A TWO-FAMILY EXTREMAL PROBLEM IN HAMMING SPACE

Rudolf AHLSWEDE*

Faculty of Mathematics, University of Bielefeld, Universitätsstrasse 1, 4800 Bielefeld, W. Germany

Abbas EL GAMAL†, King F. PANG‡

Information Systems Laboratory, Electrical Engineering Department, Stanford University, Stanford, CA 94305, USA

Received 15 March 1983

The pair \((A, B); A, B \subseteq \{0, 1\}^n\) is called an \((m, \delta)-system\), if for the Hamming distance function \(d\),

\[d(a, b) = \delta \forall a \in A, \forall b \in B.\]  \(\text{(1)}\)

Let \(S^n_{\delta}\) denote the set of those systems. We consider the function

\[M(m, \delta) \triangleq \max\{|A| | A, B \in S^n_{\delta}; |A| \leq m\}\]  \(\text{(2)}\)

and prove the following Theorem: For \(m = 1, 2, \ldots, \max_{0 < \delta \leq m} M(m, \delta) = 2^{2n}\), if \(m = 2n\) or \(m = 2n + 1\). The maximum is assumed for \(\delta = n\).

1. Introduction

The study of \((m, \delta)-systems\) is motivated by the problem of lower bounding the two-way complexity (in the sense of Yao [1]) of the Hamming distance function. Results in this direction will be contained in [2].

Two-family extremal problems have frequently been considered in the literature [3]. Replacement of (1) by

\[d(a, b) \leq \delta \forall a \in A, \forall b \in B\]  \(\text{(1')}\)

yields an extremal problem, which has been solved in [4]. However, in spite of the similarity between conditions (1) and (1'), the present proof techniques are quite different from those in [4]. Actually, we give two proofs of the Theorem. The first is by a 1-step and the second by a 2-step induction in \(m\). The examples

\[A \triangleq \{01, 10\}^n, \quad B \triangleq \{11, 00\}^n\]  \(\text{(3)}\)

\[A \triangleq \{01, 10\}^n \times \{\}, \quad B \triangleq \{11, 00\}^n \times \{\}\]  \(\text{(4)}\)

are crucial for understanding the Theorem. They immediately yield

* Work done while Visiting Professor at ISL, Stanford.
† Work partially supported by DARPA under Contract MDA-0680 and by U.S. Air Force under Contract F49620-79C-0058.
‡ Work partially supported by NSF under Contract 80-26102.

0012-365X/84/$3.00 © 1984, Elsevier Science Publishers B.V. (North-Holland)
Lemma 1.

\[ M(2n, n) \geq 2^{2n}, \quad M(2n + 1, n) \geq 2^{2n} \quad (n = 1, 2, \ldots) \]

Thus only the inequality

\[ \max_{0 \leq \delta \leq m} M(m, \delta) \leq 2^{2n} \tag{I} \]

remains to be proved.

First we show that this inequality follows from either one of the following two propositions. Actually, these derivations establish also their equivalence. The proofs for the propositions will be given subsequently.

**Proposition 1.** \( M(2n + 1, n) = M(2n, n), \ (n = 1, 2, \ldots) \).

**Proposition 2.** \( M(2n, n) \leq 2^{2n}, \ (n = 1, 2, \ldots) \).

2. Preliminaries

The operation \( \overline{\cdot} \) applied to a sequence denotes complementation, that is, component-wise exchange of 0’s and 1’s. When applied to a set of sequences, it is understood in the Minkowski sense. For ease of reference, a simple property of the Hamming distance function with respect to complementation is stated as

**Lemma 2.** (i) \( d(\overline{a}, b) = d(a, \overline{b}) = m, \ \forall a, b \in \{0, 1\}^m \),

(ii) \( (A, B) \in S^m_\delta \Rightarrow (A, B) \in S^m_{m-\delta} \),

(iii) \( M(m, \delta) = M(m, m-\delta) \).

We also adopt the following notation: For a set \( C \subseteq \{0, 1\}^m \) and \( \epsilon \in \{0, 1\} \), define

\[ C^\sharp_{\epsilon} \triangleq \{(c_1, \ldots, c_m) \in C : c_i = \epsilon\} \subseteq \{0, 1\}^m \tag{5} \]

\[ C^\sharp^*_{\epsilon} \triangleq \{(c_1, \ldots, c_{i-1}, \epsilon, c_{i+1}, \ldots, c_m) : (c_1, \ldots, c_m) \in C\} \subseteq \{0, 1\}^{m-1} \tag{6} \]

Analogously, for two components \( s, t \) we define \( C^\sharp_{rs} \subseteq \{0, 1\}^m \) and \( C^\sharp^*_{r\epsilon} \subseteq \{0, 1\}^{m-1} \).

3. Proposition 1 \( \Rightarrow \) Proposition 2

We proceed by induction in \( n \). The case \( n = 1 \) is settled by inspection. Now for \( (A, B) \in S_{n+1}^{2(n+1)} \), clearly

\[ A = A^1_1 \cup A^1_0, \quad B = B^1_1 \cup B^1_0 \tag{7} \]
and for $\varepsilon \in \{0, 1\}$
\begin{align}
|A^*_{\varepsilon}\rangle &= |A_{\varepsilon}\rangle, \\
|B^*_{\varepsilon}\rangle &= |B_{\varepsilon}\rangle; \\
|A^*_{\varepsilon}\rangle|B^*_{\varepsilon}\rangle &\leq M(2n + 1, n + 1); \\
|A^*_{\varepsilon}\rangle|B^*_{\varepsilon}\rangle &\leq M(2n + 1, n). 
\end{align}

Since by Lemma 2, $M(2n + 1, n + 1) = M(2n + 1, n)$, the relations (7)–(10) and Proposition 1 imply
\begin{equation}
|A| |B| = \sum_{\varepsilon, n \in \{0, 1\}} |A_{\varepsilon}\rangle|B_n\rangle \leq 4M(2n, n).
\end{equation}

Since $M(2n, n) \leq 2^{2n}$ by hypothesis, thus $M(2(n + 1), n + 1) \leq 4(2^{2n}) = 2^{2(n + 1)}$. \qedsymbol

4. **Proposition 2 $\Rightarrow$ (I)**

Case $m = 2n$. For $(A, B) \in S^2_{2n}$, consider $(A \times \bar{A}, B \times \bar{B})$. By Lemma 2, this is an element of $S^2_{2n}$. Therefore, by Proposition 2, $|A \times \bar{A}| |B \times \bar{B}| = (|A| |B|)^2 \leq 2^{4n}$ and hence $M(2n, \delta) \leq M(2n, n)$.

Case $m = 2n + 1$. Since $d(\bar{a}, b) + d(a, b) = m$ and $m$ is odd, necessarily $d(\bar{a}, b) \neq d(a, b)$ and thus for $(A, B) \in S^m_n$ also $A \cap \bar{A} = \emptyset$, $B \cap \bar{B} = \emptyset$. By Lemma 2,
\begin{equation}
(C, D) \triangleq (A \times \bar{A} \cup \bar{A} \times A, B \times \bar{B} \cup \bar{B} \times B) \in S^2_{2n+1}
\end{equation}
and thus by Proposition 2 $|C| |D| = 4(|A| |B|)^2 \leq 2^{2(2n+1)}$, which gives $|A| |B| \leq 2^{2n}$. Hence, we have derived (I) and in conjunction with Lemma 1, Proposition 1 is also proved.

5. **Proof of Proposition 1**

The proof is based on two key observations. For any $(A, B) \in S^2_{n+1}$:
\begin{equation}
\forall t \in \{1, \ldots, 2n + 1\} \quad \text{if} \quad A^*_{\varepsilon t} \cap \bar{A}^*_{\varepsilon t} \neq \emptyset \quad \text{then} \quad B^*_{\varepsilon t} = \emptyset. \tag{OI}
\end{equation}

Clearly for $a, \bar{a} \in A^*_{\varepsilon t}$ and $b \in B^*_{\varepsilon t}$ we have $d(a, b) = d(\bar{a}, b) = n - 1$ in contradiction to $d(\bar{a}, b) = 2n - d(a, b) = n + 1$.

\begin{equation}
\exists t \in \{1, \ldots, 2n + 1\}: |A_{\varepsilon}\rangle|B^*_{t}\rangle + |A^*_{\varepsilon}\rangle|B_{t}\rangle \geq |A_{\varepsilon}\rangle|B_{0}\rangle + |A^*_{\varepsilon}\rangle|B^*_{0}\rangle. \tag{OII}
\end{equation}

For this just notice that $\sum_{t=1}^{2n+1} (|A_{\varepsilon}\rangle|B^*_{t}\rangle + |A^*_{\varepsilon}\rangle|B_{t}\rangle)$ counts the number of identical components for all pairs of sequences $(a, b) \in A \times B$ and therefore equals $(n + 1) |A| |B|$. On the other hand $\sum_{t=1}^{2n+1} (|A_{\varepsilon}\rangle|B_{0}\rangle + |A^*_{\varepsilon}\rangle|B^*_{0}\rangle)$ counts the number of distinct components for all pairs of sequences $(a, b) \in A \times B$ and therefore equals $n |A| |B|$, a smaller quantity. The Pigeon Hole Principle gives (OII).
We can assume without loss of generality that \( t = 2n + 1 \) and omit the index \( t \). Notice that \((A_1, B_1)\) and \((A_0, B_0)\) are \((2n, n)\)-systems.

**Case 1.** \( \exists \varepsilon \in \{0, 1\} \) with \( A_\varepsilon^* \cap A_{\bar{\varepsilon}}^* \neq \emptyset \) and \( B_\varepsilon^* \cap B_{\bar{\varepsilon}}^* \neq \emptyset \). By (OI), \( A_\varepsilon^* = B_{\bar{\varepsilon}}^* = \emptyset \) and thus \( |A| \leq |A_\varepsilon^*| |B_\varepsilon^*| \leq M(2n, n) \).

**Case 2.** \( \exists \varepsilon \in \{0, 1\} \) with \( A_\varepsilon^* \cap A_{\bar{\varepsilon}}^* \neq \emptyset \) and \( B_\varepsilon^* \cap B_{\bar{\varepsilon}}^* = \emptyset \) (resp. vice versa). By (OI) \( B_{\bar{\varepsilon}}^* = \emptyset \) and thus \( |A| \leq |A_\varepsilon^*| |B_\varepsilon^*| + |A_{\bar{\varepsilon}}^*| |B_{\bar{\varepsilon}}^*| \leq 2 |A_\varepsilon^*| |B_\varepsilon^*| \) (by OII)). Replace now \( B_\varepsilon^* \) by \( D \triangleq B_\varepsilon^* \cup B_{\bar{\varepsilon}}^* \). Since \((A_\varepsilon^*, D) \in S_{2n}^2 \), we get again \( |A| \leq |A_\varepsilon^*| |D| \leq M(2n, n) \).

**Case 3.** \( \forall \varepsilon \in \{0, 1\} \): \( A_\varepsilon^* \cap A_{\bar{\varepsilon}}^* = \emptyset \), \( B_\varepsilon^* \cap B_{\bar{\varepsilon}}^* = \emptyset \). Choose now \( \varepsilon \) such that \( |A_\varepsilon^*| |B_{\bar{\varepsilon}}^*| > |A_{\bar{\varepsilon}}^*| |B_\varepsilon^*| \) and define \( C \triangleq A_\varepsilon^* \cup A_{\bar{\varepsilon}}^* \), \( D \triangleq B_\varepsilon^* \cup B_{\bar{\varepsilon}}^* \). Now \((C, D) \in S_{2n}^2 \) and
\[
|A| \leq |A_\varepsilon^*| |B_{\bar{\varepsilon}}^*| + |A_{\bar{\varepsilon}}^*| |B_\varepsilon^*| \leq 2(|A_{\bar{\varepsilon}}^*| |B_{\bar{\varepsilon}}^*| + |A_\varepsilon^*| |B_\varepsilon^*|) \quad \text{(by OII)}
\leq 4 |A_\varepsilon^*| |B_{\bar{\varepsilon}}^*| \quad \text{(by choice of} \ \varepsilon \text{)}
= |C| |D| \leq M(2n, n).
\]

6. Proof of Proposition 2

Again the proof is based on two observations.

If \((A, B) \in S_{2n}^2\), then \((\bar{A}, \bar{B}), (A, \bar{B}), (\bar{A}, B) \in S_{2n}^2\) and also \((A \cup \bar{A}, B \cup \bar{B}) \in S_{2n}^2\). We can therefore assume \( A = \bar{A}, B = \bar{B} \) and thus
\[
|A_t^\varepsilon| = |A_t^{\bar{\varepsilon}}| = \frac{1}{2} |A| \quad \text{and} \quad |B_t^\varepsilon| = |B_t^{\bar{\varepsilon}}| = \frac{1}{2} |B| \quad (1 \leq t \leq 2n, \varepsilon \in \{0, 1\}).
\]

(OIII)

Further
\[
\exists t \in \{2, 3, \ldots, 2n\} \quad \text{and} \quad \exists \eta \in \{0, 1\} \colon |A_1 \eta| |A|^{-1} \geq \frac{1}{2}
\]
\[
\text{and} \quad |B_1 \eta| |B|^{-1} \geq \frac{1}{2}. \quad \text{(OIV)}
\]

For this, notice that \( \sum_{t=1}^{2n} |A_t^1| |B_t^1| + |A_t^0| |B_t^0| \) counts the number of distinct components for all pairs of sequences \((a, b) \in A_t^1 \times B_t^1\) and therefore equals
\[
|A_1^1| |A|^{-1} \geq \frac{1}{2}
\]
and
\[
|B_1^1| |B|^{-1} \geq \frac{1}{2}. \quad \text{(OIV)}
\]

Now again we distinguish among three cases.

**Case 1.** \( A_{1n}^\varepsilon \cap A_{\bar{1n}}^{1\bar{\varepsilon}} \neq \emptyset \) and \( B_{1n}^\varepsilon \cap B_{\bar{1n}}^{1\bar{\varepsilon}} \neq \emptyset \). Notice that by the distance properties necessarily, \( A_{1n}^\varepsilon = \emptyset \) and \( B_{1n}^{1\bar{\varepsilon}} = \emptyset \). By (OIII) therefore
\[
|A| \leq 4 |A_{1n}^\varepsilon| |B_{1n}^{1\bar{\varepsilon}}| \leq 4M(2(n-1), n-1).
\]

**Case 2.** \( A_{1n}^\varepsilon \cap A_{\bar{1n}}^{1\bar{\varepsilon}} \neq \emptyset \) and \( B_{1n}^\varepsilon \cap B_{\bar{1n}}^{1\bar{\varepsilon}} = \emptyset \). By the previous argument necessarily, \( B_{1n}^{1\bar{\varepsilon}} = \emptyset \). Define now \( C \triangleq A_{1n}^\varepsilon \) and \( D \triangleq B_{1n}^\varepsilon \cup B_{\bar{1n}}^{1\bar{\varepsilon}} \). Since \( (a, b) = (d, \bar{d}) \) and obviously \( (a, \bar{d}) = (d, b) \), also \( d(a, \bar{d}) = d(d, b) \). Thus \((C, D) \in S_{2n-2}^2\) and, by (OIII) and (OIV), \( |A| \leq 4 |C| |D| \leq 4M(2n - 2, n-1) \).

**Case 3.** \( A_{0n}^\varepsilon \cap A_{\bar{0n}}^{1\bar{\varepsilon}} = \emptyset \) and \( B_{0n}^\varepsilon \cap B_{\bar{0n}}^{1\bar{\varepsilon}} = \emptyset \). Since \( \bar{A} = A \) and \( \bar{B} = B \) we have now \( A_{0n}^\varepsilon = A_{\bar{0n}}^{1\bar{\varepsilon}} \), \( B_{0n}^\varepsilon = B_{\bar{0n}}^{1\bar{\varepsilon}} \). Furthermore, since \( d(1\eta, 1\bar{\eta}) = d(0\bar{\eta}, 0\eta) = 1 \), for
$C \triangleq A_{1,n}^{\ast 11} \cup A_{1,n}^{\ast 11}$, $D \triangleq B_{1,n}^{\ast 11} \cup B_{1,n}^{\ast 11}$ we have $(C, D) \in S_{n-1}^{2^{n-2}}$ and thus again, by (OIII) and (OIV), $|A||B| \leq 4 |C||D| \leq 4M(2n-2, n-1)$.

By the inductive hypothesis for $n-1$, $M(2n-2, n-1) \leq 2^{2(n-1)}$ and hence $M(2n, n) \leq 4(2^{2(n-1)}) = 2^{2n}$.

References