New stochastic sketching methods for Big Data Ridge Regression

Cheikh Saliou Touré

Student at ENS Cachan

Tutor : Robert Gower

Inria Paris (Sierra department)

July, 2017

Abstract

//

Contents

1. *Hadamard Sketches*

1.1 Algorithm

1.2 Convergence rate

$Z = AS^T(SAS^T)^{-1}SA$

$\rho = 1 - \lambda_{min}(A^{-\frac{1}{2}}E[Z]A^{-\frac{1}{2}})$
$S_i = I_{C_i}H.$

Denote $I_C = [I_{C_1}, I_{C_2}, \ldots, I_{C_r}]$.
$\tilde{A} = \frac{1}{n}HAH^T = \frac{H}{\sqrt{n}}A\frac{H^T}{\sqrt{n}}$.

$$A^{-\frac{1}{2}}E[Z]A^{-\frac{1}{2}} = E[A^{\frac{1}{2}}S^T(SAS^T)^{-1}SA^{\frac{1}{2}}] = \sum_i p_i A^{\frac{1}{2}}H^T I_{C_i}^T(I_{C_i}HAH^T I_{C_i}^T)^{-1}I_{C_i}HA^{\frac{1}{2}}$$

$$= A^{\frac{1}{2}}H^T\frac{1}{n}E[I_C^T(I_C\tilde{A}I_C^T)^{-1}I_C]HA^{\frac{1}{2}} = H^{-1}\tilde{A}^{\frac{1}{2}}E[I_C^T(I_C\tilde{A}I_C^T)^{-1}I_C]\tilde{A}^{\frac{1}{2}}(\frac{H^T}{n})^{-1} = \frac{H^T}{n}\tilde{A}^{\frac{1}{2}}E[I_C^T(I_C\tilde{A}I_C^T)^{-1}I_C]\tilde{A}^{\frac{1}{2}}H.$$

Hence :

$\rho = 1 - \lambda_{min}(A^{-\frac{1}{2}}E[Z]A^{-\frac{1}{2}}) = 1 - \lambda_{min}(H\frac{H^T}{n}\tilde{A}^{\frac{1}{2}}E[I_C^T(I_C\tilde{A}I_C^T)^{-1}I_C]\tilde{A}^{\frac{1}{2}})$

We recognize the convergence rate in the Randomized Newton Method and then, denoting by $\rho_{Newton}(M)$ the convergence rate of the Newton method associated with the definite positive matrix M, we obtain that :
$\rho = 1 - \lambda_{min}(\tilde{A}^{\frac{1}{2}}E[I_C^T(I_C\tilde{A}I_C^T)^{-1}I_C]\tilde{A}^{\frac{1}{2}}) = 1 - (1 - \rho_{Newton}(\tilde{A})) = \rho_{Newton}(\tilde{A})$

2. *Count-min Sketches*

2.1 Algorithm

2.2 Convergence rate

$Z = AS^T(SAS^T)^{-1}SA$

3. *Conclusion*

References

[1] ROBERT GOWER AND PETER RICHTARIK, Randomized iterative methods for linear systems, SIAM, (2015).