Oded: October 30, 2016

A natural, but wrong, conjecture is that the generating matrix of a good error correcting code has high rigidity. Zeev Dvir showed me a proof that this conjecture is wrong, and I'm presenting it below.

Theorem 1 (a generating matrix of a good code may have low rigidity): *For every sufficiently small constant $\epsilon > 0$, all sufficiently large k, and every $d \in [k/4]$, there exist an $n = O(k)$ and a k-by-n matrix M such that every non-zero linear combination of the rows of M has Hamming weight $(0.5 \pm \epsilon) \cdot n$ but the matrix M has rigidity at most $O(kn/d)$ with respect to rank $10d \log(k/d)$.*

(For example, setting $d = 0.001k$ corresponds to rigidity $O(n)$ and rank $0.01k$, whereas setting $d = \sqrt{n}$ corresponds to rigidity $O(n^{3/2})$ and rank $\widetilde{O}(\sqrt{n})$.) Let me stress that the low rigidity is a property of a specific generating matrix of the code. In contrast, every linear code of constant rate has a generating matrix of high rigidity. (This is the case because the linear space defined by such a code contains k linearly independent coordinates, and so a random matrix that spans this linear space contains a random k-by-k submatrix.)

Proof: For a fixed $\epsilon > 0$, we fix a sufficiently large k, and let $r = 10d \log(k/d)$ and $m = k + r$. Consider the k-by-m matrix $G = [I|G']$, where I is the identity matrix and G' is a k-by-r matrix such that any linear combination of at most d of its rows yields a vector of weight at least d. (Indeed, a random G' satisfies the latter condition, with overwhelmingly high probability, since $r = 10d \log(k/d) \gg d$.) It follows that the code generated by G has distance at least d (since linear combinations that take at most d rows of G are handled by G' and the other linear combinations are handled by I).

Next, fixing $c = O(\log(1/\epsilon))$ and a sufficiently large $n = O(k)$, we consider a random m-by-n matrix H in which each entry is set to 1 with probability $p = c/d$ independently of all other entries. Denoting the top k rows of H by S and the remaining rows by H' we get $GH = IS + G'H' = S + G'H'$. We shall show that, with overwhelmingly high probability over the choice of H, the matrix GH generates a good code in which all non-zero codewords have weight $(0.5 \pm \epsilon) \cdot n$ although GH does not have rigidity $2p \cdot kn = 2c \cdot kn/d$ with respect to rank $r = 10d \log(k/d)$.

The first claim is proved by observing that each non-zero linear combination of the rows of GH equals a (random) vector v that is the sum of at least d rows of H (since $xGH = yH$ for $y = xG$, which has weight at least d). Now, each entry in v (i.e., each coordinate of yH) is a sum of at least d independent random variables that are each 1 with probability $p = c/d$, independently of all other entries in v. Hence, each entry in v is 1 with probability $0.5 \pm \exp(-\Omega(c))$, and the first claim follows (using the stochastic independence of the entries and $n = \Omega(m/\epsilon^2)$).

The second claim is proved by observing that, with overwhelmingly high probability, the matrix S has weight at most $2p \cdot kn = 2c \cdot kn/d$. On the other hand, G' is an k-by-r matrix, which implies that $G'H'$ has rank at most $r = 10d \log(k/d)$. Hence, $GH = S + G'H'$ does not have rigidity $2c \cdot kn/d$ with respect to rank $10d \log(k/d)$. ■

Acknowledgements. I wish to thank Avishay Tal for helpful comments regarding an earlier version of this write-up.