Stanford University

Closing the Approximation Gap in Sparse Learning Algorithms

Varun Srivastava
vsriva@stanford.edu
Department of Electrical Engineering, Stanford University

Abstract

Sparse learning algorithms are of increasingly large importance in many areas of numerical computing and optimization where they model existence of low dimensional structure (in the form of very few pairwise interactions). Sparse matrices occur in graph theory (as adjacency matrices of graphs with few edges), machine learning (in techniques such as LASSO), and genomics (where sparsity is required for interpretation of data). Pilanci et al. [3] propose a technique to impose sparsity using l_0 norm constraints by reformulating several sparse learning problems as a convex boolean program. This posters presents new theoretical results for the sparse learning setting proposed in [3] deriving tighter approximation bounds for the original independent randomized rounding as well as concentration results for dependent rounding techniques.

Objectives

1. To analyze the approximation gap under (independent) randomized rounding using Matrix Bernstein bounds
2. To analyze the approximation gap under dependent randomized rounding (Pipage Rounding [2]) using Matrix Chernoff bounds
3. Comparing the performance of the proposed scheme with other sparsity constrained Learning Algorithms such as LASSO [4] and Orthogonal Matching Pursuit.
4. Numerical Estimate of Computational Complexity

Sparse Learning Problem

The canonical l_0 (semi)-norm constrained problem is given by

$$P^* := \inf_{\substack{w\in\mathbb{R}^d \\ ||w||_0\leq k}} \sum_{i=1}^{n} f(<x_i, w>; y_i) + \frac{1}{2}\rho||w||_2^2 \tag{1}$$

Independent Randomized Rounding

In the case of independent randomized rounding, the approximation error is given by

$$G(\hat{u}) - G(\tilde{u}) \leq \frac{2}{\rho}\left\|X(D(\hat{u}) - D(\tilde{u}))X^T\right\| = \frac{2}{\rho}||Z||$$

, thus a bound on the operator norm of Z yields an *upper bound* on the approximation. Using the Matrix Bernstein concentration inequality [5], it can be shown that

$$\mathbb{P}[||Z|| \geq \sqrt{r}t] \leq 2n \cdot \mathrm{Exp}\left(-2t^2\right) \quad 0 \leq t \leq \frac{3\sqrt{r}}{4} \tag{2}$$

$$\mathbb{E}[||Z||] \leq \sqrt{\frac{r}{2}\log 2n} + \frac{1}{3}\log 2n \tag{3}$$

Dependent Randomized Rounding

While independent randomized rounding gives us concentration bounds, the same is not always true for dependent rounding techniques (which are used to better satisfy problem constraints such as the cardinality constraint for the problem in consideration). However, Pipage rounding proposed in [1] allows us to apply Matrix Chernoff bound [2].

$$\mathbb{P}[||Z|| \geq (1+\delta)\mu] \leq ne^{\frac{-\delta^2\mu}{3}} \quad \forall\, 0 < \delta < 1 \tag{4}$$

Numerical Results

1. **L0 seminorm constrained optimization**

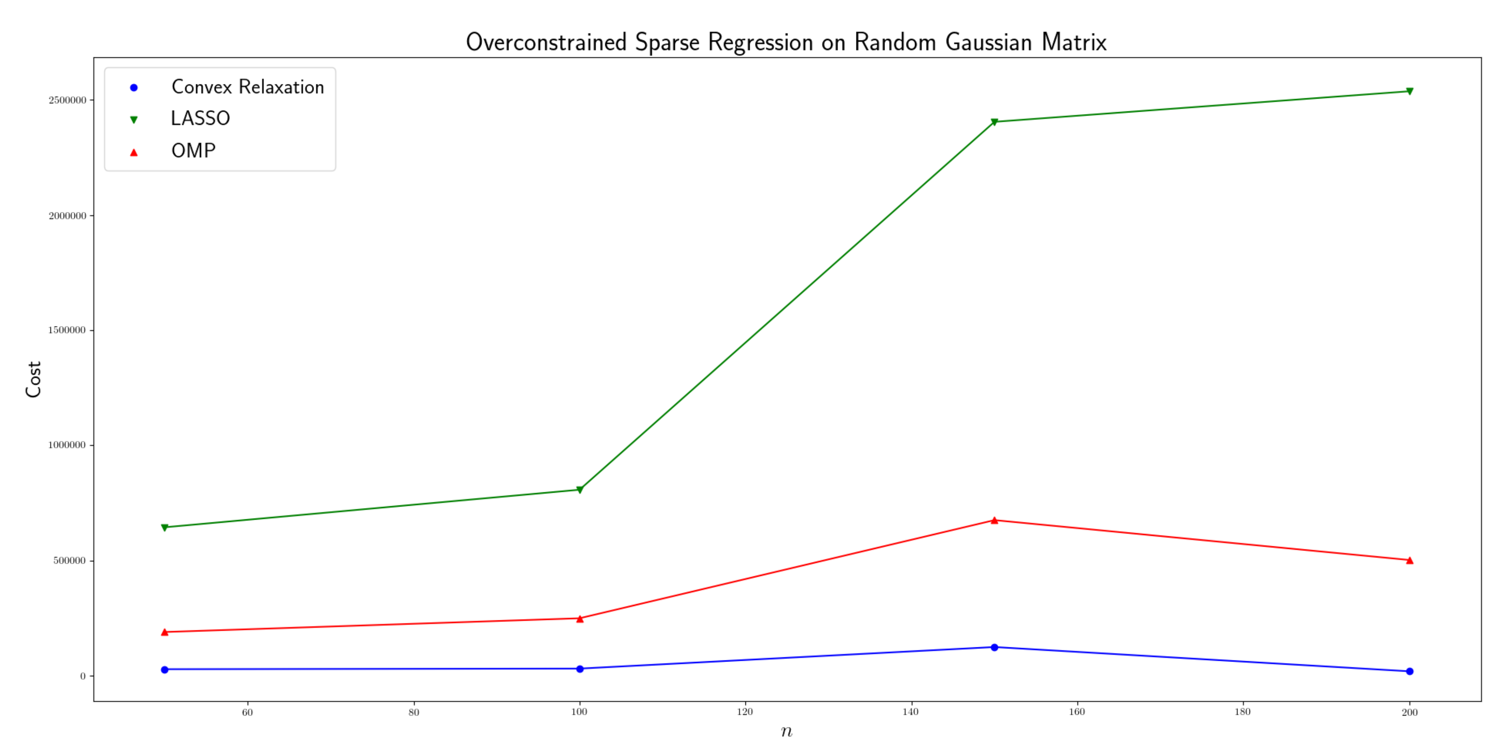

Figure 1: True Sparsity = 10, Enforced Sparsity (l0 norm) = 5, d = 20

2. **Computational Complexity**

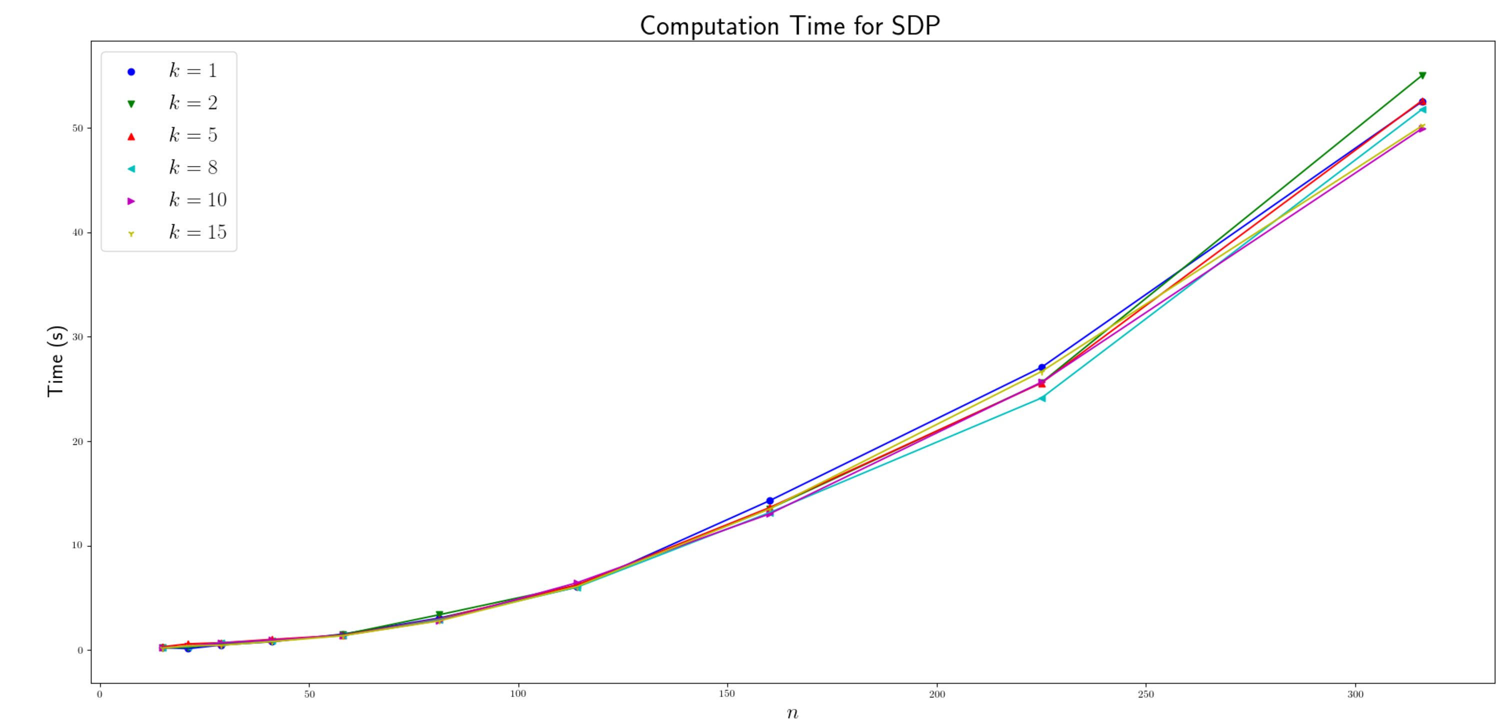

Figure 2: Computation time for different sample sizes

Discussion & Future Work

The given results have been shown for the case of Linear Regression where $X \in \mathbb{R}^{n\times d}$, and the l_0 norm constraint $||w||_0 \leq k$. Figure 1 shows the performance of LASSO, OMP and the given algorithm for sparse regression where the weight vector is forced to be sparser than the true solution. The convex relaxed version very clearly outperforms the other variants in the overconstrained setting.

However, this performance comes at the expensive cost of computation time, which from Figure 2 appears to be *exponential* in the dimension n (and independent of k for the well conditioned), which makes the SDP formulation infeasible for all but the smallest problems.

Future work can focus on alternative formulations based on subgradients, or quasi Newton Methods for optimization which could make this formulation a viable alternative to LASSO and OMP which outperform the given scheme from the perspective of computational complexity.

References

[1] A. A. Ageev and M. I. Sviridenko. Pipage rounding: A new method of constructing algorithms with proven performance guarantee. *Journal of Combinatorial Optimization*, 8(3):307–328, 2004.

[2] N. J. Harvey and N. Olver. Pipage rounding, pessimistic estimators and matrix concentration. In *Proceedings of the twenty-fifth annual ACM-SIAM symposium on Discrete algorithms*, pages 926–945. SIAM, 2014.

[3] M. Pilanci, M. J. Wainwright, and L. El Ghaoui. Sparse learning via boolean relaxations. *Mathematical Programming*, 151(1):63–87, 2015.

[4] R. Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1):267–288, 1996.

[5] J. A. Tropp et al. An introduction to matrix concentration inequalities. *Foundations and Trends® in Machine Learning*, 8(1-2):1–230, 2015.