

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	Randomized Newton Method	2
1.1	Algorithm	2
1.2	Convergence rate (draft)	2
2	Hadamard Sketches	3
2.1	Algorithm	3
2.2	Convergence rate (draft)	3
3	Count-min Sketches	4
3.1	Algorithm	4
3.2	Convergence rate	4
4	Conclusion	5

1. Randomized Newton Method

1.1 Algorithm

1.2 Convergence rate (draft)

$$Z = AI_C^T(I_C AI_C^T)^{-1}I_C A$$

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

$$A^{-\frac{1}{2}}E[Z]A^{-\frac{1}{2}} = \sum_i p_i A^{\frac{1}{2}}I_{C_i}^T(I_{C_i}AI_{C_i}^T)^{-1}I_{C_i}A^{\frac{1}{2}}$$

for any $i \in \{1, \dots, n\}$, $A^{\frac{1}{2}}I_{C_i}^T(I_{C_i}AI_{C_i}^T)^{-1}I_{C_i}A^{\frac{1}{2}}$ is a projection matrix and then its eigenvalues are a nonempty subset of $\{0, 1\}$.

Since λ_{\max} is convex, we obtain that :

$$0 \leq \lambda_{\min}(A^{-\frac{1}{2}}E[Z]A^{-\frac{1}{2}}) \leq \lambda_{\max}(A^{-\frac{1}{2}}E[Z]A^{-\frac{1}{2}}) \leq \sum_i p_i \lambda_{\max}(A^{\frac{1}{2}}I_{C_i}^T(I_{C_i}AI_{C_i}^T)^{-1}I_{C_i}A^{\frac{1}{2}}) \leq 1.$$

$$\mathbf{C} = (I_{C_1}^T, \dots, I_{C_r}^T).$$

$A^{-\frac{1}{2}}E[Z]A^{-\frac{1}{2}} = (A^{\frac{1}{2}}\mathbf{C}D)(D\mathbf{C}^T A^{\frac{1}{2}})$ where $D = \sqrt{p} \text{diag}((S_1 A S_1^T)^{-\frac{1}{2}}, \dots, (S_r A S_r^T)^{-\frac{1}{2}})$ wherein $p = \frac{1}{\binom{n}{s}}$ is the uniform probability of choosing s rows uniformly on $\{1, \dots, n\}$, knowing that s is the sketchsize.

Proposition 1.2.1

$$\lambda_{\min}(A^{-\frac{1}{2}}E[Z]A^{-\frac{1}{2}}) \geq \frac{1}{\binom{n}{s}} \frac{\lambda_{\min}(A)}{\lambda_{\max}(A)}$$

And then :

$$\lambda_{\min}(A^{-\frac{1}{2}}E[Z]A^{-\frac{1}{2}}) \geq \lambda_{\min}(\mathbf{C}^T \mathbf{A} \mathbf{C}) \lambda_{\min}(D^2)$$

[ref Rob thesis]

$$\lambda_{\min}(D^2) = p \min_i \frac{1}{\lambda_{\max}(I_{C_i} A I_{C_i}^T)} = p \min_i \frac{1}{\lambda_{\max}(I_{C_i}^T I_{C_i}) \lambda_{\max}(A)} \geq \frac{p}{\lambda_{\max}(A)}, \text{ since for any } i \in \{1, \dots, n\}, \text{ for any } x \text{ in } \mathbb{R}^n \langle I_{C_i}^T I_{C_i} x | x \rangle = \|I_{C_i} x\|^2 \leq \|x\|^2 \text{ and then } \lambda_{\max}(I_{C_i}^T I_{C_i}) \leq 1.$$

$$\text{Therefore, } \lambda_{\min}(A^{-\frac{1}{2}}E[Z]A^{-\frac{1}{2}}) \geq p \frac{\lambda_{\min}(\mathbf{C}^T \mathbf{A} \mathbf{C})}{\lambda_{\max}(A)} = p \frac{\lambda_{\min}(A) \lambda_{\min}(\mathbf{C} \mathbf{C}^T)}{\lambda_{\max}(A)} \geq p \frac{\lambda_{\min}(A)}{\lambda_{\max}(A)}$$

2. Hadamard Sketches

2.1 Algorithm

2.2 Convergence rate (draft)

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

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

$$S_i = I_{C_i}H.$$

$$\tilde{A} = \frac{1}{n}HAH^T = \frac{H}{\sqrt{n}}A\frac{H^T}{\sqrt{n}}.$$

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

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

Hence :

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

We recognize the convergence rate in the Randomized Newton Method and then, denoting by $\rho_{\text{Newton}}(M)$ the convergence rate of the Newton method associated with the definite positive matrix M , we obtain that :

$$\rho = 1 - \lambda_{\min}(\tilde{A}^{\frac{1}{2}}E[I_C^T (I_C \tilde{A} I_C^T)^{-1} I_C] \tilde{A}^{\frac{1}{2}}) = 1 - (1 - \rho_{\text{Newton}}(\tilde{A})) = \rho_{\text{Newton}}(\tilde{A})$$

3. *Count-min Sketches*

3.1 **Algorithm**

3.2 **Convergence rate**

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

4. *Conclusion*

References

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