

# EXPERIMENTAL STUDY OF RANDOM PROJECTIONS BELOW THE JL LIMIT

Xiuyi Ye

Dr. James Keller, Thesis Supervisor

## ABSTRACT

Random projection is a method used to reduce dimensionality of desired objects with pairwise distances preserved at a relatively high probability. The mathematical theory behind this is called the Johnson-Lindenstrauss (JL) lemma. So, the basic idea of the JL lemma is that a set of points in a high dimensional space  $p$  are randomly projected down to a lower dimensional space  $q$ . This  $q$  can be as low as  $q_0$  to still make sure that with a certain probability the projected pairwise distances are within  $1 \pm \epsilon$ , of the pairwise distances before the projection, where  $\epsilon$  is usually a very small value. This technique has already been used in a variety of areas like clustering, image and text data processing. Lots of researchers have already studied the properties and performance of the JL lemma above  $q_0$  ( $q_0$  is usually called the JL limit or JL bound), where  $q = p-1, p-2, \dots, q_0$ , but no research has investigated using the JL lemma below the JL limit ( $q = q_0-1, q_0-2, \dots, 2$ ). With much lower dimension, the data processing, storing almost everything is going to be so much easier. We can visualize the clustering information about data sets in 2D plots. One thing should not be forgotten is that the distance preservation is probabilistic. How well will the distances being preserved below the JL bound? Will it affect or even completely destroy the cluster structure after the projection? What is a good projection method? We are going to study and answer these questions as much as we can in this thesis.