Optimizing the overlap: $\mathrm{tr}(\rho_A\rho_B)$

In this note, we want to claim that for a single qubit circuits, it is more efficient to optimize the overlap $\mathrm{tr}(\rho_A\rho_B)$ than optimizing the Hilbert-Schmidt (HS) cost function, $C_{HS} \equiv 1-\frac{1}{2}\mathrm{tr}(\rho_A-\rho_B)^2$. Moreover, we want to claim that for a general circuit, optimizing the overlap is an efficient way of optimizing the trace distance (TD) cost function, $C_{TD} \equiv 1-\frac{1}{2}\mathrm{tr}(|\rho_A-\rho_B|)$.

1 Overlap vs HS cost function

In this section, we will prove that optimizing the overlap between two 2-dimensional density matrices automatically optimizes the HS cost function between these density matrices (see theorem 1 below). This is a useful result because optimization of the overlap take lesser time than the optimization of the HS cost function. This is becuase the computation of the gradient of the HS cost function involves the computation of the gradients of $\mathrm{tr}(\rho_A^2)$, $\mathrm{tr}(\rho_B^2)$, and $\mathrm{tr}(\rho_A\rho_B)$. Therefore, the computation of the gradient of the HS cost function will take $2-3$ times computational resources and time than the computation of the gradient of just the overlap.

Theorem 1 *Consider a 2-dimensional Hilbert space and suppose two density matrices ρ_A and ρ_B are such that their overlap is small: $tr(\rho_A\rho_B) = \epsilon$ where $\epsilon \ll 1$. Then these density matrices are almost pure, i.e. $tr(\rho_A^2) \sim tr(\rho_B^2) = 1-O(\epsilon)$. Moreover the HS cost between these matrices satisfies $\epsilon \leq C_{HS}(\rho_A,\rho_B) \leq 2\,\epsilon$.*

Proof Since ρ_A and ρ_B are 2-dimensional density matrices, we can write them as

$$\rho_A = \frac{1}{2}\big(\mathbf{1}+\mathbf{n}_A\cdot\vec{\sigma}\big) \qquad \rho_B = \frac{1}{2}\big(\mathbf{1}+\mathbf{n}_B\cdot\vec{\sigma}\big) \tag{1}$$

where $\vec{\sigma}=(\sigma_x,\sigma_y,\sigma_z)$ is a vector Pauli operator. Note that the overlap between these two density matrices is given by

$$\mathrm{tr}(\rho_A\rho_B) = \frac{1}{2}\big(1+\mathbf{n}_A\cdot\mathbf{n}_B\big)\,. \tag{2}$$

Therefore, if $\mathrm{tr}(\rho_A\rho_B) = \epsilon$, then $\mathbf{n}_A\cdot\mathbf{n}_B = -1+2\epsilon$, and hence $|\mathbf{n}_A\cdot\mathbf{n}_B| = 1-2\epsilon$. Now using the Cauchy-Schwarz inequality, we get

$$|\mathbf{n}_A|\,|\mathbf{n}_B| \geq 1-2\epsilon\,. \tag{3}$$

Moreover, note that the purity of ρ_A and that of ρ_B is given by

$$\mathrm{tr}(\rho_A^2) = \frac{1}{2}\big(1+|\mathbf{n}_A|^2\big) \qquad \mathrm{tr}(\rho_B^2) = \frac{1}{2}\big(1+|\mathbf{n}_B|^2\big)\,. \tag{4}$$

Since $\mathrm{tr}(\rho_A^2) \leq 1$ and $\mathrm{tr}(\rho_B^2) \leq 1$, we deduce that $|\mathbf{n}_A| \leq 1$ and $|\mathbf{n}_B| \leq 1$. Combining these conditions with Eq. (3), we deduce that

$$|\mathbf{n}_A| = 1 - c_A\,\epsilon + O(\epsilon^2) \qquad\qquad |\mathbf{n}_B| = 1 - c_B\,\epsilon + O(\epsilon^2), \tag{5}$$

where $c_A \geq 0$, $c_B \geq 0$, and $c_A + c_B \leq 2$. Inserting these results in Eq. (4), we find that $\mathrm{tr}(\rho_A^2) = 1 - c_A\epsilon + O(\epsilon^2)$ and $\mathrm{tr}(\rho_B^2) = 1 - c_B\epsilon + O(\epsilon^2)$.

Moreover, the HS cost between ρ_A and ρ_B becomes

$$\begin{aligned} C_{HS}(\rho_A, \rho_B) &= 1 + \mathrm{tr}(\rho_A\rho_B) - \frac{1}{2}\big(\mathrm{tr}(\rho_A^2) + \mathrm{tr}(\rho_B^2)\big)\,, && (6)\\ &= \frac{2 + (c_A + c_B)}{2}\,\epsilon + O(\epsilon^2)\,. && (7) \end{aligned}$$

Since $c_A \geq 0$ and $c_B \geq 0$, we deduce that $C_{HS} \geq \epsilon$. Also since $c_A + c_B \leq 2$, we get $C_{HS} \leq 2\epsilon$.

This finishes the proof of theorem 1.

Comment: Can something like theorem 1 be true for Hilbert spaces of more than 2 dimensions? The answer is No. The easiest way to see this is through a counter example. Consider a three dimensional Hilbert space and let $\{|0\rangle, |1\rangle, |2\rangle\}$ be an orthonormal basis. Now suppose $\rho_A = |0\rangle\langle 0|$ and $\rho_B = \frac{1}{2}|1\rangle\langle 1| + \frac{1}{2}|2\rangle\langle 2|$. Even though there is no overlap between these states, the state ρ_B is not pure.

1.1 Numerical comparisons

In this subsection, we numerically check our proposal that optimizing the overlap between two density matrices of a single qubit automatically optimizes the HS cost function between these density matrices. Moreover, we will check that optimization of the overlap takes lesser amount of time than the optimization of the HS cost as argued above. We consider two different data sets. One is the 1D data set from [1]; see Fig. (1). The other is the randomly generated linearly separable data set; see Fig. (2).

Case 1: Data set from [1]

This data set is shown in the left column of Fig. (1). We find that the 300 steps of optimization the overlap took 2.5 minutes whereas the same number of steps for the optimization of the HS cost took almost 9.0 minutes.

Case 2: Randomly generated linearly separable data set

This data set is shown in the left column of Fig. (2). We find that the 300 steps of optimization the overlap took 2.3 minutes whereas the same number of steps for the optimization of the HS cost

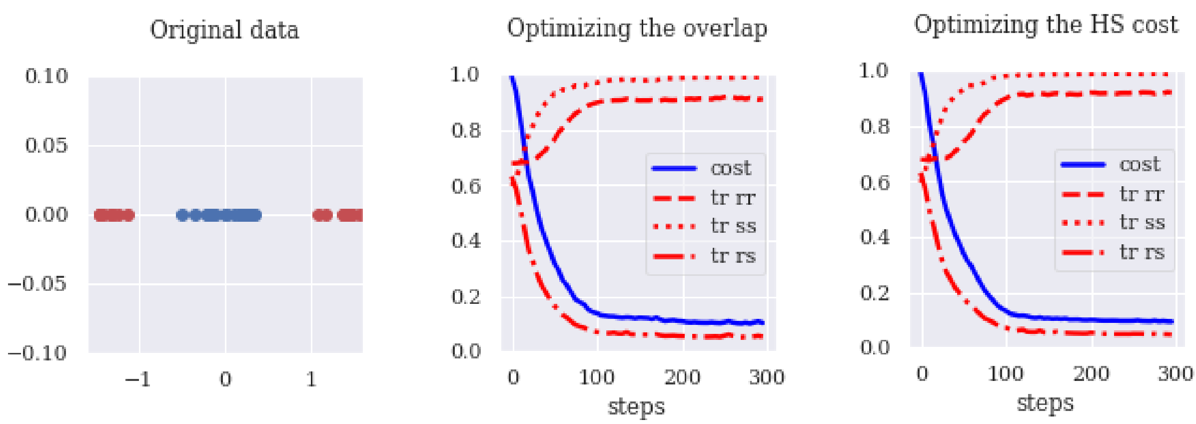

Figure 1: Data set *(left)* from [1]. The result when we optimize the overlap *(center)* and when we optimize the HS cost *(right)*.

took almost 7.2 minutes.

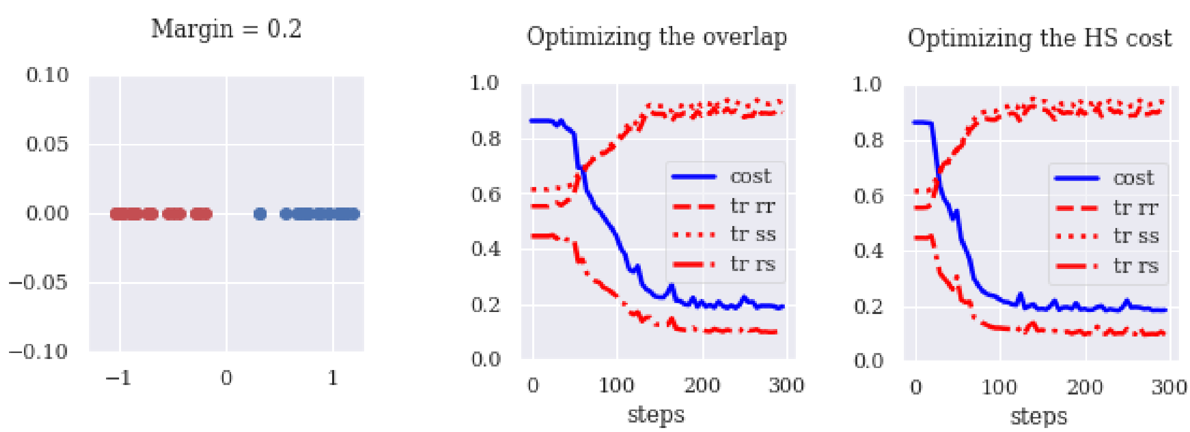

Figure 2: Randomly generated linearly separable data set *(left)*. The result when we optimize the overlap *(center)* and when we optimize the HS cost *(right)*.

2 Overlap vs TD cost function

As discussed in [1], Helstrom measurements are optimal for the task of state discrimination. However, we need to optimize the TD cost, $C_{TD} \equiv 1 - \frac{1}{2}||\rho_A - \rho_B||_1$, to minimize the risk of the Helstrom classifier. Unfortunately, it is difficult to measure the trance distance, and hence, to minimize it. Therefore, it will be useful if we can optimize the TD cost indirectly by optimizing the overlap. In other words, we need a theorem like theorem 1 but for the trace distance.

The rationale behind hoping for such a theorem is a known fact that the trace distance $\frac{1}{2}||\rho_A - \rho_B||_1$ attains its maximum value of 1 if and only if the density matrices ρ_A and ρ_B have orthogonal support. This means that the TD cost between two density matrices is zero if the overlap between them is zero, and vice versa [2]. Hence, the idea that optimizing the overlap automatically optimizes the overlap does not seem far-fetched.

To quantify the above discussion, we define the fidelity between two density matrices [2]:

$$F(\rho_A, \rho_B) \;\equiv\; \mathrm{tr}\left(\sqrt{\rho_A^{1/2}\,\rho_B\,\rho_A^{1/2}}\right). \tag{8}$$

It is known that the fidelity between the density matrices is small if the overlap between them is small, and vice versa. Moreover, it is known that the fidelity and the trace distance are related by [2]

$$1 - F(\rho_A, \rho_B) \;\leq\; \frac{1}{2}||\rho_A - \rho_B||_1 \;\leq\; \sqrt{1 - F^2(\rho_A, \rho_B)}\,. \tag{9}$$

Using these inequalities and the definition, we prove the following theorem.

Theorem 2 *Suppose two density matrices ρ_A and ρ_B are such that their overlap is small: $tr(\rho_A\rho_B) = \epsilon$ where $\epsilon \ll 1$. Then the TD cost between these matrices satisfies $\epsilon/2 \leq C_{TD}(\rho_A, \rho_B) \leq \sqrt{R\,\epsilon}$, where R is the rank of the matrix $M \equiv \rho_A^{1/2}\,\rho_B\,\rho_A^{1/2}$.*

Proof We start by defining a Hermitian and a positive semi-definite matrix $M \equiv \rho_A^{1/2}\,\rho_B\,\rho_A^{1/2}$. The trace of this matrix is equal to the overlap between ρ_A and ρ_B. That is,

$$\mathrm{tr}M \;=\; \mathrm{tr}\left(\rho_A^{1/2}\,\rho_B\,\rho_A^{1/2}\right) \;=\; \mathrm{tr}\left(\rho_A\,\rho_B\right) \;=\; \epsilon. \tag{10}$$

Comparing the definition of M with Eq. (8), we get

$$F \;= \mathrm{tr}\sqrt{M} \tag{11}$$

$$= \sum_i \sqrt{\lambda_i}\,, \tag{12}$$

where we have denoted the eigenvalues of M by $\{\lambda_i\}$. Since M is a Hermitian and positive semi-definite operator, all of its eigenvalues are non-negative, i.e. $\lambda_i \geq 0$. Using this fact, we find that $F \geq \sqrt{\epsilon}$. That is,

$$F^2 \;=\; \left(\sum_i \sqrt{\lambda_i}\right)^2 \;=\sum_{i,j} \sqrt{\lambda_i}\,\sqrt{\lambda_j} \;=\sum_i \lambda_i \;+\sum_{i\neq j} \sqrt{\lambda_i}\,\sqrt{\lambda_j} \;\geq\; \sum_i \lambda_i \;=\; \epsilon\,, \tag{13}$$

where the inequality in the second last step comes the fact that the cross-terms are non-negative.

Now our task is to find an upper bound on the fidelity. To do this, our job is to find the maximum possible value of the sum in Eq. (12). However, we also have to ensure that the trace of the M is ϵ. We can do this using the Lagrange multiplier, μ. In short, we need to find the maximum of

$$\tilde{F} = \sum_{i=1}^{R} \sqrt{\lambda_i} - \mu \left[\sum_{i}^{R} \lambda_i - \epsilon \right] , \tag{14}$$

where R is the rank of the matrix M. Solving for $\partial \tilde{F} / \partial \lambda_i = 0$, we find that all of the λ's must be equal to each other. Combining this with the trace constraint, we find the condition for the maxima is that $\lambda_i = \epsilon / R$. Inserting this in Eq. (8), we find that the maximum possible fidelity is

$$F_{max} = \sqrt{R\epsilon} . \tag{15}$$

Hence, $F \leq \sqrt{R\epsilon}$.

Now using the first inequality in Eq. (9) and using the definition of the TD cost, we find that

$$C_{TD} \leq F \leq \sqrt{R\epsilon} . \tag{16}$$

Similarly, using the second inequality in Eq. (9), we get

$$C_{TD} \geq 1 - \sqrt{1 - F^2} . \tag{17}$$

So far we have not used the condition that $\epsilon \ll 1$. Therefore, all of our results so far are valid for arbitrary $\epsilon \in [0, 1]$. Now let us assume that $\epsilon \ll 1$. In this case, fidelity is also small, and hence we can write Eq. (17) as

$$C_{TD} \geq \frac{1}{2} F^2 + O(F^4) . \tag{18}$$

Combining this with Eq. (13), we get

$$C_{TD} \geq \epsilon / 2 . \tag{19}$$

This finishes the proof of theorem 2.

References

[1] Seth Lloyd, Maria Schuld, Aroosa Ijaz, Josh Izaac, and Nathan Killoran. Quantum embeddings for machine learning. *arXiv e-prints*, page arXiv:2001.03622, January 2020.

[2] Michael A. Nielsen and Isaac L. Chuang. *Quantum Computation and Quantum Information: 10th Anniversary Edition.* Cambridge University Press, 2010.