Hilbert-Schmidt distance and entanglement witnessing

Palash Pandya, Omer Sakarya, and Marcin Wieśniak

1 Institute of Mathematics, Faculty of Mathematics, Physics, and Informatics, University of Gdańsk, 80-308 Gdańsk, Poland
2 Institute of Informatics, Faculty of Mathematics, Physics, and Informatics, University of Gdańsk, 80-308 Gdańsk, Poland

Gilbert proposed an algorithm for bounding the distance between a given point and a convex set. In this article we apply the Gilbert’s algorithm to get an upper bound on the Hilbert-Schmidt distance (HSD) between a given state and the set of separable states. While HSD does not form a proper entanglement measure, it can nevertheless be useful for witnessing entanglement. We provide here a few methods based on the Gilbert’s algorithm that can reliably qualify a given state as strongly entangled or practically separable, while being computationally efficient. The method also outputs successively improved approximations to the Closest Separable State (CSS) for the given state. We demonstrate the efficacy of the method with examples.

Entanglement is by far, the most surprising feature of quantum mechanics. For pure states, this means that the state cannot be written as tensor product of pure states as subsystems. This implies that the properties of one subsystem are defined only in reference to others, and can be established by measurements constructed on remote parts of the system. Mixed entangled states are those, which cannot be written as a mixture of pure product states. This phenomenon is the ground for advantages of various quantum information processing protocols, such as quantum communication complexity reduction protocols, or quantum games, over their classical counterparts. This advantage is often related to the Bell theorem [1], in which strictly quantum nature is contrasted not with local quantum-mechanical statistics, but with local realistic theories, in which outcomes of all local measurements are preassigned.

Entanglement can be classified in many different ways. The most obvious criterion is the number of subsystems effectively involved, i.e., how many subsystems need to be entangled to create the state of our interest. This classification can be generalized in at least two different ways, i.e., depth of entanglement, which tells us how many subsystems must be entangled at least to recreate the state, and the structure of entanglement, describing the necessary connections between individual constituents. For multipartite case, we have individual classes of entangled states, such that we cannot transform between these classes with local operations and classical communication (LOCC) [2]. We can also distinguish those entangled states, which can provide statistics violating any Bell inequalities, or those, from which maximally entangled states can be distilled by means of local operations, i.e., free entangled states, as opposed to those with bound entanglement [3]. Bound entangled states can exist for any system larger than $2 \times 3$.

This also brings us to the problem of entanglement measures [3]. It is natural to quantize non-classicality. Any proper measure of entanglement must satisfy certain axioms, such as nullification for separable states, normalization for the two-qubit maximally entangled state, additivity under tensor product, or monotonicity under local operations supported by classical communication [4]. For pure states, there is a unique measure given by the entropy of squares of moduli of the Schmidt coefficients. For generic states, two measures were operationally induced, i.e., distillable entanglement and entanglement cost [5, 6]. Subsequently, other measures were proposed, but all of them suffer from the practical impossibility of calculating the value for a given state.

In such a case, a more relaxed approach is taken. Namely, it often suffices to certify that the state is entangled. This can be conveniently done with a witness operator [7], which assumes mean values from a certain range, but has eigenvalues beyond this range. By Jamiołkowski-Choi isomorphism [9], they are related to positive, but not completely positive maps. This is a universal method to certify entanglement for any non-separable state, however, unfortunately, we do not know a general form of a witness, or all possible not completely positive maps. Likewise, quadratic entanglement criteria have been proposed, but it is also very difficult to construct a nonlinear criterion for generic states.

Here, we discuss the Gilbert algorithm for estimating and bounding from above, the Cartesian distance between a given point and a convex set with known extremal points. This is known as the Hilbert-Schmidt distance (HSD) between a given state and the set of separable states. The algorithm allows us to find an upper bound on the Cartesian distance between a given state and the set of separable states. This is done by an infinite series of corrections. Up to now, it has been successfully used to find examples falsifying local realistic models [10], as well as to truncate quantum states so that they remain in classes of SLOCC-equivalence [11]. We propose a few methods to qualify a given state as strongly entangled or practically separable by analyzing the numerical output of the procedure. To apply the Gilbert algorithm, it is only necessary to be able to reach all extremal points of the set, which is easily doable for local realistic models and separable states.

The (simplified) Gilbert algorithm is as follows:

- Parameters: dimensions of subsystems $d_1, d_2, ...,$
Input data: the state to be tested \( \rho_0 \), and any separable state \( \rho_1 \).

Output data: the closest state found \( \rho_1 \), list of values of \( D^2(\rho_0, \rho_1) = \Tr(\rho_0 - \rho_1)^2 \).

1. Increase the counter of trials \( c_t \) by 1. Draw a random pure product state \( \rho_2 \), hereafter called a trial state.

2. Run a preselection for the trial state by checking a value of a linear functional. If it fails, go back to point 1.

3. In case of successful preselection symmetrize \( \rho_1 \) with respect to all symmetries by \( \rho_0 \), which respect the separability.

4. Find the minimum of \( \Tr(\rho_0 - p\rho_1 - (1 - p)\rho_2)^2 \) with respect to \( p \).

5. If the minimum occurs for \( 0 \leq p \leq 1 \), update \( \rho_1 \leftarrow p\rho_1 + (1 - p)\rho_2 \), add the new value of \( D^2(\rho_0, \rho_1) \) to the list and increase the success \( c_s \) counter value by 1.

6. Go to step 1 until a chosen criterion HALT is met.

We will now provide remarks on the algorithm.

Our simplification with respect to the original version is that, while we take a random separable state, in the original version, we optimize the boundary point in each execution of the loop, i.e., we would vary \( \rho_2 \) to maximize \( \Tr[(\rho_2 - \rho_1)(\rho_0 - \rho_1)] \). Such a maximization is by itself a semi-definite problem, and our simplified algorithm is also able to yield precise estimations. Many decompositions into product states requires many admixtures, which requires either a large number of corrections, or symmetrizations. Additionally, this simplification also contributes to the versatility of the algorithm applications.

Ad. 1: The algorithm can actually never reach the final state, but rather improves the approximation. Rather than getting the actual Hilbert-Schmidt distance to the separable state, we get an upper bound by getting closer and closer members of this set. As a result, the Gilbert algorithm cannot be used directly to certify entanglement. Nevertheless, this information can be still useful in a few different ways. First, knowing CSS, it is straight-forward to construct an entanglement witness, which reads \( \Tr[\rho_0(\rho_0 - \rho_{CSS})] \). It may happen that a state found by the algorithm is close enough that it can be used to construct a successful entanglement criterion. In some cases, if both \( \rho_0 \) and CSS have analytical forms, it can be possible to anticipate the latter. Such an anticipation shall also be tested with the algorithm both as \( \rho_0 \) and initial \( \rho_1 \) to confirm its separability and proximity to the tested state. Finally, we can estimate \( D^2(\rho_0) \) from the decay of \( D^2(\rho_0, \rho_1) \). The examples calculated for this work show that there is a linear dependence between \( c_s \) and \( D^2(\rho_0, \rho_1) \). Namely, the sample correlation coefficient, \begin{equation}
R(x, y) = \frac{\langle x y \rangle - \langle x \rangle \langle y \rangle}{\sqrt{\langle x^2 \rangle - \langle x \rangle^2} \sqrt{\langle y^2 \rangle - \langle y \rangle^2}}
\end{equation}
between \( x \) being the series of values of \( c_s \), i.e. a sequence of consecutive positive integers, and
\begin{equation}
y_{cs} = \beta |\ln[D^2_{cs}(\rho_0, \rho_1) - a]|^b,
\end{equation}
where \( a \) and \( b \) are free parameters of maximization of \( R(x, y) \). \begin{equation}
a \approx \lim_{c_s \to \infty} D^2(\rho_0, \rho_1) \end{equation} is the estimate for \( D^2(\rho_0) \). The controversy may lie in the fact that the found values of \( b \) vary, even for individual runs of the algorithm for the same state. Still, for sufficiently large values of \( c_s \), the dependence is remarkably close to linear. At the same time, there is a strong linear dependence between \( c_s \) and \( c_s' \). Typical values in our examples were \( 6 \leq b \leq 9 \) and \( 0.35 \leq f \leq 0.51 \).

Finally, it often seems from the algorithm that the CSS commutes with \( \rho_0 \). If this conjecture is taken, then one may observe the convergence of the eigenvalues are found. Since the HS norm is invariant under any unitary, for any matrix \( A \),
\begin{align}
\Tr[UAU^\dagger] &= \Tr[UAA^\dagger U^\dagger] \\
&= \Tr[A^\dagger] \end{align}
we can easily find \( D^2(\rho_0) \) with the limit of the spectrum of \( \rho_1 \).

Since we cannot have presumptions about CSS, one shall draw states, the product of which will constitute a trial state, with the Hilbert-Schmidt distance uniform measure \[14\]. To get a \( d \)-dimensional state, one takes a list of \( 2d \) random real numbers drawn with the Normal distribution with a fixed deviation and the mean equal to 0. Consecutive pairs in the list are combined to form
a list of \( d \) complex numbers. Subsequently, we normalize the list. If the random real numbers are evenly distributed on interval \([0, 1]\), we take two of them, \( x_1 \) and \( x_2 \), and we build a normally distributed complex variable 
\[
\sqrt{-2 \ln x_2} e^{2\pi i x_1}
\]

When \( \rho_0 \) is strictly real, one may be tempted to draw real trial states. However, if one does so, misleading results are obtained. For example, for a two-qubit maximally entangled state, we get
\[
\lim_{c_x \to 0} \rho_{1,\text{real}} = \frac{1}{8} \begin{pmatrix} 3 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 3 \end{pmatrix},
\]

while if we allow complex trial states, we get the Werner state,
\[
\lim_{c_x \to 0} \rho_{1,\text{complex}} = \frac{1}{6} \begin{pmatrix} 2 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 2 \end{pmatrix}.
\]

However, there are certain states, for which taking strictly real trial states seems to suffice, which greatly reduces the complexity of the problem.

**Ad 2:** The preselection criterion is 
\[
\text{Tr}[(\rho_2 - \rho_1)(\rho_0 - \rho_1)] > 0.
\]

The geometrical interpretation is that the angle between vectors \( \rho_0 - \rho_1 \) and \( \rho_2 - \rho_1 \) is not larger than \( \frac{\pi}{2} \). This implies that the point belonging to line \( \{\rho_1, \rho_2\} \) lies towards \( \rho_2 \), i.e. \( \partial_p D^2(\rho_0, pp_1 + (1 - p)\rho_2)|_{p=0} > 0 \). Thus, an admixture of \( \rho_2 \) is certain to decrease \( D^2(\rho_0, \rho_1) \). In comparison to the original Gilbert’s algorithm, more corrections are made, but each of them may be less efficient.

**Ad 3:** If a problem possesses a symmetry, either there is a group of solutions together respecting this symmetry, or there is a unique solution possessing this symmetry. In case of the Hilbert-Schmidt distance between any given state and a convex set of separable states, we always get a unique solution. For example, consider two states \( \rho \) and \( \rho' \), which are equidistant to \( \rho_0 \), \( D^2(\rho_0, \rho) = D(\rho_0, \rho') \). Now, consider an arbitrary combination \((1 + x)\rho + (1 - x)\rho'\)/2, and we get that the second derivative with respect to \( x^2 \) is \( 2D^2(\rho, \rho') > 0 \), while, since in our case \( \rho \) and \( \rho' \) are related to each other by a unitary transformation related to the symmetry of \( \rho_0 \), the first derivative vanishes. Hence the state minimizing \( D^2 \) is \((\rho + \rho')/2\). A similar argument holds for any number of equidistant states. Since there is a unique CSS, it must possess symmetries of \( \rho_0 \) respecting separability, we also expect \( \rho_1 \) to have them. Then, with \( U \) representing a transformation associated with such a symmetry we have
\[
\text{Tr}[(U\rho_2 U^\dagger - \rho_1)(\rho_0 - \rho_1)]
\]
\[
= \text{Tr}[U(\rho_2 - \rho_1)U^\dagger(\rho_0 - \rho_1)]
\]
\[
= \text{Tr}[U(\rho_2 - \rho_1)U^\dagger U(\rho_0 - \rho_1)U^\dagger]
\]
\[
= \text{Tr}[U(\rho_2 - \rho_1)(\rho_0 - \rho_1)U^\dagger]
\]
\[
= \text{Tr}[(\rho_2 - \rho_1)(\rho_0 - \rho_1)].
\]

In the algorithm, if the state \( \rho_0 \), possesses a discrete symmetry, \( U \) of order \( k \), then the preselected \( \rho_2 \) is symmetrized according to \( \rho_2 \leftarrow \frac{1}{k} \sum_{x=1}^{k-1}[U^x \rho_2 (U^x)^\dagger] \), and likewise in the case of continuous symmetry, the summation is replaced by an integral.

Let us now present few examples of our algorithm. First, we consider the maximally entangled state of two \( d \)-dimensional system,
\[
|\phi_d \rangle = \frac{1}{\sqrt{d}} \sum_{i=0}^{d-1} |i, i \rangle.
\]

This a the special case, as we know the CSS explicitly. It is the mixture of \( |\phi_d \rangle |\phi_d \rangle \) and the white noise with respective weights \( \frac{1}{d+1} \) and \( \frac{d}{d+1} \). This give \( D^2(|\phi_d \rangle|\phi_d \rangle) = \frac{d^{-1}}{d+1} \). In Fig. 2 we present the convergence of \( D^2(\rho_0, \rho_1) \) to \( \frac{d^{-1}}{d+1} \) for \( d = 2, 3, \ldots, 9 \).

![FIG. 2. Convergence of \( D^2(\rho_0, \rho_1) \) to the analytical limit (dashed lines) as a function of \( c_x \) for \( d = 2, 3, \ldots, 9 \) (bottom to top).](image)

We made 21 runs of the algorithm for the two-ququart maximally entangled state \( (d = 4) \) with HALT condition \( c_x = 1000 \). Series of values of \( D^2(\rho_0, \rho_1) \) from each run were used to maximize the sample correlation coefficient, as described above. Each time the maximization yielded a value above 0.999. The dependence of exponent \( b \) on the limit of HSD \( a \) is shown in Fig. 3.

Another interesting case are GHZ states of \( N \) qubits,
\[
|GHZ_{N,2} \rangle = \frac{1}{\sqrt{2}} \left(|0 \rangle^N + |1 \rangle^N \right).
\]

In this case we also know the CSS, which is a mixture,
\[ \sigma_N = x_N \sigma_{1,N} + (1 - x_N) \sigma_{2,N}, \]

where

\[
\sigma_{1,N} = \frac{1}{2} \begin{pmatrix}
1 & 0 & \ldots & 0 \\
0 & 0 & \ldots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \ldots & 1
\end{pmatrix},
\]

\[
\sigma_{2,N} = \frac{1}{2^N} \begin{pmatrix}
1 & 0 & \ldots & 0 & 1 \\
0 & 1 & \ldots & 0 & 0 \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
0 & 0 & \ldots & 1 & 0 \\
1 & 0 & \ldots & 0 & 1
\end{pmatrix},
\]

\[
x_N = \frac{(2^N - 2)^2}{4 + 4^N - 2^{N+1}},
\]

\[
D^2(|GHZ_n,2\rangle\langle GHZ_n,2|) = \frac{(2^N - 2)}{-4 + 2^{3-N} + 2^{N+1}}. \tag{9}
\]

Our algorithm converges to these states, the convergence is much faster if we take a decohered GHZ state as the initial \( \rho_1 \). For \( N = 4 \) and with initial \( \rho_1 = \sigma_N \), the algorithm was unable to find a single corrections within 6,000,000 trial states.

The last example is the two-qutrit UPB bound entangled state [15] With the Gilbert algorithm, we have reached \( D^2(\rho_0, \rho_1) = 0.002 \) with \( c_s = 8300 \) and \( c_t = 50,000,000 \) (the algorithm was particularly slow in this specific case). We then used the the sequence of values of \( D^2(\rho_0, \rho_1) \) to maximize the sample correlation coefficient, getting \( R(x, y) = 0.999996 \) with list of values of \( D^2(\rho_0, \rho_1) \) for \( \text{mod}_{100} c_s = 0, a = 0.00189996, \) and \( b = 10.1345 \). We also found a fit \( c_s \propto c_t^{4.21467} \). We then ran the algorithm again with the state found previously as \( \rho_1 \) and the HALT condition \( c_t = 50000000 \). This allowed for \( c_s = 30178, D^2(\rho_0, \rho_1) \leq 4.85 \times 10^{-6}, a = 8.82 \times 10^{-7} \) and \( f = 0.605 \). Thus, Gilbert’s algorithm can distinguish between entangled and non-entangled states with very high precision. The pace of convergence is also a sign for separability or entanglement.

We also used the results for two-qubit UPB states to construct an entanglement witness. If CSS is known to be \( \rho_{CSS} \), a necessary entanglement condition for state \( \rho \) is

\[
\log_{10}(D(\rho_{0}, \rho_{1}) + 0.00189996))^{10.1345}
\]

\[
\begin{array}{ccccccc}
\text{FIG. 3. Relation between } a \text{ and } b \text{ found in maximizing } R(x, y) \\
\text{in 21 runs of the algorithm for the maximally entangled state} \\
of two ququarts (d = 4) \text{ and HALT condition } c_t = 1000.
\end{array}
\]

\[
\begin{array}{ccccccc}
\text{FIG. 4. Dependence between the number of corrections to } \rho_1 \\
\text{and } D^2(\rho_0, \rho_1) \text{ for the two-qutrit UPB BE state [15].}
\end{array}
\]

\[
\text{Tr } \rho(\rho_0 - \rho_{CSS}) > 0. \text{ In general, we do not know } \rho_{CSS}, \text{ but only } \rho_1. \text{ The optimal entanglement witness for } \rho_0 \text{ shall take form } W_{\rho_0} = (\rho_0 - \rho_1) - \max_{\phi \in SEP \langle \phi | \rho_0 - \rho_1 | \phi \rangle}. \text{ Using the data from the previous experiments, we get } \max_{\phi \in SEP \langle \phi | \rho_0 - \rho_1 | \phi \rangle} = 0.0130011 \text{ and } \text{Tr } \rho_0(\rho_0 - \rho_1) = 0.0148792.
\]

Conclusions: We have presented the use of the Gilbert algorithm to estimate the Hilbert-Schmidt distance between a given state and the set of separable states. It works by a successive construction of an optimal separable states. While the optimal state cannot be reached by the algorithm, it still gives us a lot of useful information on entanglement of a given state. The algorithm guarantees that the state will converge to a close approximation of the CSS as the number of corrections increases. It can be straightforwardly implemented on most computational platforms, does not require large amounts of memory, and is readily formulated for any system.

The algorithm algorithm can find many interesting applications. It can be combined with other algorithms, for example, for finding bound entangled states, or violation of Bell inequalities. This will greatly improve our understanding of geometry of quantum states. The algorithm can also be used to find a separable or entangled state, which could be considered the worst-case noise scenario.

We gratefully acknowledge O. Gühne for his help and support in writing this work. This work is a part of NCN grants No. UMO-2015/19/ST-2/B/01999 and UMO-2017/26/E/ST2/01008.

[1] J. S. Bell, Physics 1, 195 (1964).

[2] A. Acín, D. Bruß, M. Lewenstein, and A. Sanpera, Phys.
Rev. Lett. 87, 040401 (2001).

[3] C. H. Bennett, H. J. Bernstein, S. Popescu and B. Schumacher, Phys. Rev. A 53, 2046 (1996); C. H. Bennett, D. P. Di Vincenzo, J. Smolin and W. K. Wootters, Phys. Rev. A 54, 3814 (1997); C. H. Bennett, G. Brassard, S. Popescu, B. Schumacher, J. Smolin and W. K. Wootters, Phys. Rev. Lett. 76, 722 (1996).

[4] M. Horodecki, P. Horodecki, and R. Horodecki, Phys. Rev. Lett. 84, 2014 (2000).

[5] C.H. Bennett, D.P. DiVincenzo, J.A. Smolin, , W.K. Wootters, Phys. Rev. A 54 3824, (1996).

[6] P. M. Hayden, M. Horodecki, and B. M. Terhal, Jour. Phys. A: Math. Gen. 34, 35 (2001).

[7] M. Horodecki, P. Horodecki, and R. Horodecki, Phys. Lett. A 283, 1 (2001).

[8] M. Horodecki, P. Horodecki, and R. Horodecki, Phys. Rev. Lett. 80, 5239 (1998).

[9] M. Choi, Linear Algebra and Its Applications, 285-290 (1975); A. Jamiołkowski, Rep. Math. Phys. 3, 275 (1972).

[10] S. Brierley, M. Navasques, and T. Vertesi, arXiv:1609.05011

[11] J. Shang and O. Gühne, Phys. Rev. Lett. 120, 050506 (2018).

[12] C. Witte and M. Trucks, Phys. Lett. A 257, 14 (1999).

[13] M. Ozawa, Phys. Lett. A 268, 158 (1999).

[14] K. Życzkowski and H.-J. Sommers, Jour. Phys. A 34, 7111 (2001).

[15] C. H. Bennett. et al., Phys. Rev. Lett. 82, 5385 (1999).