3 Johnson-Type Bounds and Applications to List Decoding

This chapter, as well as the next one, explore the relation between the list decoding radius and minimum distance of a code. Understanding the relation between these parameters is useful for two reasons: (a) for several important families of codes like Reed-Solomon codes, we have precise bounds on the distance, and one can use the relation between list decoding radius and distance to understand the list decoding potential of these codes; and (b) this shows that one approach to construct good list decodable codes is to construct large distance codes, and the latter is a relatively well-studied and better understood problem. Also, historically the most significant algorithmic results on list decoding have been fueled by an attempt to decode codes whose good minimum distance highlighted their good combinatorial list decodability properties.

3.1 Introduction

In order to perform list decoding up to a certain number, say e, errors efficiently, we need the guarantee that every Hamming ball of radius e has a "small" number of codewords. This is because the list decoding algorithm will have a runtime that is at least the size of the list it outputs, and we want the algorithm to be efficient even for the worst-case error pattern. The exact size of the list can be either set to a suitably large constant (independent of the blocklength), or to a fixed polynomial function of the blocklength.

Unique decoding is based upon the fact that in a code of minimum distance d any Hamming ball of radius less than $d/2$ can have at most one codeword. For list decoding we would like upper bounds on the number of codewords in a ball of radius e for e larger than $d/2$. A classical bound in coding theory, called the Johnson bound [108, 109] (see also [132]), proves an upper bound on the number of codewords at a Hamming distance exactly e from an arbitrary word, as long as e is less than a certain function of the distance and blocklength of the code. Such a bound is of direct interest to constant-weight codes (which are codes all of whose codewords have the same Hamming weight), and is also used in the Elias-Bassalygo upper bound on the dimension of codes with certain minimum distance.

V. Guruswami: List Decoding of Error-Correcting Codes, LNCS 3282, pp. 33-44, 2004.

For purposes of list decoding, we need a Johnson-style bound for the number of codewords at a distance of at most e (not exactly e) from a received word. In this chapter, we present a very general version of such a bound. Owing to their strong resemblance to the Johnson bound, we call our bounds Johnson-type (or simply, Johnson) bounds. The main result of this chapter is the fact any q-ary code of blocklength n and distance d is list decodable with "small" lists for up to $e_J(n, d, q)$ errors, where $e_J(n, d, q)$ is a function only of n, d, q (and *not* the structure of the code). We call this quantity $e_J(n, d, q)$ the "Johnson bound on list decoding radius" or "Johnson radius" of the code, and it is always greater than $d/2$.

Proofs of the Johnson bound seem to come in one of two flavors. The original proof and some of its derivatives follow a linear algebra based argument [108, 109, 50, 73, 89], while more recent proofs, most notably [128, 53, 1] are more geometric. Our proof follows the latter spirit, extending these proofs to the case of general alphabets.

Moreover, our techniques easily allow us to extend our results and also prove a weighted version of the Johnson bound which is of interest to some questions raised by the investigations on "soft" list decoding algorithms (more details on this and the connection to soft decoding will be discussed in later chapters in Part II of the book).

3.2 Definitions and Notation

We first recall some notation. For $\mathbf{x}, \mathbf{y} \in [q]^n$ the Hamming distance between $\mathbf{x}$ and $\mathbf{y}$ is denoted $\Delta(\mathbf{x}, \mathbf{y})$. For $\mathbf{r} \in [q]^n$ and $0 \leq e \leq n$, the Hamming ball of radius e around $\mathbf{r}$ is defined by $B_q(\mathbf{r}, e) = \{\mathbf{x} \in [q]^n : \Delta(\mathbf{r}, \mathbf{x}) \leq e\}$.

The key quantity to study in our context is the following. Let $A'_q(n, d, e)$ denote the maximum number of points that may be placed in some ball $B_q(\mathbf{r}, e)$ such that all pairwise distances between the points are at least d. More formally,

$$A'_q(n, d, e) = \max\{|S| : S \subseteq B_q(\mathbf{r}, e) \text{ for some } \mathbf{r} \in [q]^n \text{ and } \forall \mathbf{x}, \mathbf{y} \in S, \Delta(\mathbf{x}, \mathbf{y}) \geq d\} \ . \quad (3.1)$$

(We use the notation $A'_q(n, d, e)$ instead of the apparently more natural choice $A_q(n, d, e)$ because the notation $A_q(n, d, e)$ in coding theory literature normally refers to the maximum number of points (with pairwise distances at least d) that may be placed **on** the surface of (instead of within) the ball $B_q(\mathbf{r}, e)$. To avoid confusion with this standard terminology, we use $A'_q(n, d, e)$ instead. We clearly have $A_q(n, d, e) \leq A'_q(n, d, e)$, and thus any upper bound we derive on $A'_q(n, d, e)$ also applies to $A_q(n, d, e)$.)

Clearly for any code $\mathcal{C} \subseteq [q]^n$ of minimum distance d, $A'_q(n, d, e)$ is an upper bound on the number of codewords of $\mathcal{C}$ that can lie in a Hamming

ball of radius e. Hence, our objective in this chapter is to obtain an upper bound on the function $A'_q(n, d, e)$.

It is common practice to denote these functions as $A(n, d, e)$ and $A'(n, d, e)$ for the binary ($q = 2$) case.

3.3 The Johnson Bound on List Decoding Radius

Theorem 3.1 ([91, 1]). *Let $\mathcal{C}$ be any q-ary code of blocklength n and minimum distance $d = (1 - 1/q)(1 - \delta)n$ for some $0 < \delta < 1$. Let $e = (1 - 1/q)(1 - \gamma)n$ for some $0 < \gamma < 1$ and let $\mathbf{r} \in [q]^n$ be arbitrary. Then, provided $\gamma > \sqrt{\delta}$, we have*

$$|B_q(\mathbf{r}, e) \cap \mathcal{C}| \leq \min\Big\{n(q-1), \frac{1-\delta}{\gamma^2 - \delta}\Big\} \ . \tag{3.2}$$

Furthermore, for the case when $\gamma = \sqrt{\delta}$, we have $|B_q(\mathbf{r}, e) \cap \mathcal{C}| \leq 2n(q-1) - 1$.

The theorem below is merely a restatement of the above result in different notation, and follows immediately from the above result (it is a straightforward calculation to check this).

Theorem 3.2. *Let q, n, d be arbitrary positive integers with $d < (1 - 1/q)n$.*

(i) Let $e \geq 1$ be any integer that satisfies the condition

$$e < e_J(n, d, q) \stackrel{\text{def}}{=} \Big(1 - \frac{1}{q}\Big)\Big(1 - \sqrt{1 - \frac{q}{q-1} \cdot \frac{d}{n}}\Big)n \ . \tag{3.3}$$

Then we have

$$A'_q(n, d, e) \leq \min\{n(q-1), \frac{nd}{nd - 2e\big(n - \frac{qe}{2(q-1)}\big)}\} \ . \tag{3.4}$$

In other words, for an integer $L \geq 1$, if

$$e \leq e_J(n, d, q, L) \stackrel{\text{def}}{=} n\Big(1 - \frac{1}{q}\Big)\Big(1 - \sqrt{1 - \frac{q}{q-1}\frac{L-1}{L}\frac{d}{n}}\Big) \ , \tag{3.5}$$

then $A'_q(n, d, e) \leq L$.

(ii) Furthermore, if $e = e_J(n, d, q)$, then $A'_q(n, d, e) \leq 2n(q-1) - 1$.

The above theorem says that a q-ary code of blocklength n and distance d can be list decoded with small lists for up to $e_J(n, d, q)$ errors. For purposes of easy future reference, we give the quantity $e_J(n, d, q)$ the label *"Johnson bound on list decoding radius"* , or simply the *"Johnson radius"* of a code.

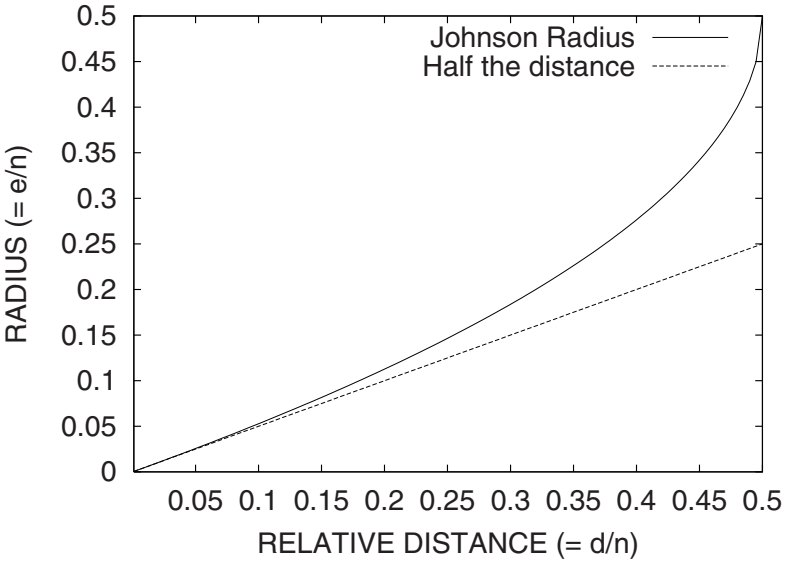

Fig. 3.1. Plot of Johnson radius as a function of relative distance for binary codes. This shows that list decoding always permits decoding beyond half the distance.

When we want to make the alphabet size explicit, we will refer to $e_J(n, d, q)$ as the *"q-ary Johnson radius"*. For decoding with lists of size L, we give the quantity $e_J(n, d, q, L)$ the label *"Johnson radius for list-of-L decoding"*.

It is easy to verify that the Johnson radius $e_J(n, d, q)$ defined in Equation (3.3) satisfies

$$e_J(n, d, q) > d/2$$

for every n, d, q with $1 \leq d \leq (1 - 1/q)n$. This captures the claim that *list decoding with polynomial-sized lists always permits one to decode beyond half the distance*. As an illustration, we plot the Johnson radius for binary codes in Figure 3.1, normalized by blocklength, as a function of the relative distance of the code. Note for any every value of the relative distance δ in the range $0 < \delta < 1/2$, the Johnson radius is strictly greater than half the minimum distance.

Before moving on to the proof of Theorem 3.1, we state the following corollary to the above statement. This gives a (weaker) version of the above bounds that ignores the alphabet size q of the code. But it has a simpler, easily stated form, and for large q approaches the above bounds.

Corollary 3.3. *Let q, n, d, e be arbitrary positive integers with $e \leq d \leq n$.*

(i) If $e < n - \sqrt{n(n-d)}$, then $A'_q(n, d, e) \leq n(q-1)$.
(ii) if $e \leq n - \sqrt{n(n-d+d/L)}$, then $A'_q(n, d, e) \leq L$.

Proof: The proof follows from Theorem 3.2 and the fact that

$$(1-\sqrt{1-x}) \leq (1-1/q)\big(1-\sqrt{1-\frac{qx}{q-1}}\big)$$

for every integer q and every x, $0 \leq x \leq (1-1/q)$. The above inequality can be proved using a straightforward calculation. Using the above inequality with $x = d/n$ and $x = \frac{L-1}{L}\frac{d}{n}$ implies that the conditions on e stated in the corollary imply the Conditions (3.3) and (3.5) respectively. □

3.3.1 Proof of Theorem 3.1

Proof Idea: The proof follows a "geometric" approach. We identify elements of $[q]^n$ with vectors in $\mathbb{R}^{nq}$ by replacing the symbol i $(1 \leq i \leq q)$ by the unit vector of length q with a 1 in position i. This allows us to embed the codewords and the "received" word $\mathbf{r}$ into $\mathbb{R}^{nq}$. Next, by appropriately shifting the set of vectors corresponding to the codewords that are close to $\mathbf{r}$, we get a set of vectors such that the inner product of any two distinct vectors from this set is non-positive. By a standard geometric upper bound on the cardinality of such a set of vectors, we get the required upper bound on the number of codewords that are "close" to $\mathbf{r}$.

Our idea extends proofs for the binary case, given by [53, 128, 1]. These works used an appropriate embedding of the binary codewords in $\mathbb{R}^n$ and an appropriate shifting of vectors to establish "Johnson-style" bounds by appealing to bounds on spherical codes, i.e., bounds on the cardinality of a set of unit vectors in real space with a specified minimum angle between any pair of vectors. It may be noted that the generalization to arbitrary alphabets is not automatic. (Of the several potential approaches, our proof hits upon the right path.)

Proof of Theorem 3.1: Assume without loss of generality that $\mathbf{r} = \langle q, q, \ldots, q \rangle$, i.e is the symbol q repeated n times. Let $C_1, C_2, \ldots, C_m$ be all the codewords of $\mathcal{C}$ that lie within $B_q(\mathbf{r}, e)$ where $e = (1-1/q)(1-\gamma)n$. Our goal is to get an upper bound on m provided γ is large enough.

We associate a vector in $\mathbb{R}^{nq}$ with $\mathbf{r}$ and with each codeword C_i. Each vector is to be viewed as having n blocks each having q components (the n blocks correspond to the n codeword positions). For $1 \leq l \leq q$, denote by $\hat{e}_l$ the q-dimensional unit vector with 1 in the lth position and 0 elsewhere. For $1 \leq i \leq m$, the vector $\mathbf{c_i}$ associated with the codeword C_i has in its jth block the components of the vector $\hat{e}_{C_i[j]}$ ($C_i[j]$ is the jth symbol of C_i, treated as an integer between 1 and q). The vector associated with the received word $\mathbf{r}$, which we also denote $\mathbf{r}$ by abuse of notation, is defined similarly. Let $\mathbf{1} \in \mathbb{R}^{nq}$ be the all 1's vector. Now define $\mathbf{v} = \alpha\mathbf{r} + \frac{(1-\alpha)}{q}\mathbf{1}$ for a parameter $0 \leq \alpha \leq 1$ to be specified later in the proof. Note that the $\mathbf{c_i}$'s and $\mathbf{v}$ all lie in the space defined by the intersection of the n "hyperplanes" $\{\ \mathcal{H}'_j : \ \sum_{\ell=1}^{q} x_{j,\ell} = 1\ \}$ for

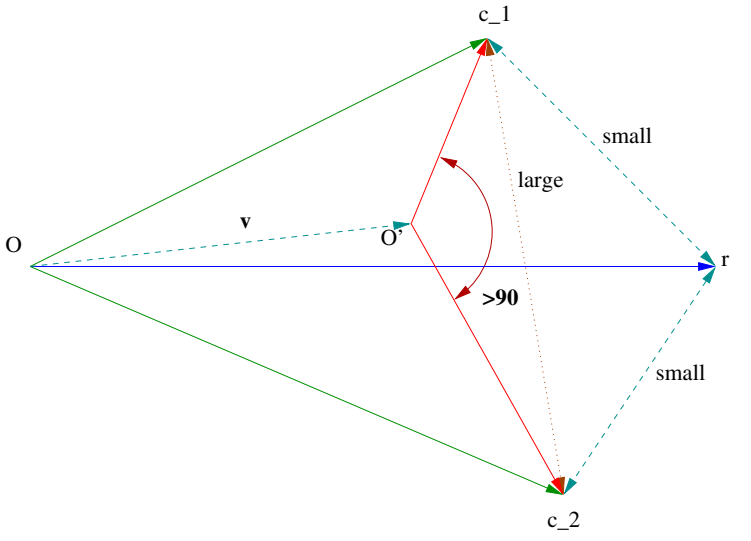

Fig. 3.2. Geometric picture behind proof of Theorem 3.1

$1 \le j \le n$. Hence the vectors $(\mathbf{c_i} - \mathbf{v})$, for $1 \le i \le m$, all lie in $\mathcal{H} = \bigcap_{j=1}^{n} \mathcal{H}_j$ where $\mathcal{H}_j = \{\mathbf{x} \in \mathbb{R}^{nq} : \sum_{\ell=1}^{q} x_{j,\ell} = 0\}$. It is easy to see that $\mathcal{H}$ is an $n(q-1)$-dimensional subspace of $\mathbb{R}^{nq}$. We thus conclude that the vectors $(\mathbf{c_i} - \mathbf{v})$, $1 \le i \le m$, all lie in an $n(q-1)$-dimensional space.

The idea behind the rest of the proof is the following. We will pick α so that the vectors $(\mathbf{c_i} - \mathbf{v})$, for $1 \le i \le m$, have all pairwise dot products less than 0. Geometrically speaking, we shift the origin O to O' where $OO' = \mathbf{v}$, and require that relative to the new origin the vectors corresponding to the codewords have pairwise angles which are greater than 90 degrees (see Figure 3.2). By a simple geometric fact (stated in Lemma 3.4 below), it will then follow that the number of codewords m is at most the dimension $n(q-1)$ of the space in which all these vectors lie.

For $1 \le i \le m$, let $e_i = \Delta(\mathbf{r}, C_i)$. Note that $e_i \le e$ for every i. Now

$$\langle \mathbf{c_i}, \mathbf{v} \rangle = \alpha \langle \mathbf{c_i}, \mathbf{r} \rangle + \frac{(1-\alpha)}{q} \langle \mathbf{c_i}, \mathbf{1} \rangle = \alpha(n - e_i) + (1-\alpha)\frac{n}{q} \tag{3.6}$$

$$\langle \mathbf{v}, \mathbf{v} \rangle = \alpha^2 n + 2(1-\alpha)\alpha\frac{n}{q} + (1-\alpha)^2\frac{n}{q} = \frac{n}{q} + \alpha^2\Big(1 - \frac{1}{q}\Big)n \tag{3.7}$$

$$\langle \mathbf{c_i}, \mathbf{c_j} \rangle = n - \Delta(C_i, C_j) \le n - d \ . \tag{3.8}$$

Using (3.6), (3.7) and (3.8), and the fact that each $e_i \le e$, we get, for $i \ne j$,

$$\langle \mathbf{c_i} - \mathbf{v}, \mathbf{c_j} - \mathbf{v} \rangle \le 2\alpha e - d + \Big(1 - \frac{1}{q}\Big)(1-\alpha)^2 n \ . \tag{3.9}$$

Using $e = (1-1/q)(1-\gamma)n$ and $d = (1-1/q)(1-\delta)n$ the above simplifies to

$$\langle \mathbf{c_i} - \mathbf{v}, \mathbf{c_j} - \mathbf{v} \rangle \leq \Big(1 - \frac{1}{q}\Big) n \Big(\delta + \alpha^2 - 2\alpha\gamma\Big) \tag{3.10}$$

Thus as long as $\gamma > \frac{1}{2}\Big(\frac{\delta}{\alpha} + \alpha\Big)$ we will have all pairwise dot products to be negative just as we wanted. We pick α to minimize $\big(\frac{\delta}{\alpha} + \alpha\big)$, or in other words we set $\alpha = \sqrt{\delta}$. Now as long as $\gamma > \sqrt{\delta}$, we will have $\langle \mathbf{c_i} - \mathbf{v}, \mathbf{c_j} - \mathbf{v} \rangle < 0$ for all $1 \leq i < j \leq m$. To complete the proof, we note that (for the choice $\alpha = \sqrt{\delta}$), for every $1 \leq i \leq m$, $\langle \mathbf{c_i} - \mathbf{v}, \mathbf{v} \rangle \geq (1-1/q)n\sqrt{\delta}(\gamma - \sqrt{\delta}) > 0$ (this is easily checked using (3.6) and (3.7)). Thus provided $\gamma > \sqrt{\delta}$, we have $\langle \mathbf{c_i} - \mathbf{v}, \mathbf{v} \rangle > 0$ for $1 \leq i \leq m$. Now applying Part (iii) of Lemma 3.4, with the setting $\mathbf{v_i} = \mathbf{c_i} - \mathbf{v}$ and $\mathbf{u} = \mathbf{v}_{|\mathcal{H}}$, the projection of $\mathbf{v}$ onto the subspace $\mathcal{H}$, implies that $m \leq n(q-1)$ (recall that the vectors $(\mathbf{c_i} - \mathbf{v})$, $1 \leq i \leq m$, all lie in $\mathcal{H}$ and $\dim(\mathcal{H}) = n(q-1)$).

We now prove that if $\gamma > \sqrt{\delta}$, then $m \leq \frac{1-\delta}{\gamma^2 - \delta}$. For this we set $\alpha = \gamma$. Now from Equation (3.10) we have

$$\langle \mathbf{c_i} - \mathbf{v}, \mathbf{c_j} - \mathbf{v} \rangle \leq (1-1/q)n(\delta - \gamma^2) \ . \tag{3.11}$$

Thus if $\gamma > \sqrt{\delta}$, we have $\langle \mathbf{c_i} - \mathbf{v}, \mathbf{c_j} - \mathbf{v} \rangle < 0$. Now for the choice $\alpha = \gamma$, we have for each i, $1 \leq i \leq m$,

$$\|\mathbf{c_i} - \mathbf{v}\|^2 = \langle \mathbf{c_i} - \mathbf{v}, \mathbf{c_i} - \mathbf{v} \rangle \leq 2\alpha e + (1-1/q)(1-\alpha)^2 n = n(1-1/q)(1-\gamma^2) \ .$$

Denote by $\mathbf{w_i}$ the unit vector $-\frac{\mathbf{c_i} - \mathbf{v}}{\|\mathbf{c_i} - \mathbf{v}\|}$. We then have

$$\langle \mathbf{w_i}, \mathbf{w_j} \rangle \leq -\frac{\gamma^2 - \delta}{1 - \gamma^2} \tag{3.12}$$

for $1 \leq i < j \leq m$ (this follows from (3.11) and (3.12)). By a well-known geometric fact (see Lemma 3.5 for the simple proof), it follows that the number of such vectors, m, is at most $\big(1 + \frac{1-\gamma^2}{\gamma^2 - \delta}\big) = \frac{1-\delta}{\gamma^2 - \delta}$, as desired.

To handle the case when $\gamma = \sqrt{\delta}$, we can choose $\alpha = \sqrt{\delta}$, and we then have $\langle \mathbf{c_i} - \mathbf{v}, \mathbf{c_j} - \mathbf{v} \rangle \leq 0$ for all $1 \leq i < j \leq m$, and also $\langle \mathbf{c_i} - \mathbf{v}, \mathbf{v} \rangle \geq 0$ for each $i = 1, 2, \ldots, m$. Now applying Part (ii) of Lemma 3.4, we get $m \leq 2n(q-1) - 1$. □

3.3.2 Geometric Lemmas

We now state and prove the geometric facts that were used in the above proof.

Lemma 3.4. *Let $\mathbf{v_1}, \ldots, \mathbf{v_m}$ be non-zero vectors in $\mathbb{R}^N$ such that $\langle \mathbf{v_i}, \mathbf{v_j} \rangle \leq 0$ for all $1 \leq i < j \leq m$. Then the following hold:*

(i) $m \leq 2N$.
(ii) *Suppose that there exists a non-zero* $\mathbf{u} \in \mathbb{R}^N$ *such that* $\langle \mathbf{u}, \mathbf{v_i} \rangle \geq 0$ *for* $i = 1, 2, \ldots, m$. *Then* $m \leq 2N - 1$.
(iii) *Suppose there exists an* $\mathbf{u} \in \mathbb{R}^N$ *such that* $\langle \mathbf{u}, \mathbf{v_i} \rangle > 0$ *for* $i = 1, 2, \ldots, m$. *Then* $m \leq N$.

A proof of Part (i) of the above lemma can be found, for instance, in [30, Chapter 10, page 71]. The proofs of the other two parts are similar. For completeness, we present a self-contained proof below.

Proof of Lemma 3.4: We first prove (iii). Suppose for contradiction that $m \geq N + 1$. Then since the vectors $\mathbf{v_1}, \ldots, \mathbf{v_m}$ all lie in $\mathbb{R}^N$, they must be linearly dependent. Let $S \subseteq [m]$ be a non-empty set of *minimum* size for which a relation of the form $\sum_{i \in S} a_i \mathbf{v_i} = \mathbf{0}$ holds with each $a_i \neq 0$. We claim that the a_i's must all be positive or all be negative. Indeed, if not, by collecting terms with positive a_i's on one side and those with negative a_i's on the other, we will have an equation of the form $\sum_{i \in T^+} a_i \mathbf{v_i} = \sum_{j \in T^-} b_j \mathbf{v_j} = \mathbf{w}$ (for some vector $\mathbf{w}$) where T^+ and T^- are *disjoint* non-empty sets with $T^+ \cup T^- = S$, and all $a_i, b_j > 0$. By the minimality of S, $\mathbf{w} \neq \mathbf{0}$ and hence $\langle \mathbf{w}, \mathbf{w} \rangle > 0$. On the other hand $\langle \mathbf{w}, \mathbf{w} \rangle = \langle \sum_{i \in T^+} a_i \mathbf{v_i}, \sum_{j \in T^-} b_j \mathbf{v_j} \rangle = \sum_{i,j} a_i b_j \langle \mathbf{v_i}, \mathbf{v_j} \rangle \leq 0$ since $a_i b_j > 0$ and $\langle \mathbf{v_i}, \mathbf{v_j} \rangle \leq 0$ for each $i \in T^+$ and $j \in T^-$. This contradiction shows that we may assume that $a_i > 0$ for all $i \in S$.

Now $\sum_{i \in S} a_i \mathbf{v_i} = \mathbf{0}$, so that $\sum_{i=1}^{s} a_i \langle \mathbf{u}, \mathbf{v_i} \rangle = 0$. But this is impossible since for each i we have $a_i > 0$ and $\langle \mathbf{u}, \mathbf{v_i} \rangle > 0$. We have thus arrived at a contradiction, and therefore such a linear dependence $\sum_{i \in S} a_i \mathbf{v_i} = \mathbf{0}$ does not exist. Thus the vectors $\mathbf{v_1}, \mathbf{v_2}, \ldots, \mathbf{v_m}$ are linearly independent and we must have $m \leq N$.

To prove (ii), we use induction on N. The statement clearly holds for $N = 1$. For $N > 1$, we proceed exactly as above. If $m \leq N$, we have nothing to prove, so assume $m > N$ so that $\mathbf{v_1}, \ldots, \mathbf{v_m}$ are linearly independent, and as above, let $S \subseteq [m]$ be a non-empty set of minimum size for which a relation of the form $\sum_{i \in S} a_i \mathbf{v_i} = \mathbf{0}$ holds with each $a_i \neq 0$. Arguing as above, we may assume that $a_i > 0$ for every $i \in S$. Assume for definiteness that $S = \{1, 2, \ldots, s\}$. We thus have the linear dependence $\sum_{i=1}^{s} a_i \mathbf{v_i} = \mathbf{0}$ with each $a_i > 0$, and since this is a minimum sized linear dependence, $\mathbf{v_1}, \ldots, \mathbf{v_s}$ must span a subspace W of $\mathbb{R}^N$ of dimension $(s - 1)$.

Since $\sum_{i=1}^{s} a_i \mathbf{v_i} = \mathbf{0}$, we have $\sum_{i=1}^{s} a_i \langle \mathbf{v_i}, \mathbf{v}_\ell \rangle = 0$ for each $\ell = s + 1, \ldots, m$. Since $a_i > 0$ for $1 \leq i \leq s$ and $\langle \mathbf{v_i}, \mathbf{v}_\ell \rangle \leq 0$, it must be therefore be the case that $\mathbf{v_i}$ is orthogonal to $\mathbf{v}_\ell$ for all i, ℓ with $1 \leq i \leq s$ and $s < \ell \leq m$. A similar argument shows $\mathbf{u}$ is orthogonal to $\mathbf{v_i}$ for each $i = 1, 2, \ldots, s$. Thus the vectors $\mathbf{v_{s+1}}, \ldots, \mathbf{v_m}$ and $\mathbf{u}$ all lie in $W^\perp$ which has dimension equal to $(N - s + 1)$. Since $s > 1$, the induction hypothesis applied to these vectors implies that $m - s \leq 2(N - s + 1) - 1$, or in other words $m \leq 2N - s + 1 \leq 2N - 1$, as desired.

Finally (i) follows immediately from (ii). Indeed, apply (ii) with vectors $\mathbf{v_1}, \ldots, \mathbf{v_{m-1}}$ and $-\mathbf{v_m}$ playing the role of $\mathbf{u}$. This implies $m - 1 \leq 2N - 1$, or in other words $m \leq 2N$. □

Lemma 3.5. *Let $\varepsilon > 0$ be a positive real and let $\mathbf{w_1}, \mathbf{w_2}, \ldots, \mathbf{w_m}$ be m unit vectors such that $\langle \mathbf{w_i}, \mathbf{w_j} \rangle \leq -\varepsilon$ for all $1 \leq i < j \leq m$. Then $m \leq 1 + \frac{1}{\varepsilon}$.*

Proof: We have

$$0 \leq \langle \sum_{i=1}^{m} \mathbf{w_i}, \sum_{i=1}^{m} \mathbf{w_i} \rangle = \sum_{i=1}^{m} \langle \mathbf{w_i}, \mathbf{w_i} \rangle + 2 \sum_{1 \leq i < j \leq m} \langle \mathbf{w_i}, \mathbf{w_j} \rangle \leq m - m(m-1)\varepsilon \ ,$$

which gives $m \leq 1 + 1/\varepsilon$. □

3.4 Generalization in Presence of Weights

For applications to "soft" list decoding algorithms which will be discussed in Part II of the book, it is of interest to prove a version of the Johnson bound in the presence of weights on codeword symbols. Such a bound is also of independent interest, since it covers the case of decoding under errors-and-erasures and the case when for each position one receives a small list of candidate symbols one of which is the correct one, all under a uniformly applicable bound.

We next state the weighted version of the Johnson bound that follows from our proof technique. The bound in Part (i) of the theorem generalizes the result of Theorem 3.2. The result from Part (ii) applies under a more general condition than Condition (3.3) (or even Condition (3.13)), but the upper bound itself is slightly weaker (since it is $(nq - 1)$ instead of $n(q - 1)$). The result of Part (iii) generalizes the result of Theorem 3.2, Condition 3.5.

Theorem 3.6. *Let $\mathcal{C} \subseteq [q]^n$ be a code of blocklength n and minimum distance d. Let $\{w_{i,j} : 1 \leq i \leq n; 1 \leq j \leq q\}$ be an arbitrary set of non-negative real weights. Define $W_i = \sum_{j=1}^{q} w_{i,j}$ and $W_i^{(2)} = \sum_{j=1}^{q} w_{i,j}^2$, $W_{\mathrm{tot}} = \sum_{i,j} w_{i,j}$, and $W_{\mathrm{tot}}^{(2)} = \sum_{i,j} w_{i,j}^2$. Then:*

(i) The number of codewords $C \in \mathcal{C}$ that satisfy

$$\sum_{i=1}^{n} \frac{w_{i,C_i}}{W_i} > \frac{n}{q} + \sqrt{\left(n\Big(1 - \frac{1}{q}\Big) - d\right)\left(\sum_{i=1}^{n} \frac{W_i^{(2)}}{W_i^2} - \frac{n}{q}\right)} \ . \tag{3.13}$$

is at most $n(q - 1)$.

(ii) The number of codewords $C \in \mathcal{C}$ that satisfy

$$\sum_{i=1}^{n} w_{i,C_i} > \frac{W_{\mathrm{tot}}}{q} + \sqrt{\left(n\Big(1 - \frac{1}{q}\Big) - d\right)\left(W_{\mathrm{tot}}^{(2)} - \frac{(W_{\mathrm{tot}})^2}{nq}\right)} \tag{3.14}$$

is at most $(nq - 1)$.

(iii) For any integer $L \geq 2$, the number of codewords $C \in \mathcal{C}$ that satisfy

$$\sum_{i=1}^{n} w_{i,C_i} \geq \frac{W_{\mathrm{tot}}}{q} + \sqrt{\left(n\Big(1-\frac{1}{q}\Big) - d + \frac{d}{L}\right)\left(W_{\mathrm{tot}}^{(2)} - \frac{(W_{\mathrm{tot}})^2}{nq}\right)} \quad (3.15)$$

is at most L.

Proof: We do not give a full proof here, rather we indicate the only changes that must be made to the proof of Theorem 3.1 in order to prove our claim. For Part (i), the only modification required in the proof of Theorem 3.1 is to pick $\mathbf{r}$ so that its (i,j)'th component, for $1 \leq i \leq n$ and $1 \leq j \leq q$, equals $\frac{w_{i,j}}{W_i}$. The vector $\mathbf{v}$ is defined as before to be $\alpha\mathbf{r} + \frac{(1-\alpha)}{q}\mathbf{1}$ for

$$\alpha = \sqrt{\frac{n(1-1/q)-d}{\sum_i \frac{W_i^{(2)}}{W_i^2} - n/q}}\ .$$

Once once again all the vectors $(\mathbf{c_i} - \mathbf{v})$ lie in an $n(q-1)$-dimensional subspace of $\mathbb{R}^{nq}$. It can be proved as in the proof of Theorem 3.1 that these vectors have pairwise non-positive dot products, which gives the desired $n(q-1)$ upper bound on the number of codewords.

For Parts (ii) and (iii), we pick $\mathbf{r}$ so that its (i,j)'th component for $1 \leq i \leq n$ and $1 \leq j \leq q$, equals $\frac{nw_{i,j}}{W_{\mathrm{tot}}}$, and the rest of the proof follows that of Theorem 3.1. Note that W_{tot}/q is the expected value of $\sum_i w_{i,r_i}$ for a random vector $\mathbf{r} \in [q]^n$, and $\big(W_{\mathrm{tot}}^{(2)} - \frac{(W_{\mathrm{tot}})^2}{nq}\big)$ is proportional to the variance of the $w_{i,j}$'s. Thus, the above theorem states that the number of codewords which have weighted agreement bounded away from the expectation by a certain number of standard deviations is small. The upper bound of $(nq-1)$ (instead of $n(q-1)$) in Part (ii) of above theorem arises since we are only able to ensure that the vectors $(\mathbf{c_i} - \mathbf{v})$ all lie in an $(nq-1)$-dimensional subspace (namely that defined by $\sum_{i,j} x_{i,j} = 0$), and not an $n(q-1)$-dimensional subspace as in Part (i). □

We now state a corollary similar to Corollary 3.3 that ignores the alphabet size in the decoding condition. The proof again follows because it can be verified (after a straightforward but tedious calculation) that the stated conditions in fact imply the Conditions (3.14) and (3.15) above.

Corollary 3.7. *Let $\mathcal{C} \subseteq [q]^n$ be a code of blocklength n and minimum distance d. Let $\{w_{i,j} : 1 \leq i \leq n; 1 \leq j \leq q\}$ be an arbitrary set of non-negative real weights.*

(i) The number of codewords $C \in \mathcal{C}$ that satisfy

$$\sum_{i=1}^{n} w_{i,C_i} > \Big((n-d)\sum_{i,j} w_{i,j}^2\Big)^{1/2} \quad (3.16)$$

is at most $(nq-1)$.

(ii) For any integer $L \geq 2$, the number of codewords $C \in \mathcal{C}$ that satisfy

$$\sum_{i=1}^{n} w_{i,C_i} \geq \left(\left(n - d + \frac{d}{L} \right) \sum_{i,j} w_{i,j}^2 \right)^{1/2} \tag{3.17}$$

is at most L.

A bound similar to Corollary 3.7 above can also be worked out for the case when the different codeword positions have different contributions towards the minimum distance. Such a bound is of interest for certain codes like the Chinese Remainder Code and will be stated and formally proved in the form of Theorem 7.10 in Section 7.6.1 of the book. We refer the reader interested in seeing a full proof of Corollary 3.7 above to the proof of Theorem 7.10.

3.5 Notes

The quantity $A(n, d, w)$ for constant-weight binary codes has a rich history and has been studied for almost four decades, and its study remains one of the most basic questions in coding theory. The first upper bounds on the quantity $A(n, d, w)$ for constant-weight codes appear in the work of Johnson [108, 109]. Since then several proofs have appeared in the literature, including generalizations of the bound to the case of q-ary alphabets for $q > 2$ (cf. [34] for a discussion and detailed bibliography).

The quantity $A'(n, d, e)$, which is of more direct interest to list decoding, seems to have received much less explicit attention. It must be said that several proofs that provide upper bounds on $A(n, d, e)$ work with little or no modification to yield upper bounds on $A'(n, d, e)$ as well. This was made explicit for example in [50, 22]. Upper bounds on $A'_q(n, d, e)$ identical to the second upper bound in (3.4) of this chapter are stated in [34]. Proofs of such bounds that follow a linear algebra based argument appear, for instance, in [73, 89].

The contribution of the results in this chapter is that we extend the more recent upper bounds for the binary case from [1] (which are based on geometric arguments) to bounds on $A'_q(n, d, e)$, and furthermore we obtain some elegant weighted generalizations of the Johnson bound. In particular, the upper bound $A'_q(n, d, e) \leq n(q - 1)$ for $e < e_J(n, d, q)$ that we proved in Theorem 3.2 appears to be new. For the case $q = 2$, this result was known. Specifically, Elias [48] proved that if d is *odd*, then $A'(n, d, e) \leq n$ as long as e is at most the binary Johnson radius $e_J(n, d, 2)$. For even d, however, $A'(n, d, e) = O(n^2)$ was the best known bound that was made explicit till the recent work of Agrell, Vardy and Zeger [1], who showed that $A'(n, d, e) \leq n$ whenever $e < e_J(n, d, 2)$. (Actually, Agrell et al claim their result only for $A(n, d, e)$, but their proof works for the case of $A'(n, d, e)$ as well.)

Combinatorial results of a flavor similar to this chapter appear in two other parts of the book: (a) in Section 7.6.1 where a bound similar to Corollary 3.7 is proved for the case when the minimum distance is measured with a non-uniform weight on the codeword positions, and (b) in Section 8.5.1 where we prove a result along the lines of Theorem 3.2, but instead of bounding the number of codewords in a Hamming ball of certain radius, we establish a more general result concerning the coset weight distribution of a code, purely as a function of its minimum distance.

The material in this chapter appears in [91].