Properties of Binary Pearson Codes

Weber, Jos; Schouhamer Immink, KA

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Properties of Binary Pearson Codes

Jos H. Weber
Delft University of Technology
Delft, The Netherlands
Email: j.h.weber@tudelft.nl

Kees A. Schouhamer Immink
Turing Machines Inc.
Rotterdam, The Netherlands
Email: immink@turing-machines.com

Abstract—We consider the transmission and storage of data that use coded symbols over a channel, where a Pearson-distance-based detector is used for achieving resilience against unknown channel gain and offset, and corruption with additive noise. We discuss properties of binary Pearson codes, such as the Pearson noise distance that plays a key role in the error performance of Pearson-distance-based detection. We also compare the Pearson noise distance to the well-known Hamming distance, since the latter plays a similar role in the error performance of Euclidean-distance-based detection.

I. INTRODUCTION

In mass data storage devices, the user data are translated into physical features that can be either electronic, magnetic, optical, or of other nature. Due to process variations, the magnitude of the physical effect may deviate from the nominal values, which may effect the reliable read-out of the data. For example, over a long time, the charge in memory cells, which represents the stored data, may fade away, and as a result the main physical parameters change resulting in channel mismatch and increased error rate. It has been found that such *retention* errors to be dominant errors in solid-state memories.

The detector's ignorance of the exact value of the channel's main physical parameters [1], [2], [3], [4], a phenomenon called *channel mismatch*, may seriously degrade the error performance of a storage or transmission medium [5]. Researchers have been seeking for methods that can withstand channel mismatch. Immink and Weber [5] advocated a novel data detection method based on the *Pearson distance* that offers invariance, "immunity", to offset and gain mismatch. They also showed the other side of the medal of Pearson-distance-based detection, namely that it is less resilient to additive noise than conventional Euclidean-distance-based detection.

In this paper, we investigate the relationship between the noise resilience of Euclidean distance versus Pearson distance detection. The outline of the paper is as follows. In Section II, we present the channel model under consideration, we re-capitulate the relevant prior art of minimum Euclidean and Pearson distance detection, and we review the definition of Pearson codes. In Section III, we discuss properties of binary Pearson codes, such as lower and upper bounds to the error performance of detectors based on the Pearson distance, and the difference in noise resilience of minimum Euclidean versus minimum Pearson distance detection. Section IV concludes our paper.

II. BACKGROUND AND PRELIMINARIES

In this section we present some prior art, mainly from [5], and set the scene for the results of this paper.

A. Pearson distance

We start with the definition of two quantities of an n-vector of reals, $\boldsymbol{z}$, namely the average of $\boldsymbol{z}$ by $\overline{z} = \frac{1}{n} \sum_{i=1}^{n} z_i$ and the (unnormalized) variance of $\boldsymbol{z}$ by

$$\sigma_z^2 = \sum_{i=1}^{n} (z_i - \overline{z})^2. \tag{1}$$

The *Pearson distance* between the vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ in $\mathbb{R}^n$ is defined by

$$\delta_{\mathrm{p}}(\boldsymbol{x}, \boldsymbol{y}) = 1 - \rho_{\boldsymbol{x},\boldsymbol{y}}, \tag{2}$$

where the *(Pearson) correlation coefficient* [6] is defined by

$$\rho_{\boldsymbol{x},\boldsymbol{y}} = \frac{\sum_{i=1}^{n} (x_i - \overline{x})(y_i - \overline{y})}{\sigma_x \sigma_y}. \tag{3}$$

It is immediate that the Pearson distance $\delta_{\mathrm{p}}(\boldsymbol{x}, \boldsymbol{y})$ is undefined if $\boldsymbol{x}$ or $\boldsymbol{y}$ has variance zero, i.e., is a 'constant' vector $(c, c, \ldots, c)$ with $c \in \mathbb{R}$. Further, note that the Pearson distance is not a regular *metric*, but a measure of similarity between the vectors $\boldsymbol{x}$ and $\boldsymbol{y}$. It can easily be verified that the *triangle inequality* condition, $\delta_{\mathrm{p}}(\boldsymbol{x}, \boldsymbol{z}) \leq \delta_{\mathrm{p}}(\boldsymbol{x}, \boldsymbol{y}) + \delta_{\mathrm{p}}(\boldsymbol{y}, \boldsymbol{z})$, is not always satisfied. For example. let $n = 4$ and let $\boldsymbol{x} = (0001)$, $\boldsymbol{y} = (0011)$, and $\boldsymbol{z} = (0010)$, then $\delta_{\mathrm{p}}(\boldsymbol{x}, \boldsymbol{z}) = 1.3333$, $\delta_{\mathrm{p}}(\boldsymbol{x}, \boldsymbol{y}) = 0.4226$, and $\delta_{\mathrm{p}}(\boldsymbol{y}, \boldsymbol{z}) = 0.4226$.

B. Channel model

We assume a simple linear channel model where the sent codeword $\boldsymbol{c}$, taken from a finite codebook $\mathcal{S} \subset \mathbb{R}^n$, is received as the real-valued vector

$$\boldsymbol{r} = a(\boldsymbol{c} + \boldsymbol{\nu}) + b\mathbf{1}, \tag{4}$$

where $\mathbf{1}$ is the all-one vector $(1, 1, \ldots, 1)$ of length n, while a, $a > 0$, is an unknown gain, $b \in \mathbb{R}$ is an unknown offset, and $\boldsymbol{\nu} = (\nu_1, \ldots, \nu_n)$ is additive noise with $\nu_i \in \mathbb{R}$ being zero-mean independent and identically distributed (i.i.d) noise samples with Gaussian distribution $\mathcal{N}(0, \sigma^2)$, where $\sigma^2 \in \mathbb{R}$ denotes the variance. We assume that the parameters a and b vary slowly, so that during the transmission of the n symbols in a codeword the parameters a and b are fixed, but that these values may be different for the next transmitted codeword.

C. Detection

A *minimum Pearson distance detector* outputs a codeword according to the *minimum distance decision rule*

$$\boldsymbol{c}_{\mathrm{p}} = \arg\min_{\hat{\boldsymbol{c}} \in \mathcal{S}} \delta_{\mathrm{p}}(\boldsymbol{r}, \hat{\boldsymbol{c}}). \qquad (5)$$

Due to the properties of the Pearson correlation coefficient such a detector is immune to gain and offset mismatch [5]. However, it is more sensitive to noise than the well-known *minimum Euclidean distance detector* which outputs

$$\boldsymbol{c}_{\mathrm{e}} = \arg\min_{\hat{\boldsymbol{c}} \in \mathcal{S}} \delta^2(\boldsymbol{r}, \hat{\boldsymbol{c}}), \qquad (6)$$

with

$$\delta^2(\boldsymbol{x}, \boldsymbol{y}) = \sum_{i=1}^{n} (x_i - y_i)^2 \qquad (7)$$

being the squared Euclidean distance between $\boldsymbol{x}$ and $\boldsymbol{y}$.

The computation of the probability that a minimum Pearson distance detector errs has been investigated in [5]. A principal finding is that it is not the (minimum) Pearson distance, $\delta_{\mathrm{p}}(\boldsymbol{x}, \boldsymbol{y})$, between codewords $\boldsymbol{x}$ and $\boldsymbol{y}$, that governs the error probability, but a quantity called *Pearson noise distance*, which is denoted by $d(\boldsymbol{x}, \boldsymbol{y})$. The squared Pearson noise distance, $d^2(\boldsymbol{x}, \boldsymbol{y})$, between the vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ is given by

$$d^2(\boldsymbol{x}, \boldsymbol{y}) = 2\sigma_x^2 \delta_{\mathrm{p}}(\boldsymbol{x}, \boldsymbol{y}) = 2\sigma_x^2 (1 - \rho_{\boldsymbol{x}, \boldsymbol{y}}). \qquad (8)$$

The union bound estimate of the word error rate (WER) is

$$\mathrm{WER}_{\mathrm{Pear}} \le \frac{1}{|\mathcal{S}|} \sum_{\boldsymbol{x} \in \mathcal{S}} \sum_{\boldsymbol{y} \in \mathcal{S}, \boldsymbol{x} \ne \boldsymbol{y}} Q\left(\frac{d(\boldsymbol{x}, \boldsymbol{y})}{2\sigma}\right), \qquad (9)$$

where $Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-\frac{u^2}{2}} du$ is the well-known Q-function. Note the similarity with the union bound for the WER in case of a Euclidean detector, which reads

$$\mathrm{WER}_{\mathrm{Eucl}} \le \frac{1}{|\mathcal{S}|} \sum_{\boldsymbol{x} \in \mathcal{S}} \sum_{\boldsymbol{y} \in \mathcal{S}, \boldsymbol{x} \ne \boldsymbol{y}} Q\left(\frac{\delta(\boldsymbol{x}, \boldsymbol{y})}{2\sigma}\right) \qquad (10)$$

for an additive Gaussian noise channel, i.e., a channel as in (4) with the values of a and b being known to the receiver. We emphasize again that for WER performance of the Pearson distance detector it does not matter whether the gain and offset values are known to the receiver or not, while the performance of the Euclidean distance detector quickly deteriorates when the gain and offset drift away from their ideal values $a = 1$ and $b = 0$ while being unknown to the receiver [5].

For small σ, the WER is dominated by the term with the smallest distance between any $\boldsymbol{x}$ and any different $\boldsymbol{y} \in \mathcal{S}$, so that

$$\mathrm{WER}_{\mathrm{Pear}} \approx N_{p,\mathrm{min}} Q\left(\frac{d_{\mathrm{min}}}{2\sigma}\right), \quad \sigma \ll 1, \qquad (11)$$

where $d_{\mathrm{min}} = \min_{\boldsymbol{x}, \boldsymbol{y} \in \mathcal{S}, \boldsymbol{x} \ne \boldsymbol{y}} d(\boldsymbol{x}, \boldsymbol{y})$ and $N_{p,\mathrm{min}}$ is the number of codewords $\boldsymbol{y}$ (called *nearest neighbors*) at minimum Pearson noise distance d_{min} from $\boldsymbol{x}$, averaged over all $\boldsymbol{x} \in \mathcal{S}$.

D. Pearson codes

In order to allow easy encoding and decoding operations, it is common to use a q-ary codebook $\mathcal{S}$, i.e., $S \subseteq \mathcal{Q}^n$ with $\mathcal{Q} = \{0, 1, \ldots, q - 1\}$. Since a minimum Pearson distance detector cannot deal with codewords $\boldsymbol{c}$ with $\sigma_c = 0$ and cannot distinguish between the words $\boldsymbol{c}$ and $c_1 \boldsymbol{1} + c_2 \boldsymbol{c}$, $c_2 > 0$, well-chosen words must be barred from $\mathcal{Q}^n$ to guarantee unambiguous detection. Weber *et al.* [7] coined the name *Pearson code* for a set of codewords that can be uniquely decoded by a minimum Pearson distance detector. Codewords in a Pearson code $\mathcal{S}$ satisfy two conditions, namely

- *Property A:* If $\boldsymbol{c} \in \mathcal{S}$ then $c_1 \boldsymbol{1} + c_2 \boldsymbol{c} \notin \mathcal{S}$ for all $c_1, c_2 \in \mathbb{R}$ with $(c_1, c_2) \ne (0, 1)$ and $c_2 > 0$;
- *Property B:* $c\boldsymbol{1} \notin \mathcal{S}$ for all $c \in \mathbb{R}$.

For a binary Pearson code, i.e., $q = 2$, this implies that only two vectors must be barred, namely the all-'0' vector $\boldsymbol{0}$ and all-'1' vector $\boldsymbol{1}$. Hence, the largest binary Pearson code of length n is

$$\mathcal{P}_n = \{0, 1\}^n \setminus \{\boldsymbol{0}, \boldsymbol{1}\}. \qquad (12)$$

However, in order to improve the error performance, it may be necessary to further restrict the codebook, particularly by avoiding codeword pairs with a small Pearson noise distance. In the next section we investigate properties of the Pearson (noise) distance and detector that provide more insight and as such could be useful in the process of designing good Pearson codes.

III. PROPERTIES OF BINARY PEARSON CODES

In this section, we study the important binary case, $q = 2$. Particularly, we will determine bounds on the Pearson noise distance and make comparisons with the Hamming distance.

First we give some notation. Let $\boldsymbol{x}$ and $\boldsymbol{y}$ be two n-vectors taken from the code $S \subset \{0, 1\}^n$. We define the integers

$$w_x = \sum_{i=1}^{n} x_i, \quad w_y = \sum_{i=1}^{n} y_i, \quad w_{xy} = \sum_{i=1}^{n} x_i y_i, \qquad (13)$$

where w_x and w_y are the *weights* of the vectors $\boldsymbol{x}$ and $\boldsymbol{y}$, respectively, and w_{xy}, the *index of '1'-coincidence* (or *overlap*) of the vectors $\boldsymbol{x}$ and $\boldsymbol{y}$, denotes the number of indices i where $x_i = y_i = 1$. Note that all additions and multiplications in (13) are over the real numbers.

For clerical convenience, we define the real-valued function

$$\varphi_n(w_x, w_y, w_{xy}) = d(\boldsymbol{x}, \boldsymbol{y}). \qquad (14)$$

Using (8) and the above definitions, we have

$$\varphi_n^2(w_x, w_y, w_{xy}) = 2\sigma_x^2 \left(1 - \frac{w_{xy} - \frac{w_x w_y}{n}}{\sigma_x \sigma_y}\right), \qquad (15)$$

where

$$\sigma_x^2 = w_x - \frac{w_x^2}{n} \text{ and } \sigma_y^2 = w_y - \frac{w_y^2}{n}. \qquad (16)$$

For all $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{P}_n$, the integer variables w_x, w_y, and w_{xy} satisfy

$$1 \le w_x, w_y \le n - 1, \qquad (17)$$

$$\max\{w_x + w_y - n, 0\} \le w_{xy} \le \min\{w_x, w_y\}, \text{ and} \quad (18)$$

$$w_{xy} \le w_x - 1 \text{ if } \boldsymbol{x} \ne \boldsymbol{y} \text{ and } w_x = w_y. \quad (19)$$

In the next subsections we present the main results of this paper.

A. Bounds on the Pearson noise distance

Since the Pearson noise distance $d(\boldsymbol{x}, \boldsymbol{y})$ plays a crucial role in the performance of a Pearson code, we should investigate which values it can take. We start with a simple upper bound.

Theorem 1: For any two codewords $\boldsymbol{x}$ and $\boldsymbol{y}$ in $\mathcal{P}_n$, $n \ge 2$, it holds that

$$d^2(\boldsymbol{x}, \boldsymbol{y}) \le 4\sigma_x^2 \le \begin{cases} n & \text{if } n \text{ is even,} \\ n - 1/n & \text{if } n \text{ is odd,} \end{cases}$$

where equality holds in the first inequality if and only if $\boldsymbol{y} = \boldsymbol{1} - \boldsymbol{x}$, while equality holds in the second inequality if and only if $w_x = \lfloor n/2 \rfloor$ or $w_x = \lceil n/2 \rceil$.

Proof. It is a well-known property of the Pearson correlation coefficient, $\rho_{\boldsymbol{u}, \boldsymbol{v}}$, of any two real-valued non-constant vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ of the same length, that $|\rho_{\boldsymbol{u}, \boldsymbol{v}}| \le 1$ and also that $\rho_{\boldsymbol{u}, \boldsymbol{v}} = -1$ if and only if $\boldsymbol{v} = c_1 \boldsymbol{1} + c_2 \boldsymbol{u}$, where the coefficients c_1 and c_2, $c_2 < 0$, are real numbers [6, Sec. IV.4.6]. Hence, for any $\boldsymbol{x} \in \mathcal{P}_n$, $d^2(\boldsymbol{x}, \boldsymbol{y})$ is maximized over all $\boldsymbol{y} \in \mathcal{P}_n$ if and only if $\boldsymbol{y} = \boldsymbol{1} - \boldsymbol{x}$, i.e., by setting $\boldsymbol{y}$ as the inverse of $\boldsymbol{x}$. The results as stated in the theorem now easily follow by observing that

$$d^2(\boldsymbol{x}, \boldsymbol{1} - \boldsymbol{x}) = \varphi_n^2(w_x, n - w_x, 0) = 4\sigma_x^2 = 4\left(w_x - \frac{w_x^2}{n}\right)$$

and that the last expression is maximized if and only if $w_x = \lfloor \frac{n}{2} \rfloor$ or $w_x = \lceil \frac{n}{2} \rceil$. ∎

In case two codewords have equal weight, we have the following useful observation.

Lemma 1: For any two codewords $\boldsymbol{x}$ and $\boldsymbol{y}$ in $\mathcal{P}_n$, $n \ge 2$, of equal weight, it holds that

$$d^2(\boldsymbol{x}, \boldsymbol{y}) = 2(w_x - w_{xy}).$$

Proof. From (14)-(16) and the fact $w_x = w_y$ it follows that

$$
\begin{aligned}
d^2(\boldsymbol{x}, \boldsymbol{y}) &= 2\sigma_x^2\left(1 - \frac{w_{xy} - \frac{w_x^2}{n}}{\sigma_x^2}\right) = 2\left(\sigma_x^2 - w_{xy} + \frac{w_x^2}{n}\right) \\
&= 2(\sigma_x^2 - w_{xy} + w_x - \sigma_x^2) = 2(w_x - w_{xy}),
\end{aligned}
$$

which shows the stated result. ∎

The minimum Pearson noise distance, $d_{\min}$, between any two different codewords plays a key role in the evaluation of the error performance of the minimum Pearson detector, see (11). The next theorem shows that $d_{\min}$ of $\mathcal{P}_n$ equals $\varphi_n(1, 2, 1)$. This was already conjectured in [5], but is now formally proved.

Theorem 2: For any two different codewords $\boldsymbol{x}$ and $\boldsymbol{y}$ in $\mathcal{P}_n$, $n \ge 3$, it holds that

$$d^2(\boldsymbol{x}, \boldsymbol{y}) \ge \varphi_n^2(1, 2, 1) = \frac{2n-2}{n}\left(1 - \sqrt{\frac{n-2}{2n-2}}\right),$$

where equality holds if and only if $w_x = w_{xy} = 1$, $w_y = 2$ or $w_x = n - 1$, $w_y = w_{xy} = n - 2$.

Proof. Our strategy is to look for three integers, w_x, w_y, and w_{xy}, that minimize the function $\varphi_n(w_x, w_y, w_{xy})$, under the constraints (17)-(19). Any two different codewords $\boldsymbol{x}$ and $\boldsymbol{y}$ having the found parameters will then minimize $d(\boldsymbol{x}, \boldsymbol{y})$. Since $\rho_{\boldsymbol{x}, \boldsymbol{y}} = \rho_{\boldsymbol{y}, \boldsymbol{x}}$, it follows from (8) that it holds for such $\boldsymbol{x}$ and $\boldsymbol{y}$ that $\sigma_x^2 \le \sigma_y^2$, i.e., $w_x \le w_y \le n - w_x$. Further, we may and will assume $w_x \le n/2$ since

$$d(\boldsymbol{x}, \boldsymbol{y}) = d(\boldsymbol{1} - \boldsymbol{x}, \boldsymbol{1} - \boldsymbol{y}) \quad (20)$$

for all $\boldsymbol{x}$ and $\boldsymbol{y}$ in $\mathcal{P}_n$.

With regard to the selection of the integer w_{xy}, it is straightforward from (15) that we should choose it as large as possible for any values of w_x and w_y. We distinguish between the cases $w_x = w_y$ and $w_x < w_y$.

In case $w_x = w_y$, the value of w_{xy} is at most $w_x - 1$ since $\boldsymbol{x} \ne \boldsymbol{y}$. Hence, from Lemma 1, we find $d^2(\boldsymbol{x}, \boldsymbol{y}) = 2(w_x - w_{xy}) \ge 2$. Note that the expression in the theorem is clearly smaller than 2.

In case $w_x < w_y$, the maximum value of w_{xy} is w_x. Note that $w_x < w_y$ implies that $1 \le w_x \le \lfloor (n-1)/2 \rfloor$. We proceed with the selection of w_y. From (14)-(16), we have

$$\varphi_n^2(w_x, w_y, w_x) = 2\sigma_x^2(1 - \alpha), \quad (21)$$

where

$$\alpha^2 = \frac{\left(w_x - \frac{w_x w_y}{n}\right)^2}{\sigma_x^2 \sigma_y^2} = \left(\frac{1}{w_y} - \frac{1}{n}\right)\frac{w_x^2}{\sigma_x^2}, \alpha > 0. \quad (22)$$

It is immediate from (21) and (22) that, for any value of w_x, the function $\varphi_n(w_x, w_y, w_x)$ is at a minimum when the factor $\frac{1}{w_y} - \frac{1}{n}$ is at a maximum. We conclude that, for all w_x, the choice $w_y = w_x + 1$ minimizes (21). Subsequently, we substitute $w_y = w_x + 1$, and analyze the function

$$\psi_n(w_x) = \varphi_n^2(w_x, w_x + 1, w_x) = 2\sigma_x^2(1 - \beta) \quad (23)$$

in the single (integer) variable, w_x, where, using (22), we write

$$\beta^2 = \left(\frac{1}{w_x + 1} - \frac{1}{n}\right)\frac{w_x^2}{\sigma_x^2} = \frac{w_x(n - 1 - w_x)}{(w_x + 1)(n - w_x)}, \beta > 0. \quad (24)$$

In order to determine the value of $w_x \in \{1, 2, \ldots, \lfloor (n-1)/2 \rfloor\}$ minimizing $\psi_n(w_x)$, we consider the function $f_n(w)$ which is obtained by replacing the discrete variable w_x in $\psi_n(w_x)$ by the continuous variable w, with $w \in [1, \lfloor (n-1)/2 \rfloor]$. We replace w_x by w in (24) as well and then express w in β, obtaining

$$w = \frac{n-1}{2} - \frac{n}{2}\sqrt{g_n(\beta)}, \quad (25)$$

where

$$g_n(\beta) = 1 + \frac{1}{n^2} + \frac{2(\beta^2 + 1)}{n(\beta^2 - 1)}, \quad 0 < \beta_0 \le \beta \le \beta_1 < 1,$$
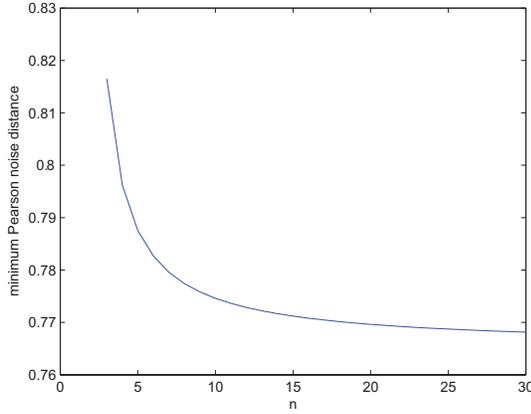
$$\beta_0^2 = \beta^2|_{w=1} = \frac{n-2}{2n-2},$$

Fig. 1. Minimum Pearson noise distance of $\mathcal{P}_n$.

and

$$\beta_1^2 = \beta^2|_{w=\lfloor\frac{n-1}{2}\rfloor} = \frac{\lfloor\frac{n-1}{2}\rfloor\lceil\frac{n-1}{2}\rceil}{\lfloor\frac{n+1}{2}\rfloor\lceil\frac{n+1}{2}\rceil}.$$

Note that $g_n(\beta)$ is strictly decreasing with β on the interval under consideration, and thus w is a strictly increasing function of β. Next, we substitute (25) in $f_n(w)$, and the resulting function with variable β is

$$\begin{aligned} h_n(\beta) &= \left(\frac{n^2-1}{2n} - \frac{ng(\beta)}{2} - \sqrt{g(\beta)}\right)(1-\beta) \\ &= \frac{\beta-1}{n} + \frac{\beta^2+1}{\beta+1} + (\beta-1)\sqrt{g(\beta)}. \end{aligned} \quad (26)$$

It is not hard to show that the three terms in (26) are all strictly increasing with β in the range $\beta_0 \leq \beta \leq \beta_1$, so that $h_n(\beta)$ is at a minimum for $\beta = \beta_0$. Thus, $\varphi_n(w_x, w_y, w_x)$ achieves a minimum when $w_x = w_{xy} = 1$ and $w_y = 2$. The expression stated in the theorem follows by substituting these parameters into (15). Because of (20), also the choice $w_x = n - 1$ and $w_y = w_{xy} = n - 2$ achieves this minimum. Finally, it follows from the strict monotonicity of the functions used in the above derivation that no other choices achieve the minimum Pearson noise distance. ∎

Note that it follows from this theorem that, for large values of n, the minimum Pearson noise distance of the code $\mathcal{P}_n$ approaches $\sqrt{2-\sqrt{2}} \approx 0.765$. A graphical representation is provided in Figure 1.

We conclude this subsection with a look at the number of codeword pairs having a certain Pearson noise distance between each other. In this respect, note that the number of pairs $(\boldsymbol{x}, \boldsymbol{y})$ with given values for w_x, w_y, and w_{xy} is

$$\frac{n!}{w_{xy}!(w_x-w_{xy})!(w_y-w_{xy})!(n-w_x-w_y+w_{xy})!}, \quad (27)$$

which easily follows from standard combinatorial arguments. For example, it follows from this result and Theorem 2 that the number of codeword pairs $(\boldsymbol{x}, \boldsymbol{y})$ in $\mathcal{P}_n$ at minimum Pearson noise distance is $2 \times \frac{n!}{1!0!1!(n-2)!} = 2n(n-1)$. Hence, dividing

this expression by the number of codewords gives $N_{p,\min}$, which can be used, together with the minimum distance result from Theorem 2, in (11) to obtain an approximate value for the WER of a Pearson distance based detector.

B. Hamming versus squared Pearson noise distance

The *Hamming distance* between two vectors is an essential notion in coding theory, and a comparison between the properties of Hamming and Pearson distance is therefore relevant. Since $x_i, y_i \in \{0, 1\}$, the Hamming distance equals the squared Euclidean distance, i.e.,

$$d_{\mathrm{H}}(\boldsymbol{x}, \boldsymbol{y}) = \sum_{i=1}^{n}(x_i - y_i)^2 = w_x + w_y - 2w_{xy}. \quad (28)$$

It is essential that we define a fair yardstick for quantifying the noise resilience of minimum Euclidean and Pearson distance detection. To that end, we consider the ratio between the squared Pearson noise distance and the Hamming distance, denoted by $g_{\boldsymbol{x},\boldsymbol{y}}$, i.e.,

$$g_{\boldsymbol{x},\boldsymbol{y}} = \frac{d^2(\boldsymbol{x}, \boldsymbol{y})}{d_{\mathrm{H}}(\boldsymbol{x}, \boldsymbol{y})}. \quad (29)$$

It follows from the WER analysis in Subsection II-C that this ratio being smaller than one implies that the Euclidean detector is more resilient to noise than the Pearson detector in case $\boldsymbol{x}$ is transmitted and $\boldsymbol{y}$ is considered as an alternative for $\boldsymbol{x}$ in the decoding process. Vice versa for this ratio being larger than one.

As a first observation, note that it follows from Lemma 1 and (28) that $d^2(\boldsymbol{x}, \boldsymbol{y}) = d_{\mathrm{H}}(\boldsymbol{x}, \boldsymbol{y})$ and thus $g_{\boldsymbol{x},\boldsymbol{y}} = 1$ in case $\boldsymbol{x}$ and $\boldsymbol{y}$ are of equal weight. Evidently, there is no error performance difference between minimum Pearson and Euclidean detectors for codewords drawn from a constant weight set.

In the remainder of this subsection, we consider vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ from $\mathcal{P}_n$ with the weight of $\boldsymbol{x}$ being fixed at $w_x \in \{1, 2, \ldots, n-1\}$ and the Hamming distance $d_{\mathrm{H}}(\boldsymbol{x}, \boldsymbol{y})$ being fixed at $d_H \in \{1, 2, \ldots, n\}$. Since the overlap w_{xy} of $\boldsymbol{x}$ and $\boldsymbol{y}$ is expected to have a high impact on $g_{\boldsymbol{x},\boldsymbol{y}}$, we consider two extreme options for w_{xy} in our analysis: 1) we choose $w_y \in \{1, 2, \ldots, n-1\}$ such that w_{xy} is as small as possible, 2) we choose w_y such that w_{xy} is as large as possible, in both cases under the constraints of the fixed values for the weight of $\boldsymbol{x}$ and the Hamming distance between $\boldsymbol{x}$ and $\boldsymbol{y}$.

Case 1: It follows in a straightforward way that the minimal overlap of $\boldsymbol{x}$ and $\boldsymbol{y}$ is

$$w_{xy} = \begin{cases} w_x - d_{\mathrm{H}} & \text{if } 1 \leq d_{\mathrm{H}} \leq w_x - 1, \\ 1 & \text{if } d_{\mathrm{H}} = w_x, \\ 0 & \text{if } w_x + 1 \leq d_{\mathrm{H}} \leq n, \end{cases} \quad (30)$$

achieved for

$$w_y = \begin{cases} w_x - d_{\mathrm{H}} & \text{if } 1 \leq d_{\mathrm{H}} \leq w_x - 1, \\ 2 & \text{if } d_{\mathrm{H}} = w_x, \\ d_{\mathrm{H}} - w_x & \text{if } w_x + 1 \leq d_{\mathrm{H}} \leq n. \end{cases} \quad (31)$$
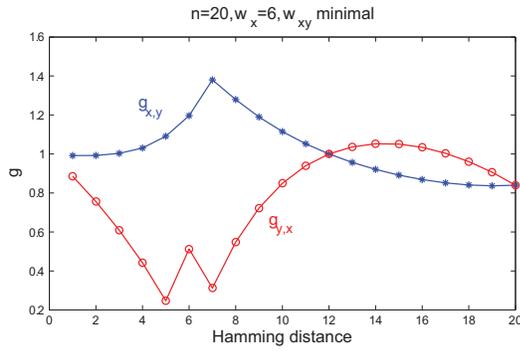
Fig. 2. The ratios $g_{\boldsymbol{x},\boldsymbol{y}}$ and $g_{\boldsymbol{y},\boldsymbol{x}}$ in case $n = 20$, $w_x = 6$, $1 \le d_{\mathrm{H}} \le 20$, and $w_y \in \{1, 2, \ldots, 19\}$ chosen such that w_{xy} is minimized (Case 1).

Fig. 3. The ratios $g_{\boldsymbol{x},\boldsymbol{y}}$ and $g_{\boldsymbol{y},\boldsymbol{x}}$ in case $n = 20$, $w_x = 6$, $1 \le d_{\mathrm{H}} \le 20$, and $w_y \in \{1, 2, \ldots, 19\}$ chosen such that w_{xy} is maximized (Case 2).
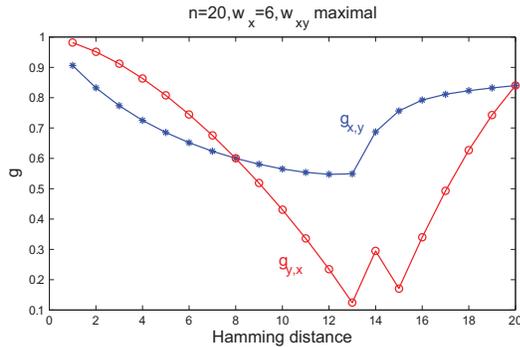
Case 2: Similarly, we have that the maximal overlap of $\boldsymbol{x}$ and $\boldsymbol{y}$ is

$$w_{xy} = \begin{cases} w_x & \text{if } 1 \le d_{\mathrm{H}} \le n - w_x - 1, \\ w_x - 1 & \text{if } d_{\mathrm{H}} = n - w_x, \\ n - d_{\mathrm{H}} & \text{if } n - w_x + 1 \le d_{\mathrm{H}} \le n, \end{cases} \quad (32)$$

achieved for

$$w_y = \begin{cases} w_x + d_{\mathrm{H}} & \text{if } 1 \le d_{\mathrm{H}} \le n - w_x - 1, \\ n - 2 & \text{if } d_{\mathrm{H}} = n - w_x, \\ 2n - d_{\mathrm{H}} - w_x & \text{if } n - w_x + 1 \le d_{\mathrm{H}} \le n. \end{cases} \quad (33)$$

The g-ratios can now be obtained from (29) by applying (14), (15), and (28). Figures 2 and 3 show, for Cases 1 and 2, respectively, the resulting $g_{\boldsymbol{x},\boldsymbol{y}}$ and $g_{\boldsymbol{y},\boldsymbol{x}}$ values for $n = 20$, $w_x = 6$, and $1 \le d_{\mathrm{H}} \le 20$.

Several interesting observations can be made from these figures. First of all, note that there are 'irregularities' for the $g_{\boldsymbol{y},\boldsymbol{x}}$ curves at $d_{\mathrm{H}} = w_x = 6$ (Case 1) and $d_{\mathrm{H}} = n - w_x = 14$ (Case 2). For Case 1 this can be explained as follows. From (30) we see that w_{xy} equals $\max\{0, w_x - d_{\mathrm{H}}\}$, except when $d_{\mathrm{H}} = w_x = 6$, because this would imply $w_y = 0$ (impossible since $\boldsymbol{0} \notin \mathcal{P}_n$). Hence, $w_{xy} = 1 > 0$ for $d_{\mathrm{H}} = 6$, which leads to the observed notch. Similarly for Case 2 (using $\boldsymbol{1} \notin \mathcal{P}_n$).

Further, we observe that all curves end at the same point. This is due to the fact that for $d_{\mathrm{H}} = n$ the only possible

options for w_y and w_{xy} when w_x is given read $w_y = n - w_x$ and $w_{xy} = 0$. The resulting value is

$$g_{\boldsymbol{x},\boldsymbol{1}-\boldsymbol{x}} = g_{\boldsymbol{1}-\boldsymbol{x},\boldsymbol{x}} = \frac{4\sigma_x^2}{n} = \frac{4w_x\left(1 - \frac{w_x}{n}\right)}{n} \quad (34)$$

in general, and thus 0.84 for the example under consideration.

Finally, note that, as expected, the largest g-ratios are found in Case 1. Most strikingly, we see that these ratios may even exceed the value one (see Figure 2), suggesting that the noise resistance of the Pearson detector is higher than the noise resistance of the Euclidean detector for these cases. Of course, this cannot be true, since a Euclidean detector is well-known to be optimal in case of Gaussian noise. Indeed, we observe that in all cases that $g_{\boldsymbol{x},\boldsymbol{y}}$ exceeds one, its counterpart $g_{\boldsymbol{y},\boldsymbol{x}}$ is smaller than one. Similarly, $g_{\boldsymbol{y},\boldsymbol{x}} > 1$ implies $g_{\boldsymbol{x},\boldsymbol{y}} < 1$. Since codeword pairs with smaller distances are dominant with respect to contributions to the WER, the overall result is still that from the noise perspective Euclidean detectors are superior to Pearson detectors, which is the price to be paid for the immunity of the latter detectors to gain and offset mismatches. The analysis as done in this paper can be exploited in the design of new Pearson codes, i.e., subsets of $\mathcal{P}_n$, with a noise performance closer to the Euclidean case, by avoiding the selection of codeword pairs with small Pearson noise distances. It is clear that in order to increase the Pearson noise distance, the focus should not only be on Hamming distance increase, since these two distance measures are certainly not growing proportionally. Rather, also the codeword weights must be taken into account.

IV. CONCLUSIONS

We have investigated various properties of Pearson-distance-based detection and Pearson codes. For binary codes, we have derived upper and lower bounds on the Pearson noise distance and studied relations with the Hamming distance.

As possibilities for future work we identify (i) application of the findings in order to construct codes with an increased minimum Pearson noise distance and (ii) extension of the results to q-ary codes.

REFERENCES

[1] N. Papandreou, T. Parnell, H. Pozidis, T. Mittelholzer, E. Eleftheriou, C. Camp, T. Griffin, G. Tressler, and A. Walls, "Using Adaptive Read Voltage Thresholds to Enhance the Reliability of MLC NAND Flash Memory Systems," Proc. 24th Great Lakes Symp. on VLSI, ACM, pp. 151-156, 2014

[2] Y. Emre, C. Yang, K. Sutaria, Y. Cao, and C. Chakrabarti, "Enhancing the reliability of STT-RAM through circuit and system level techniques," in *Tech. Dig. IEEE Workshop on Signal Processing Systems (SiPS)*, Quebec City, Canada, pp. 125-130, Oct. 2012.

[3] A. Jiang, R. Mateescu, M. Schwartz, and J. Bruck, "Rank Modulation for Flash Memories," *IEEE Trans. Inform. Theory,* vol. IT-55, no. 6, pp. 2659-2673, June 2009.

[4] K. A. S. Immink, "A Survey of Codes for Optical Disk Recording," *IEEE J. Select. Areas Commun.,* vol. 19, no. 4, pp. 756-764, April 2001.

[5] K. A. S. Immink and J. H. Weber, "Minimum Pearson Distance Detection for Multi-Level Channels with Gain and/or Offset Mismatch," *IEEE Trans. Inform. Theory,* vol. IT-60, pp. 5966-5974, Oct. 2014.

[6] A. M. Mood, F. A. Graybill, and D. C. Boes, *Introduction to the Theory of Statistics*, 3rd ed. New York, NY, USA: McGraw-Hill, 1974.

[7] J. H. Weber, K. A. S. Immink, and S. Blackburn, "Pearson Codes," *IEEE Trans. Inform. Theory,* vol. IT-62, no. 1, pp. 131-135, Jan. 2016.