Annals of the Institute of Statistical Mathematics (2022) 74:559–579
https://doi.org/10.1007/s10463-021-00809-z

A high-dimensional M-estimator framework for bi-level variable selection

Bin Luo[1] **· Xiaoli Gao**[2]

Received: 7 November 2020 / Revised: 12 May 2021 / Accepted: 17 August 2021 /
Published online: 9 September 2021

Abstract

In high-dimensional data analysis, bi-level sparsity is often assumed when covariates function group-wisely and sparsity can appear either at the group level or within certain groups. In such cases, an ideal model should be able to encourage the bi-level variable selection consistently. Bi-level variable selection has become even more challenging when data have heavy-tailed distribution or outliers exist in random errors and covariates. In this paper, we study a framework of high-dimensional M-estimation for bi-level variable selection. This framework encourages bi-level sparsity through a computationally efficient two-stage procedure. In theory, we provide sufficient conditions under which our two-stage penalized M-estimator possesses simultaneous local estimation consistency and the bi-level variable selection consistency if certain non-convex penalty functions are used at the group level. Both our simulation studies and real data analysis demonstrate satisfactory finite sample performance of the proposed estimators under different irregular settings.

Keywords Bi-level variable selection · Estimation consistency · High dimensionality · M-estimation · Non-convexity

Gao is partially supported by Simons Foundation Grant: SF359337.

✉ Bin Luo
bin.luo2@duke.edu

[1] Department of Biostatistics and Bioinformatics, Duke University, 2424 Erwin Road, Durham, NC 27705, United States

[2] Department of Mathematics and Statistics, The University of North Carolina at Greensboro, 116 Petty Building, Greensboro, NC 27402, United States