Reduced-Rank Regression with Operator Norm Error

Praneeth Kacham David P. Woodruff

Carnegie Mellon University

Introduction

We give an algorithm for the reduced-rank regression problem:

$$\min_{\text{rank-}k\ X} \|AX - B\|_2$$

given matrices $A \in \mathbb{R}^{n\times c}$ and $B \in \mathbb{R}^{n\times d}$.

- Just the usual multi-response linear regression but with **Operator Norm** error instead of **Frobenius Norm** error in addition to having a **rank** constraint.
- Can also be thought of as finding the best rank k approximation for matrix B with all the columns being spanned by columns of A.
- No algorithms with running time depending on input-sparsity were known earlier.
- For $k \ll \min(c, d)$, the running time of our algorithm is dominated by
$$(\text{nnz}(A) + \text{nnz}(B) + c^2) \cdot k/\varepsilon^{1.5} + c^\omega$$
up to logarithmic factors. Here nnz$(\cdot)$ denotes the number of nonzero entries.
- We give an analysis for **Block Krylov Iteration** algorithm to obtain a low rank approximation with **approximate** matrix-vector product oracle.
- We reduce the regression problem to low rank approximation of an appropriate matrix, construct an approximate matrix-vector product oracle for this matrix, and then obtain a $1 + \varepsilon$ approximate solution to the regression problem

Previous Work

- **Sou and Rantzer**, in control theory literature, studied the problem and gave the following equivalence:

  There is a rank k matrix X with $\|AX - B\|_2 < \beta$ if and only if
$$\sigma_{k+1}(AA^+B(\beta^2 I - \Delta)^{-1/2}) < 1.$$
  Here $\Delta = B^\mathsf{T}(I - AA^+)B$ and A^+ is the **Moore-Penrose** pseudoinverse.
- From their result, given a rank k matrix Y with
$$\|Y - AA^+B(\beta^2 I - \Delta)^{-1/2}\|_2 < 1, \quad (1)$$
  we can compute a rank k matrix X satisfying $\|AX - B\|_2 < \beta$.

A Key Takeaway

- **Surprisingly**, Sou and Rantzer also show that
$$\inf_{\text{rank-}k\ X} \|AX - B\|_2 =: \text{Opt} = \min(\sigma_{k+1}(B), \|(I - AA^+)B\|_2)$$
- So, Opt is **equal** to a simple lower bound on its value

Our Idea

- We can find a Y that satisfies (1) by computing **Singular Value Decomposition** (SVD) of
$$M := AA^+B(\beta^2 I - \Delta)^{-1/2}.$$
- Computing the matrix M and then its SVD is **very slow** and **cannot** make use of sparsity of the matrices A and B.
- We relax the requirement on Y and show that if Y is a rank k matrix satisfying
$$\|Y - AA^+B(\beta^2 I - \Delta)^{-1/2}\|_2 < 1 + \varepsilon$$
then we can find a rank k matrix X satisfying
$$\|AX - B\|_2 < (1 + C\varepsilon)\beta$$
for an absolute constant C.
- So, for $\beta = (1 + \varepsilon)$Opt and $M = AA^+B(\beta^2 I - \Delta)^{-1/2}$ we have $\sigma_{k+1}(M) < 1$ and if Y is a rank k matrix with
$$\|Y - M\|_2 \le (1 + \varepsilon)\sigma_{k+1}(M) < 1 + \varepsilon,$$
we can construct a $1 + O(\varepsilon)$ approximation.
- So we just have to find a $1 + \varepsilon$ approximate **low rank approximation** of matrix M.
- Can use Block Krylov Iteration algorithm of Musco and Musco to compute such a matrix Y.
- Issue: Block Krylov Iteration algorithm needs **exact** matrix-vector products with matrices M and M^T—**very slow** to compute.
- We resolve this by showing that **Block Krylov** iteration algorithm works even with **approximate** matrix-vector products.
- This is the **first** analysis for Block Krylov Iteration algorithm that works with **worst-case** approximate matrix-vector products.

Block Krylov Iteration

We prove the following result that shows Block Krylov Iteration algorithm can be run with approximate matrix-vector products.

Theorem

Given any vectors x and y, if we can compute x' and y' with
$$\|x' - Mx\|_2 \le \alpha\|M\|_2\|x\|_2$$
$$\|y' - M^\mathsf{T}y\|_2 \le \alpha\|M^\mathsf{T}\|_2\|y\|_2$$
in time $T(\alpha)$, then in time
$$\approx T\left(\frac{\varepsilon}{\kappa^q \cdot \text{poly}(k)}\right) qk$$
for $q \approx 1/\sqrt{\varepsilon}$, we can compute a rank k matrix M' with
$$\|M - M'\|_2 \le (1 + \varepsilon)\sigma_{k+1}(M).$$

Wrap up

- There is still the issue that we cannot compute even approximate matrix-vector products with the matrices M and M^T fast.
- We instead approximate $(1 - x)^{-1/2}$ with a suitable low degree polynomial and define another matrix $\tilde{M}$ using this polynomial.
- We show that a low rank approximation of $\tilde{M}$ can also be used to construct a solution for the regression problem.
- Using **High Precision Regression** techniques, we show that *fast* approximate matrix-vector product oracles can be constructed for matrix $\tilde{M}$.

References

[1] Cameron Musco and Christopher Musco. Randomized block krylov methods for stronger and faster approximate singular value decomposition. *Advances in Neural Information Processing Systems*, 2015:1396–1404, 2015.

[2] Kin Cheong Sou and Anders Rantzer. On generalized matrix approximation problem in the spectral norm. *Linear algebra and its applications*, 436(7):2331–2341, 2012.