List and local list decoding of error-correcting codes

Satyadev Nandakumar

September 11, 2026

Contents

1 The need for list decoding

Using unique decoding - that is, if we want to ensure that there is exactly one original message corresponding to the erroneous received message, then the fractional Hamming distance can be at most 1/2. This means that at most 1/4 fraction of the message can be erroneous, if we want to uniquely decode.

This is a strict barrier. However, a breakthrough was achieved when we realized the following - suppose we want to recover from a message that has 40 percent error, there are *not many* original codewords corresponding to that received message. So we can output a *short list* of possible original messages. This approach is called *list decoding*.

First, we have the formal bound on how the number of possible messages.

Theorem 1.1. *Johnson bound, 1962 If $E : \{0,1\}^n \to \{0,1\}^m$ is an error-correcting code with distance at least $\frac{1}{2} - \epsilon$ for some $\epsilon > 0$ and $\delta > \sqrt{epsilon}$, then for every $w \in \{0,1\}^m$, there are at most $\frac{1}{2\delta^2}$ vectors $y_1, \ldots, y_\ell$ such that for every i, $\Delta(y_i, x) \leq 1/2 - \delta$.*

For example, let $\epsilon = \frac{1}{n^2}$ and $\delta = \frac{1}{n}$. Then the number of possible strings in the list is at most $\frac{1}{2}n^2$ according to the above bound. In general, if ϵ is inverse polynomial in n, the length of the list is some polynomial in n.

Proof. Suppose w, $y_1, \ldots, y_\ell \in \{0,1\}^m$ satisfy this condition. Now, define ℓ vectors $z_1, \ldots, z_\ell \in \{0,1\}^m$ as follows. For each $1 \leq i \leq \ell$ and position $1 \leq j \leq n$ in the vector z_i, let $z_i[k] = 1$ if $y_i[k] = x[k]$ and set $z_i[k] = -1$ otherwise.

By assumption, each y_i agrees with x on $1/2 + \delta$ coordinates. Hence for every $1 \leq i \leq \ell$, we have

$$z_i[1] + \cdots + z_i[m] = |\{k \mid y_i[k] = w[k]\}| \tag{1}$$
$$\geq 2\delta m. \tag{2}$$

Also, E is a code with fractional Hamming distance at least $1/2 + \epsilon$. Hence, for every distinct i, j between 1 and ℓ, we have

$$\langle z_i, z_j \rangle = \sum_{k=1} m z_i[k] z_j[k] \leq 2\epsilon m \leq 2\delta^2 m. \tag{3}$$

We now show that the above two bounds imply the required bound on ℓ.

Set $v = z_1 + \cdots + z_\ell$. By the first estimate, $v[1] + \ldots v[m] \geq 2\ell\delta m$. Hence $\langle v, v \rangle \geq \frac{|\sum_{k=1}^{m} v[k]|^2}{m} \geq 4\delta^2 m \ell^2$.

Now, consider $\langle v, v \rangle$. By the second estimate above, we have

$$\langle v, v \rangle = \sum_{i=1}^{ell} \langle z_i, z_i \rangle + \sum_{i=1}^{\ell} \sum_{j=1}^{\ell} \delta_{ij} \langle z_i z_j \rangle \leq \ell m + \binom{\ell}{2} 2\delta^2 m \leq \ell m + \ell^2 2\delta^2 m. \tag{4}$$

By putting these two bounds together, we get the required bound on ℓ. $\square$