Yeonjong Shin
Optimal least squares method for polynomial regression

Department of Mathematics
University of Utah
155S 1400E Rm 233
Salt Lake City
Utah 84112 USA
`shin@math.utah.edu`
Dongbin Xiu

We present a quasi-optimal sample set for least squares regression. The quasi-optimal set is designed in such a way that, for a given number of samples, it can deliver the regression result as close as possible to the result obtained by a (much) larger set of candidate samples. The quasi-optimal set is determined by maximizing a quantity measuring the mutual column orthogonality and the determinant of the model matrix. This procedure is non-adaptive, in the sense that it does not depend on the sample data. This is useful in practice, as it allows one to determine, prior to the potentially expensive data collection procedure, where to sample the underlying system. In addition to presenting the theoretical motivation of the quasi-optimal set, we also present its efficient implementation via a greedy algorithm, along with several numerical examples to demonstrate its efficacy. Since the quasi-optimal set allows one to obtain a near optimal regression result under any affordable number of samples, it notably outperforms other standard choices of sample sets.