Combinatorics, Probability and Computing (2000) 9, 465–479. Printed in the United Kingdom

Discrete Isoperimetric Inequalities and the Probability of a Decoding Error

JEAN-PIERRE TILLICH[1] and GILLES ZÉMOR[2]

[1] LRI, bâtiment 490,
Université Paris-Sud, 91405 Orsay, France
(e-mail: tillich@lri.fr)

[2] École Nationale Supérieure des Télécommunications,
75 634 Paris 13, France
(e-mail: zemor@infres.enst.fr)

Received 20 April 1999; revised 19 January 2000

We derive improved isoperimetric inequalities for discrete product measures on the n-dimensional cube. As a consequence, a general theorem on the threshold behaviour of monotone properties is obtained. This is then applied to coding theory when we study the probability of error after decoding.

1. Introduction

Consider the n-cube, or binary Hamming space $\boldsymbol{H}^n = \{0,1\}^n$ of dimension n, and denote by $|x|$ the weight $\sum_{i=1}^n x_i$ of a binary vector $x = (x_1, x_2, \ldots, x_n) \in \boldsymbol{H}^n$. For $0 < p < 1$ let μ_p denote the product measure on $\boldsymbol{H}^n$ defined for any subset $\Omega \subset \boldsymbol{H}^n$ by

$$\mu_p(\Omega) = \sum_{x \in \Omega} p^{|x|}(1-p)^{n-|x|}.$$

Let us write $x \preceq y$ if for any $i = 1, 2, \ldots, n$ we have $x_i \leqslant y_i$. We shall say that Ω is increasing if, for any $x \in \Omega$, $x \preceq y$ implies that y is also in Ω.

The theory of random graphs has been concerned with many increasing sets Ω and with the behaviour of the function $f(p) = \mu_p(\Omega)$. Quite often a threshold phenomenon is observed: $f(p)$ jumps from near 0 to near 1 in a short interval that shrinks as n grows.

In many cases this threshold behaviour can be proved by a direct study of $f(p)$. This has not always been successful, however, and the following indirect strategy has been investigated by a number of authors, including [4, 8, 12, 13, 14, 15, 16]: find conditions on Ω which are easy to check and which imply that $\mu_p(\Omega)$ satisfies a differential inequality

of the form

$$\frac{d\mu_p(\Omega)}{dp} \geqslant a(n)b(p)c(\mu_p(\Omega)) \tag{1.1}$$

where b and c are positive and continuous functions on $(0,1)$, and $a(n) \to \infty$ when $n \to \infty$. Then, the integration of such a differential inequality shows that $\mu_p(\Omega)$ behaves like a threshold function.

One example of a famous problem that has long eluded the direct approach and has recently been solved by techniques of this kind is the phase transition phenomenon for the k-SAT problem for $k \geqslant 3$ [7].

In this paper we are concerned with the *isoperimetric* method for obtaining inequalities of type (1.1). This originates in [12] and involves the quantity

$$\begin{aligned} h_\Omega(x) &= 0, && \text{if } x \notin \Omega, \\ h_\Omega(x) &= \operatorname{card}\{y \notin \Omega, d(x,y) = 1\}, && \text{if } x \in \Omega, \end{aligned}$$

where $d(x,y)$ denotes the Hamming distance between x and y, that is, the number of coordinates i such that $x_i \neq y_i$.

Crucial to the isoperimetric method is the Margulis–Russo identity for increasing sets (see [12]):

$$\frac{d\mu_p(\Omega)}{dp} = \frac{1}{p}\int_\Omega h_\Omega(x)d\mu_p(x). \tag{1.2}$$

This means that to obtain a differential inequality of type (1.1) one need only lower-bound $\int_\Omega h_\Omega(x)d\mu_p(x)$ by a function of $\mu_p(\Omega)$. Such an inequality can be named isoperimetric, because the above integral can be thought of as measure of the 'boundary' of Ω, and $\mu_p(\Omega)$ is its 'volume'.

Margulis brought in the quantity

$$\Delta = \inf_{\omega \in \partial\Omega} h_\Omega(\omega),$$

where $\partial\Omega = \{\omega, h_\Omega(\omega) \neq 0\}$, and noticed that any increasing set satisfies a differential inequality of form (1.1) with $a = \sqrt{\Delta}$. In words, increasing sets with large Δ have a sharp threshold. Talagrand improved Margulis' original isoperimetric inequalities in [14] and these were further refined by Bobkov and Goetze in [2].

In this paper our purpose is twofold.

1. We shall further improve the isoperimetric inequalities of Margulis, Talagrand, Bobkov and Goetze. This will yield improved criteria for the threshold behaviour of monotone sets.
2. We shall extend the scope of the method which Margulis originally devised to prove the threshold behaviour of the probability of disconnecting a graph. We apply it to coding theory, namely we prove and measure the threshold behaviour of the probability of a decoding error. This approach to coding was initiated in [17], and is significantly improved here.

The next section highlights the main results.

2. Main results

2.1. Discrete isoperimetric inequalities

The additional condition on Ω (*i.e*, $h_\Omega(x)$ is either zero or greater than Δ) brought in by Margulis might seem somewhat mysterious. Let us remind the reader of the idea behind this condition.

Assume we have an increasing set Ω. The issue is: what would be an additional constraint on Ω that would make the measure of its boundary $\int h_\Omega(x)d\mu_p$ 'large'? It turns out that among increasing sets Ω with measure equal to that of a given Hamming ball B centred on $(1,1,\ldots,1)$, the largest boundary is achieved by B, that is, $\int h_\Omega(x)d\mu_p \leqslant \int h_B(x)d\mu_p$ (see Lemma 5.1 in [8] for instance). So, one way of forcing Ω to have a large boundary is to make it 'look like' a Hamming ball, and this is exactly what Margulis' condition achieves. There are other conditions for which Ω tends to 'look like' a Hamming ball and which force the boundary of Ω to be large, for instance the fact that Ω is invariant under a subgroup of S_n (see [3, 4, 7, 11, 15, 16]). This is relevant to the probability of a decoding error for cyclic codes.

Unfortunately, given that Ω is increasing and that $h_\Omega(x)$ is either zero or greater than Δ, proving a sharp lower bound on the quantity $\int h_\Omega$ seems difficult to obtain directly by induction on the dimension n.

One of the ideas brought in by Talagrand in [14] is to use instead a modified measure of the boundary of Ω, namely the quantity $\int \sqrt{h_\Omega}d\mu_p$, then prove (by induction on the dimension n) a general isoperimetric inequality (which holds for any subset Ω), and notice that by the Cauchy–Schwartz inequality any inequality of the form

$$\int \sqrt{h_\Omega}d\mu_p \geqslant f(\mu_p(\Omega))$$

implies that

$$\int h_\Omega d\mu_p \geqslant \sqrt{\Delta} f(\mu_p(\Omega)).$$

This follows from the chain of inequalities

$$\begin{aligned}\frac{1}{\sqrt{\Delta}}\int h_\Omega d\mu_p = \sqrt{\int h_\Omega d\mu_p}\sqrt{\frac{\int h_\Omega d\mu_p}{\Delta}} &\geqslant \sqrt{\int h_\Omega d\mu_p}\sqrt{\int_{\partial\Omega} d\mu_p} \\ &\geqslant \int \sqrt{h_\Omega}d\mu_p \geqslant f(\mu_p(\Omega)).\end{aligned}$$

His isoperimetric inequalities were improved by Bobkov and Goetze [2], who obtain, for any increasing Ω:

$$\int \sqrt{h_\Omega}d\mu_p \geqslant \frac{1}{12\sqrt{\ln\frac{1}{p(1-p)}}} J(\mu_p(\Omega)) \tag{2.1}$$

where

$$J(x) = x(1-x)\sqrt{\ln\left(\frac{1}{x(1-x)}\right)}.$$

In this paper we further improve on this by considering another function on the right-

hand side of this inequality, namely $\Psi(x)$, which is defined on $(0,1)$ by $\Psi(x) = \phi(\Phi^{-1}(x))$ and extended by continuity on $[0,1]$ with $\Psi(0) = \Psi(1) = 0$, where

- ϕ denotes the normal density, *i.e.*, $\phi(t) = \frac{1}{\sqrt{2\pi}} e^{-t^2/2}$,
- Φ stands for the Gaussian cumulative distribution, *i.e.*, $\Phi(x) = \int_{-\infty}^{x} \phi(t)dt$.

We obtain the following result.

Theorem 2.1. *For any increasing set Ω we have*

$$\int \sqrt{h_\Omega} d\mu_p \geqslant \frac{1}{\sqrt{2 \ln 1/p}} \Psi(\mu_p(\Omega)). \tag{2.2}$$

Comments.

1. When $\mu_p(\Omega)$ tends to 0, it can be checked that the lower bound in the above theorem is equivalent to
$$\frac{1}{\sqrt{\ln 1/p}} J(\mu_p(\Omega))$$
(see Lemma 3.1 below). This improvement is more significant than it looks: see the comment that follows Theorem 2.2. Also, replacing J by Ψ will make the integration of the inequality more precise.
2. The isoperimetric inequality in Theorem 2.1 is quite sharp for small sets, and this is true for *any* p. This can be tested on subsets of small size of the form
$$\Omega = \{x | x_1 = \cdots = x_k = 1\}$$
(in other words subcubes of codimension k that contain the point $(1,1,\ldots,1)$). First of all let us note that $\mu_p(\Omega) = p^k$. We choose k as an increasing function of n such that $p^k = o(1)$. Moreover $h_\Omega(x) = k$ for every $x \in \Omega$. Therefore $\int \sqrt{h_\Omega(x)} d\mu_p = \sqrt{k}\mu_p(\Omega)$. On the other hand, $\Psi(\mu_p(\Omega))$ is asymptotically equivalent (as n goes to infinity) to $\mu_p(\Omega)\sqrt{-2\ln \mu_p(\Omega)} = \mu_p(\Omega)\sqrt{-2k \ln p}$ by property (iv) of Lemma 3.1 in Section 3. In other words, for these sets the right-hand side and the left-hand side of the isoperimetric inequality of Theorem 2.1 are asymptotically equivalent. However, this inequality is by no means sharp for sets of measure $1/2$, for instance. For these sets it might well be that the increasing sets Ω with the smallest boundary (measured by $\int \sqrt{h_\Omega}$) are Hamming balls of measure $1/2$ centred around $(1,1,\ldots,1)$. We are not aware of any isoperimetric inequality that would prove this conjecture.

Theorem 2.1 can be 'integrated' to yield the following.

Theorem 2.2. *Let $\Omega \subset \boldsymbol{H}^n$ be an increasing set, let $f(p) = \mu_p(\Omega)$ and let θ be defined by $f(\theta) = 1/2$. Then $f(p)$ satisfies*

$$f(p) \leqslant \Phi\left(\sqrt{2\Delta}(\sqrt{-\ln\theta} - \sqrt{-\ln p})\right) \quad \textit{for} \quad 0 < p < \theta, \tag{2.3}$$

$$f(p) \geqslant \Phi\left(\sqrt{2\Delta}(\sqrt{-\ln\theta} - \sqrt{-\ln p})\right) \quad \textit{for} \quad \theta < p < 1. \tag{2.4}$$

Comment. For fixed $p < \theta$ and for large Δ the lower bound is equivalent to

$$\frac{1}{\sqrt{2\pi}u} e^{-u^2/2}$$

where $u = \sqrt{2\Delta}(\sqrt{-\ln\theta} - \sqrt{-\ln p})$. It should be noted that applying the aforementioned isoperimetric inequalities of [14, 2] would also yield similar exponential bounds of the form $e^{-\alpha u^2/2}$ (see [17]). However, the constant α we get in the exponent in this case turns out to be quite small. For instance, the constant α obtained by using inequality (2.1) is equal to $\frac{1}{144}$. This comes from the fact that after integration any constant K in the right-hand side of the isoperimetric inequality gets squared in the exponent of the corresponding lower bound of $f(p)$.

2.2. Applications to the probability of a decoding error

Let $C \subset \boldsymbol{H}^n$ be a linear code of minimal Hamming distance

$$d = \min_{c \in C, c \neq 0} |x|.$$

Suppose a codeword c is transmitted over the binary symmetric channel with transition probability p. This means that the received vector $v = (v_1, v_2, \ldots, v_n)$ is such that $v_i = c_i + e_i \bmod 2$ where the e_i are independent $\{0, 1\}$ random variables with $P(e_i = 1) = p$. The decoder decodes v by choosing the closest codeword x for the Hamming distance; and if there are several codewords equally distant from v he picks one according to some predefined scheme. This is a maximum-likelihood decoding scheme. We define it in this way so as to have a fixed set of error vectors for which decoding fails. The decoder succeeds if $x = c$. The associated *decoding region* is the set $\Omega \subset \boldsymbol{H}^n$ of those vectors ω such that the vector $v = c + \omega$ is decoded back into c. We are interested in the probability that a decoding error occurs, which can be expressed as the function

$$f_e(p) = 1 - \mu_p(\Omega).$$

Define the *threshold probability* as the transition probability θ_e such that $f_e(\theta_e) = 1/2$. In words, this is the channel error probability for which the 'maximum-likelihood decoder' defined above fails with probability $1/2$. Our main result is as follows.

Theorem 2.3. *Let C be a binary linear code of any length, and minimum distance d. Over the binary symmetric channel with transition probability p, the probability of decoding error $f_e(p)$ associated with C and any transmitted codeword c satisfies*

$$f_e(p) \leqslant 1 - \Phi\left[\sqrt{d}\left(\sqrt{-\ln(1-\theta_e)} - \sqrt{-\ln(1-p)}\right)\right] \quad \text{for} \quad 0 < p < \theta_e,$$

$$f_e(p) \geqslant 1 - \Phi\left[\sqrt{d}\left(\sqrt{-\ln(1-\theta_e)} - \sqrt{-\ln(1-p)}\right)\right] \quad \text{for} \quad \theta_e < p < 1.$$

Comments.

1. Theorem 2.3 displays the threshold behaviour of $f_e(p)$: the larger the minimum distance, the sharper the jump from almost zero to almost one.
2. The upper bound in Theorem 2.3 is of the form $f_e(p) \leqslant \exp(-dg(\theta_e, p))$ where $g(\theta_e, p) > 0$ for $p < \theta_e$, that is, $f_e(p)$ is exponentially small in d. In particular, families

of codes with minimal distance growing linearly with their length n have a probability of decoding error which decreases exponentially with n, as long as $\theta_e - p$ stays bounded below by some $\varepsilon > 0$. Such a behaviour for $f_e(p)$ is known to hold asymptotically for *almost all* codes; Theorem 2.3 holds for *all* linear codes.

3. This seems to be the first upper bound on the decoding error probability involving only p, the minimum distance of the code and θ_e, which is exponential in d when p is bounded away from θ_e. This result is essentially best possible up to numerical constants in the exponent.
4. Theorem 2.3 is really an application of the very general upper bound on the probability of a monotone property stated in Theorem 2.2. It is quite surprising that such a general approach yields bounds with reasonable constants! It seems to us that there should be other interesting consequences. We will give another application, also in the field of coding theory, but which concerns the erasure channel.

The paper is organized as follows. In Section 3 we translate Theorem 2.1 into a general result on the threshold behaviour of monotone properties. Then we show how this leads to Theorem 2.3. In Section 4 we prove Theorem 2.1. In Section 5 we discuss results of a similar nature for the erasure channel.

3. From Theorem 2.1 to Theorem 2.3

Let us first gather here a few properties on Φ and Ψ which are very useful for proving several facts and propositions of this paper (for a proof of these statements see, for instance, [5, Lemma 5.2, p. 88]).

Lemma 3.1.

(i) Ψ *is a positive and concave function on* $(0,1)$, *and* $\Psi(x) = \Psi(1-x)$ *for every* $x \in (0,1)$,
(ii) $\Psi' = -\Phi^{-1}$,
(iii) $\Psi\Psi'' = -1$ *on* $(0,1)$,
(iv) $\lim_{s\to 0^+} \frac{\Psi(s)}{s\sqrt{-2\ln s}} = \lim_{s\to 0^+} \frac{-\Phi^{-1}(s)}{\sqrt{-2\ln s}} = 1$,
(v) $\lim_{s\to -\infty} \frac{-s\Phi(s)}{\phi(s)} = \lim_{s\to \infty} \frac{s(1-\Phi(s))}{\phi(s)} = 1$.

These properties can be used to derive Theorem 2.2 from Theorem 2.1.

Proof of Theorem 2.2. First note that the Cauchy–Schwartz inequality gives us

$$\int \sqrt{h_\Omega(x)}\, d\mu_p \leqslant \left(\mu_p(\partial\Omega) \int h_\Omega(x) d\mu_p \right)^{1/2}$$

and, since $\int h_\Omega(x) d\mu_p \geqslant \int_{\partial\Omega} \Delta d\mu_p = \Delta \mu_p(\partial\Omega)$ by definition of Δ, we get

$$\int h_\Omega(x) d\mu_p \geqslant \sqrt{\Delta} \int \sqrt{h_\Omega(x)}\, d\mu_p. \tag{3.1}$$

Next apply Margulis and Russo's formula (1.2) which, together with (3.1) and Theorem 2.1, gives us

$$f'(p) \geqslant \frac{\sqrt{\Delta}}{p\sqrt{2\ln 1/p}}\Psi(f(p)).$$

Apply property (iii) of Lemma 3.1, $-\frac{1}{\Psi(s)} = \Psi''(s)$, to obtain

$$-\Psi''(f(p))f'(p) \geqslant \frac{\sqrt{\Delta}}{p\sqrt{2\ln 1/p}}. \tag{3.2}$$

Next, multiply by -1 and integrate: we get, for $p < \theta$,

$$\int_p^\theta \Psi''(f(s))f'(s)ds \leqslant \int_p^\theta \frac{-\sqrt{\Delta}}{s\sqrt{-2\ln s}}ds,$$

that is,

$$\Psi'(\theta) - \Psi'(p) \leqslant \left[\sqrt{-2\Delta\ln s}\right]_p^\theta.$$

Then we use the the fact that $\Psi'(f(\theta)) = \Psi'(1/2) = 0$ and $-\Psi'(f(p)) = \Phi^{-1}(f(p))$ by property (ii) of Lemma 3.1. The left-hand side of the last inequality is therefore simply $\Phi^{-1}(f(p))$. Since Φ is increasing, apply Φ to obtain (2.3). To obtain (2.4), integrate (3.2) between θ and p. □

Maximum-likelihood decoding

Let $C \subset \boldsymbol{H}^n$ be a linear code of dimension k and minimum distance d. Let $r = n - k$. Let H be a parity-check matrix for C and for any $x \in \boldsymbol{H}^n$ define its syndrome $\sigma(x) = H.{}^t x$. To every one of the 2^r possible syndromes s associate an $\omega \in \boldsymbol{H}^n$ of minimum weight such that $\sigma(\omega) = s$. Let Ω be the set of all those ω s, so that σ is a one-to-one correspondence between Ω and the set S of syndromes. The set Ω is a *decoding region* for the zero codeword, that is, a set of *correctable error-patterns*. A maximum-likelihood decoding scheme consists of adding to the received vector v the vector $\omega \in \Omega$, such that $\sigma(\omega) = \sigma(v)$. A decoding error occurs if the codeword thus obtained is not the original codeword, that is, if the error vector is not in Ω. This happens with probability

$$f_e(p) = 1 - \mu_p(\Omega).$$

Remark. The set Ω is *decreasing*, that is, $x \in \Omega$ and $y \preceq x$ implies $y \in \Omega$.

We have the following result.

Proposition 3.2. *If Ω is a decoding region for the zero codeword of C, and if $\Delta = \inf_{\omega\in\partial\Omega} h_\Omega(\omega)$, then*

$$\Delta \geqslant d/2.$$

Proof. Let $\omega \in \partial\Omega$. This means that no codeword is nearer to ω than the zero codeword ($\omega \in \Omega$), and that there exists $c \in C$ such that changing one '0' coordinate of ω to '1' will

change ω into a vector closer to c than to zero (ω is on the frontier). But then there must be at least $|c|/2$ '0' coordinates of ω that, when changed to '1', change ω into a vector closer to c than to zero. Otherwise $\omega + c$ would be a vector of weight strictly less than ω and with the same syndrome. This contradicts the definition of Ω. □

For any vector $x = (x_i)_{1 \leqslant i \leqslant n}$ of $\boldsymbol{H}^n$, let $\bar{x} = (1 - x_i)_{1 \leqslant i \leqslant n}$. Note that $\overline{\Omega}$ is an increasing set, that $h_\Omega(x) = h_{\overline{\Omega}}(\bar{x})$, and that

$$\mu_p(\Omega) = \mu_{1-p}(\overline{\Omega}).$$

Theorem 2.3 follows therefore from Theorem 2.2 applied to $\overline{\Omega}$.

4. Proof of Theorem 2.1

To prove Theorem 2.1 we proceed as in [2] and first prove an inequality for increasing functions on $\boldsymbol{H}^n$ which implies Theorem 2.1 when applied to the characteristic function 1_Ω of an increasing set Ω. The point is that this more general inequality can be proved by induction on n (whereas we do not know how to prove Theorem 2.1 by induction on n). As in [2, 14] we will work with the quantity

$$Mf(x) = \sqrt{\sum_{d(x,y)=1} ((f(x) - f(y))^+)^2},$$

which is defined for any real function f on $\boldsymbol{H}^n$, and where $a^+ = \max(a, 0)$. Note that for any subset Ω we have $M1_\Omega = \sqrt{h_\Omega}$.

Here and henceforth we denote by $\boldsymbol{E}f$ the quantity $\int f d\mu_p$ and by $\mathscr{F}_n$ the set of functions defined over $\boldsymbol{H}^n$ which take on values only on $[0, 1]$ and which are increasing with respect to the partial order $\preceq$: whenever $x \preceq y$ we have $f(x) \leqslant f(y)$. Note that any characteristic function of an increasing set of $\boldsymbol{H}^n$ is in $\mathscr{F}_n$.

Lemma 4.1. *For any function f in $\mathscr{F}_n$:*

$$\boldsymbol{E}\left(\sqrt{2 \ln 1/p (Mf)^2 + \Psi(f)^2}\right) \geqslant \Psi(\boldsymbol{E}f). \tag{4.1}$$

The proof of this lemma is by induction on n and borrows many ideas from [1, 2]. Before we give its proof let us show how it implies Theorem 2.1.

Proof of Theorem 2.1. Let $f = 1_\Omega$; then $\boldsymbol{E}f = \mu_p(\Omega)$. Moreover, since $\Psi(0) = \Psi(1) = 0$ and $M1_\Omega = \sqrt{h_\Omega}$, we have

$$\boldsymbol{E}\left(\sqrt{2 \ln 1/p (M1_\Omega)^2 + \Psi(1_\Omega)^2}\right) = \sqrt{2 \ln 1/p}\, \boldsymbol{E}(\sqrt{h_\Omega}) = \sqrt{2 \ln 1/p} \int \sqrt{h_\Omega} d\mu_p,$$

and this gives Theorem 2.1. □

We will now prove Lemma 4.1 by induction on n. The first step is to prove that it holds for $n = 1$. In this case it boils down to the following.

Lemma 4.2. *Let $q = 1 - p$. For any x in $[0,1]$ and h in $[0, 1-x]$ we have*

$$(\Psi(x+ph) - q\Psi(x))^2 - 2p^2 \ln(1/p)h^2 - p^2 (\Psi(x+h))^2 \leqslant 0.$$

Indeed, when $n = 1$ we should prove that for any function $f \in \mathscr{F}_1$ we have

$$q\Psi(f(0)) + p\sqrt{\Psi(f(1))^2 + 2\ln 1/p(f(1) - f(0))^2} \geqslant \Psi\left(qf(0) + pf(1)\right). \tag{4.2}$$

Let $x = f(0)$, and $h = f(1) - f(0)$. Note that x and h both belong to $[0,1]$, and so does $x + h = f(1)$. This gives the aforementioned range for x and h. Moreover $qf(0) + pf(1) = x + ph$. An equivalent form for (4.2) is therefore

$$p\sqrt{\Psi(x+h)^2 + 2\ln(1/p)h^2} \geqslant \Psi(x+ph) - q\Psi(x). \tag{4.3}$$

Note that Ψ is concave and nonnegative (property (i) of Lemma 3.1) and therefore $\Psi(x+ph) - q\Psi(x) \geqslant p\Psi(x+h) \geqslant 0$. We can square both sides of (4.3) to get an equivalent inequality, and rearrange terms to obtain the inequality of Lemma 4.2.

Proof of Lemma 4.2. Fix x and let $F(h) = (\Psi(x+ph) - q\Psi(x))^2 - 2p^2 \ln(1/p)h^2 - p^2\Psi(x+h)^2$. We will prove Lemma 4.2 by noticing that for any choice of x we have $F(0) = F'(0) = 0$ and for any h in the range $(0, 1-x]$: $F''(h) \leqslant 0$. Note that $F(0) = 0$ and that, for $h \in (0, 1-x]$,

$$F'(h)/2 = p\Psi'(x+ph)\left(\Psi(x+ph) - q\Psi(x)\right) - 2p^2 \ln(1/p)h - p^2\Psi'(x+h)\Psi(x+h).$$

F is a continuous function and $\lim_{h\to 0^+} F'(h) = 0$, which implies $F'(0) = 0$. Moreover,

$$\begin{aligned} F''(h)/2 &= p^2\Psi''(x+ph)\left(\Psi(x+ph) - q\Psi(x)\right) + p^2\Psi'(x+ph)^2 \\ &\quad -2p^2 \ln 1/p - p^2\Psi''(x+h)\Psi(x+h) - p^2\Psi(x+h)^2 \\ &= p^2\left(\Psi'(x+ph)^2 - \Psi'(x+h)^2\right) - 2p^2 \ln 1/p - p^2 q\Psi(x)\Psi''(x+ph). \end{aligned}$$

We have used here property (iii) of Lemma 3.1. By using this property again we obtain

$$\begin{aligned} \Psi'(x+ph)^2 - \Psi'(x+h)^2 &= 2\int_{x+h}^{x+ph} \Psi'(t)\Psi''(t)dt \\ &= -2\int_{x+h}^{x+ph} \frac{\Psi'(t)}{\Psi(t)}dt \\ &= 2\ln\frac{\Psi(x+h)}{\Psi(x+ph)}. \end{aligned}$$

Hence, by substituting this expression into the calculation of $F''(h)$ and using property (iii) to get rid of $\Psi''(x+ph)$, we obtain

$$\frac{F''(h)}{2p^2} = 2\ln\left(\frac{p\Psi(x+h)}{\Psi(x+ph)}\right) + \frac{q\Psi(x)}{\Psi(x+ph)}. \tag{4.4}$$

Let $u = \frac{q\Psi(x)}{\Psi(x+ph)}$. We substitute for u into (4.4) and obtain

$$\begin{aligned} \frac{F''(h)}{2p^2} &= 2\ln\left(\frac{p\Psi(x+h) + q\Psi(x)}{\Psi(x+ph)} - u\right) + u \\ &\leqslant 2\ln(1-u) + u. \end{aligned}$$

The last inequality follows from the fact that ln is increasing and

$$0 \leqslant \frac{p\Psi(x+h)+q\Psi(x)}{\Psi(x+ph)} \leqslant 1.$$

The last inequality is just a consequence of the positivity of Ψ on $(0,1)$ and its concavity:

$$\Psi(x+ph) \geqslant q\Psi(x)+p\Psi(x+h).$$

By the same arguments we also have

$$0 \leqslant u = \frac{q\Psi(x)}{\Psi(x+ph)} \leqslant 1.$$

Note that $g(u) = 2\ln(1-u)+u$ is decreasing on $[0,1)$ and that $g(0)=0$: this implies that $F''(h) \leqslant 0$ on $(0,1-x]$ and we are done. □

It remains now to finish the proof of Lemma 4.1 by induction on n to prove that, if Lemma 4.1 holds for $n=1$, it holds for every $n \geqslant 1$.

Lemma 4.3. *If* (4.1) *holds for any function belonging to* $\mathscr{F}_n$*, then it also holds for every function* $f \in F_n$ *and any* $n \geqslant 1$.

Proof. The proof follows an idea due to Bobkov and is basically the same as the proof of Lemma 2.3 given in [2], with a slight modification of the induction hypothesis. We assume that the lemma holds up to a given $n \geqslant 1$. Consider now a function $f \in \mathscr{F}_{n+1}$. We put $f_0(x) = f(x,0)$ and $f_1(x) = f(x,1)$ where $x \in \boldsymbol{H}^n$. For $g \in \mathscr{F}_n$ we use the notation $\boldsymbol{E}_n g = \int g d\mu^n$. Note that

$$\boldsymbol{E}_{n+1} f = (1-p)\boldsymbol{E}_n f_0 + p\boldsymbol{E}_n f_1.$$

We apply this to $\sqrt{\Psi(f)^2+(Mf)^2}$ and we obtain

$$\begin{aligned}
&\boldsymbol{E}_{n+1}\sqrt{\Psi(f)^2+(Mf)^2} && (4.5)\\
&= (1-p)\boldsymbol{E}_n\sqrt{\Psi(f(x,0))^2+(Mf(x,0))^2} + p\boldsymbol{E}_n\sqrt{\Psi(f(x,1))^2+(Mf(x,1))^2}\\
&= (1-p)\boldsymbol{E}_n\sqrt{\Psi(f_0)^2+(Mf_0)^2} + p\boldsymbol{E}_n\sqrt{\Psi(f_1)^2+(Mf_1)^2+(f_1-f_0)^2}\\
&\geqslant (1-p)\boldsymbol{E}_n\sqrt{\Psi(f_0)^2+(Mf_0)^2} +\\
&\quad p\sqrt{\left(\boldsymbol{E}_n\sqrt{\Psi(f_1)^2+(Mf_1)^2}\right)^2+(\boldsymbol{E}_n(f_1-f_0))^2} && (4.6)\\
&\geqslant (1-p)\Psi(\boldsymbol{E}_n f_0) + p\sqrt{\Psi(\boldsymbol{E}_n f_1)^2+(\boldsymbol{E}_n f_1-\boldsymbol{E}_n f_0)^2} && (4.7)\\
&\geqslant \Psi(\boldsymbol{E}_{n+1} f). && (4.8)
\end{aligned}$$

- (4.6) is a consequence of the triangle inequality

$$\int\sqrt{u^2+v^2} \geqslant \sqrt{\left(\int u\right)^2+\left(\int v\right)^2}$$

  applied to $u = \sqrt{\Psi(f_1)^2+(Mf_1)^2}$ and $v = f_1 - f_0$.
- Inequality (4.7) follows from the induction assumption applied to f_0 and f_1.

- Inequality (4.8) follows from the same assumption applied to g which is defined by $g(0) = \boldsymbol{E}_n f_0$ and $g(1) = \boldsymbol{E}_n f_1$ which clearly belongs to $\mathscr{F}_1$ and verifies $\boldsymbol{E}_1 g = \boldsymbol{E}_{n+1} f$.

□

5. The erasure channel

In this section we derive another application of Theorem 2.2 to coding in the context of the erasure channel. Let $C \subset \mathbb{F}_q^n$ be a linear code over the finite field $\mathbb{F}_q$ with q elements. For $x \in \mathbb{F}_q^n$ we shall denote by x^H the binary vector of $\boldsymbol{H}^n$ obtained from x by changing its nonzero coordinates to '1'. We shall say that the binary vector v *covers* the q-ary vector x if $x^H \preceq v$.

The *erasure channel* does not corrupt codewords by changing symbols but simply by erasing them. For example, the message $(3, 5, 1, 1, 2, 4, 4, 3, 1)$ is sent but what is received is $(3, -, -, 1, 2, -, 4, -, 1)$. More precisely, the erasure vector is a random binary vector $(e_1, e_2, \ldots, e_n)$ of $\boldsymbol{H}^n$ where the e_i are independent and equal to 1 with probability p: the ith coordinate of a codeword $x \in C$ is erased if and only if $e_i = 1$. When the original codeword x is not the only one that coincides with the partially erased message on the set of non-erased coordinates, we shall say that ambiguous reception occurs. Because of the linearity of C, an ambiguity occurs if and only if the erasure vector equals some $\omega \in \boldsymbol{H}^n$ such that

$$c^H \preceq \omega$$

for some nonzero codeword $c \in C$. We see therefore that the probability of ambiguous decoding equals

$$f_a(p) = \mu_p(\Omega)$$

where $\Omega = \{\omega \mid \exists c \in C, c \neq 0, c^H \preceq \omega\}$. Clearly the set Ω is increasing so that Theorem 2.2 will apply. Furthermore, we have the following.

Lemma 5.1. *Let* $\partial\Omega = \{\omega, h_\Omega(\omega) \neq 0\}$ *and let* $\Delta = \inf_{\omega \in \partial\Omega} h_\Omega(\omega)$. *We have* $\Delta = d$, *where* d *is the minimum distance of code* C.

Proof. Let $\omega \in \partial\Omega$. This means that there exists $v \in \boldsymbol{H}^n \setminus \Omega$ such that $d(\omega, v) = 1$. Let i be the coordinate in which v and ω differ. Because Ω is increasing we have $\omega_i = 1$ and $v_i = 0$. Suppose that ω covers two linearly independent codewords c and c'; in other words, all codewords of a subcode C' of C of dimension 2. Then the linear mapping

$$\begin{aligned} C' &\to \mathbb{F}_q, \\ x &\mapsto x_i \end{aligned}$$

has nonzero kernel, the codewords of which must be covered by v. Therefore the set of codewords covered by ω can only be a subcode of dimension 1. Therefore there are at least d ways of changing a '1' coordinate of ω to zero so that the resulting binary vector covers no nonzero codeword. □

Lemma 5.1 together with Theorem 2.2 yields the following.

Theorem 5.2. *Over the erasure channel with erasure probability p, the probability $f_a(p)$ of ambiguous reception of a codeword belonging to a q-ary code C with minimum distance d satisfies:*

$$f_a(p) \leqslant \Phi\left(\sqrt{2d}(\sqrt{-\ln\theta_a} - \sqrt{-\ln p})\right) \quad \text{for} \quad 0 < p < \theta_a, \tag{5.1}$$

$$f_a(p) \geqslant \Phi\left(\sqrt{2d}(\sqrt{-\ln\theta_a} - \sqrt{-\ln p})\right) \quad \text{for} \quad \theta_a < p < 1, \tag{5.2}$$

where θ_a is defined by $f_a(\theta_a) = 1/2$.

It might be argued that, if the erasure vector covers a subcode of dimension m, then the receiver knows that the original codeword belongs to a certain coset of a subcode of dimension m: in particular, he can recover it with probability at least $1/q^m$. If m is small, then maybe that is not so bad, so that the receiver may still recover something even if $p > \theta_a$. Actually, this almost never happens: before giving this a precise meaning we need a lemma.

Lemma 5.3. *Define the sequence $\Omega_1, \Omega_2, \ldots, \Omega_t \ldots$ of subsets of $\boldsymbol{H}^n$ by $\Omega_1 = \Omega$, $\Omega_2 = \Omega \backslash \partial\Omega$, and inductively, $\Omega_{t+1} = \Omega \backslash \partial\Omega_t$. Then Ω_t equals the set of binary vectors that cover a subcode of dimension t.*

Proof. The proof of Lemma 5.1 has already proved the result for $t = 2$ and the same argument generalizes inductively. Indeed, if $\omega \in \Omega_t$ covers a subcode C' of dimension $t+1$, then because the linear mapping

$$\begin{aligned} C' &\to \mathbb{F}_q, \\ x &\mapsto x_i \end{aligned}$$

must have a kernel of dimension at least t, any $v \in \boldsymbol{H}^n$ at Hamming distance 1 from ω must stay in Ω_t. □

The tth *generalized Hamming weight* of C is defined to be the smallest support d_t of a subcode of dimension t. Notice that $\Delta(\Omega_t) \geqslant d_t \geqslant d$.

Let f_t be the function defined by $f_t(p) = \mu_p(\Omega_t)$ and let θ_t be such that $f_t(\theta_t) = 1/2$. We have the following result.

Proposition 5.4. *The quantity $\theta_{t+1} - \theta_t$ is bounded above by a function of the minimum distance d which goes to zero as d grows to infinity.*

Proof. Suppose the contrary. Then there exists $\gamma > 0$ and some sequence of codes with d growing to infinity for which we always have $\theta_{t+1} - \theta_t \geqslant \gamma$. By Theorem 2.2, we have, when d grows to infinity, $f_t(\theta_t + \gamma/4) \to 1$, $f_{t+1}(\theta_{t+1} - \gamma/4) \to 0$; and, because $\mu_p(\partial\Omega_t) = \mu_p(\Omega_t) - \mu_p(\Omega_{t+1})$, for all p such that $\theta_t + \gamma/4 \leqslant p \leqslant \theta_{t+1} - \gamma/4$,

$$\mu_p(\partial\Omega_t) \to 1$$

independently of p because f_t and f_{t+1} are increasing functions.

Now, Margulis and Russo's formula (1.2) and $\Delta(\Omega_t) \geqslant d_t \geqslant d$ imply

$$f_t'(p) \geqslant \frac{d}{p}\mu_p(\partial\Omega_t)$$

for all $\theta_t + \gamma/4 \leqslant p \leqslant \theta_{t+1} - \gamma/4$. But then $f_t(p) \geqslant d\gamma/2(1 + \varepsilon(d))$, where $\varepsilon(d) \to 0$ when $d \to \infty$. This contradicts $f_t(p) \leqslant 1$. □

Let $g(p)$ be the probability of error if the receiver chooses at random one of the codewords that coincides with the received message on the set of non-erased positions. What the above discussion shows is that, when d grows to infinity (however slowly), not only does $f_a(p)$ jump suddenly from almost zero to almost one, but so does $g(p)$.

Proposition 5.4 has another interesting consequence. Theorem 2.2 gives bounds on $\mu_p(\Omega)$ involving only the parameters Δ and θ, but it is usually difficult to make them explicit when only Δ is known. Since the ball centred on zero and of radius $\Delta - 1$ must lie totally outside Ω, one can argue that, as n grows, θ must stay at least as large as $\liminf_{n\to\infty} \Delta/n$. This can not be improved without further information because Ω might very well be the set of vectors of weight $\geqslant \Delta$. In the present case, however, it is possible to derive a better asymptotic bound on θ_a, namely the following.

Proposition 5.5. *For a linear q-ary code C of length n and minimum distance d we have*

$$\theta_a \geqslant \frac{q}{q-1}\delta(1 + \varepsilon(d)),$$

where $\varepsilon(d) \to 0$ when $d \to \infty$.

Proof. Denote by $\delta_t = d_t/n$ the normalized generalized Hamming weights of C. Because $\Delta(\Omega_t) = d_t$ we must have $\theta_t \geqslant \delta_t(1 + \varepsilon(d))$. But Griesmer's bound implies $d_t \geqslant d_1(1 + \frac{1}{q} + \cdots + \frac{1}{q^{t-1}})$; hence the result, by Proposition 5.4. □

6. Concluding remarks

Generalization to nonlinear codes

Linearity is not really crucial to Theorem 2.3, but expressing the result gets messy without this hypothesis. Linearity is crucial in Theorem 5.2, though.

Other decoding schemes

We have focused on maximum-likelihood decoding. But Theorem 2.2 could be applied in principle to any decoding scheme with monotone decoding regions.

Locating θ_e

Since it is not always clear what the value of θ_e is for a given code, it would be interesting to determine an asymptotic lower bound on θ_e as a function of the relative minimum distance $\delta = d/n$. It is not clear to us what the best lower bound is, but let us sketch one possible argument. Let v be a vector of weight $z = \zeta n$. For $\alpha \leqslant 2$, let $N(v, \alpha)$ be the

number of codewords of weight αz and at distance z from v. Let $N(\alpha)$ be the average value of $N(v,\alpha)$ when v runs over all vectors of weight z, that is,

$$N(\alpha) = \binom{n}{z}^{-1} \sum_{|v|=z} N(v,\alpha).$$

Denote by $A(n,w,d)$ the maximum size of a binary code of length n, constant weight w and minimum distance d. The number of vectors of weight z and at distance z from a codeword of weight αz is less than $2^{\alpha z}\binom{n-\alpha z}{(1-\alpha/2)z}$; hence $N(\alpha) \leqslant B(\alpha)$, where

$$B(\alpha) = A(n,\alpha z,d)\binom{n}{z}^{-1} 2^{\alpha z}\binom{n-\alpha z}{(1-\alpha/2)z}.$$

Whenever $B(\alpha)$ is exponentially small for all possible values of α, then a random vector of weight z will, with overwhelming probability, be closer to the zero codeword than to any other; therefore we must have $\theta_e > \zeta$ for n large enough. Denoting by

$$R(\omega,\delta) = \liminf_{n\to\infty} \frac{1}{n} \log_2 A(n,\omega n,\delta n),$$

and taking exponents, we get that $\theta_e > \zeta$ whenever

$$0 > \max_{\alpha\leqslant 2} \left[R(\alpha\zeta,\delta) + \alpha\zeta + (1-\alpha\zeta) H\left(\frac{(1-\alpha/2)\zeta}{1-\alpha\zeta}\right) - H(\zeta) \right],$$

where $H(x) = -x\log_2 x - (1-x)\log_2(1-x)$ is the binary entropy function.

References

[1] Bobkov, S. (1997) An isoperimetric inequality on the discrete cube, and an elementary proof of the isoperimetric inequality in Gauss space. *Ann. Probab.* **25** 206–214.

[2] Bobkov, S. and Goetze, F. (1996) Discrete isoperimetric and Poincaré-type inequalities. Technical report SFB 343, University of Bielefeld 96-086.

[3] Bourgain, J., Kahn, J., Kalai, G., Katznelson, Y. and Linial, N. (1992) The influence of variables in product spaces. *Israel J. Math.* **77** 55–64.

[4] Bourgain, J. and Kalai, G. (1997) Influences of variables and threshold intervals under group symmetries. *Geometric and Functional Analysis* **7** 438–461.

[5] Csiszàr, I. and Körner, J. (1981) *Information Theory Coding Theorems for Discrete Memoryless Systems*, Academic Press.

[6] Elias, P. (1956) Coding for two noisy channels. In *Information Theory*, Academic Press, pp. 61–74.

[7] Friedgut, E. (1999) Sharp thresholds of graph properties, and the k-sat problem. Appendix of J. Bourgain, *J. Amer. Math. Soc.* **12** 1017–1054.

[8] Friedgut, E. and Kalai, G. (1996) Every monotone graph property has a sharp threshold. *Proc. Amer. Math. Soc.* **124** 2993–3002.

[9] Gallager, R. G. (1965) A simple derivation of the coding theorem and some applications. *IEEE Trans. Inform. Theory* **11** 3–18.

[10] Gallager, R. G. (1968) *Information Theory and Reliable Communications*, Wiley.

[11] Kahn, J., Kalai, G. and Linial N. (1988) The influence of variables on Boolean functions. In *Proc. 29th Ann. Symp. on Foundations of Comput. Sci.*, IEEE Press, pp. 68–80.

[12] Margulis, G. (1974) Probabilistic characteristics of graphs with large connectivity. *Problemy Peredachi Informatsii* **10** 101–108.

[13] Russo, L. (1982) An approximative zero-one law. *Zeit. Warsch. und Verwandte Gebiete* **61** 129–139.
[14] Talagrand, M. (1993) Isoperimetry, logarithmic Sobolev inequalities on the discrete cube, and Margulis' graph connectivity theorem. *Geometric and Functional Analysis* **3** 295–314.
[15] Talagrand, M. (1997) On boundaries and influences. *Combinatorica* **17** 275–285.
[16] Talagrand, M. (1994) On Russo's approximate zero-one law. *Ann. Probab.* **22** 1576–1587.
[17] Zémor, G. (1994) Threshold effects in codes. In *Algebraic Coding*, Vol. 781 of *Springer Lecture Notes in Computer Science*, pp. 278–286.