5 List Decodability Vs. Rate

Once you eliminate the impossible, whatever remains,
no matter how improbable, must be the truth.
Sherlock Holmes (by Sir Arthur Conan Doyle)

5.1 Introduction

In the previous two chapters, we have seen on the one hand that any code of distance d can be list decoded up to its Johnson radius (which is always greater than $d/2$). On the other hand, we have seen that, in general, the list decoding radius (for polynomial-sized lists), purely as a function of the distance of the code, cannot be larger than the Johnson radius. Together these pose limitations to the performance of list decodable codes if one *only* appeals to the distance-LDR relation of the code in order to bound its list decoding radius. To present a concrete example, these imply that one can use a binary code family of relative distance δ to list decode a fraction $(1 - \sqrt{1-2\delta})/2$ of errors, but no better (in general). Hence, to list decode a fraction $(1/2 - \varepsilon)$ of errors, one needs binary codes of relative distance $(1/2 - O(\varepsilon^2))$. The best known explicit constructions of code families of such high relative distance achieve a rate of only $O(\varepsilon^6)$ [6, 164], and there is an upper bound of $O(\varepsilon^4 \log(1/\varepsilon))$ for the rate of such code families [139].

This raises several natural questions. Can one achieve rate better than $\Omega(\varepsilon^4)$ for binary codes that have list decoding radius $(1/2 - \varepsilon)$? Note that the limitation discussed above comes in part from the rate vs. distance trade-off of codes, and in part from bounding the list decoding radius purely as a function of the distance of the code (via the Johnson bound). If one is interested in list-of-L decoding for some large constant L, the parameters that are directly relevant to the problem are list-of-L decoding radius and the rate of the code. Note that the distance of the code does not (at least directly) appear to be relevant to the problem at all. Since we are only interested in list-of-L decoding, why should one use codes optimized for the minimum distance (i.e., the list-of-1 decoding radius)? A closer examination of this question suggests the possibility that by "directly" optimizing the rate as a function of the list decoding radius, one might be able to do better than the two-step method that goes via the distance of the code.

V. Guruswami: List Decoding of Error-Correcting Codes, LNCS 3282, pp. 79-92, 2004.

This indeed turns out to be the case, as results in this chapter demonstrate. We will exhibit codes that achieve trade-offs between list decodability and rate which are provably beyond what can be achieved by going via the distance of the code. While the rate vs. distance trade-off is one of the central problems in coding theory and has received lots of attention, the list decoding radius vs. rate question has received much less attention. This chapter studies this trade-off and proves non-trivial lower bounds on the rate of certain list decodable codes. The basic approach is to use the probabilistic method to show the existence of certain codes. The results of this chapter highlight the potential and limits of list decoding, which in turn sets up the stage for the algorithmic results of Part II by indicating the kind of parameters one can hope for in *efficiently* list decodable codes. Furthermore, some of the results provide "good" inner codes for some of our later concatenated code constructions.

5.2 Definitions

The aim of this chapter is to study the trade-offs between list decoding radius and the rate of code families. In order to undertake such a study systematically, we first develop some definitions and notation. It might be of help to the reader to recall the definition of list decoding radius from Section 2.1.4.

Definition 5.1. *For an integer q, real p with $0 \leq p \leq (1 - 1/q)$, and list size function $\ell : \mathbb{Z}^+ \to \mathbb{Z}^+$, the* rate function *for q-ary codes with list-of-ℓ decoding radius p, denoted $R_{\ell,q}(p)$, is defined to be*

$$R_{\ell,q}(p) = \sup_{\mathcal{C}:\mathsf{LDR}_\ell(\mathcal{C}) \geq p} R(\mathcal{C}) \ . \tag{5.1}$$

where the supremum is taken over all q-ary code families $\mathcal{C}$ with $\mathsf{LDR}_\ell(\mathcal{C}) \geq p$. When ℓ is the constant function which takes on the value L for some integer $L \geq 1$, we denote the above quantity as simple $R_{L,q}(p)$.
For a family of integer-valued functions $\mathcal{F}$, one defines the quantity

$$R_{\mathcal{F},q}(p) = \sup_{\ell \in \mathcal{F}} R_{\ell,q}(p) \ .$$

Remark: We have the restriction $p \leq (1-1/q)$ in the above definition, since it is easy to see that a q-ary code family of non-vanishing rate can never be list decoded from beyond a fraction $(1-1/q)$ of errors with polynomial-sized lists. We will often omit the subscript q when the alphabet size is clear from context, or when referring to the binary case. Whether the list size subscript is a constant, an integer-valued function, or a family of integer-valued functions will be clear from the context.

Note that $R_{L,q}(p)$ is the best (largest) rate of a q-ary code family which can list decoded up to a fraction p of errors using lists of size L. We next define the rate function for list decoding with arbitrary constant-sized lists.

Definition 5.2. *For an integer q and real p, $0 \leq p \leq (1 - 1/q)$, the rate function for list decoding by constant-sized lists, denoted $R_q^{\text{const}}(p)$, is defined to be*

$$R_q^{\text{const}}(p) = \lim_{L \to \infty} \sup \{R_{L,q}(p)\} \ .$$

We will also be interested in the analogous rate function $R_{L,q}$ when the codes in consideration are restricted to be linear. This is an interesting case to consider both combinatorially and because linear codes are much easier to represent, encode and operate with.

Definition 5.3. *We define the analogous rate functions $R_{\ell,q}$, $R_{L,q}$ and $R_{\mathcal{F},q}$ when restricted to* linear codes *by $R_{\ell,q}^{\text{lin}}$, $R_{L,q}^{\text{lin}}$ and $R_{\mathcal{F},q}^{\text{lin}}$, respectively. Likewise the function R_q^{const}, when restricted to linear codes, is denoted by $R_q^{\text{const,lin}}$.*

5.3 Main Results

With the definitions of the previous section in place, we now move on to studying the properties of the rate functions $R_{L,q}$ and the like. Firstly, note that $R_{1,q}(p)$ is precisely the best asymptotic rate of a q-ary code family of relative distance $2p$, and its study is one of the most important and still widely open problems in coding theory. Similarly, while a precise understanding of $R_{L,q}$ seems hopeless at this point, we can nevertheless focus on obtaining good upper and lower bounds on this function. And, as the result of Theorem 5.4 below states, the function R_q^{const} is in fact precisely known.

5.3.1 Basic Lower Bounds

We remark here that the results in this section are proved by analyzing the performance of random codes and showing that a random code of a certain rate has the desired list decodability properties with *very high probability*. In other words, *"most"* codes have the rate vs. list decodability trade-off claimed in this section. The following result was implicit in [203] and was explicitly stated and proved in [50].

Theorem 5.4 ([203, 50]). *For every q and every p, $0 \leq p \leq (1 - 1/q)$, we have*

$$R_q^{\text{const,lin}}(p) = R_q^{\text{const}}(p) = 1 - H_q(p) \tag{5.2}$$

(recall that $H_q(x) = x \log_q(q-1) - x \log_q x - (1-x) \log_q(1-x)$ is the q-ary entropy function).

We will defer the proof of the above result to later in this section. It is easy to verify that $H_q(1 - 1/q - \varepsilon) \simeq 1 - O(\varepsilon^2)$ for small $\varepsilon > 0$, and hence the above result implies, in particular, that for each fixed q, the best rate for families of linear q-ary codes list decodable up to a fraction $(1 - 1/q - \varepsilon)$ of

errors is $\Theta(\varepsilon^2)$. Recall that the best rate one could hope for via the "distance and Johnson bound" based approach was about ε^4. The conclusion therefore is that there exist codes which are list decodable well beyond their Johnson radius with small lists, and in fact most codes have this property!

"Capacity-Theoretic" Interpretation of Theorem 5.4 There is a very nice interpretation of the result of Theorem 5.4 by comparing it with Shannon's theorem on capacity of noisy channels, when applied to the specific case of the q-ary symmetric channel, call it qSC_p. The channel qSC_p transmits a q-ary symbol without distortion with probability $(1-p)$, and with the remaining probability, distorts it to one of the other $(q-1)$ symbols, picked uniformly at random. In other words, the probability that symbol α is distorted to symbol β equals $\frac{p}{q-1}$ if $\alpha \neq \beta$, and equals $(1-p)$ if $\alpha = \beta$. The Shannon capacity of such a channel equals $1 - H_q(p)$. Therefore, one can communicate reliably over this channel at a rate as close to $1 - H_q(p)$ as one seeks, but not at any rate greater than $1 - H_q(p)$.

The channel qSC_p makes an expected fraction p of errors, and in fact for all sufficiently large blocklengths, the fraction of errors will be close to p with overwhelming probability (by the Chernoff-Hoeffding bounds for i.i.d. events). However, Shannon's theorem relies on the fact the (close to) p fraction of errors will be randomly distributed. The result of Theorem 5.4 states that by using list decoding with list size a sufficiently large constant, we can communicate at a rate arbitrarily close to the "capacity" $1 - H_q(p)$, even if the channel corrupts an *arbitrary p fraction* of symbols in an *adversarial manner*.

Thus, list decoding allows us to approach the Shannon capacity *even if the errors are adversarially effected*, provided we use lists of large enough size in the decoding. This view indicates that list decoding can achieve the best performance one can hope for under a standard probabilistic error model even under the much stronger adversarial error model.

Proof of Theorem 5.4 In order to prove Theorem 5.4, we first focus on results that obtain lower bounds on the rate function for list decoding with a fixed list size L. We will then apply these results in the limit of large L to deduce Theorem 5.4. We first prove a lower bound on $R_{L,q}(p)$ for general codes, and will then prove a result for linear codes.

Theorem 5.5 ([50]). *For every q and every p, $0 \leq p \leq (1 - 1/q)$, we have*

$$R_{L,q}(p) \geq 1 - H_q(p)\Big(1 + \frac{1}{L}\Big) \ . \tag{5.3}$$

Proof: Fix a large enough blocklength n and set $e = \lfloor np \rfloor$. The idea is to pick a *random* code consisting of $2M$ codewords, where M is a parameter that will be fixed later in the proof. We will show that with high probability by removing at most M of the codewords the resulting code will be (e, L)-list

decodable. This is a fairly standard method in coding theory and is called "random coding with expurgation".

The probability that a fixed set of $(L+1)$ codewords all lie in a fixed Hamming sphere (in the space $[q]^n$) of radius e equals $(V_q(n,e)/q^n)^{L+1}$ where $V_q(n,e)$ is the volume of a Hamming sphere of radius e in $[q]^n$. It is well-known that $V_q(n,e) \leq q^{H_q(e/n)n} \leq q^{H_q(p)n}$ (see for example [193, Chapter 1]). Hence this probability is at most $q^{-(L+1)(1-H_q(p))n}$.

Therefore, the expected number N_{bad} of sets of $(L+1)$ codewords which all lie in *some* Hamming sphere of radius e is at most

$$\binom{2M}{L+1} \cdot q^n \cdot q^{-(L+1)(1-H_q(p))n} \leq (2M)^{L+1} \cdot q^{-Ln+(L+1)H_q(p)n} \ . \tag{5.4}$$

Let us pick M so that it is at least the upper bound in (5.4). For example, we can pick

$$M = \lceil q^{(1-(1+1/L)H_q(p))n} 2^{1+1/L} \rceil \geq q^{(1-(1+1/L)H_q(p))n} \ . \tag{5.5}$$

Then the expected value of N_{bad} is at most M, and therefore there exists a code with $2M$ codewords that has at most M sets of $(L+1)$ codewords that lie in a Hamming ball of radius e. Now, we can remove one codeword from each of these (at most M) subsets of $(L+1)$ codewords that lies in a ball of radius e. This process reduces the size of the code by at most M codewords. After this expurgation, we have a code with at least M codewords which is (e,L)-list decodable. Since $e = \lfloor pn \rfloor$, using (5.5) we get the desired lower bound on $R_{L,q}(p)$. □

We next prove the analog of the above result when restricted to linear codes; the lower bound is much weaker than that for general codes in that one needs very large lists to get close to the limiting rate $R_q^{\text{const}}(p) = 1 - H_q(p)$. The result first appeared implicitly in the work of Zyablov and Pinsker [203].

Theorem 5.6. *For every q and every p, $0 \leq p \leq (1-1/q)$, we have*

$$R_{L,q}^{\text{lin}}(p) \geq 1 - H_q(p)\Big(1 + \frac{1}{\log_q(L+1)}\Big) \ . \tag{5.6}$$

Proof: The idea is to once again pick a random code (specifically a linear code of blocklength n and dimension k) and then argue that with high probability it will have the required (e,L)-list decodability property (as before we set $e = \lfloor pn \rfloor$).

The main problem in applying the argument from the proof of Theorem 5.5 is that a subset of L codewords of a random linear code are no longer mutually independent. A random $[n,k]_q$ linear code $\mathbf{C}$ is picked by picking a random $n \times k$ matrix A over $\mathbb{F}_q$, and the code is given by $\{A\mathbf{x} : \mathbf{x} \in \mathbb{F}_q^k\}$. Define $J = \lceil \log_q(L+1) \rceil$. Now every set of L distinct non-zero messages in $\mathbb{F}_q^k$ contain a subset of at least J messages which are linearly independent over

$\mathbb{F}_q$. It is easily verified that such linearly independent J-tuples are mapped to J mutually independent codewords by a random linear code. We can then apply estimates similar to the proof of Theorem 5.5 applied to this subset of J codewords.

We now bound from above the probability that a random linear code $\mathbf{C}$ is *not* (e, L)-list decodable. We first make the following useful observation: A linear code $\mathbf{C}$ is (e, L)-list decodable iff no one of the Hamming balls of radius e around points in $B_q(\mathbf{0}, e)$ contain L or more *non-zero* codewords. The condition is clearly necessary; its also sufficient by linearity. Indeed, suppose there is some $\mathbf{y} \in \mathbb{F}_q^k$ with $|B_q(\mathbf{y}, e) \cap \mathbf{C}| \geq L + 1$. Let $\mathbf{c} \in B_q(\mathbf{y}, e) \cap \mathbf{C}$. By linearity, we have $|B_q(\mathbf{y} - \mathbf{c}, e) \cap \mathbf{C}| \geq L + 1$ as well. But $\mathbf{w} = \mathbf{y} - \mathbf{c}$ has Hamming weight at most e, and $B_q(\mathbf{w}, e)$ has at least L *non-zero* codewords.

The probability that codewords corresponding to a fixed J-tuple of linearly independent messages all lie in a fixed Hamming ball $B_q(\mathbf{w}, e)$ is at most $\left(q^{(H_q(p)-1)n}\right)^J$. Multiplying this by the number of such linearly independent J-tuples of messages and the number of choices for the center $\mathbf{w} \in B_q(\mathbf{0}, e)$, we get that the probability that some J-tuple of linearly independent messages all lie in some Hamming ball of radius e is at most

$$q^{kJ} \cdot q^{H_q(p)n} \cdot q^{(H_q(p)-1)Jn} = q^{-nJ\left(1-(1+1/J)H_q(p)-k/n\right)} \ . \tag{5.7}$$

Since every set of L non-zero codewords has a subset of J codewords corresponding to the encodings of linearly independent messages, the above also gives an upper bound on the probability that $\mathbf{C}$ is not (e, L)-list decodable. Picking the dimension to be, say, $k = \lfloor (1 - (1 + 1/J)H_q(p))n - \sqrt{n} \rfloor$, we get exponentially small failure probability for random linear codes with rates approaching $1 - (1 + 1/J)H_q(p)$. Hence there exists a linear code family of rate $1 - (1 + 1/J)H_q(p)$ and $\mathsf{LDR}_{L,q} \geq p$, as desired. □

Proof of Theorem 5.4: The lower bounds in both Theorems 5.5 and 5.6 approach $1 - H_q(p)$ as the list size $L \to \infty$. It remains to prove the upper bounds. Clearly $R_q^{\text{const,lin}}(p) \leq R^{\text{const}}(p)$, so it suffices to prove $R^{\text{const}}(p) \leq 1 - H_q(p)$. This is quite straightforward. Let C be a q-ary code of blocklength n and rate $r > 1 - H_q(p)$. Pick a *random* $\mathbf{x} \in [q]^n$ and consider the random variable $X = |B_q(\mathbf{x}, pn) \cap C|$. The expected value of X is clearly $|C| \cdot |B_q(\mathbf{0}, pn)|/q^n$ which is at least $q^{(r+H_q(p)-1)n-o(n)}$. If $r > 1 - H_q(p)$, this quantity is of the form $q^{\Omega(n)}$. Hence a random ball of radius pn has exponentially many codewords. We must therefore have $R^{\text{const}}(p) \leq 1 - H_q(p)$. □

We also record the following result which is obtained by combining the Gilbert-Varshamov bound (for rate vs. distance trade-off) with the Johnson bound on list decoding radius (which gives a certain LDR vs. distance trade-off). Such a result was made explicit for binary codes in [50] — below we state it for general alphabets.

Theorem 5.7. *For every prime power q and every p, $0 \leq p \leq (1-1/q)$, and every integer $L \geq 1$, we have*

$$R^{\mathrm{lin}}_{L,q}(p) \geq 1 - H_q\left(\left(1-\frac{1}{q}\right)\frac{L}{L-1}\left(1-\left(1-\frac{qp}{q-1}\right)^2\right)\right) . \tag{5.8}$$

Proof (Sketch): The Gilbert-Varshamov bound (see, for instance, [193, Chapter 5]) implies that there exist q-ary linear code families of relative distance δ and rate R where

$$R \geq 1 - H_q(\delta) . \tag{5.9}$$

(In fact a random linear code achieves this trade-off with high probability.) The result of Theorem 3.1 on the Johnson radius for list decodability implies that a q-ary code of relative distance δ and blocklength n is (pn, L)-list decodable for

$$p = \left(1-\frac{1}{q}\right)\left(1-\left(1-\frac{q}{q-1}\frac{L-1}{L}\delta\right)^{1/2}\right) . \tag{5.10}$$

Combining (5.9) and (5.10) gives us the desired result. □

5.3.2 An Improved Lower Bound for Binary Linear Codes

Consider the result of Theorem 5.6 for the case of binary linear codes and when $p = 1/2 - \varepsilon$ (i.e. we wish to correct close to the "maximum" possible fraction of errors). For this case it implies that there exist rate $\Theta(\varepsilon^2)$ families which are list decodable to a fraction $(1/2 - \varepsilon)$ of errors with lists of size $2^{O(\varepsilon^{-2})}$. While the list size is a constant, it is exponential in $1/\varepsilon$ and it is desirable to reduce it to polynomial in $1/\varepsilon$. By appealing to the Johnson radius based bound of Theorem 5.7, one can achieve a list size of $O(1/\varepsilon^2)$ for decoding up to a fraction $(1/2 - \varepsilon)$ of errors, but the rate goes down to $O(\varepsilon^4)$.

Next, we present an improved result for *binary linear* codes which combines the optimal $\Omega(\varepsilon^2)$ rate with $O(1/\varepsilon^2)$ list size. Recall that the result of Theorem 5.5 already implies this for general, non-linear codes, and the following result closes the gap between linear and non-linear codes for list decoding up to a fraction $(1/2 - \varepsilon)$ of errors (closing this disparity was highlighted by Elias [50] as an open question).

As we shall show in Section 5.3.4, a list size of $\Omega(1/\varepsilon^2)$ is really necessary (even for general, non-linear codes), and thus this result is optimal up to constant factors for the case $p = (1/2 - \varepsilon)$.

Theorem 5.8. *For each fixed integer $L \geq 1$, and $0 \leq p \leq 1/2$, we have*

$$R^{\mathrm{lin}}_{L}(p) \geq 1 - H(p) - \frac{1}{L} , \tag{5.11}$$

where $H(x) = -x \lg x - (1-x)\lg(1-x)$ denotes the binary *entropy function.*

Proof: For each fixed integer $L \geq 1$ and $0 \leq p < 1/2$ and for all large enough n, we use the probabilistic method to guarantee the existence of a binary linear code $\mathbf{C}$ of blocklength n that is (e, L)-list decodable for $e = pn$, and whose dimension is $k = \lfloor (1 - H(p) - 1/L)n \rfloor$. This clearly implies the lower bound on the rate function for binary linear codes claimed in (5.11).

The code $\mathbf{C} = C_k$ will be built iteratively in k steps by randomly picking the k basis vectors in turn. Initially the code C_0 will just consist of the all-zeroes codeword $b_0 = 0^n$. The code C_i, $1 \leq i \leq k$, will be successively built by picking a random (non-zero) basis vector b_i that is linearly independent of $b_1, \ldots, b_{i-1}$, and setting $C_i = \mathrm{span}(b_1, \ldots, b_i)$. Thus $\mathbf{C} = C_k$ is an $[n, k]_2$ linear code. We will now analyze the list of L decoding radius of the codes C_i, and the goal is to prove that the list of L decoding radius of $\mathbf{C}$ is at least e.

The key to analyzing the list of L decoding radius is the following potential function S_C defined for a code C of blocklength n:

$$S_C = \frac{1}{2^n} \sum_{x \in \{0,1\}^n} 2^{\frac{n}{L} \cdot |B(x,e) \cap C|} \ . \tag{5.12}$$

For notational convenience, we denote S_{C_i} be S_i. Also denote by T_x^i the quantity $|B(x, e) \cap C_i|$, so that $S_i = 2^{-n} \sum_x 2^{nT_x^i/L}$.

Let $B = |B(\mathbf{0}, e)| = |B(\mathbf{0}, pn)|$; then $B \leq 2^{H(p)n}$ (see for example Theorem (1.4.5) in [193, Chapter 1]). Clearly

$$S_0 = \frac{(2^n - B) + B \cdot 2^{n/L}}{2^n} \leq 1 + B \cdot 2^{-n(1-1/L)} \leq 1 + 2^{n(H(p)-1+1/L)} \ . \tag{5.13}$$

Now once C_i has been picked with the potential function S_i taking on some value, say $\hat{S}_i$, the potential function S_{i+1} for $C_{i+1} = \mathrm{span}(C_i \cup \{b_{i+1}\})$ is a random variable depending upon the choice of b_{i+1}. We consider the expectation $\mathbf{E}[S_{i+1} | S_i = \hat{S}_i]$ taken over the random choice of b_{i+1} chosen uniformly from outside $\mathrm{span}(b_1, \ldots, b_i)$. For better readability, below we sometimes use $\exp_2(z)$ to denote 2^z.

$$\begin{aligned}
&\mathbf{E}[S_{i+1} | S_i = \hat{S}_i] \\
&= 2^{-n} \sum_x \mathbf{E}[\exp_2(n/L \ \cdot T_x^{i+1})] \\
&= 2^{-n} \sum_x \mathbf{E}[\exp_2(n/L \cdot \big(|B(x,e) \cap C_i| + |B(x,e) \cap (C_i + b_{i+1})|\big))] \\
&= 2^{-n} \sum_x \Big(\exp_2(n/L \ \cdot T_x^i) \mathop{\mathbf{E}}_{b_{i+1}} [\exp_2(n/L \ \cdot T_{x+b_{i+1}}^i)] \Big)
\end{aligned} \tag{5.14}$$

where in the second and third steps we used the fact that if $z \in B(x, e) \cap C_{i+1}$, then either $z \in B(x, e) \cap C_i$, or $z + b_{i+1} \in B(x, e) \cap C_i$. To estimate the quantity (5.14), we use the fact that the expectation of a positive random variable taken over b_{i+1} chosen randomly from outside $\mathrm{span}(b_1, \ldots, b_i)$ is at

most $(1-2^{i-n})^{-1}$ times the expectation taken over b_{i+1} chosen uniformly at random from $\{0,1\}^n$. Using (5.14) we therefore get:

$$\begin{aligned}\mathbf{E}[S_{i+1}|S_i=\hat{S}_i] &\le (1-2^{i-n})^{-1}2^{-n}\sum_x\Big(2^{n/L\ \cdot T_x^i}\cdot\Big(\frac{1}{2^n}\sum_{y\in\{0,1\}^n}2^{n/L\ \cdot T_{x+y}^i}\Big)\Big)\\ &= (1-2^{i-n})^{-1}\hat{S}_i\ \cdot 2^{-n}\sum_x 2^{n/L\ \cdot T_x^i}\\ &= \frac{\hat{S}_i^2}{(1-2^{i-n})}\ . \end{aligned} \tag{5.15}$$

Applying (5.15) repeatedly for $i=0,1,\ldots,k-1$, we conclude that there exists an $[n,k]$ binary linear code $\mathbf{C}$ with

$$\begin{aligned}S_{\mathbf{C}}=S_k &\le \frac{S_0^{2^k}}{\prod_{i=0}^{k-1}(1-2^{i-n})^{2^{k-i}}}\\ &\le \frac{S_0^{2^k}}{(1-2^{k-n})^k}\le\frac{S_0^{2^k}}{1-k2^{k-n}}\end{aligned} \tag{5.16}$$

since $(1-x)^a\ge 1-ax$ for $x,a\ge 0$. Combining (5.16) with (5.13), we have

$$S_k\le(1-k2^{k-n})^{-1}\big(1+2^{n(H(p)-1+1/L)}\big)^{2^k}$$

and using $(1+x)^a\le(1+2ax)$ for $ax\ll 1$, this gives

$$S_k\le 2\cdot(1+2\cdot 2^{k+(H(p)-1+1/L)n})\le 6 \tag{5.17}$$

(the last inequality follows since $k=\lfloor(1-H(p)-1/L)n\rfloor$). By the definition of the potential S_k from Equation (5.12), this implies that $2^{n/L\cdot|B(x,e)\cap\mathbf{C}|}\le 6\cdot 2^n<2^{n+3}$, or $|B(x,e)\cap\mathbf{C}|\le(1+\frac{3}{n})L$ for every $x\in\{0,1\}^n$. If $n>3L$, this implies $|B(x,e)\cap\mathbf{C}|<L+1$ for every x, implying that $\mathbf{C}$ is (e,L)-list decodable, as desired. □ *(Theorem 5.8)*

Remark: One can also prove Theorem 5.8 with the additional property that the relative distance $\delta(\mathbf{C})$ of the code (in addition to its list -of-L decoding radius) also satisfies $\delta(\mathbf{C})\ge p$. This can be done, for example, by conditioning the choice of the random basis vector b_{i+1} in the above proof so that $\mathrm{span}(b_1,b_2,\ldots,b_{i+1})$ does not contain any vector of weight less than pn. It is easy to see that with this modification, Equation (5.15) becomes

$$\mathbf{E}[S_{i+1}|\hat{S}_i]\le\frac{\hat{S}_i^2}{(1-2^{i+H(p)n-n})}.$$

Using exactly similar calculations as in the above proof, we can then guarantee that there exists a code $\mathbf{C}$ of dimension $k=\lfloor(1-H(p)-1/L)n\rfloor$ and minimum distance at least pn that satisfies $S_{\mathbf{C}}=O(1)$, and consequently satisfies $\mathsf{LDR}_L(\mathbf{C})\ge p$.

Note that Theorem 5.8, as with the results from the previous section, is a non-constructive result, in that it only proves the existence of a code with the desired properties, and does not give an explicit or polynomial time construction. In fact, unlike the results of Theorems 5.4, 5.5 or 5.6, it does not even give a high probability result. (For those who might be aware of such terminology on the probabilistic method, the technique used to prove Theorem 5.8 is called the *semirandom method*.) Also the proof seems to work for the binary case and does not generalize, at least in any obvious fashion, to the q-ary case for $q > 2$. The following specific questions, therefore, remain open:

Question 5.9. Does a random binary linear code have the property claimed in Theorem 5.8 with high probability ?

Question 5.10. Does an analogous result to Theorem 5.8 hold for q-ary linear codes for $q > 2$? Specifically, does $R^{\mathrm{lin}}_{L,q} \geq 1 - H_q(p) - \frac{1}{L}$ hold for every prime power q ?

We believe that the answer to both of the questions above is yes. Finally, we note the following capacity-theoretic consequence of Theorem 5.8: there exist binary linear codes of rate within ε of the Shannon capacity of the binary symmetric channel with cross-over probability p, namely within ε of $1 - H(p)$, even when the fraction p of errors are effected *adversarially* as opposed to randomly, provided we use list decoding with lists of size $1/\varepsilon$.

5.3.3 Upper Bounds on the Rate Function

So far, all of our results concerning the rate functions R_L and R^{lin}_L established lower bounds on these functions. In other words they proved that codes with a certain list-of-L decoding radius and certain rate exist. We now turn to the questions of upper bounds on these functions, namely results which demonstrate that codes of certain rate and list decodability do not exist. We focus on binary codes for this section.

The result of Theorem 5.4 shows that one can achieve a rate arbitrarily close to the optimum rate $1 - H(p)$ for codes with list decoding radius p, provided one allows the list size L to grow beyond any finite bound (i.e. by letting $L \to \infty$). This raises the question whether one can attain the rate $1 - H(p)$ with any finite list size L. The following result, due to Blinovsky [27, 28], proves that the unbounded list size is in fact necessary to approach a rate of $1-H(p)$; in other words, it proves that $R_L(p)$ is strictly smaller than $1-H(p)$ for any finite L and $0 < p < 1/2$. The proof of the result is quite complicated and we refer the interested reader to [27, Theorem 3] (see also [28, Chapter 2]).

Theorem 5.11 ([27]). *For every integer $L \geq 1$, and each p, $0 \leq p \leq 1/2$, we have*

$$R_L(p) \leq 1 - H(\lambda) \ , \tag{5.18}$$

where λ, $0 \le \lambda \le 1/2$, is related to p by

$$p = \sum_{i=1}^{\lceil L/2 \rceil} \binom{2i-2}{i-1} \frac{(\lambda(1-\lambda))^i}{i} \ . \tag{5.19}$$

Corollary 5.12. *For every $L \ge 1$ and every p, $0 < p < 1/2$, we have $R_L(p) < 1 - H(p)$.*

Proof: It is not difficult to see that, for $0 \le y \le 1/2$,

$$\sum_{i=1}^{\infty} \binom{2i-2}{i-1} \frac{(y(1-y))^i}{i} = y \ . \tag{5.20}$$

Indeed, this follows from the fact that the generating function $C(x) = \sum_{n \ge 0} c_n x^n$ for *Catalan numbers*, defined by $c_n = \frac{1}{n+1}\binom{2n}{n}$ for $n \ge 0$, equals $C(x) = (1 - \sqrt{1-4x})/2$. Equation (5.20) above follows with the setting $x = y(1-y)$ in the generating function for Catalan numbers. We therefore have that the λ which satisfies Condition (5.19) is strictly greater than p. Hence, $H(\lambda) > H(p)$, and thus $R_L(p) \le 1 - H(\lambda) < 1 - H(p)$. □

5.3.4 "Optimality" of Theorem 5.8

Consider the case of list decoding radius close to $1/2$, i.e., the case when $p = 1/2 - \varepsilon$. In this case, Theorem 5.8 implies the existence of binary linear code families $\mathcal{C}$ of rate $\Omega(\varepsilon^2)$ and $\mathsf{LDR}_L(\mathcal{C}) \ge 1/2 - \varepsilon$ for list size $L = O(1/\varepsilon^2)$ (Theorem 5.5 showed the same result for general, non-linear codes). We now argue that in light of Theorem 5.11, this result for binary codes for the case $p = 1/2 - \varepsilon$ is in fact asymptotically optimal. That is, the rate and list size guaranteed by Theorems 5.5 and 5.8 are the best possible up to a constant factor.

By Theorem 5.4, $R_L(p) \le 1 - H(p)$ for any finite L, and hence for $p = 1/2 - \varepsilon$, we get that the rate can be at most $O(\varepsilon^2)$. It remains to show that in order to have list-of-L decoding radius $(1/2 - \varepsilon)$ and a positive rate, one needs $L = \Omega(\varepsilon^{-2})$. To do this we make use of the result of Theorem 5.11.

The λ that satisfies Condition (5.19) must be at least p. Hence if $p = (1/2 - \varepsilon)$, we have $1/2 \ge \lambda \ge (1/2 - \varepsilon)$. Therefore $\lambda(1-\lambda) \ge 1/4 - \varepsilon^2$.

Now for any integer $\ell \ge 0$ we have

$$\begin{aligned}
&\sum_{i=\ell+1}^{\infty} \binom{2i-2}{i-1} \frac{(\lambda(1-\lambda))^i}{i} \\
&\ge \binom{2\ell}{\ell} \frac{(\lambda(1-\lambda))^{\ell+1}}{\ell+1} \sum_{j=0}^{\infty} (\lambda(1-\lambda))^j \Big(\frac{2(2\ell+1)}{\ell+2}\Big)^j \\
&= \frac{\ell+2}{\ell+1} \binom{2\ell}{\ell} \frac{(\lambda(1-\lambda))^{\ell+1}}{(\ell+2) - 2(2\ell+1)\lambda(1-\lambda)}
\end{aligned} \tag{5.21}$$

where in the first step we use the fact that if $i = \ell + 1 + j$,

$$\frac{\binom{2i-2}{i-1}\frac{1}{i}}{\binom{2\ell}{\ell}\frac{1}{\ell+1}} = 2^j \prod_{s=\ell+1}^{\ell+j} \frac{2s-1}{s+1} \geq \Big(\frac{2(2\ell+1)}{\ell+2}\Big)^j .$$

Together with Condition (5.19) and Equation (5.20) applied with the choice $y = \lambda$, Equation (5.21) above implies

$$\lambda \geq p + \frac{\ell+2}{\ell+1}\binom{2\ell}{\ell}\frac{(\lambda(1-\lambda))^{\ell+1}}{(\ell+2) - 2(2\ell+1)\lambda(1-\lambda)} ,$$

where $\ell = \lceil L/2 \rceil$. Plugging in the above into the bound of Theorem 5.11 and using $\lambda(1-\lambda) \geq 1/4 - \varepsilon^2$, we get, after some straightforward algebraic manipulations,

$$\lambda \geq p + \Omega\Big(\frac{(1-4\varepsilon^2)^{\ell+1}}{\ell^{3/2}\varepsilon^2}\Big) .$$

Since $R_L(p) \leq 1 - H(\lambda)$ by Theorem 5.11, we get

$$R_L(p) \leq 1 - H\Big(p + \Omega\Big(\frac{(1-4\varepsilon^2)^{\ell+1}}{\ell^{3/2}\varepsilon^2}\Big)\Big) . \tag{5.22}$$

In order to have positive rate, the argument to the entropy function $H(\cdot)$ in the above bound must be at most $1/2$. When $p = 1/2 - \varepsilon$, this requires $1/(\ell^{3/2}\varepsilon^2) = O(\varepsilon)$, or $\ell = \Omega(\varepsilon^{-2})$. Since $\ell = \lceil L/2 \rceil$, we needs list size $L = \Omega(\varepsilon^{-2})$, as we desired to show. We record this fact in the following result:

Theorem 5.13. *Let $\varepsilon > 0$ be a sufficiently small constant and let $\mathcal{C}$ be a binary code family of rate r that satisfies $\mathsf{LDR}_L(\mathcal{C}) \geq (1/2 - \varepsilon)$. Then we must have $r = O(\varepsilon^2)$ and $L = \Omega(1/\varepsilon^2)$.*

5.4 Prelude to Pseudolinear Codes

For $q > 2$, the lower bound on rate we know for list decodable q-ary codes is much weaker for linear codes (Theorem 5.6) than for general codes (Theorem 5.5). We conjecture that there exists an answer to open question 5.10 in the affirmative, however a proof of this fact has been elusive.

Linear codes have the advantage of succinct representation and efficient encoding (for example, using the generator matrix). Thus, they are very attractive from a complexity view-point. This is particularly important for us later on when we will use the codes guaranteed by the results of the previous two sections as inner codes in concatenated schemes. In light of the fact that the existential results are weaker for linear codes, we introduce the notion of "pseudolinear" codes, which albeit non-linear, still have succinct representations and admit efficient encoding.

The basic idea behind pseudolinear codes is the following: to encode a message $\mathbf{x} \in \mathbb{F}_q^k$, first "map" it into a longer string $\mathbf{h}_\mathbf{x} \in \mathbb{F}_q^{k'}$ and then encode $\mathbf{h}_\mathbf{x}$ using a suitable $n \times k'$ "generator" matrix A into $A\mathbf{h}_\mathbf{x}$. The name pseudolinear comes from the fact that the non-linear part of the mapping is confined to the first step which maps $\mathbf{x}$ to $\mathbf{h}_\mathbf{x}$. Of course, to make this useful the mapping $\mathbf{x} \mapsto \mathbf{h}_\mathbf{x}$ must be easy to specify and compute – this will be the case; in fact the mapping will be explicitly specified.

The crucial property of pseudolinear codes for purposes of list decodability will be that by taking $k' = O(kL)$, we can ensure that under the mapping $\mathbf{x} \mapsto \mathbf{h}_\mathbf{x}$, every set of L distinct non-zero $\mathbf{x}$'s are mapped into a set of L *linearly independent* vectors in $\mathbb{F}_q^{k'}$. Then if we pick a "random" pseudolinear code by picking a random $n \times k'$ matrix A, we will have the property that the codewords corresponding to any set of L non-zero messages will be mutually independent. This "L-wise independence property" can then be used to analyze the list-of-L decoding properties of the random code, in a manner similar to the analysis of a general, random code.

In a nutshell, the above allows us to translate the list-of-L decoding performance of general codes into similar bounds for L-wise independent pseudolinear codes. The big advantage of pseudolinear codes over general codes is their succinct representation (since one only needs to store the "generator" matrix A) and their efficient encoding. They are thus attractive for use as inner codes in concatenated schemes.

To avoid burdening the reader at this stage, the formal definitions relating to pseudolinear codes and the analog of Theorem 5.5 and related results for pseudolinear codes are deferred to Chapter 9 (pseudolinear codes will not be used in the book until that point). For now, the reader can take comfort in the fact there is a way to achieve the list decoding performance of general codes with the more structured pseudolinear codes.

5.5 Notes

Initial works [48, 199, 162, 2] on list decoding investigated the notion on probabilistic channels, and used random coding arguments to explore the average decoding error probability of block codes for the binary symmetric and more general discrete memoryless channels. Combinatorial questions of the nature investigated in this chapter (and in this book in general), on the other hand, are motivated by worst-case, not average, error-correcting behavior.

The study of the maximum rate of (e, L)-list decodable codes in the limit of large blocklength n with e/n and L fixed originated in the work of Zyablov and Pinsker [203] who were interested mainly in the use of such codes as inner codes in concatenated schemes. The study of the relation between rate and list decodability was undertaken systematically for the first time by Blinovsky [27] (see also [28]), where non-trivial upper and lower bounds on $R_L(p)$

are obtained. The paper of Elias [50] is a very useful resource on this topic as it presents a nice, limpid survey of the relevant results together with some new results.

The result of Theorem 5.4 was first implicitly observed in [203]. The result of Theorem 5.5 and its proof are from [50]. Theorem 5.6 was first observed in [203]; the proof in this chapter follows the presentation in [50]. The result of Theorem 5.7 is the generalization to the q-ary case of a similar result for binary codes that was observed in [50].

Elias [50] was the first to note the disparity between the results for linear and non-linear codes, and posed the open question whether the requirement of very large lists in Theorem 5.6 for linear codes was inherent or, as he correctly suspected, was an artifact of the proof techniques. The result of Theorem 5.8 for binary linear codes can be viewed as a positive resolution of this question. This result appears in a joint paper of the author with Håstad, Sudan and Zuckerman [80].

Recently, Wei and Feng [195] obtained rather complicated lower bounds for the function $R_L(p)$ as well as its linear counterpart $R_L^{\text{lin}}(p)$. Their bounds are hard to state and do not have simple closed forms. They conjecture that their lower bounds for the linear and non-linear case are identical for every value of the list size. However, they are able to prove this only for list size at most 3.

Upper bounds on the rate function $R_L(p)$ have been studied by Blinovsky [27], and he obtained some non-trivial bounds which were mentioned in Section 5.3.3. For the case of list size $L = 2$, an improvement to the upper bound from Theorem 5.11 appears in [16]. A recent paper by Blinovsky [29] revisits the bounds for the linear and non-linear case from [27], and shows that the lower bound proved for linear codes is weaker than the one for non-linear codes for a list size as small as 5.

The notion of pseudolinear codes was defined and basic combinatorial results concerning them were proven by the author in joint work with Indyk [81].

Combinatorial results of a similar flavor to those discussed in this chapter appear in three other places in this book: in Chapter 8 where a generalization of Theorem 5.8 is proven, in Chapter 9 where pseudolinear codes are discussed, and in Chapter 10 where we discuss analogous questions for the case of erasures (instead of the errors case discussed in this chapter). Due to the local nature of the use of these results, we chose not to present them in this chapter, but instead postpone them to the relevant chapters where they are needed.