Randomized Kaczmarz Algorithm

Xiaowen Li

University of Kansas
Department of Mathematics
lee_li@ku.edu

MATH 881 Final Project
Dec. 7th
Lawrence, KS, USA

Introduction

- The Kaczmarz method is a well-known iterative algorithm for solving a linear system of equations Ax = b.

$$x_{k+1} = x_k + \frac{b_i - \langle a_i, x_k \rangle}{\|a_i\|_2^2} a_i$$

- The classical Kaczmarz's method sweeps through the rows of A in a cyclic manner. (i = k mod m+1)
- One variation of Kaczmarz's method consists of randomly choosing in each iteration the row for the projection. (We will call this method the simple randomized method in this project)

Our randomized version of Kaczmarz

- We will choose the rows for the projection steps randomly according to probabilities given by the row-norms of A.

Algorithm 1 (Random Kaczmarz algorithm) Let $Ax = b$ be a linear system of equations as in (1), and let x_0 be arbitrary initial approximation to the solution of (1). For $k = 0, 1, \dots,$ compute

$$x_{k+1} = x_k + \frac{b_{r(i)} - \langle a_{r(i)}, x_k \rangle}{\|a_{r(i)}\|_2^2} a_{r(i)},$$

where $r(i)$ is chosen from the set $\{1, 2, \dots, m\}$ at random, with probability proportional to $\|a_{r(i)}\|_2^2$.

Why choosing the probabilities according to the row-norms?

- Choosing the probabilities according to the row-norms is related to the idea of **preconditioning a matrix by row-scaling**.
- From the viewpoint of preconditioning, it is clear that other methods of choosing a diagonal preconditioner will in general perform better.
- However, finding the optimal diagonal preconditioner for the system $Ax = b$ can be very expensive while inverting the entire matrix A when A is large in size.
- Therefore, a cheaper, suboptimal alternative is needed. Scaling by the inverses of the squared row norms has been shown to be an efficient means to **balance computational costs with optimality**.

PROS

- This algorithm converges **exponentially fast** to the solution from classical Kaczmarz's method, and it also preforms better than the simple randomized method to general matrices (and not just to very restricted special cases)
- The rate of convergence is expressible in terms of **standard quantities** in numerical analysis (condition numbers of matrices)
- The average error will not be worse than the given expression:

$$\mathbb{E}\|x_k - x\|_2^2 \leq \left(1 - \kappa(A)^{-2}\right)^k \cdot \|x_0 - x\|_2^2.$$

CONS

- **Not optimal** - As from above we know the average error has a worst case, it also has a 'best case' - the estimate cannot be improved beyond a constant factor:

$$\mathbb{E}\|x_k - x\|_2^2 \geq \left(1 - 2k/\kappa(A)^2\right) \cdot \|x_0 - x\|_2^2$$

- **Still kind of expensive** - One may not even be willing to spend the computational effort to compute the row-norms of A, since the cost is still in the order of mn operations for an m $\times$ n matrix A

Experiment - input

- We followed the same structure of the input as introduced in the paper section 4.1:
- In order to show this method is efficient in general matrices. we pick our input randomly by a nonuniform sampling distribution - we consider trigonometric polynomials.
- Let r = 50 and m = 700; n = 2r+1
- We define $f(t) = \sum_{l=-r}^{r} x_l e^{2\pi i l t}$; where $x = \{x_l\}_{l=-r}^{r} \in \mathbb{C}^{2r+1}$
- And taking nodes $\{t_k\}_{k=1}^{m}$ to be non-uniformly spaced by generating the sampling points t_j randomly between [0,1] and then sort them from smallest to largest.
- With all of the above we arrive the linear system of equations: $Ax = b$; where $A_{j,k} = \sqrt{w_j} e^{2\pi i l t_j}, b_j = \sqrt{w_j} f(t_j), w_j = \frac{t_{j+1} - t_{j-1}}{2}$

Experiment - functions

- Classical Kaczmarz's method

```
pickedi = mod(i,m)+1;
```

Where i is between 1 and m, m is the number of rows in matrix A

- Simple Randomized Kaczmarz's method

```
pickedi = randi(m);
```

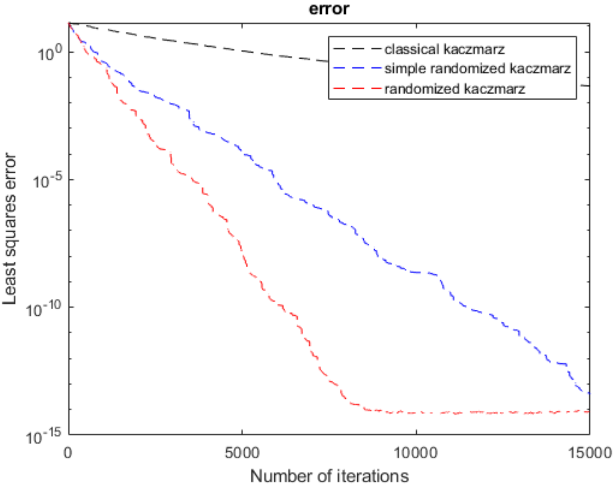
- Randomized Kaczmarz's method

```
for i = 1:m
  normrow = [normrow,norm(A(i,:))];
  index = [index,i];
end

weight = normrow/sum(normrow);

for i = 1:maxit
  %randsample to generate weighted random number from given vector
  pickedi = randsample(index,1,true,weight);
```

Experiment - outcomes



Experiment - conclusion

- Our randomized Kaczmarz's method significantly outperforms the other Kaczmarz methods, demonstrating not only the power of choosing the projections at random but also the importance of choosing the projections according to their relevance.

Future work

- How to choose a relaxation parameter to improve the convergence?
- An update on this randomized Kaczmarz's method was made by Yonina C. Eldar and Deanna Needell in their paper named *'Acceleration of randomized kaczmarz method via the johnson–lindenstrauss lemma'* published in the year 2011. They modified the method to have each iteration selecting the optimal projection from a randomly chosen set, and in most cases their algorithm significantly improves the convergence rate.

Bibliography

Yonina C Eldar and Deanna Needell.
Acceleration of randomized kaczmarz method via the johnson–lindenstrauss lemma.
Numerical Algorithms, 58(2):163–177, 2011.

Karlheinz Gröchenig.
Irregular sampling, toeplitz matrices, and the approximation of entire functions of exponential type.
Mathematics of Computation, 68(226):749–765, 1999.

Thomas Strohmer and Roman Vershynin.
Comments on the randomized kaczmarz method.
Journal of Fourier Analysis and Applications, 15(4):437–440, 2009.

Thomas Strohmer and Roman Vershynin.
A randomized kaczmarz algorithm with exponential convergence.
Journal of Fourier Analysis and Applications, 15(2):262, 2009.

Thank you