

# Differentially Private Best Subset Selection Via Integer Programming

Kayhan Behdin, Peter Prastakos, Rahul Mazumder

Massachusetts Institute of Technology

May 11, 2024

# Problem Definition and Motivation

- ▶ Best Subset Selection (BSS):

$$\min_{\beta \in \mathbb{R}^p} \|\mathbf{y} - \mathbf{X}\beta\|_2^2 \quad \text{s.t.} \quad \|\beta\|_0 \leq s, \quad \|\beta\|_2^2 \leq r^2 \quad (1)$$

- ▶ An important methodological problem
  - ▶ Can be computationally challenging
  - ▶ Recent work uses Mixed Integer Programming (MIP) to solve large BSS instances [1,2].
- 
- ▶  $(\epsilon, \delta)$ -Differentially Private (DP) Algorithm  $\mathcal{A}$ :

$$\mathbb{P}(\mathcal{A}(\mathcal{D}) \in K) \leq e^\epsilon \mathbb{P}(\mathcal{A}(\mathcal{D}') \in K) + \delta$$

for any measurable event  $K \subset \text{range}(\mathcal{A})$  and for any pair of neighboring datasets  $\mathcal{D}$  and  $\mathcal{D}'$ .

# Problem Definition and Motivation

- ▶ **Goal:** Designing an  $(\epsilon, 0)$ -DP algorithm for variable selection (i.e., optimal location of nonzeros) in the BSS.
- ▶ **Current Algorithms for DP-BSS:**
  - ▶ convex relaxations, private Lasso [3, 4, 5, 6]
  - ▶ Markov chain mixing [7]
- ▶ Recent work has shown that (non-private) BSS can have favorable practical and theoretical properties over its convex relaxations under certain settings [8,9].
- ▶ **Our Proposal:** A new DP-variable selection method for the **original BSS problem** (1). We use techniques from MIP to scale-up our selection procedure.

# Exponential Mechanism

Lemma (From [10])

*The exponential mechanism  $\mathcal{A}_E(\cdot)$  that follows*

$$\mathbb{P}(\mathcal{A}_E(\mathcal{D}) = o) \propto \exp\left(-\frac{\varepsilon \mathcal{R}(o, \mathcal{D})}{2\Delta}\right), \quad \forall o \in \mathcal{O} \quad (2)$$

*ensures  $(\varepsilon, 0)$ -DP.*

# Problem Formulation

Define our outcome set as  $\mathcal{O} = \{S \subseteq [p] : |S| = s\}$  and the objective for each  $S$  as:

$$\mathcal{R}(S, \mathcal{D}) = \min_{\beta \in \mathbb{R}^{|S|}} \|\mathbf{y} - \mathbf{X}_S \beta\|_2^2 \quad \text{s.t.} \quad \|\beta\|_2^2 \leq r^2$$

The global sensitivity is

$$\Delta = \max_{S \in \mathcal{O}} \max_{\mathcal{D}, \mathcal{D}' \text{ are neighbors}} \mathcal{R}(S, \mathcal{D}) - \mathcal{R}(S, \mathcal{D}').$$

Lemma (\*): If  $|y| \leq b_y$  for  $y \in \mathcal{Y}$ , and  $\|\mathbf{x}\|_\infty \leq b_x$  for  $\mathbf{x} \in \mathcal{X}$ .  
Then,  $\Delta \leq 2b_y^2 + 2b_x^2 r^2 s$ .

## MIP Setup

- ▶  $\forall k \in [R]$  where  $R \ll \binom{p}{s}$ , define

$$\hat{S}_k(\mathcal{D}) \in \arg \min_S \mathcal{R}(S, \mathcal{D}) \quad \text{s.t.} \quad S \subseteq [p], |S| = s,$$

$$S \neq \hat{S}_i(\mathcal{D}), \forall i \in [k-1]$$

- ▶  $\hat{S}_k(\mathcal{D})$  can be obtained by solving a series of MIPs:

$$\min_{\mathbf{z}^{(k)}, \boldsymbol{\beta}^{(k)}, \boldsymbol{\theta}^{(k)}} \|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}^{(k)}\|_2^2$$

$$\text{s.t.} \quad \boldsymbol{\beta}^{(k)}, \boldsymbol{\theta}^{(k)} \in \mathbb{R}^p, \mathbf{z}^{(k)} \in \{0, 1\}^p, \boldsymbol{\theta}^{(k)} \geq 0, \sum_{i=1}^p z_i^{(k)} = s,$$

$$\sum_{i=1}^p \theta_i^{(k)} \leq r^2, (\beta_i^{(k)})^2 \leq \theta_i^{(k)} z_i^{(k)} \quad \forall i \in [p]$$

$$\sum_{i \in \hat{S}_j(\mathcal{D})} z_i^{(k)} \leq s - \frac{1}{2}, \quad j = 1, \dots, k-1.$$

where  $\hat{S}_k(\mathcal{D}) = \{i : \hat{z}_i^{(k)} \neq 0\}$ .

## DP Algorithm

Define the following probability distribution:

$$\mathbb{P}_0(k) \propto \begin{cases} \exp\left(-\varepsilon \mathcal{R}(\hat{S}_k(\mathcal{D}), \mathcal{D}) / (2\Delta)\right) & \text{if } k \leq R \\ \left(\binom{p}{s} - R\right) \exp\left(-\varepsilon \mathcal{R}(\hat{S}_R(\mathcal{D}), \mathcal{D}) / (2\Delta)\right) & \text{if } k = R + 1. \end{cases}$$

---

### Algorithm BSS with DP guarantees

---

- 1: **procedure**  $\mathcal{M}(\mathcal{D}, b_x, b_y, r, R, T)$
  - 2:     Clip  $\mathbf{X}, \mathbf{y}$  to  $b_x, b_y$ , respectively, as in (\*). Take  $\Delta$  as in (\*).  
      Form  $\mathbb{P}_0$ .
  - 3:     Draw  $a(\mathcal{D}) \sim \mathbb{P}_0$
  - 4:     **if**  $a(\mathcal{D}) \leq R$  **then**
  - 5:         **return**  $\hat{S}_{a(\mathcal{D})}(\mathcal{D})$
  - 6:     **else**
  - 7:         **return** a uniform draw from  $\{\hat{S}_k : k > R\}$
-

# DP Guarantees

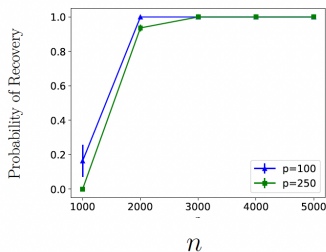
**Theorem 1:** Suppose  $1 < R < \binom{p}{s}$ . The procedure  $\mathcal{M}$  is  $(\varepsilon, 0)$ -DP. Moreover,  $\mathbb{P}(\mathcal{M}(\mathcal{D}) = \hat{\mathcal{S}}_1(\mathcal{D})) \geq \mathbb{P}_0(1)$ .

- ▶ No need to sample from a non-uniform distribution with exponentially large support.
- ▶ Intuition: “Flatten” the tail of exponential mechanism [10].

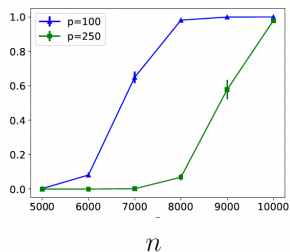


# Numerical Experiments

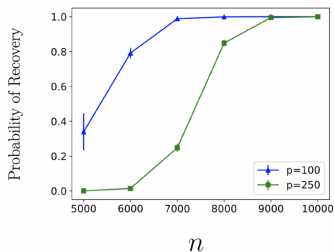
SNR = 2,  $\varepsilon = 5$ ,  $s = 5$



SNR = 5,  $\varepsilon = 2$ ,  $s = 7$



SNR = 5,  $\varepsilon = 1$ ,  $s = 5$



# Conclusion

- ▶ A new pure-DP algorithm for variable selection in BSS (1).
- ▶ We use MIP techniques to develop our DP variable selection algorithm.
- ▶ Good statistical performance and scalable to  $p \approx 250$ .

## References

- [1] Dimitris Bertsimas and Bart Van Parys. Sparse high-dimensional regression: Exact scalable algorithms and phase transitions. 2020.
- [2] Hussein Hazimeh, Rahul Mazumder, and Ali Saab. Sparse regression at scale: Branch-and-bound rooted in first-order optimization. *Mathematical Programming*,
- [3] Abhradeep Guha Thakurta and Adam Smith. Differentially private feature selection via stability arguments, and the robustness of the lasso. In *Conference on Learning Theory*.
- [4] Weidong Liu, Jiyuan Tu, Xiaojun Mao, and Xi Chen. Majority vote for distributed differentially private sign selection. *arXiv preprint arXiv:2209.04419*, 2022.
- [5] Daniel Kifer, Adam Smith, and Abhradeep Thakurta. Private convex empirical risk minimization and high-dimensional regression. In *Conference on Learning Theory*, pp. 25–1. *JMLR Workshop and Conference Proceedings*, 2012.

## References

- [6] Kobbi Nissim, Sofya Raskhodnikova, and Adam Smith. Smooth sensitivity and sampling in private data analysis. In Proceedings of the thirty-ninth annual ACM symposium on Theory of computing, pp. 75–84, 2007.
- [7] Saptarshi Roy and Ambuj Tewari. On the computational complexity of private high-dimensional model selection via the exponential mechanism. arXiv preprint arXiv:2310.07852, 2023.
- [8] Yongyi Guo, Ziwei Zhu, and Jianqing Fan. Best subset selection is robust against design dependence. arXiv preprint arXiv:2007.01478
- [9] Yilin Guo, Haolei Weng, and Arian Maleki. Signal-to-noise ratio aware minimaxity and higher-order asymptotics. IEEE Transactions on Information Theory, 2023.
- [10] Frank McSherry and Kunal Talwar. Mechanism design via differential privacy. In 48th Annual IEEE Symposium on Foundations of Computer Science (FOCS'07), pp. 94–103. IEEE, 2007.