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 \( \beta \in \mathbb{R}^d \). For \( X \in \mathbb{R}^{m \times p} \) and \( W \in \mathbb{R}^{m \times n} \) 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 \( \beta \).

Regularity Assumptions and a Challenge

To achieve \( \beta \), we need structural assumptions on \( \beta \).

- Sparsity! \( k \leq \eta \) non zero coordinates.
- Vast literature.
- For \( X \) with iid \( N(0,1) \) entries and \( W \in \mathbb{R}^{m \times n} \) entries \( \left< x, \beta \right> \) we need \( k \log \left( \frac{p}{\eta} \right) \) samples (Compressed Sensing).
- Other assumptions: Block-sparsity [Baron et al. 05], Tree-sparsity [He et al. 09]

Output of a Generative Model [Bora et al. 17]

Similar issue: can achieve some \( \eta \) but not always small.

This Work

New efficient algorithm for recovering \( \beta \) from \((Y, X)\) under a new regularity assumption (\( \eta \)-rationality assumption)

based on a connection with lattice-based algorithms.

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

The \( \eta \)-Rationality Assumption

Every entry of \( \beta \) is a rational number with denominator \( \eta \).

Alternatively: For \( \beta = [\beta_1, \ldots, \beta_d] \), the goal is to solve: \( \min_{\lambda \in \mathbb{Z}^d} \left< \lambda, \beta \right> \).

Well-studied in Integer Programming and Cryptography.

Why \( \eta \)-rational?

- Large \( Q \): large but finite domain for the coefficients.
- Small \( Q \): standard in wireless communication.

Example: Linear models for GPS [Broid, Hassibi ‘98].

Physics laws imply integer coordinates \( \left< \lambda, \beta \right> \leq 1 \).

Previous Computational Results

Sample size needs to grow!

- Convex Relaxations for \( \beta \in \{-1,1\}^p \). \( \alpha = 0 \) consider \( \min \left< \beta, x \right> \text{ s.t. } Y = X \beta \).

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 \( \alpha > 0 \) and any \( \beta \), but we only know \( \alpha = 0 \) (could be any sublinear quantity) and needs delicate choice of \( \lambda \) (not ideal).

Main Results

Theorem 1 (Efficient Recovery with \( \eta \)-rationality assumption)

Let \( \beta \) samples, \( \eta < p \), and \( \sigma \leq \frac{\sqrt{p} \eta}{\sqrt{\log p}} \).

There exists a polynomial-in-\( n, p, \log Q \) time algorithm which with \( (Y, X) \) outputs \( \beta \) w.h.p. as \( p \to \infty \).

- The algorithm 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 (Efficient \( \eta \)-rational recovery)

Let \( \eta \) samples, \( \eta < p \), and \( \sigma > \frac{\eta}{\sqrt{\log p}} \).

Then if \( X \) has iid \( N(0,1) \) entries and \( W \) iid \( N(0, \sigma^2) \), its impossible to w.h.p. recover correctly any \( \eta \)-rational \( \beta \).

- 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 \( b_1, \ldots, b_d \)) the goal is to solve: \( \min_{x \in \mathcal{L}} \|x\|_2 \).

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 \( \beta \); finds \( x \in \mathcal{L} \) with \( \|x\|_2 \leq \sqrt{d} \min_{x \in \mathcal{L}} \|x\|_2 \). \( \|x\|_2 \) Time poly in \( p, \log m, \|h\|_1 \).

Using LLL for HDLR (General Scheme)

Step 1: Create a lattice \( \mathcal{L} = \langle Y, X \rangle \) such that approximately ‘shortest vectors of \( \mathcal{L} \) are multiples of \( \beta \).

Step 2: Use LLL and recover a multiple of \( \beta \).

Step 2: Recover \( \beta \) from a multiple (needs special structure!)

Note: Step 1 is inspired by the use of LLL in cryptography (Lagarias, Odlyzko ‘88; [Frieze ‘96]).

Preliminary Experiments

Integer Data

Assume \( X \) iid uniform in \([0,1]^p \), \( \beta \) iid uniform in \([0,1]^n \) and \( n \) noise.

Preliminary Success. Running time against input size \( n \).

Real-valued Data

Assume \( X \) iid \( U(0,1) \), \( W \) iid \( (\sigma e_i, -\sigma e_i) \) and \( \beta \) iid uniform in \([0,1]^n \).

Success is exact recovery.

Plot: Avg Success against noise level \( \sigma \) and truncation level.

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

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 \( \beta = \{X_i, \beta_i\} \) (many applications in Crystallography and MRI).

- Can we tolerate higher noise levels for smaller \( Q \)?