

# High Dimensional Linear Regression using Lattice Basis Reduction

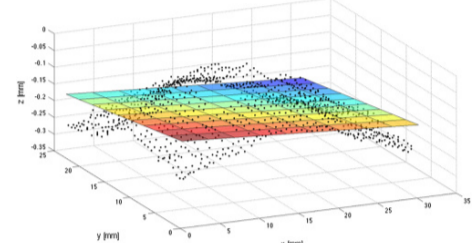
Ilias Zadik, joint work with David Gamarnik

Operations Research Center, Massachussets Insititute of Technology (MIT)

## High Dimensional Linear Regression (HDLR)

Recovering unknown coefficients  $\beta^*$  from few noisy observations and large number of features arises in a broad variety of contexts including

- pricing of a product in the digital economy (**econometrics**)
- GPS modeling and signal denoising (**telecommunications**)
- MRI analysis (**compressive sensing**)
- Generative Models and GANs (**neural networks**)



## The Model

**Setup:** Let  $\beta^* \in \mathbb{R}^p$ . For  $X \in \mathbb{R}^{n \times p}$  and  $W \in \mathbb{R}^n$  we get  $n$  noisy linear samples of  $\beta^*$ ,  $Y \in \mathbb{R}^n$ , given by,  $Y := X\beta^* + W$ .

**Goal:** Given data  $(Y, X)$  with  $Y := X\beta^* + W$  recover  $\beta^*$  with  $n \ll p$  and  $p \rightarrow +\infty$ .

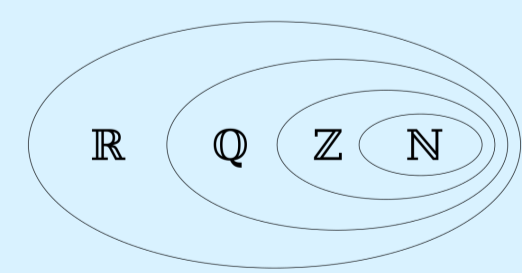
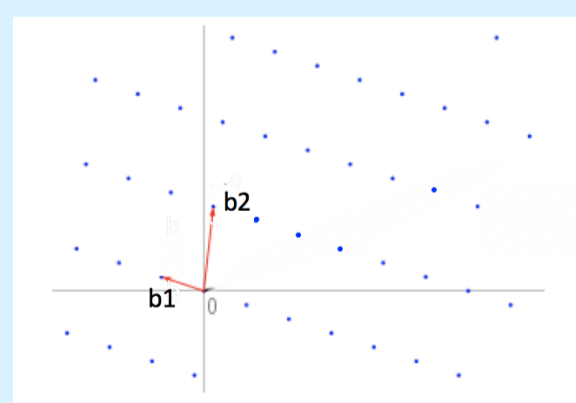
## Regularity Assumptions and a Challenge

To achieve  $n \ll p$  we need structural assumption on  $\beta^*$ .

- Sparsity!**  $k \leq p$  non zero coordinates. Vast literature. For  $X$  with iid  $N(0,1)$  entries and  $W$  with iid  $N(0, \sigma^2)$  entries ( $\sigma^2 \ll k$ ) we need  $k \log(\frac{p}{k})$  samples (Compressed Sensing)
- Issue:**  $k \log(\frac{p}{k})$  can still be **too large** for applications.
- Other assumptions:** Block-sparsity [Baron et al'05], Tree-Sparsity [He et al '09] Ouput of a Generative Model [Bora et al '17]
- Similar issue:** can achieve some  $n < p$  but not always small.

## This Work

**New efficient algorithm** for recovering  $\beta^*$  from  $(Y, X)$  under a new regularity assumption (**Q-rationality assumption**) based on a connection with **lattice-based algorithms**.



**Guarantees:** works for any  $n$  (even  $n = 1$ ) given sufficiently small noise!

## The Q-Rationality Assumption

Every entry of  $\beta^*$  is a **rational number** with fixed denominator  $Q$ . **Alternatively:** For  $Q = 2^M$ ,  $\log Q = M$  bits after zero position per entry.

## Why Q-rational?

- Large  $Q$ : **Large** but **finite domain** for the coefficients.
- Small  $Q$ : Standard in wireless communication. *Example:* Linear models for GPS ([Boyd, Hassibi '98]), **Physics laws** imply **integer coordinates** [ $Q = 1$ ].



## Under Q-rationality, One Sample Suffices

**Lemma 1** Assume  $X$  with iid  $N(0,1)$  entries and  $W$  with iid  $N(0, \sigma^2)$  entries. Given one sample  $n = 1$  and small  $\sigma$  we can recover **exactly** the  $Q$ -rational  $\beta^*$ .

**Intuition for  $\sigma = 0$ :** Each row of  $X$ ,  $X_1$ , has iid  $N(0,1)$  entries and therefore linearly independent entries over rationals. Hence, from  $(Y_1 = \langle X_1, \beta^* \rangle, X_1)$  we can recover  $\beta^*$ .

## Previous Computational Results

*Sample size needs to grow!*

- Convex Relaxations** For  $\beta^* \in \{-1, 1\}^p$  ( $Q = 1$ ),  $\sigma = 0$  consider  $\min \|\beta\|_\infty$ , s.t.  $Y = X\beta$ . Works if and only if  $n > p/2$ , i.e. needs **linear samples** ([Chandrasekaran et al '10], [Amelunxen et al '13])
- Statistical-Physics-based algorithm (AMP)** [Donoho et al '11] works for some  $n = o(p)$  and any  $Q$  but
  - we only know  $n = o(p)$  (could be any sublinear quantity) and
  - needs delicate choice of  $X$  (not iid!)

## Main Results

**Theorem 1 (Efficient Recovery with  $n = 1$ )** Let  $n$  samples,  $n \ll p$ , and  $0 \leq \sigma \leq \exp\left(-\frac{p \max\{p, \log Q\}}{n}\right)$ .

There exists a **polynomial-in- $n, p, \log Q$**  time algorithm which with input  $(Y, X)$  outputs exactly  $\beta^*$  w.h.p. as  $p \rightarrow +\infty$ .

- The theorem works with **any  $X$  with iid well-behaved continuous** entries (or uniform iid integer in a large domain) and any  $W$  with  $\|W\|_\infty \leq \sigma$ !

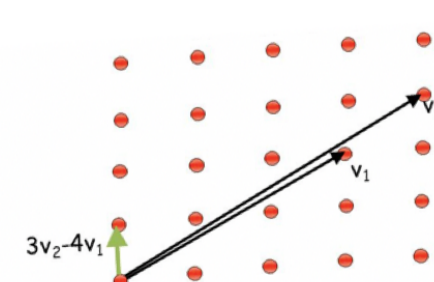
**Theorem 2** Let  $n$  samples,  $n \ll p$ , and  $\sigma > \exp\left(-\frac{p \log Q}{n}\right)$ .

Then if  $X$  has iid  $N(0,1)$  entries and  $W$  iid  $N(0, \sigma^2)$  entries, its impossible to w.h.p. recover correctly any  $Q$ -rational  $\beta^*$  with **any algorithm** with only access to  $(Y, X)$ .

- If  $\log Q > p$ : our algorithm has **optimal noise-tolerance!**

## Shortest Vector Problem (SVP)

For a lattice  $\mathcal{L}$  (integer linear combinations of some vectors  $b_1, \dots, b_m \in \mathbb{Z}^p$ ) the goal is to solve:  $\min_{x \in \mathcal{L} \setminus \{0\}} \|x\|_2$



Well-studied in **Integer Programming** and **Cryptography**.

## The LLL Algorithm for SVP

*Lattice Basis Reduction!*

SVP is NP-Hard but **Lenstra-Lenstra-Lovasz (LLL)**, algorithm efficiently *approximates* it; finds  $\hat{x} \in \mathcal{L} \setminus \{0\}$  with

$$\|\hat{x}\|_2 \leq 2^{\frac{p}{2}} \min_{x \in \mathcal{L} \setminus \{0\}} \|x\|_2.$$

Time poly in  $p, \log \max_i \{\|b_i\|_\infty\}$ .

## Using LLL for HDLR (General Scheme)

**Step 1:** Create a lattice  $\mathcal{L} = \mathcal{L}(Y, X)$  such that

**“approximately” shortest vectors of  $\mathcal{L} \leftrightarrow$  multiples of  $\beta^*$**

**Step 2:** Use LLL and recover a multiple of  $\beta^*$ .

**Step 3:** Recover  $\beta^*$  from a multiple (needs special structure!)

**Note:** *Step 1* is Inspired by the use of LLL in cryptography ([Lagarias, Odlyzko '83], [Frieze '86])

## The Algorithm: Special Case

- $n = 1$ ,  $\sigma = 0$ ,  $\beta^*$  binary,  $Y_1 = \langle X_1, \beta^* \rangle$ .
- $X_1 \in \mathbb{Z}^p$  with iid **uniform** in  $[2^N]$  entries for large  $N$

**Step 1:** For  $M$  sufficiently large enough set  $\mathcal{L}_M(Y_1, X_1)$  produced by the columns of

$$A_M := \begin{bmatrix} MX_1 & -MY_1 \\ I_{p \times p} & 0 \end{bmatrix}$$

**Lemma:** Each  $z \in \mathcal{L}_M$ ,  $\|z\|_2 < M$  is a multiple of  $\begin{bmatrix} 0 \\ \beta^* \end{bmatrix}$ , w.h.p.

**Intuition:**

$$z = A_M \begin{bmatrix} \beta \\ \lambda \end{bmatrix} = \begin{bmatrix} M \langle X_1, \beta \rangle - M \lambda Y_1 \\ \beta \end{bmatrix} = \begin{bmatrix} 0 \\ \beta \end{bmatrix},$$

Either  $|z_1| \geq M \Rightarrow \|z\|_2 \geq M$  or

$z_1 = 0 \Rightarrow \langle X_1, \beta - \lambda \beta^* \rangle = 0$ , **low probability with  $\beta \neq \lambda \beta^*$ !**

**Step 2:** Choose  $M$  appropriately so that LLL outputs a multiple of  $\beta^*$ .

- Choose  $M = \lceil 2^{\frac{p}{2}} \sqrt{p} \rceil + 1$ .
- We know  $A_M \begin{bmatrix} \beta^* \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ \beta^* \end{bmatrix} \in \mathcal{L}$ .
- LLL outputs  $\hat{x}$  with norm at most  $2^{\frac{p}{2}} \left\| \begin{bmatrix} 0 \\ \beta^* \end{bmatrix} \right\|_2 \leq 2^{\frac{p}{2}} \sqrt{p} < M$ .
- Using the lemma we are done!

**Step 3:** Rescale to get  $\beta^*$ .

## Special Case $\rightarrow$ General Case

- One sample  $n = 1 \rightarrow$  many samples  $n > 1$ . (Way: **Redesign** the Lattice)
- Noiseless  $\sigma = 0 \rightarrow$  noisy  $\sigma > 0$ . (Way: **Redesign** the Lattice)
- Integer  $Y, X \rightarrow$  real  $Y, X$ . (Way: **Truncate first bits and Rescale** the data  $(Y, X)$ )
- Binary coefficients  $\beta^* \rightarrow Q$ -rational  $\beta^*$ . (Way: **Translate and Rescale** the samples  $Y$ )

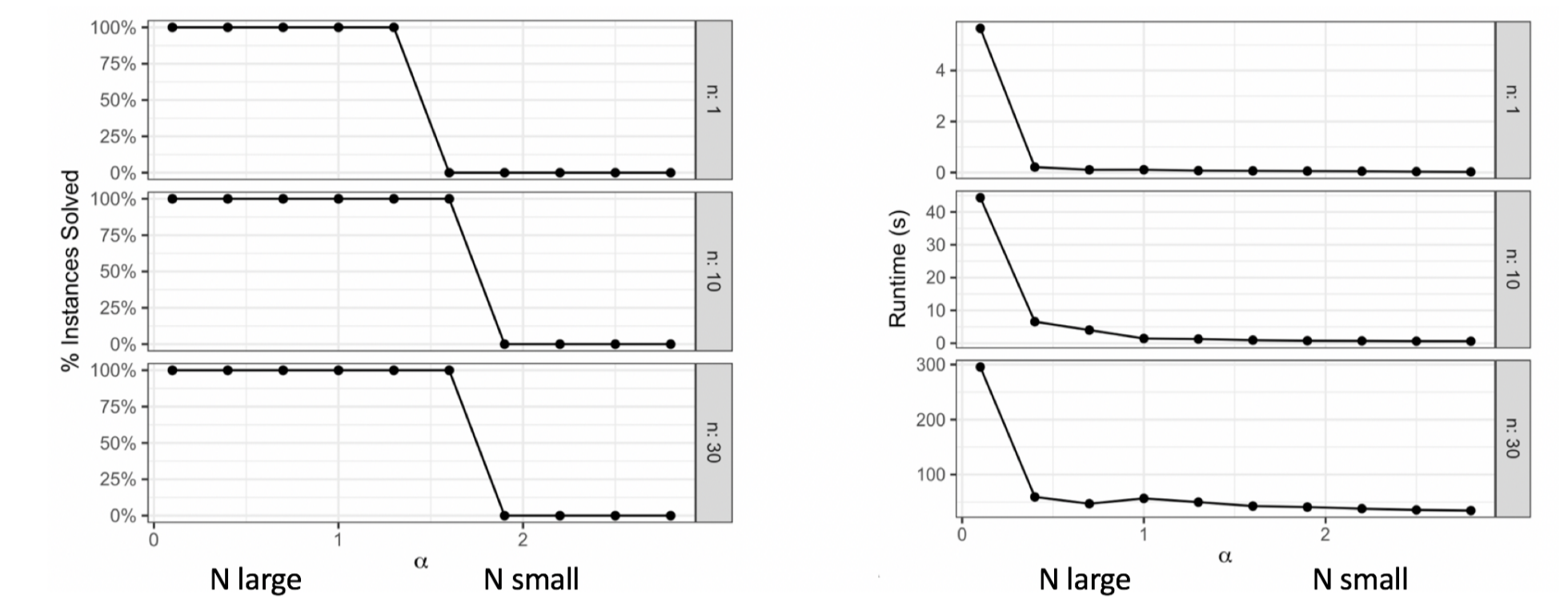
## (Preliminary) Experiments

(joint work with Patricio Foncea and Andrew Zheng)

### Integer Data

Assume  $X$  iid uniform in  $[2^N]$ ,  $\beta^*$  iid uniform in  $[100]$  and no noise. Success is exact recovery.

*Plot:* Avg Success/ Running Time against input size  $N$ .

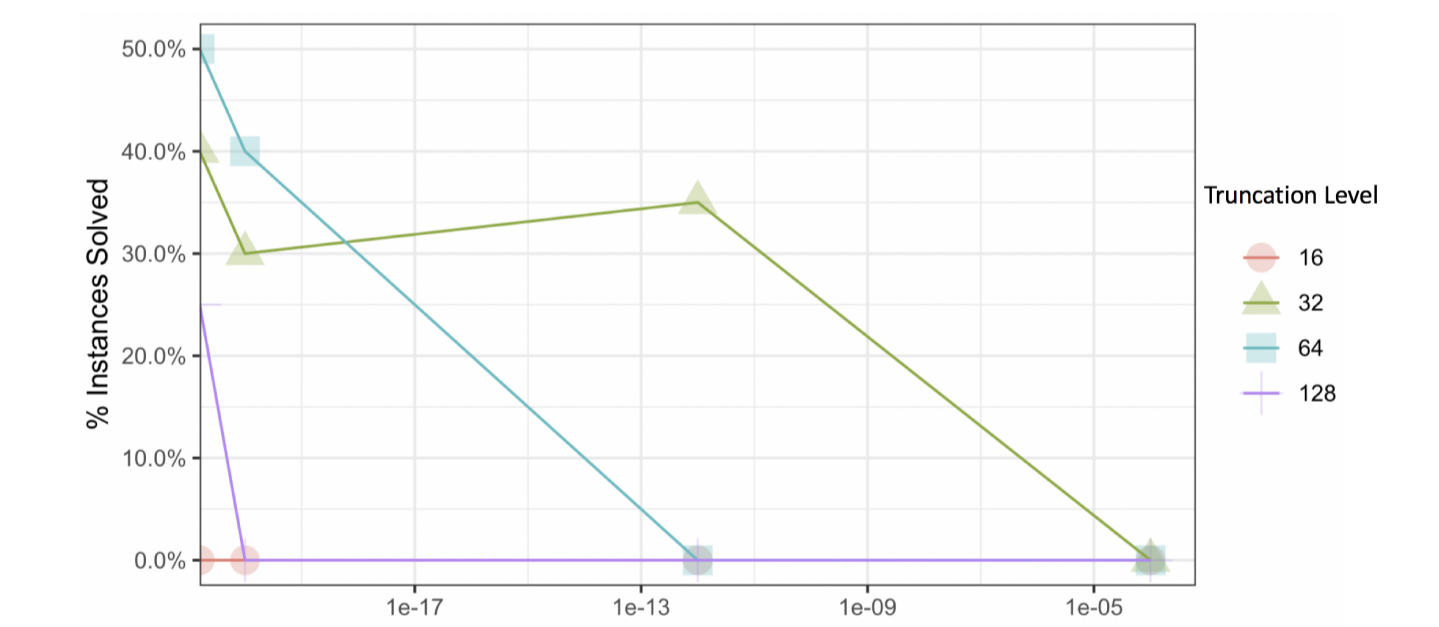


**Figure 5:** (20 instances per dot)  $p = 30$ ,  $n = 1, 10, 30$ ,  $\alpha \sim 1/N$ .

### Real-valued Data

Assume  $X$  iid  $U(0,1)$ ,  $W$  iid  $U(-\sigma, \sigma)$  and  $\beta^*$  iid uniform in  $[100]$ . Success is exact recovery.

*Plot:* Avg Success against noise level  $\sigma$  and truncation level.



**Figure 6:** (20 instances per dot)  $p = 30$  and  $n = 10$ .

## Conclusion

- High dimensional linear regression** with **rational coefficients** can be **efficiently solved with one sample  $n = 1$** , under small noise!
- New algorithm for **high dimensional linear regression** using **lattice based methods (LLL algorithm)**.
- The algorithm has **guarantees for large  $p$** , but also **works well for small  $p$** .

## Open Questions

- Can lattice-based methods also be used for **non-linear inference** problems?  
Example: *Phase-Retrieval* where  $Y_i = |\langle X_i, \beta^* \rangle|$  (many applications in Crystallography and MRI).
- Can we tolerate **higher noise levels** for smaller  $Q$ ?