Solution to Citadel Problem

Vansh Ramani and Akarsh Gupta

July 15, 2024

Introduction

The motivation for this solution is derived from solving systems of linear equations under constraints. We especially use Non-Negative Linear System and optimize it to include the criteria for "well behaved" as defined in 4.3 . The problem also shares similarities with approaches in quantum mechanics. This perspective has provided a unique framework for addressing the problem at hand.

In Summary, our algorithm will use general logic to build constraints and give formal mathematical proofs to consolidate the ideas used.

1 Exercise 1

To show that the given probability distribution marginalizes to the probability table, we first define our notations:

- P_0 represents the probability of the outcome 0000 for $a_1b_1a_2b_2$
- P_1 represents the probability of the outcome 0001 for $a_1b_1a_2b_2$
- $\vdots$
- P_{15} represents the probability of the outcome 1111 for $a_1b_1a_2b_2$

We see that with general logic, the relation between total probability distribution can be related by general logic in a set of 16 equations.

Given the global probability distribution:

$$
\begin{array}{cccc}
P_0 = \dfrac{1}{8} & P_1 = 0 & P_2 = 0 & P_3 = \dfrac{1}{4} \\
P_4 = \dfrac{1}{4} & P_5 = 0 & P_6 = 0 & P_7 = 0 \\
P_8 = \dfrac{1}{4} & P_9 = 0 & P_{10} = \dfrac{1}{8} & P_{11} = 0 \\
P_{12} = 0 & P_{13} = 0 & P_{14} = 0 & P_{15} = 0
\end{array}
$$

Now, we can express the relationships between this global distribution and the probability table:

$$
\begin{aligned}
P_0 + P_1 + P_2 + P_3 &= P(0,0|a_1,b_1) = \frac{3}{8} \\
P_8 + P_9 + P_{10} + P_{11} &= P(1,0|a_1,b_1) = \frac{3}{8} \\
P_4 + P_5 + P_6 + P_7 &= P(0,1|a_1,b_1) = \frac{1}{4} \\
P_{12} + P_{13} + P_{14} + P_{15} &= P(1,1|a_1,b_1) = 0 \\
P_0 + P_1 + P_4 + P_5 &= P(0,0|a_1,b_2) = \frac{3}{8} \\
P_8 + P_9 + P_{12} + P_{13} &= P(1,0|a_1,b_2) = \frac{3}{8} \\
P_2 + P_3 + P_6 + P_7 &= P(0,1|a_1,b_2) = \frac{1}{4} \\
P_{10} + P_{11} + P_{14} + P_{15} &= P(1,1|a_1,b_2) = 0 \\
P_0 + P_2 + P_4 + P_6 &= P(0,0|a_2,b_1) = \frac{3}{8} \\
P_8 + P_{10} + P_{12} + P_{14} &= P(1,0|a_2,b_1) = \frac{3}{8} \\
P_1 + P_3 + P_5 + P_7 &= P(0,1|a_2,b_1) = \frac{1}{4} \\
P_9 + P_{11} + P_{13} + P_{15} &= P(1,1|a_2,b_1) = 0 \\
P_0 + P_2 + P_8 + P_{10} &= P(0,0|a_2,b_2) = \frac{5}{8} \\
P_4 + P_6 + P_{12} + P_{14} &= P(1,0|a_2,b_2) = \frac{1}{8} \\
P_1 + P_3 + P_9 + P_{11} &= P(0,1|a_2,b_2) = 0 \\
P_5 + P_7 + P_{13} + P_{15} &= P(1,1|a_2,b_2) = \frac{1}{4}
\end{aligned}
$$

To verify that the global distribution marginalizes to the given probability table, we need to substitute the values of P_i into these equations and check if they hold true.

On a simple check from Table 1, we can see that all of these equations hold.

Therefore, we have shown that the given global probability distribution indeed marginalizes to the specified probability table.

A B	(0, 0)	(1, 0)	(0, 1)	(1, 1)
a_1 b_1	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{4}$	0
a_1 b_2	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{4}$	0
a_2 b_1	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{4}$	0
a_2 b_2	$\frac{5}{8}$	$\frac{1}{8}$	0	$\frac{1}{4}$

Table 1: Marginalized Probability Table Ex.1

2 Exercise 2

2.1 Problem Description

We represent the system of equations as $Ax = B$, where:

- A is a 16×16 matrix
- x is a 16×1 vector representing the probabilities $P_0, P_1, \ldots, P_{15}$
- B is a 16×1 vector representing the probabilities in the table

The matrix A is constructed as a sparse matrix, where each row contains exactly four 1's and twelve 0's. We can represent this efficiently by listing the indices of the 1's in each row:

$$
\begin{aligned}
\text{Row 1:}&\{0, 1, 2, 3\} \\
\text{Row 2:}&\{8, 9, 10, 11\} \\
\text{Row 3:}&\{4, 5, 6, 7\} \\
\text{Row 4:}&\{12, 13, 14, 15\} \\
\text{Row 5:}&\{0, 1, 4, 5\} \\
\text{Row 6:}&\{8, 9, 12, 13\} \\
\text{Row 7:}&\{2, 3, 6, 7\} \\
\text{Row 8:}&\{10, 11, 14, 15\} \\
\text{Row 9:}&\{0, 2, 4, 6\} \\
\text{Row 10:}&\{8, 10, 12, 14\} \\
\text{Row 11:}&\{1, 3, 5, 7\} \\
\text{Row 12:}&\{9, 11, 13, 15\} \\
\text{Row 13:}&\{0, 2, 8, 10\} \\
\text{Row 14:}&\{4, 6, 12, 14\} \\
\text{Row 15:}&\{1, 3, 9, 11\} \\
\text{Row 16:}&\{5, 7, 13, 15\}
\end{aligned}
$$

Each set of indices represents the positions of the 1's in that row of the matrix A. All other entries in each row are 0.

Now, the first thought that came to my mind was to apply a simple $Ax = b$ solver (which was my first hint towards this approach being wrong; it could not possibly be *this* straightforward!)

The dimensions of the matrix A are pretty large to perform row-reduction and find the rank of the matrix manually. Therefore I used the *linalg* library of NumPy to calculate the rank, which turned out to be 9.

But I have 16 variables! 16 probabilities for each of the combinations possible in the global probability distribution.

Therefore, this is an **under-determined system of linear equations**

Along with that, I need to consider two more constraints: all the variables I end up with must be positive, and the sum of the variables must add up to 1. If the solution of $Ax = b$ exists, then it must automatically cause sum of variables to add up to 1.

By summing all the equations, we get:

$$
\begin{aligned}
&(P_0 + P_1 + P_2 + P_3) + (P_8 + P_9 + P_{10} + P_{11}) + (P_4 + P_5 + P_6 + P_7) + (P_{12} + P_{13} + P_{14} + P_{15}) \\
&+ (P_0 + P_1 + P_4 + P_5) + (P_8 + P_9 + P_{12} + P_{13}) + (P_2 + P_3 + P_6 + P_7) + (P_{10} + P_{11} + P_{14} + P_{15}) \\
&+ (P_0 + P_2 + P_4 + P_6) + (P_8 + P_{10} + P_{12} + P_{14}) + (P_1 + P_3 + P_5 + P_7) + (P_9 + P_{11} + P_{13} + P_{15}) \\
&+ (P_0 + P_2 + P_8 + P_{10}) + (P_4 + P_6 + P_{12} + P_{14}) + (P_1 + P_3 + P_9 + P_{11}) + (P_5 + P_7 + P_{13} + P_{15}) \\
&= \sum_{i,j=0,1} (P(0,0|a_i,b_j) + P(1,0|a_i,b_j) + P(0,1|a_i,b_j) + P(1,1|a_i,b_j)) \\
&= 4\,(P(0,0|a_1,b_1) + P(1,0|a_1,b_1) + P(0,1|a_1,b_1) + P(1,1|a_1,b_1)) \\
&= 4
\end{aligned}
$$

Thus, we can see that:

$$\sum_{i=0}^{15} P_i = 1$$

This system must be solved under the constraint that all values in x are non-negative. To achieve this, we use a Non-Negative Least Squares (NNLS) solver

2.2 Well-Behaved Classically Deterministic NNLS Algorithm

The NNLS algorithm solves the optimization problem:

$$\min_{x} \|Ax - b\|_2^2 \quad \text{subject to} \quad x \geq 0$$

Algorithm 1 Complete NNLS with CHSH and Well-Behaved Checks

function $\textsc{CompleteNNLS}(A,\, b)$
    if $\text{G}\textsc{lobalDistributionCheck}(A,\, b)$ **then**
        return "CHSH inequality not satisfied - No Global DIstribution"
    end if
    if $\textsc{CheckNotWellBehaved}(A,\, b)$ **then**
        return "System is not well-behaved"
    end if
    return x
    $x \leftarrow \text{NNLS}(A, b)$
end function
function $\text{NNLS}(A,\, b)$
    Initialize $x = 0$, $P = \{\}$, $R = \{1, \ldots, n\}$
    while $R \neq \{\}$ and $\exists i \in R$ such that $(A^T(b - Ax))_i > 0$ **do**
        $j \leftarrow_{i \in R} (A^T(b - Ax))_i$
        Move j from R to P
        $x_P \leftarrow_z \|A_P z - b\|_2$
        while $\exists i \in P$ such that $(x_P)_i \leq 0$ **do**
            $\alpha \leftarrow \min_{i \in P:(x_P)_i \leq 0} \frac{x_i}{x_i - (x_P)_i}$
            $x \leftarrow x + \alpha(x_P - x)$
            Move from P to R all indices i where $x_i = 0$
            $x_P \leftarrow_z \|A_P z - b\|_2$
        end while
        $x \leftarrow x_P$
    end while
    return x
end function
function $\textsc{NoGlobalDistributionCheck}(A,\, b)$
    // Implementation CHSH check as described in subsection 5.2
    return true if CHSH inequality fails, otherwise true
end function
function $\textsc{CheckNotWellBehaved}(A,\, b)$
    // Implementation of well-behaved check as described in subsection 4.3
    return true if the system is not well-behaved, otherwise false
end function

Here, P is the set of positive components, and R is the set of zero components. The NNLS algorithm terminates when all values in the dual vector $(A^T(b - Ax))$ are ≤ 0.

2.2.1 CHSH Inequality Check

Before performing NNLS, we check if the state distribution satisfies the CHSH inequality, which means a global probability distribution may exists. If it does and system is well behaved, NNLS is guaranteed to converge.
Another implementation may be, by checking if Probabilities sum to 1 after NNLS converges. This will also be sufficient to show that no solution to $Ax = b$ exists, hence no global distribution.

2.2.2 Implementation

The implementation uses the `nnls` function from the `scipy.optimize` library in Python. The maximum number of iterations is set to 100 to handle extreme cases, although the optimum is typically reached in 10 to 15 iterations.

2.2.3 Example Global Probability Distribution

The following table shows an example of a reconstructed global probability distribution:

$a_1b_1a_2b_2$	P
0000	$\frac{1}{8}$
0011	$\frac{1}{4}$
0100	$\frac{1}{4}$
1000	$\frac{1}{4}$
1010	$\frac{1}{8}$
Other	0

3 Exercise 3

Application of the algorithm to the given probability tables:

3.1 Table 1

3.1.1 The CHSH Inquality

The CHSH inequality states that if any system can account for the local hidden variables interacting within it and give a classical solution for the system , then :

$$|E(a_1, b_1) + E(a_1, b_2) + E(a_2, b_1) - E(a_2, b_2)| \leq 2$$

where $E(a_i, b_j)$ is the expectation value of the product of outcomes for measurement detector settings a_i and b_j.

3.1.2 Calculating Expectation Values

Given the probability table:

A	B	(0,0)	(1,0)	(0,1)	(1,1)
a_1	b_1	1/2	0	0	1/2
a_1	b_2	3/8	1/8	1/8	3/8
a_2	b_1	3/8	1/8	1/8	3/8
a_2	b_2	1/8	3/8	3/8	1/8

We can calculate the expectation values as follows:

$$\begin{aligned}
E(a_i, b_j) &= P(0, 0|a_i, b_j) + P(1, 1|a_i, b_j) - P(0, 1|a_i, b_j) - P(1, 0|a_i, b_j) \\
E(a_1, b_1) &= \frac{3}{4} + \frac{1}{4} - 0 - 0 = 1 \\
E(a_1, b_2) &= \frac{3}{4} + \frac{1}{4} - 0 - 0 = 1 \\
E(a_2, b_1) &= \frac{3}{4} + \frac{1}{4} - 0 - 0 = 1 \\
E(a_2, b_2) &= \frac{3}{4} + \frac{1}{4} - 0 - 0 = 1
\end{aligned}$$

3.1.3 Applying the CHSH Inequality

Now, let's substitute these values into the CHSH inequality:

$$
\begin{aligned}
|E(a_1,b_1)+E(a_1,b_2)+E(a_2,b_1)-E(a_2,b_2)| &= |1+1+1-1| \\
&= |1+1+1-1| \\
&= |2| \\
&= 2
\end{aligned}
$$

3.1.4 Conclusion

We find that $2 \geq 2$, which does not violate the CHSH inequality.
Couple this with the fact that there is no dependence of outcome of any detector setting for one of the researchers on the choice of detector setting of the other researcher

This tells us that it is possible to classically obtain a distribution for this case and that the distribution will not be affected by contextuality and therefore a definite global probability distribution can be obtained

We use our algorithm devised above to solve for such a global probability distribution

$a_1b_1a_2b_2$	P
0000	$\frac{3}{4}$
1111	$\frac{1}{4}$
Other	0

3.2 Table 2

Here it turns out, by our CHSH solver, that the CHSH parameter is $0.5 \leq 2$; and therefore, a global distribution is likely to exist

Also, our well-behaved solver shows the matrix is well-behaved, and we can expect no interference due to contextuality

$a_1b_1a_2b_2$	P
0001	$\frac{1}{2}$
0010	$\frac{1}{8}$
0100	$\frac{1}{8}$
1000	$\frac{1}{4}$
Other	0

3.3 Table 3

Here it turns out, by our CHSH solver, that the CHSH parameter is $0.875 \leq 2$; and therefore, a global distribution is likely to exist

Also, our well-behaved solver shows the matrix is well-behaved, and we can expect no interference due to contextuality

$a_1b_1a_2b_2$	P
0001	$\frac{3}{16}$
0010	$\frac{1}{8}$
0111	$\frac{1}{32}$
1000	$\frac{3}{16}$
1100	$\frac{1}{16}$
1101	$\frac{7}{32}$
1110	$\frac{3}{32}$
1111	$\frac{3}{32}$
Other	0

4 Exercise 4

4.1 Demonstrating Dependency of Alice's Measurement on Bob's Choice

To show that Bob's choice of measurement affects Alice's probabilities for measurement a_1, we need to demonstrate that:

$$P(a_1 = 0|\text{Bob chooses } b_1) \neq P(a_1 = 0|\text{Bob chooses } b_2)$$

and similarly for $a_1 = 1$.

Let's calculate these probabilities using the given probability table:

$$\begin{aligned} P(a_1 = 0|\text{Bob chooses } b_1) &= P(0,0|a_1,b_1) + P(0,1|a_1,b_1) \\ &= \frac{1}{2} + 0 = \frac{1}{2} \end{aligned}$$

$$\begin{aligned} P(a_1 = 0|\text{Bob chooses } b_2) &= P(0,0|a_1,b_2) + P(0,1|a_1,b_2) \\ &= \frac{1}{2} + \frac{1}{4} = \frac{3}{4} \end{aligned}$$

Clearly, $\frac{1}{2} \neq \frac{3}{4}$, which demonstrates that Bob's choice of measurement affects the probability of Alice's outcome for a_1.

Similarly, we can show this for $a_1 = 1$:

$$\begin{aligned} P(a_1 = 1|\text{Bob chooses } b_1) &= P(1,0|a_1,b_1) + P(1,1|a_1,b_1) \\ &= \frac{1}{4} + \frac{1}{4} = \frac{1}{2} \end{aligned}$$

$$\begin{aligned} P(a_1 = 1|\text{Bob chooses } b_2) &= P(1,0|a_1,b_2) + P(1,1|a_1,b_2) \\ &= 0 + \frac{1}{4} = \frac{1}{4} \end{aligned}$$

Again, $\frac{1}{2} \neq \frac{1}{4}$, confirming the dependency.

4.2 Understanding Why There is No Solution

On one hand, if we apply the CHSH inequality as applied above in 3.1, we find that the value of S comes out to be only 1/2; which means that the system can possibly attain a classical solution, and so indeed happens at first as a probability distribution appears! However, it leaves a significant residue and the minimization does not end up at 0; implying that the vector b does not exist in the column space of matrix A

Why does this happen, in spite of the CHSH inequality not being violated? Because of the interactions between the measurement detector setting of Bob and the measurement outcome of one particular detector setting of Alice

This is the concept of contextuality: Contextuality in quantum mechanics refers to the idea that the outcome of a measurement can depend on the context when specific set of other measurements that are performed alongside it

We can define the interaction between Bob's choice of detector setting and Alice's outcome on the first detector setting as a local hidden variable

Now, this local hidden variable cannot be clearly defined because it is contextual (as we have shown in 4.1, the two probabilities are not equal !) and therefore is open to any possible scenario, which renders the explanation of the local hidden variable indeterministic

The values can depend on a specific set of measurements, which obstructs the formation of a global probability distribution

This is the physical, intuitive understanding I achieved of the idea; Since both math and physics lead to the same conclusion that a global probability distribution cannot exist, I will rest my case here.

4.3 Mathematical Definition of Well-Behaved Nature

We can generalize this concept to state that for a well-behaved probability table, the choice of measurement by one party should not affect the outcome probabilities of the other party. Mathematically, for all $i, j \in \{1, 2\}$ and $k \in \{0, 1\}$:

$$P(a_i = k|\text{Bob chooses } b_1) = P(a_i = k|\text{Bob chooses } b_2)$$
$$P(b_j = k|\text{Alice chooses } a_1) = P(b_j = k|\text{Alice chooses } a_2)$$

These conditions can be expressed in terms of the probabilities in our table:

$$P(k, 0|a_i, b_1) + P(k, 1|a_i, b_1) = P(k, 0|a_i, b_2) + P(k, 1|a_i, b_2)$$
$$P(0, k|a_1, b_j) + P(1, k|a_1, b_j) = P(0, k|a_2, b_j) + P(1, k|a_2, b_j)$$

4.3.1 Incorporation in Matrix Solving

These conditions are implicitly incorporated in our matrix solving approach under constraints. Recall our system $Ax = B$, where x represents the probabilities of the global distribution. The structure of matrix A ensures that the sum of probabilities for each measurement choice is consistent, regardless of the other party's choice.

For example, the rows in A corresponding to $P(0, 0|a_1, b_1)+P(0, 1|a_1, b_1)$ and $P(0, 0|a_1, b_2) + P(0, 1|a_1, b_2)$ would be identical if the table were well-behaved. In our case, they are not, which is why we cannot find a consistent global distribution.

The non-negative least squares approach attempts to find a solution that best satisfies these constraints. When no exact solution exists (as in this case), the algorithm minimizes the discrepancy, highlighting the inconsistency in the probability table.

5 Exercise 5

5.1 To show that the system is well behaved

Consider the matrix A:

$$a = \begin{pmatrix} \frac{1}{2} & 0 & 0 & \frac{1}{2} \\ \frac{3}{8} & \frac{1}{8} & \frac{1}{8} & \frac{3}{8} \\ \frac{3}{8} & \frac{1}{8} & \frac{1}{8} & \frac{3}{8} \\ \frac{1}{8} & \frac{3}{8} & \frac{3}{8} & \frac{1}{8} \end{pmatrix}$$

We need to check if this matrix is "well-behaved" according to the conditions we have defined above in 4.3

1. Condition for a_1:

$$a[0,0] + a[0,2] = \frac{1}{2} + 0 = \frac{1}{2}$$

$$a[1,0] + a[1,2] = \frac{3}{8} + \frac{4}{32} = \frac{3}{8} + \frac{1}{8} = \frac{4}{8} = \frac{1}{2}$$

Since both sums are equal, this condition is satisfied.

2. Condition for a_2:

$$a[2,0] + a[2,2] = \frac{3}{8} + \frac{4}{32} = \frac{3}{8} + \frac{1}{8} = \frac{4}{8} = \frac{1}{2}$$

This condition is trivially true as it compares the same elements.

3. Condition for b_1:

$$a[0,0] + a[0,1] = \frac{1}{2} + 0 = \frac{1}{2}$$

$$a[2,0] + a[2,1] = \frac{3}{8} + \frac{1}{8} = \frac{4}{8} = \frac{1}{2}$$

Since both sums are equal, this condition is satisfied.

4. Condition for b_2:

$$a[1,0] + a[1,1] = \frac{3}{8} + \frac{4}{32} = \frac{3}{8} + \frac{1}{8} = \frac{4}{8} = \frac{1}{2}$$

$$a[3,0] + a[3,1] = \frac{1}{8} + \frac{12}{32} = \frac{1}{8} + \frac{3}{8} = \frac{4}{8} = \frac{1}{2}$$

Since both sums are equal, this condition is satisfied.

Since all conditions are satisfied, the matrix A is "well-behaved."

5.2 Formal mathematical proof that this matrix does not admit any global distribution

5.2.1 Proof using the NNLS approach

I first consider an algorithmic approach to solve this problem

I had to find solutions for linear systems of equations through quadratic programming only because my system was under-determined, and I could only hope for minimizations leading me to valid answers. I figured that the termination condition of the NNLS algorithm is if there is no such variable which is set at 0 *OR* if $A^T(b\text{-}Ax)$ had non-positive values for all of it's indices; which is essentially the dual vector and the non-positivity of the dual vector helps us satisfy complementary slackness which leads to an optimal solution

Let us suppose that there exists an x ≥ 0 such that $Ax = b$

Then

$$\min_x \|Ax - B\|_2^2 \quad = 0$$

for any non-negative x

Since we know that NNLS converges in finite iterations, we can say that if $Ax = b$ possesses a solution for a non-negative x, then on termination of the algorithm, we must find an x such that $Ax = b$

However, in our case, we find that termination of the algorithm is achieved when the condition does not hold

Since this is a convex optimization problem, we have achieved the optimal solution and cannot proceed further to improve!

And therefore,

$$\min_x \|Ax - B\|_2^2 \quad \neq 0$$

for any non-negative x

Using this outline we can draw the following mathematical proof:

We consider the problem of finding solutions for a linear system of equations through quadratic programming, specifically using the Non-Negative Least Squares (NNLS) algorithm. This approach is necessitated by the under-determined nature of our system, where we seek minimizations that lead to valid answers.

Step 1: NNLS Algorithm Termination Condition

The NNLS algorithm terminates when either:

- There is no variable set to 0, or
- $A^T(b - Ax)$ has non-positive values for all of its indices.

Note that $A^T(b - Ax)$ is the dual vector, and its non-positivity ensures the satisfaction of complementary slackness, leading to an optimal solution.

Step 2: Existence of Non-negative Solution

Let us suppose that there exists an $x \geq 0$ such that $Ax = b$. Then:

$$\min_x \|Ax - b\|_2^2 = 0$$

for any non-negative x.

Step 3: NNLS Convergence

Since NNLS converges in finite iterations, we can assert that if $Ax = b$ possesses a solution for a non-negative x, then upon termination of the algorithm, we must find an x such that $Ax = b$.

Step 4: Termination Condition in Our Case

In our specific case, we observe that the termination of the algorithm is achieved when the aforementioned condition does not hold.

Step 5: Optimality of Solution

Given that we have met our conditions of feasibility and complementary slackness, we have achieved the optimal solution and cannot proceed further to improve it.

Step 6: Non-zero Minimum

We have minimized the objective function and it's value does not turn out to be 0, which means it can never attain the value of 0 for the given constraints
Therefore, we can conclude:

$$\min_x \|Ax - b\|_2^2 \neq 0$$

for any non-negative x.

And from this we can conclude that $Ax = b$ has no solution for the given constraints and that a global probability distribution for this can never be obtained

Conclusion: Despite the well-behaved nature of the state probability matrix, we have shown that it does not admit any global probability distribution. This is evidenced by the non-existence of a non-negative solution x that satisfies $Ax = b$ exactly, as demonstrated by the non-zero minimum of the objective function in the NNLS problem.

5.2.2 Using the CHSH Inequality Approach

To prove that the probability table in question 5 has no global probability distribution, we can use the CHSH inequality. This inequality provides a bound that must be satisfied by any local hidden variable theory, which is equivalent to having a global probability distribution.

The CHSH Inequality

The CHSH inequality states that for any local hidden variable theory:

$$|E(a_1, b_1) + E(a_1, b_2) + E(a_2, b_1) - E(a_2, b_2)| \leq 2$$

where $E(a_i, b_j)$ is the expectation value of the product of outcomes for measurements a_i and b_j.

Calculating Expectation Values

Given the probability table:

A	B	(0,0)	(1,0)	(0,1)	(1,1)
a_1	b_1	1/2	0	0	1/2
a_1	b_2	3/8	1/8	1/8	3/8
a_2	b_1	3/8	1/8	1/8	3/8
a_2	b_2	1/8	3/8	3/8	1/8

We can calculate the expectation values as follows:

$$\begin{aligned}
E(a_i, b_j) &= P(0, 0|a_i, b_j) + P(1, 1|a_i, b_j) - P(0, 1|a_i, b_j) - P(1, 0|a_i, b_j) \\
E(a_1, b_1) &= \frac{1}{2} + \frac{1}{2} - 0 - 0 = 1 \\
E(a_1, b_2) &= \frac{3}{8} + \frac{3}{8} - \frac{1}{8} - \frac{1}{8} = \frac{1}{2} \\
E(a_2, b_1) &= \frac{3}{8} + \frac{3}{8} - \frac{1}{8} - \frac{1}{8} = \frac{1}{2} \\
E(a_2, b_2) &= \frac{1}{8} + \frac{1}{8} - \frac{3}{8} - \frac{3}{8} = -\frac{1}{2}
\end{aligned}$$

Applying the CHSH Inequality

Now, let's substitute these values into the CHSH inequality:

$$\begin{aligned}
|E(a_1, b_1) + E(a_1, b_2) + E(a_2, b_1) - E(a_2, b_2)| &= |1 + \frac{1}{2} + \frac{1}{2} - (-\frac{1}{2})| \\
&= |1 + \frac{1}{2} + \frac{1}{2} + \frac{1}{2}| \\
&= |2.5| \\
&= 2.5
\end{aligned}$$

Conclusion

We find that $2.5 > 2$, which violates the CHSH inequality. This violation proves that the given probability table cannot be explained by any local hidden variable theory, which is equivalent to saying that there exists no global probability distribution that can produce these correlations.

This result is consistent with our earlier findings using the matrix method. The CHSH inequality provides a more direct and general way to prove the impossibility of a global probability distribution for this specific set of correlations.

5.3 Interpretation and Potential Real Life Examples

5.3.1 Interpretation

The violation of the CHSH inequality suggests that the correlations described by this probability table exhibit quantum-like behavior. Such correlations can be realized in quantum systems but cannot be explained by classical probability theory. This is a hallmark of quantum entanglement and demonstrates the non-local hidden nature of quantum correlations.

5.3.2 Real World Applications

Now, I am *unable* to access a solution due to the presence of hidden variables which I cannot compute or work on with since they are not deterministic; they cannot be determined conclusively. I would be pretty disappointed if I was waging a cyberwar on my opponent and they had this as a lock :)

So my guess is this would have real world implications in cryptography and cybersecurity; though my guess is achieveing those quantum states would be very difficult and computationally expensive so we are not there yet for universal security but I am sure a lot of scientists are working on it

Another application where I feel this would work, would be in federated learning which I feel is future of model training in AI, By never forming a global probability distribution, individual data privacy is maintained that would address concerns over data breaches and misuse.

I think these make for a really impactful application.

A Proofs

A.1 Proof of CHSH Inequality

Bell's 1971 derivation assumes the "Objective Local Theory," which was later used by Clauser and Horne. Here, any hidden variables associated with the detectors themselves are assumed to be independent on the two sides and can be averaged out from the start.

We start with the standard assumption of independence of the two sides, enabling us to obtain the joint probabilities of pairs of outcomes by multiplying the separate probabilities, for any selected value of the hidden variable λ. λ is assumed to be drawn from a fixed distribution of possible states of the source, the probability of the source being in the state λ for any particular trial being given by the density function $\rho(\lambda)$, the integral of which over the complete hidden variable space is 1. We thus assume we can write:

$$E(a,b) = \int A(a,\lambda)B(b,\lambda)\rho(\lambda)d\lambda$$

where A and B are the outcomes. Since the possible values of A and B are 0 or 1, it follows that:

$$|A| \leq 1 \quad |B| \leq 1$$

Then, if $a = a_1$, $a' = a_2$, $b = b_1$, and $b' = b_2$ are alternative settings for the detectors,

$$E(a,b) - E(a,b') = \int [A(a,\lambda)B(b,\lambda) - A(a,\lambda)B(b',\lambda)]\rho(\lambda)d\lambda$$

$$= \int [A(a,\lambda)B(b,\lambda) - A(a,\lambda)B(b',\lambda) \pm A(a,\lambda)B(b,\lambda)A(a',\lambda)B(b',\lambda)$$
$$\mp A(a,\lambda)B(b,\lambda)A(a',\lambda)B(b',\lambda)]\rho(\lambda)d\lambda$$

$$= \int A(a,\lambda)B(b,\lambda)[1 \pm A(a',\lambda)B(b',\lambda)]\rho(\lambda)d\lambda$$
$$- \int A(a,\lambda)B(b',\lambda)[1 \pm A(a',\lambda)B(b,\lambda)]\rho(\lambda)d\lambda$$

Taking absolute values of both sides, and applying the triangle inequality to the right-hand side, we obtain:

$$|E(a,b) - E(a,b')| \leq \left| \int A(a,\lambda)B(b,\lambda)[1 \pm A(a',\lambda)B(b',\lambda)]\rho(\lambda)d\lambda \right|$$
$$+ \left| \int A(a,\lambda)B(b')[1 \pm A(a',\lambda)B(b,\lambda)]\rho(\lambda)d\lambda \right|$$

We use the fact that $[1 \pm A(a',\lambda)B(b',\lambda)]\rho(\lambda)$ and $[1 \pm A(a',\lambda)B(b,\lambda)]\rho(\lambda)$ are both non-negative to rewrite the right-hand side as:

$$\int |A(a,\lambda)B(b,\lambda)||[1 \pm A(a',\lambda)B(b',\lambda)]\rho(\lambda)d\lambda|$$
$$+ \int |A(a,\lambda)B(b',\lambda)||[1 \pm A(a',\lambda)B(b,\lambda)]\rho(\lambda)d\lambda|$$

By the earlier bound, this must be less than or equal to:

$$\int [1 \pm A(a', \lambda)B(b', \lambda)]\rho(\lambda)d\lambda + \int [1 \pm A(a', \lambda)B(b, \lambda)]\rho(\lambda)d\lambda$$

which, using the fact that the integral of $\rho(\lambda)$ is 1, is equal to:

$$2 \pm [E(a', b') + E(a', b)]$$

Putting this together with the left-hand side, we have:

$$|E(a, b) - E(a, b')| \leq 2 \pm [E(a', b') + E(a', b)]$$

which means that the left-hand side is less than or equal to both $2 + [E(a', b') + E(a', b)]$ and $2 - [E(a', b') + E(a', b)]$. That is:

$$|E(a, b) - E(a, b')| \leq 2 - |E(a', b') + E(a', b)|$$

from which we obtain:

$$2 \geq |E(a, b) - E(a, b')| + |E(a', b') + E(a', b)| \geq |E(a, b) - E(a, b') + E(a', b') + E(a', b)|$$

which is the CHSH inequality.

PS : The CHSH inequality was derived considering two observational variables, each having two possible detector settings, and each variable could attain states, either "ON" or "OFF" ; which were corresponded by the numbers 1 and - 1

Now, of course, the structural similarity of the problem led me to explore this for solving our system. However, I had to take care of one thing; I changed the "OFF" state from 0 to -1 while considering calculations in our example for correct results and interpretations. This will help make sense of the calculations that we have performed to calculate the CHSH inequality parameter

A.2 Proof of How NNLS Algorithm Terminates

The NNLS problem can be formulated as follows:

$$\min_{x \geq 0} \|Ax - b\|_2^2$$

where $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$.

Dual Vector

The dual vector y is defined as:

$$y = A^T(b - Ax)$$

Optimality Conditions

The optimality conditions for the NNLS problem are derived from the Karush-Kuhn-Tucker (KKT) conditions. These conditions are:

1. **Primal Feasibility:** $x \geq 0$
2. **Dual Feasibility:** $y \leq 0$
3. **Complementary Slackness:** $x_i y_i = 0$ for all i

Primal Feasibility

The primal feasibility condition ensures that the solution x is non-negative:

$$x \geq 0$$

Dual Feasibility

The dual feasibility condition requires that each component of the dual vector must be non-positive:

$$y \leq 0$$

Since $y = A^T(b - Ax)$, this condition checks whether the gradient of the objective function with respect to x, projected onto the constraints, is directed in a non-positive direction.

Complementary Slackness

Complementary slackness ensures that for each component i, either $x_i = 0$ or $y_i = 0$:

$$x_i y_i = 0 \quad \text{for all } i$$

This condition means that if x_i is positive, then the corresponding dual variable y_i must be zero, indicating that there is no further gradient in the direction that would increase x_i. Conversely, if y_i is negative, then x_i must be zero to satisfy the condition.

Conclusion

If the dual vector $y = A^T(b - Ax)$ satisfies the dual feasibility condition, combined with primal feasibility and complementary slackness, it indicates that x is an optimal solution to the NNLS problem. Therefore, having all negative values in the dual vector implies that the current solution x is optimal, as any potential decrease in the objective function would require violating the non-negativity constraint on x. I have checked for all of these conditions in my Python Notebook.