

Title: **Assessing Linearity in High Dimensions**

Author(s): **Art B. Owen**

Technical Report number (Dept. of Statistics, Stanford Univ.): **1999-24**

Date: **October 1999**

Summary:

Some standard numerical problems become intractable in high dimensions. Yet successes are often achieved in practice. This may be explained in terms of the underlying target function being somehow simpler than assumed in the intractability arguments. A prototypical form of simplicity is approximate linearity. In moderate dimensions linearity can be investigated by linear regression. In very high dimensions this becomes computationally inefficient and eventually infeasible, as the cost of regression for n observations in d dimensions grows as nd^2 . This paper presents a quasi-regression method for determining the degree of linearity in a function, where the cost grows only as nd . A bias corrected version of quasi-regression is able to estimate the degree of linearity with a sample size of order $d^{2/3}$. An example is given of a function on $[0, 1]^{1,000,000}$, for which the amount of linear variability is accurately estimated from only 100,000 observations.