# Batch List-Decodable Linear Regression via Higher Moments

Ilias Diakonikolas (University of Wisconsin Madison),
Daniel M. Kane (University of California San Diego),
Sushrut Karmalkar (Microsoft Research), Sihan Liu
(University of California San Diego), Thanasis Pittas
(University of Wisconsin Madison)

## Problem Statement

- Linear Model $D_{\beta*}$ over $(X, y) \in (\mathbb{R}^d \times \mathbb{R})$:

(1) Feature Vector: $X \sim \mathcal{G}$.

(2) Target Variable: $y = \langle X, \beta^* \rangle + \xi$,

where $\xi \sim \mathcal{N}(0, \sigma^2)$.

- $\alpha$-Corrupted Batched Sample Access:

(1) With probability $\alpha$, draw a batch of $n$ samples from $D_{\beta*}$.

(2) Otherwise, return a corrupted batch of $n$ adversarially chosen samples.

**Goal**: Given $\alpha$-corrupted batched sample access to $D_{\beta*}$ for some $\alpha < 1/2$,

return a short list of candidate regressors L

such that $\min_{\beta \in L} \|\beta - \beta^*\|$ is small .

## Prior Works

- In the **non-batched** setting i.e., batch size n=1, [Diakonikolas-Kane-Pensia-Pittas-Stewart'21] provides evidence that one must either:

  1. need **super-polynomial** sample size of $d^{\text{poly}(1/\alpha)}$.

  2. or run in **nearly exponential** time in d.

- Subsequent work [Das-Jain-Kong-Sen'23] designs an **efficient** algorithm in **super-linear batch-size** setting, i.e., batch-size n = $\tilde{\Omega}(1/\alpha)$: returns a list of size $O(1/\alpha^2)$ such that

$$\min_{\beta \in L} \|\beta - \beta^*\| \leq \tilde{O}(\sigma\alpha^{-1/2}/\sqrt{n}).$$

## Our Results

*What's the minimum batch-size that permits polynomial time/sample learning algorithm?*

***Message***: *The runtime/sample needed interpolates smoothly from polynomial to super-polynomial as the batch size n shrinks.*

- $n = 1$: no polynomial time/sample algorithm exists.

- $n \approx \alpha^{-1/k}$: runtime/sample size scales with $(dn)^k$

- $n \approx \alpha^{-1}$: polynomial time/sample algorithm exists.

**Theorem**. For any constant $k \in \mathbb{Z}_+$, if batch size $n \geq \Omega_k\left(\alpha^{-1/k}\right)$ and the first degree-$\Theta(k)$ moments of the feature vector distribution $\mathcal{G}$ are *SoS certifiably bounded*, there exists an algorithm with sample-size and runtime poly $\left((dn)^k, \alpha^{-1}\right)$: returns a list of size $O(1/\alpha)$ such that

$$\min_{\beta \in L} \|\beta - \beta^*\| \leq \tilde{O}(\sigma\alpha^{-1/(2k)}/\sqrt{n}).$$

**Remark.** We provide evidence that if $n \ll \log(1/\alpha)$ then **super-polynomial** computation resources may be **required**.

## Sketch of the Algorithm

**Step 1**: **Estimate** $\hat{\beta} \approx \mathbf{E}[yX]$ for $(X, y) \sim D_{\beta*}$.

Use list-decodable mean estimation algorithm [Kothari-Steinhardt'17] to estimate the mean of the random variable $W = \dfrac{1}{n} \sum_{(X,y) \in B} yX$,

where $B$ is an inlier batch.

Yield a list $L$ of size $O(1/\alpha)$ including some $\hat{\beta}$ with error $O_k\left(\alpha^{-3/k}\|\beta_*\|_2/n\right) \ll \|\beta^*\|_2/2$.

**Step 2: Iterative improvements.**

If we adjust the label via the transformations $y' = y - \langle \hat{\beta}, X \rangle$, this reduces the problem into another linear regression problem with $\|\beta'^*\|_2 \leq \|\beta^*\|_2/2$.

For each $\beta \in L$, adjust the labels based on $\beta$ and repeat **Step 1**.

Yield a list $L'$ of size $O(1/\alpha^2)$ including some $\hat{\beta}'$ with error $\ll \|\beta^*\|_2/4$.

**Step 3: List Size Pruning.**

Goal: Given the list L', returns a refined list of size $L^* = O(1/\alpha)$ such that $\min_{\beta \in L^*} \|\beta^* - \beta\| \approx \min_{\beta \in L'} \|\beta^* - \beta\|$.

Keep only $\beta$ such that there exists a soft "cluster" of batches ensuring that any distant regressors cannot have smaller $\ell_2$ loss under the soft cluster.