



# High-dimensional sign-constrained feature selection and grouping

Shanshan Qin<sup>1</sup> · Hao Ding<sup>1</sup> · Yuehua Wu<sup>1</sup> · Feng Liu<sup>2</sup>

Received: 10 January 2020 / Revised: 4 September 2020 / Accepted: 8 September 2020 /  
Published online: 12 October 2020

© The Institute of Statistical Mathematics, Tokyo 2020

## Abstract

In this paper, we propose a non-negative feature selection/feature grouping (nnFSG) method for general sign-constrained high-dimensional regression problems that allows regression coefficients to be disjointly homogeneous, with sparsity as a special case. To solve the resulting non-convex optimization problem, we provide an algorithm that incorporates the difference of convex programming, augmented Lagrange and coordinate descent methods. Furthermore, we show that the aforementioned nnFSG method recovers the oracle estimate consistently, and that the mean-squared errors are bounded. Additionally, we examine the performance of our method using finite sample simulations and applying it to a real protein mass spectrum dataset.

**Keywords** Difference convex programming · Feature grouping · Feature selection · High-dimensional · Non-negative

---

✉ Hao Ding  
dinghaostat@gmail.com

Shanshan Qin  
ssqin267@yorku.ca

Yuehua Wu  
wuyh@yorku.ca

Feng Liu  
feng.liu@uts.edu.au

<sup>1</sup> Department of Mathematics and Statistics, York University, 4700 Keele Street, Toronto, ON M3J 1P3, Canada

<sup>2</sup> Australian Artificial Intelligence Institute, University of Technology Sydney, Sydney, NSW 2007, Australia