Statistical Science
2012, Vol. 27, No. 4, 447–449
DOI: 10.1214/12-STS409

Introduction to the Special Issue on Sparsity and Regularization Methods

Jon Wellner and Tong Zhang

1. INTRODUCTION

Traditional statistical inference considers relatively small data sets and the corresponding theoretical analysis focuses on the asymptotic behavior of a statistical estimator when the number of samples approaches infinity. However, many data sets encountered in modern applications have dimensionality significantly larger than the number of training data available, and for such problems the classical statistical tools become inadequate. In order to analyze high-dimensional data, new statistical methodology and the corresponding theory have to be developed.

In the past decade, sparse modeling and the corresponding use of sparse regularization methods have emerged as a major technique to handle high-dimensional data. While the data dimensionality is high, the basic assumption in this approach is that the actual estimator is sparse in the sense that only a small number of components are nonzero. On the practical side, the sparsity phenomenon has been ubiquitously observed in applications, including signal recovery, genomics, computer vision, etc. On the theoretical side, this assumption makes it possible to overcome the problem associated with estimating more parameters than the number of observations which is impossible to deal with in the classical setting.

There are a number of challenges, including developing new theories for high-dimensional statistical estimation as well as new formulations and computational procedures. Related problems have received a lot of attention in various research fields, including applied math, signal processing, machine learning, statistics and optimization. Rapid advances have been made in recent years. In view of the growing research activities and their practical importance, we have organized this special issue of *Statistical Science* with the goal of providing overviews of several topics in modern sparsity analysis and associated regularization methods. Our hope is that general readers will get a broad idea of the field as well as current research directions.

Jon Wellner is Professor of Statistics and Biostatistics, Department of Statistics, University of Washington, Seattle, Washington 98112-2801, USA (e-mail: jaw@stat.washington.edu). Tong Zhang is Professor of Statistics, Department of Statistics, Rutgers University, New Jersey, USA (e-mail: tzhang@stat.rutgers.edu).

2. SPARSE MODELING AND REGULARIZATION

One of the central problem in statistics is linear regression, where we consider an $n \times p$ design matrix X and an n-dimensional response vector $Y \in \mathbb{R}^n$ so that

$$Y = X\bar{\beta} + \varepsilon, \tag{1}$$

where $\bar{\beta} \in \mathbb{R}^p$ is the true regression coefficient vector and $\varepsilon \in \mathbb{R}^n$ is a noise vector. In the case of $n < p$, this problem is ill-posed because the number of parameters is more than the number of observations. This ill-posedness can be resolved by imposing a sparsity constraint: that is, by assuming that $\|\bar{\beta}\|_0 \le s$ for some s, where the ℓ_0-norm of $\bar{\beta}$ is defined as $\|\bar{\beta}\|_0 = |\mathrm{supp}(\bar{\beta})|$, and the support set of $\bar{\beta}$ is defined as $\mathrm{supp}(\bar{\beta}) := \{j : \bar{\beta}_j \ne 0\}$. If $s \ll n$, then the effective number of parameters in (1) is smaller than the number of observations.

The sparsity assumption may be viewed as the classical model selection problem, where models are indexed by the set of nonzero coefficients. The classical model selection criteria such as AIC, BIC or Cp [1, 7, 11] naturally lead to the so-called ℓ_0 regularization estimator:

$$\hat{\beta}^{(\ell_0)} = \arg\min_{\beta \in \mathbb{R}^p} \left[\frac{1}{n}\|X\beta - Y\|_2^2 + \lambda\|\beta\|_0\right]. \tag{2}$$

The main difference of modern ℓ_0 analysis in high-dimensional statistics and the classical model selection methods is that the choice of λ will be different, and the modern analysis requires choosing a larger λ than that considered in the classical model selection setting because it is necessary to compensate for the effect of considering many models in the high-dimensional setting. The analysis for ℓ_0 regularization in the high-dimensional setting (e.g., [15] in this issue) employs different techniques and the results obtained are also different from the classical literature.