Matrix approximation by sampling
Lecture 12

F. M. Faulstich

22/02/2024

Why matrix approximation

- Matrix approximations can be used to speed up certain operations
  i) Matrix multiplication
  ii) Solving linear systems
  iii) Computing matrix decompositions

  ⋮
- We have already seen some sort of matrix approximations
  i) RSVD
  ii) Matrix sketching

Empirical approximation

- Let $\mathbf{B} \in \mathbb{F}^{m \times n}$ be the target matrix
- We wish to approximate $\mathbf{B}$ with a matrix that has more “structure”

Empirical approximation

- Let $\mathbf{B} \in \mathbb{F}^{m \times n}$ be the target matrix
- We wish to approximate $\mathbf{B}$ with a matrix that has more "structure"
- Let's consider

$$\mathbf{B} = \sum_{i=1}^{I} \mathbf{B}_i$$

where $\mathbf{B}_i \in \mathbb{F}^{m \times n}$ are more "strucutre"
$\rightarrow$ Sparse
$\rightarrow$ Low-rank
$\vdots$

Empirical approximation

- Let $\mathbf{B} \in \mathbb{F}^{m \times n}$ be the target matrix
- We wish to approximate $\mathbf{B}$ with a matrix that has more "structure"
- Let's consider

$$\mathbf{B} = \sum_{i=1}^{I} \mathbf{B}_i$$

where $\mathbf{B}_i \in \mathbb{F}^{m \times n}$ are more "strucutre"
$\rightarrow$ Sparse
$\rightarrow$ Low-rank
$\vdots$

- Let $\{p_i \mid i = 1, ..., I\}$ be a (given) probability distribution and

$$\mathbf{X} := \frac{1}{p_i} \mathbf{B}_i \qquad \text{(setting } 0/0 = 0\text{)}$$

Then $\mathbf{X}$ is an unbiased estimator for $\mathbf{B}$:

$$\mathbb{E}(\mathbf{X}) = \mathbf{B}$$

Empirical approximation

- The random matrix $\mathbf{X}$ has the same structures are $\mathbf{B}_i$
- A single draw of $\mathbf{X}$ rarely gives a good approximation of $\mathbf{B}$

Empirical approximation

- The random matrix $\mathbf{X}$ has the same structures are $\mathbf{B}_i$
- A single draw of $\mathbf{X}$ rarely gives a good approximation of $\mathbf{B}$
- Define the estimator

$$\bar{\mathbf{X}}_k = \frac{1}{k} \sum_{i=1}^{k} \mathbf{X}_i \qquad \text{where } \mathbf{X}_i \sim \mathbf{X} \ \text{ i.i.d.}$$

Empirical approximation

- The random matrix $\mathbf{X}$ has the same structures are $\mathbf{B}_i$
- A single draw of $\mathbf{X}$ rarely gives a good approximation of $\mathbf{B}$
- Define the estimator

$$\bar{\mathbf{X}}_k = \frac{1}{k} \sum_{i=1}^{k} \mathbf{X}_i \qquad \text{where } \mathbf{X}_i \sim \mathbf{X} \text{ i.i.d.}$$

  By linearity $\bar{\mathbf{X}}_k$ is also an unbiased estimator of $\mathbf{B}$
- If k is small we inherited some structure of $\mathbf{B}_i$

$\Rightarrow$ How many samples do we need?

Matrix Monte Carlo

Let $\mathbf{B} \in \mathbb{F}^{m \times n}$ be a fixed matrix.
Construct a random matrix $\mathbf{X} \in \mathbb{F}^{m \times n}$ that satisfies

$$\mathbb{E}(\mathbf{X}) = \mathbf{B} \quad \text{and} \quad \|\mathbf{X}\| \leq R$$

Define the per-sample second moment:

$$v(\mathbf{X}) := \max\{\|\mathbb{E}(\mathbf{X}\mathbf{X}^*)\|, \|\mathbb{E}(\mathbf{X}^*\mathbf{X})\|\}$$

Form the matrix sampling estimator

$$\bar{\mathbf{X}}_k = \frac{1}{k}\sum_{i=1}^{k} \mathbf{X}_i \qquad \text{where } \mathbf{X}_i \sim \mathbf{X} \text{ i.i.d.}$$

Then

$$\mathbb{E}(\|\bar{\mathbf{X}}_k - \mathbf{B}\|) \leq \sqrt{\frac{2v(\mathbf{X})\log(m+n)}{k}} + \frac{2R\log(m+n)}{3k}$$

The matrix Bernstein inequality

Consider a finite sequence $\{\mathbf{S}_k\}$ of i.i.d. random matrices of size $d_1 \times d_2$. Assume that

$$\mathbb{E}(\mathbf{S}_k) = \mathbf{0} \quad \text{and} \quad \|\mathbf{S}_k\| \leq L \;\forall k$$

Introduce

$$\mathbf{Z} = \sum_k \mathbf{S}_k$$

Let $v(z)$ be the matrix variance statistic of the sum:

$$v(\mathbf{Z}) = \max\{\|\mathbb{E}(\mathbf{Z}\mathbf{Z}^*)\|, \|\mathbb{E}(\mathbf{Z}^*\mathbf{Z})\|\}$$

Then

$$\mathbb{E}(\mathbf{Z}) \leq \sqrt{2v(\mathbf{Z})\log(d_1 + d_2)} + \frac{L\log(d_1 + d_2)}{3}$$

Interpretation

How large should k be to ensure that the expected approximation error lies below a ε?