CS 6220 (Fall 2017) Data-Sparse Matrix Computations **Lecture: Sep 28th**

Randomized Range Finding Algorithm

Lecturer: Anil Damle *Scribe: Shuang Tao*

1 Fixed-rank approximation problem

We would like to solve the Fixed-rank approximation problem: given $A \in \mathbb{R}^{m\times m}$, target rank k, and oversampling parameter p, find $Q \in \mathbb{R}^{m\times(k+p}$ with $Q^\top Q = I$, such that

$$\|(I - QQ^\top)A\| \leq c_1\sigma_{k+1} + c_2 \left(\sum_{j>k} \sigma_j^2\right)^{1/2}$$

with high probability. Here σ_j denote the jth largest singular value of A.

1.1 A prototype algorithm

We now state the prototype algorithm that solve the Fixed-rank approximation problem,

1) Draw $\Omega \in \mathbb{R}^{n\times(k+p)}$ with independent Gaussian random entries
2) Form the matrix product $Y = A\Omega$.
3) Construct a matrix Q whose columns form an orthonormal basis for the range of Y through QR factorization: $Y = QR$ where $Q \in \mathbb{R}^{m\times(k+p)}$ and $R \in \mathbb{R}^{(k+p)\times n}$

The computational cost of this algorithm is:

$$(k+p)n \times \underbrace{T_{rand}}_{\text{generate a } N(0,1) r.v.} + (k+p) \times \underbrace{T_{mult}}_{\text{cost of } AX} + (k+p)^2 m$$

Empirically, we have found that the performance of this algorithm depends very little on the quality of the random number generator used. The actual cost depends substantially on the matrix A and the computational environment that we are working in.

1.2 Error bound via linear algebra

Our analysis consist of two parts,

1) Given Ω, build a deterministic bound.
2) introduce probability and get w.h.p bounds.

We will now develop a deterministic error analysis for the prototype algorithm disccussed above.

Since $QQ^\top = P_Y$: spectral projector onto $range(Y)$, we have

$$\|A - QQ^\top\| = \|(I - P_Y)A\|$$

To begin, wemust introduce some notation. Let A be an $m \times n$ matrix that has a singular value decomposition $A = U\Sigma V^\top$. Roughly speaking, the proto-algorithm tries to approximate the subspace spanned by the first k left singular vectors, where k is now a fixed number. To perform the analysis, it is appropriate to partition the singular value decomposition as follows.

$$A = U \begin{bmatrix} \overbrace{\Sigma_1}^{k} & \overbrace{0}^{\min(m,n)-k} \\ 0 & \Sigma_2 \end{bmatrix} \begin{bmatrix} \overbrace{V_1^\top}^{n} \\ V_2^\top \end{bmatrix} \begin{matrix} k \\ \min(n,m)-k \end{matrix} = \underbrace{U_1\Sigma_1 V_1^\top}_{\text{rank } k} + \underbrace{U_2\Sigma_2 V_2^\top}_{\|\|_2 = \sigma_{k+1}}$$

where $U = [U_1, U_2]$.

Let Ω be an $n \times (k+p)$ test matrix, then the sample matrix Y is expressed as

$$Y = A\Omega = U \begin{bmatrix} \Sigma_1\Omega_1 \\ \Sigma_2\Omega_2 \end{bmatrix}$$

with $\Omega_1 = V_1^\top\Omega$ and $\Omega_2 = V_2^\top\Omega$. It is a useful intuition that the block $\Sigma_1\Omega_1$ reflects the gross behavior of A, while the block $\Sigma_2\Omega_2$ represents a perturbation.

Theorem 1.1. (Deterministic error bound). A is an $m \times n$ matrix with singular value decomposition $U\Sigma V^\top$. Fix $k > 0$ and Ω, let $Y = A\Omega$, then

$$\|(I - P_Y)A\|_2^2 \leq \|\Sigma_2\|_2^2 + \|\Sigma_2\Omega_2\Omega_1^+\|_2^2$$

where Ω_1^+ represent the pseudo-inverse matrix of Ω_1. The inequality also holds for Frobenius norm $\|\|_F$.
Proof. See Halko et al. [1] for details.

1.3 Gaussian test matrices

Our analysis requires detailed information about the properties of Gaussian matrices. In particular, we must understand how the norm of a Gaussian matrix and its pseudoinverse vary. We summarize the relevant results and citations here.

Proposition 1.2. (Expected norm of a scaled Gaussian matrix). For two fixed matrices S, T and a Gaussian random matrix G ,

$$\left(\mathbb{E}\|SGT\|_F^2\right)^{1/2} = \|S\|_F\|T\|_F$$
$$\mathbb{E}\|SGT\|_2 \leq \|S\|_2\|T\|_F + \|S\|_F\|T\|_2$$

Proposition 1.3. (Norm bounds of a pseudo-inverted Gaussian matrix). Expected norm of psuedo-inverse for a $k \times k + p$ Gaussian random matrix G is

$$\mathbb{E}\|G^+\|_2 \leq \frac{e\sqrt{k+p}}{p}$$

Furthermore, for $p \geq 4$ and given only $t \geq 1$.

$$\mathbb{P}\{\|G^+\|_2 \geq \frac{e\sqrt{k+p}}{p+1}t\} \leq t^{-(p+1)}$$

1.4 Probabilistic error bounds of Algorithm

Here we present tail bounds for the approximation error, which demonstrate that the average performance of the algorithm is representative of the actual performance.

Theorem 1.4. (Deviation bounds for the Frobenius error). For algorithm with $p \geq 4$ and all $u, t \geq 1$, then

$$\|(I - P_Y)A\|_2 \leq \left[\left(1 + \sqrt{\frac{3k}{p+1}}\right)\sigma_{k+1} + \frac{\sqrt{k+p}}{p+1}\left(\sum_{j>k}\sigma_j\right)^{1/2}\right] + ut\frac{e\sqrt{k+p}}{p+1}\sigma_{k+1}$$

with failure probability at most $2t^{-p} + e^{-u^2/2}$. A simplified version is

$$\|(I - P_Y)A\|_2 \leq \left(1 + \sigma\sqrt{(k+p)p\log p}\right)\sigma_{k+1} + 3\sqrt{k+p}\left(\sum_{j>k}\sigma_j\right)^{1/2}$$

with failure probability $3p^{-p}$.

It is been demonstrated in Halko et al. [1] that the Fixed-rank approximation can be adapted to solve Fixed-accuracy approximation problem. We can get some information about $\|(I - P_Y)A\|_2$ from $\|(I - P_Y)Aw\|_2$ where w is random.

In fact, we know from previous lecture that given r, Gaussian random vectors $w^{(i)}$, $i = 1, ..., r$, we have

$$\|(I - P_Y)A\|_2 \leq 10\sqrt{\frac{2}{\pi}}\max_{i=1,\ldots,r}\|(I - P_Y)Aw^{(i)}\|_2$$

with prbability $1 - 10^{-r}$, where r is a small integer, for example, 10. See Halko et al. [1] for details.

References

[1] Nathan Halko, Per-Gunnar Martinsson, and Joel A Tropp. Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions. *SIAM review*, 53(2):217–288, 2011.