STAT*6801

Proposal: Computational Methods for LASSO Regression

Vinay Joshy

November 26, 2025

Introduction and Motivation

LASSO (Least Absolute Shrinkage and Selection Operator) regression is a cornerstone of sparse statistical learning, particularly relevant in high-dimensional settings where both interpretability and variable selection are crucial [1]. The LASSO optimization problem is formulated as:

$$\min_{\beta} \frac{1}{2n}\|y - X\beta\|^2 + \lambda\|\beta\|_1$$

where $X \in \mathbb{R}^{n\times p}$ is the design matrix, $\beta \in \mathbb{R}^p$ is the vector of regression coefficients, $y \in \mathbb{R}^n$ is the response vector, and $\lambda > 0$ is the regularization parameter controlling the strength of penalization. My motivation for this project stems from my thesis research on sparse finite mixture regression models. While my thesis work involves more complex sparse group penalties in mixture settings, understanding the computational foundations of basic LASSO optimization is essential groundwork. Through my Scientific Computing course (MATH*6020), I've developed a strong interest in the theoretical and practical aspects of optimization algorithms, and I want to understand how these methods work under the hood rather than treating them as black boxes.

This project will compare coordinate descent [2] and proximal gradient methods [3] for LASSO regression, examining their convergence behavior and computational efficiency under different data conditions. I'm particularly interested in how dimensionality, sparsity patterns, and correlation structures affect algorithmic performance, and whether we can identify principled guidelines for method selection based on problem characteristics. Additionally, I will explore whether strong screening rules [4] can meaningfully reduce computational burden by safely eliminating variables guaranteed to have zero coefficients at the solution.

Methodology

This project will focus on LASSO regression for exponential family responses [5], with primary emphasis on Gaussian responses where $y = X\beta + \epsilon$ with $\epsilon \sim N(0, \sigma^2)$ and sparse β. I will begin by deriving the subgradient of the L_1 penalty and stating the Karush-Kuhn-Tucker (KKT) optimality conditions for the LASSO problem [6]. These theoretical results will provide the mathematical foundation for the optimization algorithms and serve as convergence criteria for implementations.

The simulation framework will systematically vary problem characteristics to understand algorithmic performance. I will consider sample sizes of $n \in \{100, 500\}$ and dimensionality $p \in \{10, 100, 500\}$, with true coefficient sparsity levels of 5%, 10%, and 20%. To assess robustness to correlation, I will generate data with both independent features and block diagonal covariance structures at varying signal-to-noise ratios.

I will implement three approaches from scratch in R. The coordinate descent algorithm will cyclically optimize each coefficient using soft-thresholding, while proximal gradient methods will alternate between gradient steps on the smooth least squares term and proximal operators for the L_1 penalty. I will also implement strong rules for screening [4], which safely discard predictors guaranteed to have zero coefficients at the solution, reducing effective dimensionality during optimization.

Each implementation will be evaluated on computational performance (iterations to convergence, wall-clock time) and statistical performance (variable selection accuracy via precision, recall, and F1-score). All implementations will be validated against the `glmnet` package [2] to ensure correctness of solutions. The code will be made publicly available in a GitHub repository.

Broader Context and Significance

This project connects computational algorithms and convex optimization with statistical methodology, while building foundations for my thesis research on sparse methods for mixture models. By implementing these algorithms from scratch and systematically comparing their performance, I will develop intuition about their strengths and limitations. The insights about algorithmic performance under different data conditions will inform principled method selection in applied work and deepen my understanding of the computational and algorithmic trade-offs inherent in sparse estimation problems.

References

[1] Robert Tibshirani. "Regression Shrinkage and Selection via the Lasso". In: *Journal of the Royal Statistical Society. Series B (Methodological)* 58.1 (1996), pp. 267–288. ISSN: 00359246. URL: `http://www.jstor.org/stable/2346178`.

[2] Jerome Friedman, Trevor Hastie, and Robert Tibshirani. "Regularization Paths for Generalized Linear Models via Coordinate Descent". In: *Journal of Statistical Software* 33.1 (2010). ISSN: 1548-7660. DOI: `10.18637/jss.v033.i01`.

[3] Amir Beck and Marc Teboulle. "A Fast Iterative Shrinkage-Thresholding Algorithm for Linear Inverse Problems". In: *SIAM Journal on Imaging Sciences* 2.1 (2009), pp. 183–202. DOI: `10.1137/080716542`.

[4] Robert Tibshirani et al. "Strong Rules for Discarding Predictors in Lasso-Type Problems". In: *Journal of the Royal Statistical Society Series B: Statistical Methodology* 74.2 (Mar. 1, 2012), pp. 245–266. ISSN: 1369-7412, 1467-9868. DOI: `10.1111/j.1467-9868.2011.01004.x`.

[5] Morton Kupperman. "Probabilities of hypotheses and information-statistics in sampling from exponential-class populations". In: *The Annals of Mathematical Statistics* 29.2 (1958). Publisher: Institute of Mathematical Statistics, pp. 571–575. DOI: `10.1214/aoms/1177706633`.

[6] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The elements of statistical learning: data mining, inference and prediction*. 2nd ed. Springer, 2009. URL: `http://www-stat.stanford.edu/~tibs/ElemStatLearn/`.