

## Contents

<b>1</b>	<b>Optimal Iterative Sketching with the Subsampled Randomized Hadamard Transform</b>	<b>2</b>
1.1	Background . . . . .	2
1.1.1	Problem and method . . . . .	2
1.1.2	Random matrix theory . . . . .	2

# 1 Optimal Iterative Sketching with the Subsampled Randomized Hadamard Transform

Based on [1].

The performance of iterative Hessian sketch (IHS) has only been studied empirically in existing literature. Lacotte et al. [1] show that for IHS with random matrices projected via refreshed (i.i.d.) truncated Haar matrices or subsampled randomized Hadamard transform (SRHT), the limiting rate of convergence is expected to be better than that of IHS with Gaussian random projections. Their other theoretical contributions include a closed form optimal (limiting) step size for IHS with Haar sketches, showing that momentum does not improve performance of IHS with refreshed Haar sketches, and an explicit formula for the second inverse moment of Haar sketches.

## 1.1 Background

### 1.1.1 Problem and method

Consider overdetermined least-squares problems of the form

$$\mathbf{b}^* = \arg \min_{\mathbf{b} \in \mathbb{R}^d} \left\{ f(\mathbf{b}) = \frac{1}{2} \|\mathbf{X}\mathbf{b} - \mathbf{y}\|^2 \right\}$$

where  $\mathbf{X} \in \mathbb{R}^{n \times d}$  is a given data matrix with  $n \geq d$  and  $\text{rank}(\mathbf{X}) = d$  and  $\mathbf{y} \in \mathbb{R}^n$  is a vector of observations. Iterative Hessian sketch is one iterative method for solving the problem where given step sizes  $\{\alpha_t\}$  and momentum  $\{\beta_t\}$ , the solutions are iteratively updated by

$$\mathbf{b}_{t+1} = \mathbf{b}_t - \alpha_t H_t^{-1} \nabla f(\mathbf{b}_t) + \beta_t (\mathbf{b}_t - \mathbf{b}_{t-1}).$$

The matrix  $H_t$  is an approximation of the Hessian  $H = \mathbf{X}^\top \mathbf{X}$  and is given by  $H_t = \mathbf{X}^\top \mathbf{S}_t^\top \mathbf{S}_t \mathbf{X}$  where  $\mathbf{S}_0, \dots, \mathbf{S}_t, \dots$  are refreshed (i.i.d.)  $m \times n$  sketching (random) matrices with  $m \ll n$ . The types of sketches discussed by Lacotte et al. [1] include

1. Gaussian sketches where  $(\mathbf{S}_t)_{ij} \stackrel{\text{i.i.d.}}{\sim} N(0, m^{-1})$ . Computing the matrix product  $\mathbf{S}\mathbf{X}$  is  $O(mnd)$  in general, which is larger than the cost of  $O(nd^2)$  for direct method solvers when  $m \geq d$ .
2. truncated Haar sketches using Haar matrices  $\mathbf{S}_t$  where the rows are orthonormal (TODO: other qualifications? Truncated?). Generating the matrix requires  $O(nm^2)$  using a Gram-Schmidt procedure which is larger than  $O(nd^2)$ .
3. subsampled randomized Hadamard transform where the sketch  $\mathbf{S}\mathbf{X}$  can be obtained in  $O(nd \log m)$  time. Like other orthogonal embeddings, the performance tends to be better than random projections with i.i.d. entries.

### 1.1.2 Random matrix theory

Let  $\{\mathbf{M}_n\}_n$  be a sequence of  $n \times n$  Hermitian random matrices. The empirical spectral distribution (e.s.d.) of  $\mathbf{M}_n$  is the CDF of its eigenvalues  $\lambda_1, \dots, \lambda_n$  given by  $F_{\mathbf{M}_n}(x) = \frac{1}{n} \sum_{j=1}^n \mathbb{1}[\lambda_j \leq x]$  for  $x \in \mathbb{R}$ . The eigenvalues are random and so  $F_{\mathbf{M}_n}$  is also random. TODO

## References

- [1] Jonathan Lacotte, Sifan Liu, Edgar Dobriban, and Mert Pilanci. Optimal iterative sketching methods with the subsampled randomized hadamard transform. In H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, and H. Lin, editors, *Advances in Neural Information Processing Systems*, volume 33, pages 9725–9735. Curran Associates, Inc., 2020. URL <https://proceedings.neurips.cc/paper/2020/file/6e69ebbfad976d4637bb4b39de261bf7-Paper.pdf>.