Coresets for Multiple ℓ_p Regression

David P. Woodruff Taisuke (Tai) Yasuda

Carnegie Mellon University
Computer Science Department

ICML
International Conference
On Machine Learning

Coresets for Multiple ℓ_p Regression

Coresets for ℓ_p Regression

- ℓ_p linear regression problem
  - Let $\mathbf{A}$ be an $n \times d$ matrix containing n examples with d features
  - Let $\mathbf{b}$ be a vector of n labels
  - Solve $\min_{\mathbf{x} \in \mathbb{R}^d} \|\mathbf{A}\mathbf{x} - \mathbf{b}\|_p^p = \sum_{i=1}^{n} |\langle \mathbf{a}_i, \mathbf{x} \rangle - \mathbf{b}_i|^p$
    - $p = 2$: least squares linear regression
    - $p = 1$: least absolute deviations regression
    - $p = \infty$: Chebyshev regression

Coresets for Multiple ℓ_p Regression

Coresets for ℓ_p Regression

- Coreset: small weighted subset of a dataset
  - Weighted subset $\mathbf{S}$ s.t. $\|\mathbf{S}(\mathbf{A}\mathbf{x}-\mathbf{b})\|_p^p = (1 \pm \varepsilon)\|\mathbf{A}\mathbf{x}-\mathbf{b}\|_p^p$ for every $\mathbf{x} \in \mathbb{R}^d$
  - Goal: want the coreset size $s = \mathrm{nnz}(\mathbf{S})$ to be as small as possible

Coresets for Multiple ℓ_p Regression

Coresets for ℓ_p Regression

- For single response ℓ_p regression, we know how to compute coresets nearly optimally!
  - ℓ_p Lewis weight sampling [Cohen—Peng 2015, Woodruff—Yasuda 2023]
  - We can guarantee $s = \begin{cases} \tilde{O}(\varepsilon^{-2} d^{p/2}) & p > 2 \\ \tilde{O}(\varepsilon^{-2} d) & p \leq 2 \end{cases}$
- Question in this work: what if we have multiple responses?
  - m responses specified by a $n \times m$ matrix $\mathbf{B}$
  - Goal: $\|\mathbf{S}(\mathbf{AX} - \mathbf{B})\|_{p,p}^p = (1 \pm \varepsilon)\|\mathbf{AX} - \mathbf{B}\|_{p,p}^p$ for every $d \times m$ matrix $\mathbf{X}$
  - Challenge: achieve s independent of m

Coresets for Multiple ℓ_p Regression

Main results

Theorem (Strong coreset). Lewis weight sampling gives a sampling matrix $\mathbf{S}$ such that $\|\mathbf{S}(\mathbf{AX} - \mathbf{B})\|_{p,p}^p = (1 \pm \varepsilon)\|\mathbf{AX} - \mathbf{B}\|_{p,p}^p$ for every $d \times m$ matrix $\mathbf{X}$ with

$$s = \begin{cases} \tilde{O}(\varepsilon^{-p} d^{p/2}) & p > 2 \\ \tilde{O}(\varepsilon^{-2} d) & p \leq 2 \end{cases}.$$

Furthermore, this is nearly optimal.

- **Techniques**
  - Generalization of techniques for active ℓ_p regression [Musco—Musco—Woodruff—Yasuda 2022] to achieve a polylogarithmic dependence on m
  - Averaging argument to completely remove m dependence

Coresets for Multiple ℓ_p Regression

Main results

Theorem (Weak coreset). Lewis weight sampling gives a sampling matrix $\mathbf{S}$ such that $\tilde{\mathbf{X}} = \arg\min_{\mathbf{X}} \|\mathbf{S}(\mathbf{AXG} - \mathbf{B})\|_{p,p}^p$ is a $(1+\varepsilon)$-approximate minimizer with

Embedding matrix $\mathbf{G}$

$$s = \begin{cases} \tilde{O}(\varepsilon^{1-p} d^{p/2}) & p > 2 \\ \tilde{O}(\varepsilon^{-1} d) & 1 < p \le 2 \end{cases}.$$

Furthermore, this is nearly optimal.

Coresets for Multiple ℓ_p Regression

Applications

- Nearly optimal sublinear algorithms for Euclidean power means
- Nearly optimal spanning coresets for ℓ_p subspace approximation

Coresets for Multiple ℓ_p Regression

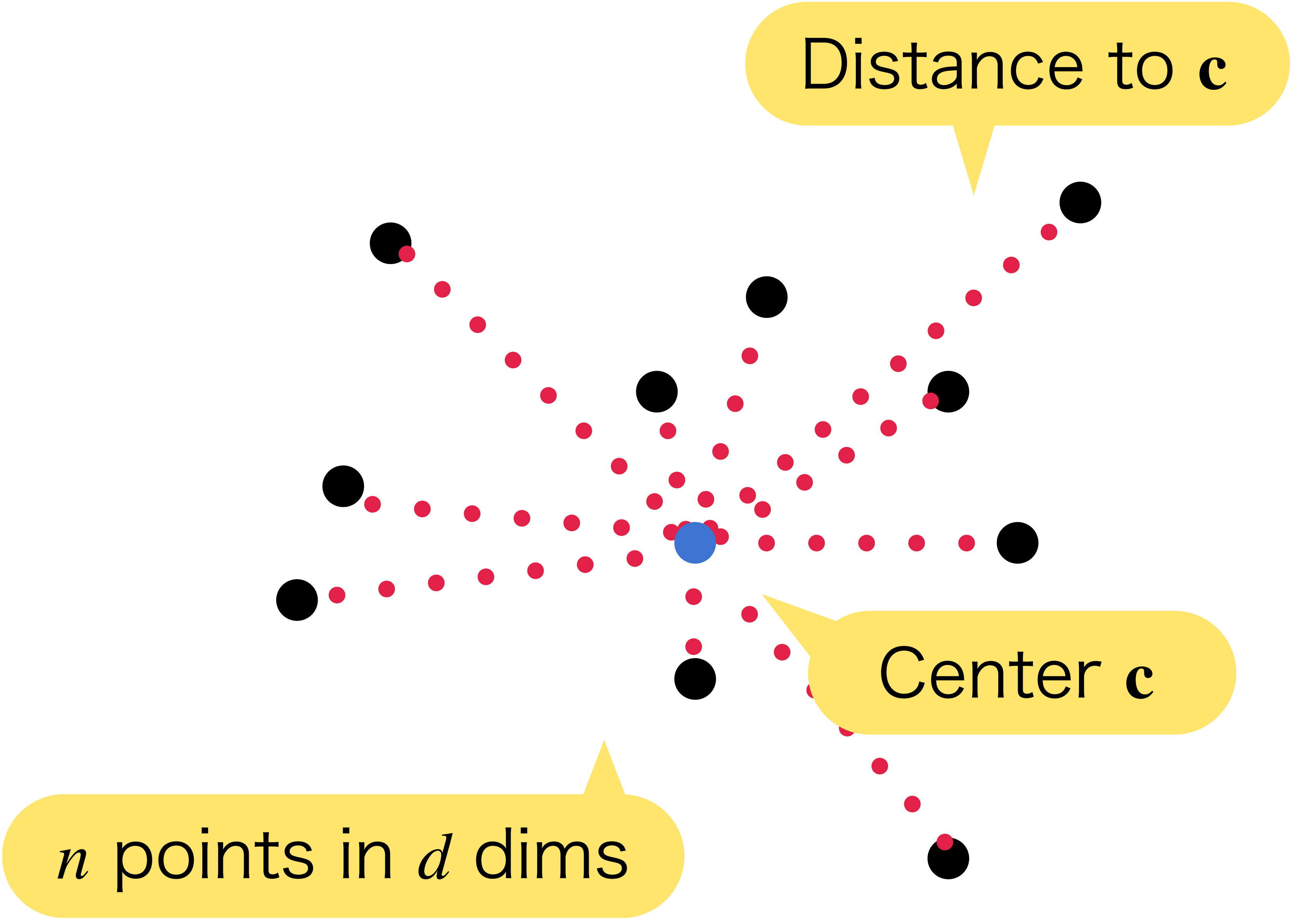
Sublinear Euclidean Power Means

Theorem. In a set of n vectors, a uniform sample of s points is sufficient for a $(1+\varepsilon)$-approximate of the Euclidean p-power mean, for

$$s = \begin{cases} \tilde{O}(\varepsilon^{-2}) & p = 1 \\ \tilde{O}(\varepsilon^{-1}) & p \in (1,2) \\ \tilde{O}(\varepsilon^{1-p}) & p \in (2,\infty) \end{cases}$$

Furthermore, these bounds are nearly optimal.

- Resolves an open Q of Cohen-Addad—Saulpic—Schwiegelshohn
- Think of $\mathbf{B}$ as the n vectors, $\mathbf{A}$ as all ones
- Dvoretzky's theorem to embed ℓ_2 into ℓ_p

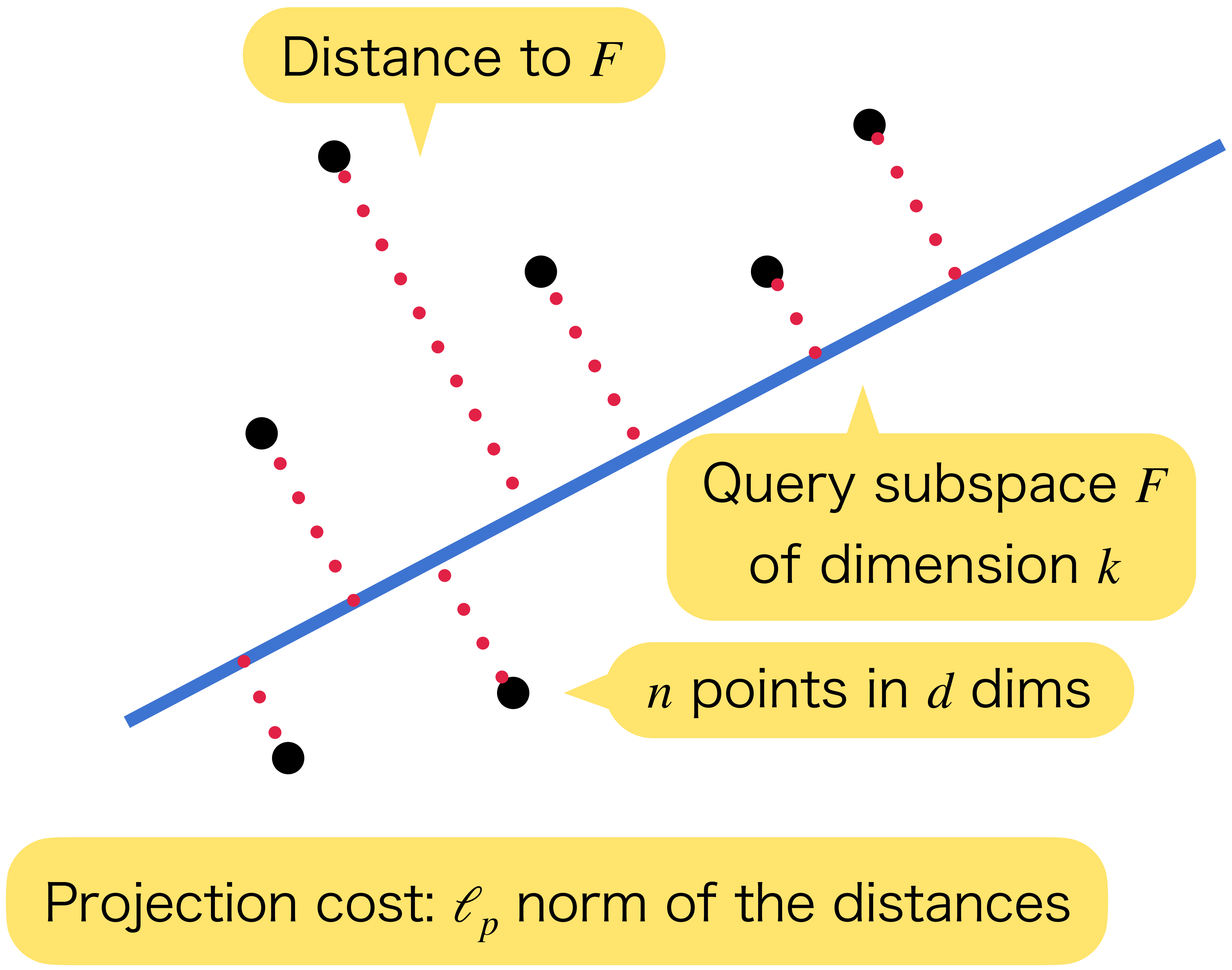


Power mean cost: ℓ_p norm of the distances

How many uniform samples do we need?

Coresets for Multiple ℓ_p Regression

Spanning Coresets for ℓ_p Subspace Approximation



How many input points are needed to span a $(1+\varepsilon)$-approximate solution?

Theorem. For $p \in (1,2)$, there is always a subset of s points that spans a $(1+\varepsilon)$-approximate solution, where $s = \tilde{O}(\varepsilon^{-1}k)$. Furthermore, these bounds are nearly optimal.

- Improves a $s = \tilde{O}(\varepsilon^{-1}k^2)$ bound of Shyamalkumar—Varadarajan
- Compute a coreset to the optimal solution

Coresets for Multiple ℓ_p Regression

Conclusion

- We study the problem of constructing coresets for multiple ℓ_p regression
- We construct coresets with nearly optimal size independent of the # responses
- Two applications:
  - Nearly optimal sublinear algorithms for Euclidean power means
  - Nearly optimal spanning coresets for ℓ_p subspace approximation