: Monographs, 31, Springer-Verlag, Berlin, 1995.

[4] Busch, P., Lahti, P.J., Pellonpää, J. P., and Ylinen, K., *Quantum measurement*. Springer, 2016.

[5] Busch, P., Lahti, P.J., and Werner, R., *Colloquium: Quantum root-mean-square error and measurement uncertainty relations.*, Reviews of Modern Physics 86 (2014): 1261.

[6] Charles, L., and Polterovich, L., *Sharp correspondence principle and quantum measurements*, Preprint, arXiv:1510.02450, to appear in St. Petersburg Math. Journal, volume dedicated to Yu.D. Burago.

[7] Coifman, R.R., and Lafon, S., *Diffusion maps*, Appl. Comput. Harmon. Anal. 21 (2006),5-30.

[8] Eldar, Y.C., and Forney, G.D. Jr., *Optimal tight frames and quantum measurement*, IEEE Trans. Inform. Theory 48 (2002), 599-610.

[9] Farebrother, R.W., *Fitting Linear Relationships: A History of the Calculus of Observations 1750–1900*, Springer Science and Business Media, 1999.

[10] Flaschka, H., and Millson, J., *Bending flows for sums of rank one matrices*, (Canad. J. Math. 57 (2005), 114-158.

[11] Giannopoulos, A.A., and Milman, V.D., *Extremal problems and isotropic positions of convex bodies*, Israel J. Math. 117 (2000), 29-60.
[12] Hayashi, M., *Quantum information. An introduction*, Springer-Verlag, Berlin, 2006.

[13] Landsman, N.P., *Mathematical topics between classical and quantum mechanics*. Springer Monographs in Mathematics. Springer-Verlag, New York, 1998.

[14] Le Floch, Y., *A brief introduction to Berezin-Toeplitz operators on compact Kähler manifolds*, manuscript, 2016.

[15] Lebeau, G., and Michel, L., *Semi-classical analysis of a random walk on a manifold*, The Annals of Probability 38(2010), 277–315.

[16] Lawler, G.F., and Sokal, A.D., *Bounds on the $L^2$ spectrum for Markov chains and Markov processes: a generalization of Cheeger's inequality*, Transactions of the AMS, 309 (1988), 557–580.

[17] Levin, D. A., Peres, Y., and Wilmer, E. L. *Markov chains and mixing times*, American Mathematical Soc., 2009.

[18] Nielsen, M.A., and Chuang, I.L., *Quantum computation and Quantum information*, Cambridge University Press, 2000.

[19] Oreshkov, O., and Calsamiglia, J., *Distinguishability measures between ensembles of quantum states*, Physical Review A, 79(2009), p.032336.

[20] Schlichenmaier, M., *Berezin-Toeplitz quantization for compact Kähler manifolds. A review of results*, Adv. Math. Phys. (2010), 927280.

[21] Zhang, G., *Berezin transform on compact Hermitian symmetric spaces*, Manuscripta Mathematica 97(1998), 371–388.

School of Mathematical Sciences
Tel Aviv University
Israel