Today’s topics

**Dimension reduction in $\ell_2$** The following result is known as the Johnson-Lindenstrauss lemma.

**Theorem** (Johnson and Lindenstrauss, 1984). *For every $n$-point subset $X \subset \ell_2$ and every $0 < \varepsilon < 1$, there exists an embedding $f : X \rightarrow \ell_2^k$ with distortion $1 + \varepsilon$ and dimension $k = O(\varepsilon^{-2} \log n)$.*

We will see two (related) proofs, based on measure concentration on the unit sphere, and on Gaussian random variables.

**Reading material.** For more details, see Matousek’s book, and the lecture notes by Goemans and by Barvinok. The course webpage will contain exact details and links for these references.

**Homework**

The first question refers to the April 23 presentation about NNS in high-dimensional space (Gionis, Indyk, and Motwani, VLDB 1999). See the webpage for more details.

1. Recall that the space (storage) required for the bucketing method contained a term of the form $(1/\varepsilon)^d$ (for dimension $d$ and approximation $1 + \varepsilon$). Explain what is the correct bound for $\varepsilon \geq 1$, namely, whether increasing $\varepsilon$ by (say) factor 2 decreases the space requirement.

2. Prove that the $\log(n)$ term in the J-L lemma is necessary, namely: For every integer $n$ there is an $n$-point subset $X \subset \ell_2$ such that every embedding with distortion 2 of $X$ into $\ell_2^k$ requires $k \geq \Omega(\log n)$. 