EE378B Inference, Estimation, and Information Processing

Randomized numerical linear algebra

Andrea Montanari *Lecture 6-7 - Due on 2/8/2021*

Homework should be submitted via Gradescope, by Monday afternoon: the code will be communicated by an announcement on Canvas. This homework requires some

For getting credit for the class, you are required to present solutions of some of these homeworks during the first 15 minutes of class starting on 1/20. Please, sign up for (at least) one slot, and be sure that your explanation lasts 15 minutes (or less). For these presentations, you are free to choose whatever format you prefer (slides, typed notes, handwriting, ...).

This week, the presentations will be:

- Monday 2/8: Questions (a), (b)
- Wednesday 2/10: Questions (c).

Problem

In randomized numerical linear algebra people use random algorithms to accelerate linear algebra calculations (at the price of an error). One basic primitive are algorithms that take as input a matrix $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ and return a (description of) a rank-k approximation of $\boldsymbol{A}$, in time that is much smaller than mn (the time to read matrix $\boldsymbol{A}$). In what follows $\boldsymbol{a}_i$ denotes the i-th column of $\boldsymbol{A}$.

A first algorithm to achieve this goal is *column sampling* which uses a probability distribution $\boldsymbol{p} = (p_i)_{i\le n}$ over the integers $[n] := \{1, \dots, n\}$, and prooceeds as follows: (1) Sample c integers $i(1), \dots, i(c) \in [n]$ with distribution $\boldsymbol{p}$; (2) Construct a matrix $\boldsymbol{B} \in \mathbb{R}^{m\times \ell}$ whose j-th column is $\boldsymbol{a}_{i(j)}/\sqrt{\ell p_{i(\ell)}}$; (3) Take the singular value decomposition of $\boldsymbol{B}$: $\boldsymbol{B} = \boldsymbol{L}\boldsymbol{S}\boldsymbol{R}^{\mathsf{T}}$, and return $\boldsymbol{L}_k$ (the matrix containing the top k left singular vectors of $\boldsymbol{B}$.

We will denote by $\boldsymbol{A}_k$ the best rank-k approximation of $\boldsymbol{A}$ in operator (or Frobenius) norm. In particular, $\|\boldsymbol{A} - \boldsymbol{A}_k\|_{\rm op} = \sigma_{k+1}(\boldsymbol{A})$.

(a) Prove that, for any vector $\boldsymbol{z} \in \mathbb{R}^m$,

$$\boldsymbol{L}_k^{\mathsf{T}}\boldsymbol{z} = 0 \;\;\Rightarrow\;\; \|\boldsymbol{A}^{\mathsf{T}}\boldsymbol{z}\|_2^2 \le \sigma_{k+1}(\boldsymbol{A}\boldsymbol{A}^{\mathsf{T}}) + 2\|\boldsymbol{A}\boldsymbol{A}^{\mathsf{T}} - \boldsymbol{B}\boldsymbol{B}^{\mathsf{T}}\|_{\rm op}\,. \tag{1}$$

Deduce that

$$\|\boldsymbol{A} - \boldsymbol{L}_k\boldsymbol{L}_k^{\mathsf{T}}\boldsymbol{A}\|_{\rm op}^2 \le \|\boldsymbol{A} - \boldsymbol{A}_k\|_{\rm op}^2 + 2\|\boldsymbol{A}\boldsymbol{A}^{\mathsf{T}} - \boldsymbol{B}\boldsymbol{B}^{\mathsf{T}}\|_{\rm op}\,. \tag{2}$$

(b) Consider the case $\boldsymbol{p} = \boldsymbol{1}/n$ (columns are sampled uniformly at random, and define $\|\boldsymbol{A}\|_{1\to 2} := \max_{i\le n}\|\boldsymbol{a}_i\|_2$. Prove that there exists an abslute constant C such that, wih probabiliy at least $1-m^{-2}$,

$$\|\boldsymbol{A} - \boldsymbol{L}_k\boldsymbol{L}_k^{\mathsf{T}}\boldsymbol{A}\|_{\rm op}^2 \le \|\boldsymbol{A} - \boldsymbol{A}_k\|_{\rm op}^2 + C\, n\|\boldsymbol{A}\|_{\rm op}\|\boldsymbol{A}\|_{1\to 2}\sqrt{\frac{\log m}{\ell}}\,. \tag{3}$$

(c) Now assume you have access to a vector $\boldsymbol{w} \in \mathbb{R}^n$, where $w_i := \|\boldsymbol{a}_i\|_2$. Prove that uou can choose the vector of probabilities $\boldsymbol{p}$, such that, wih probabiliy at least $1 - m^{-2}$,

$$\|\boldsymbol{A} - \boldsymbol{L}_k\boldsymbol{L}_k^{\mathsf{T}}\boldsymbol{A}\|_{\rm op}^2 \le \|\boldsymbol{A} - \boldsymbol{A}_k\|_{\rm op}^2 + C\,\|\boldsymbol{A}\|_F^2\sqrt{\frac{\log m}{\ell}}\,. \tag{4}$$