

# Group SCAD Regression Analysis for Microarray Time Course Gene Expression Data

Hongzhe Li

Department of Biostatistics and Epidemiology,  
University of Pennsylvania School of Medicine

## Abstract

Since many important biological systems or processes are dynamic systems, it is important to study the gene expression patterns over time in a genomic scale in order to capture the dynamic behavior of gene expression. Microarray technologies have made it possible to measure the gene expression levels of essentially all the genes during a given biological process. In order to determine the transcriptional factors involved in gene regulation during a given biological process, we propose to develop a functional response model with varying coefficients in order to model the transcriptional effects on gene expression levels and to develop a group smoothly clipped absolute deviation (gSCAD) regression procedure for selecting the transcriptional factors with varying coefficients that are involved in gene regulation during a biological process. With a proper choice of the regularization parameter, the proposed estimators possess an oracle property. Error bound of the estimates is also derived. Simulation studies indicated that such a procedure is quite effective in selecting the relevant variables with time-varying coefficients and in estimating the coefficients. Application to the yeast cell cycle microarray time course gene expression data set identified 19 of 21 known transcriptional factors related to the cell cycle process. In addition, we have identified another 52 TFs that also have periodic transcriptional effects on gene expression during the cell cycle process. Compared to simple linear regression analysis at each time point, our procedure identified more known cell cycle related transcriptional factors.