CS6220: Data-sparse Matrix Computations
Lecture 17: LSRN and Sparse Sketching Matrices

Lecturer: Anil Damle
Scriber: Johan Björck

December 15, 2017

1 LSRN

We are interested in computing $\hat{x} = A^{+}b$ to solver the problem $\min_x \|Ax - b\|_2$. We assume $A \in R^{m\times n}$ with $m >> n$ and $\text{rank}(A) = n$. At a high level, we use the following strategy

LSRN Algorithm

Starting with input matrix $A \in \mathbb{R}^{m\times n}$, rank parameter k, and error parameter ϵ.

1. Pick $\gamma > 1$ and set $s = \lceil \gamma n \rceil$.
2. Generate $G \in R^{s\times m}$ with i.i.d entries $\sim N(0,1)$.
3. Let $\hat{A} = GA$.
4. Generate the SVD-decomposition $\hat{A} = \hat{U}\hat{\Sigma}\hat{V}^T$.
5. Let $N = \hat{V}\hat{\Sigma}^{-1}$.
6. Find solution $\hat{y}$ to $\min_y \|ANy - b\|_2$.
7. Return $N\hat{y}$.

To prove that this works we need to assert that N can be used as a preconditioner and that $\kappa(AN)$ is controllable. If we solve $\|ANy - b\|_2$ and use $x^*_{right} = N\hat{y}$, for x^* and x^*_{right} to match we need $\text{range}(A^+) = \text{range}(N)$. In fact, this holds almost surely

Proof. $\text{range}(N) = \text{range}(\hat{V}) = \text{range}(A^+) = \text{range}(A^TG^T) = \text{range}(V\Sigma(GU)^T)$. Due to how G is constructed, $(GU)^T$ is full rank almost surely.

We now aim to bound $\kappa(AN)$, for which we use the two following lemmas

Lemma 1. *The spectra of AN is the same as that of $(GU)^+$, independently of A.*

Lemma 2. *Consider a $s \times n$ gaussian random matrix, with $\sigma_1 \geq \sigma_2 ... \geq \sigma_2$, then*

$$\max\left[P(\sigma_1 \geq \sqrt{s} + \sqrt{n} + t), P(\sigma_n \leq \sqrt{s} - \sqrt{n} - t)\right] \leq e^{-t^2/2}$$

Using these two lemmas it becomes possible to bound the condition number of AN, that in turns allows us to use standard conjugate gradient convergence analysis to prove convergence rate of the LSRN. These results are presented in the two theorems below.

Theorem 1. *For any $\alpha \in (0, 1 - \sqrt{n/2})$ we have* $P\left(\kappa(AN) \leq \frac{1+\alpha+\sqrt{n/s}}{1-\alpha-\sqrt{n/s}}\right) \geq 1 - e^{-\alpha^2 s/2}$

Theorem 2. *In exact arithmetic, given $\epsilon > 0$, using LSQE to solve $\min_y \|ANy - b\|_2$ converges in $\log(2/\epsilon)/log(\alpha + \sqrt{n/s})$ in the sense $\|\hat{y} - y^*\|_{(AN)^T(AN)} \leq \epsilon \|y^*\|_{(AN)^T(AN)}$ with probability at least $1 - 2e^{-\alpha^2 s/2}$ for any alpha $\in (0, 1 - \sqrt{n/2})$.*

2 Sparse Sketching Matrices

If A is sparse, the above strategy is wasteful as the random matrices will be dense. We are instead interested in finding a sparse sketching matrix $S \in R^{t \times m}$ so that we can capture AS in $\mathcal{O}(\#$ nonzeros in $A)$. Embeddings are common techniques in data analysis and primitives in algorithm design, where the typical use is dimensionality reduction. The key is finding a S such that for any fixed $m \times n$ matrix A with rank r, we have with constant probability

$$(1-\epsilon)\|Ax\|_2 \leq \|SAx\|_2 \leq (1+\epsilon)\|Ax\|_2 \qquad \forall x \in R^n$$

There are two principal types of embeddings, subspace embedding where we embed range(A), and affine embeddings where we want $\|Ax-b\| \approx \|S(Ax-b)\|$. If there was no sparsity, we could use JL-transforms which are randomized linear transformations that preservy geometry with high probability. These transformations are typically dense however, and thus not necessarily useful in the sparse context. Another alternative would be subsampled, randomized, hadamard transform. These have no dependence on A but yield $t = \mathcal{O}(n/\epsilon^2)$ at a cost $\mathcal{O}(mn \log n)$.

References

[1] Xiangrui Meng, Michael A. Saunders, and Michael W. Mahoney *LSRN: A Parallel Iterative Solver for Strongly Over- or Underdetermined Systems.* SIAM J. SCI. COMPUT. c 2014

[2] Kenneth L. Clarkson, David P. Woodruff *Low Rank Approximation and Regression in Input Sparsity Time.* ArXiv 2012