CSCI 6971 Project Report

Iterative Hessian Sketch: Fast and Accurate Solution Approximation for Constrained Least-Squares (2014)

by Mert Pilanci Martin J. Wainwright

Jiajia Yu

Department of Mathematics

April 22, 2019

Outline

Least Square Problems

- Data: $y \in \mathbb{R}^n$, $A \in \mathbb{R}^{n\times d} (n \gg d)$, Ground truth: $x^* \in \mathbb{R}^d$,

$$y = Ax^* + \omega, \quad \omega \sim N(\mathbf{0}, I_d).$$

- Goal: recover the truth x^* from data (y, A)

$$x^{LS} := \arg\min_{x \in C} \|Ax - y\|_2^2. \tag{1}$$

- Example: linear regression (unconstrained),
  SVM (l_1 constrained).
- Guarantee:
  $$\|x^{LS} - x^*\|_A := \frac{1}{\sqrt{n}} \|A(x^{LS} - x^*)\|_2 = \mathcal{O}\left(\sqrt{\frac{d}{n}}\right).$$
- Issue: when $n \gg d$, it is expensive to directly solve (1).

Least Square Problems

- Data: $y \in \mathbb{R}^n$, $A \in \mathbb{R}^{n\times d}(n \gg d)$, Ground truth: $x^* \in \mathbb{R}^d$,

$$y = Ax^* + \omega, \quad \omega \sim N(\mathbf{0}, I_d).$$

- Goal: recover the truth x^* from data (y, A)

$$x^{LS} := \arg\min_{x\in C} \|Ax - y\|_2^2. \tag{1}$$

- Example: linear regression (unconstrained),
  SVM (l_1 constrained).
- Guarantee:

$$\|x^{LS} - x^*\|_A := \frac{1}{\sqrt{n}}\|A(x^{LS} - x^*)\|_2 = \mathcal{O}\left(\sqrt{\frac{d}{n}}\right).$$

- Issue: when $n \gg d$, it is expensive to directly solve (1).

Least Square Problems

- Data: $y \in \mathbb{R}^n$, $A \in \mathbb{R}^{n \times d} (n \gg d)$, Ground truth: $x^* \in \mathbb{R}^d$,

$$y = Ax^* + \omega, \quad \omega \sim N(\mathbf{0}, I_d).$$

- Goal: recover the truth x^* from data (y, A)

$$x^{LS} := \arg\min_{x \in C} \|Ax - y\|_2^2. \tag{1}$$

- Example: linear regression (unconstrained),
  SVM (l_1 constrained).
- Guarantee:
  $$\|x^{LS} - x^*\|_A := \frac{1}{\sqrt{n}} \|A(x^{LS} - x^*)\|_2 = \mathcal{O}\left(\sqrt{\frac{d}{n}}\right).$$
- Issue: when $n \gg d$, it is expensive to directly solve (1).

Classical Sketching Method

- Ground truth: $y = Ax^* + \omega, \quad \omega \sim N(\mathbf{0}, I_d)$.
- Original problem: $y \in \mathbb{R}^n$, $A \in \mathbb{R}^{n\times d}(n \gg d)$,

$$x^{LS} := \arg\min_{x\in C} \|Ax - y\|_2^2. \tag{1}$$

- Classical sketching method: $S \in \mathbb{R}^{m\times n}$,

$$\tilde{x} := \arg\min_{x\in C} \|SAx - Sy\|_2^2. \tag{2}$$

- Guarantee: if $m \gtrsim \frac{d}{\epsilon^2}$, then

$$\|\tilde{x} - x^{LS}\|_A \leq \epsilon \|Ax^{LS} - y\|_2, \quad \text{w.h.p.}$$

Classical Sketching Method

- Ground truth: $y = Ax^* + \omega, \quad \omega \sim N(\mathbf{0}, I_d)$.
- Original problem: $y \in \mathbb{R}^n$, $A \in \mathbb{R}^{n\times d}(n \gg d)$,

$$x^{LS} := \arg\min_{x\in C} \|Ax - y\|_2^2. \tag{1}$$

- Classical sketching method: $S \in \mathbb{R}^{m\times n}$,

$$\tilde{x} := \arg\min_{x\in C} \|SAx - Sy\|_2^2. \tag{2}$$

- Guarantee: if $m \gtrsim \frac{d}{\epsilon^2}$, then

$$\|\tilde{x} - x^{LS}\|_A \leq \epsilon\|Ax^{LS} - y\|_2, \quad \text{w.h.p.}$$

Classical Sketching Method

- Classical sketching method: $S \in \mathbb{R}^{m \times n}$,

$$\tilde{x} := \arg\min_{x \in C} \|SAx - Sy\|_2^2. \tag{2}$$

- Guarantee: if $m \gtrsim \frac{d}{\epsilon^2}$, then

$$\|\tilde{x} - x^{LS}\|_A \leq \epsilon \|Ax^{LS} - y\|_2, \quad \text{w.h.p.}$$

- Goal: $\|\tilde{x} - x^{LS}\|_A \approx \|x^{LS} - x^*\|_A = \mathcal{O}\left(\sqrt{\frac{d}{n}}\right)$.
- Issue: to achieve desired accuracy, we need $m = \mathcal{O}(n)$.

Classical Sketching Method

- Classical sketching method: $S \in \mathbb{R}^{m \times n}$,

$$\tilde{x} := \arg\min_{x \in C} \|SAx - Sy\|_2^2. \tag{2}$$

- Guarantee: if $m \gtrsim \frac{d}{\epsilon^2}$, then

$$\|\tilde{x} - x^{LS}\|_A \leq \epsilon \|Ax^{LS} - y\|_2, \quad \text{w.h.p.}$$

- Goal: $\|\tilde{x} - x^{LS}\|_A \approx \|x^{LS} - x^*\|_A = \mathcal{O}\left(\sqrt{\frac{d}{n}}\right)$.
- Issue: to achieve desired accuracy, we need $m = \mathcal{O}(n)$.

Hessian Sketch

$$x^{LS} := \arg\min_{x \in C} \|Ax - y\|_2^2 = \arg\min_{x \in C} \left\{ \frac{1}{2} \|Ax\|_2^2 - \langle y, Ax \rangle \right\}$$

$$\tilde{x} := \arg\min_{x \in C} \|SAx - Sy\|_2^2 = \arg\min_{x \in C} \left\{ \frac{1}{2} \|SAx\|_2^2 - \langle S^T y, Ax \rangle \right\}$$

- Hessian Sketch:

$$\hat{x} := \arg\min_{x \in C} \left\{ \frac{1}{2} \|SAx\|_2^2 - \langle y, A^T x \rangle \right\} \quad (3)$$

Hessian Sketch

$$x^{LS} := \arg\min_{x \in C} \|Ax - y\|_2^2 = \arg\min_{x \in C} \left\{ \frac{1}{2} \|Ax\|_2^2 - \langle y, Ax \rangle \right\}$$

$$\tilde{x} := \arg\min_{x \in C} \|SAx - Sy\|_2^2 = \arg\min_{x \in C} \left\{ \frac{1}{2} \|SAx\|_2^2 - \langle S^T y, Ax \rangle \right\}$$

- Hessian Sketch:

$$\hat{x} := \arg\min_{x \in C} \left\{ \frac{1}{2} \|SAx\|_2^2 - \langle y, A^T x \rangle \right\} \tag{3}$$

Guarantees for Hessian Sketch

$$\hat{x} := \arg\min_{x \in C} \left\{ \frac{1}{2} \|SAx\|_2^2 - \langle A^T y, x \rangle \right\} \tag{3}$$

Consider transformed tangent cone

$$\mathcal{K}^{LS} := \left\{ v \in \mathbb{R}^d \middle| v = tA(x - x^{LS}) \text{ for some } t \geq 0 \text{ and } x \in C \right\}$$

and unit sphere $\mathcal{S}^{n-1} := \left\{ v \in \mathbb{R}^d \middle| \|v\|_2 = 1 \right\}$.

Illustration of transform tangent cone

$$\mathcal{K}^{LS} := \left\{ v \in \mathbb{R}^d \middle| v = tA(x - x^{LS}) \text{ for some } t \geq 0 \text{ and } x \in C \right\}$$

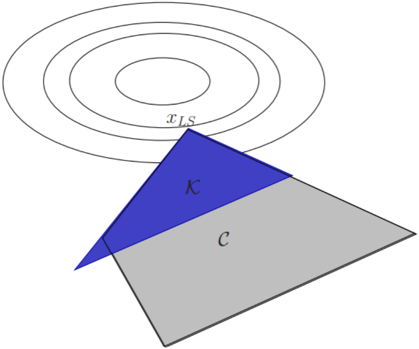


Figure: [Wainwright 2015] Tangent cone

Guarantees for Hessian Sketch

$$\hat{x} := \arg\min_{x \in C} \left\{ \frac{1}{2} \|SAx\|_2^2 - \langle A^T y, x \rangle \right\} \tag{3}$$

Consider transformed tangent cone $\mathcal{K}^{LS}$ and unit sphere $\mathcal{S}^{n-1}$.
And define

$$Z_1(S) := \inf_{v \in \mathcal{K}^{LS} \cap \mathcal{S}^{n-1}} \|Sv\|_2^2,$$

$$Z_2(S) := \sup_{v \in \mathcal{K}^{LS} \cap \mathcal{S}^{n-1}} \left| \langle u, (S^T S - I_n) v \rangle \right| \text{ where } \|u\|_2 = 1.$$

Guarantees for Hessian Sketch

$$\hat{x} := \arg\min_{x \in C} \left\{ \frac{1}{2} \|SAx\|_2^2 - \langle A^T y, x \rangle \right\} \tag{3}$$

$$\mathcal{K}^{LS} := \left\{ v \in \mathbb{R}^d \,\middle|\, v = tA(x - x^{LS}) \text{ for some } t \geq 0 \text{ and } x \in C \right\}$$

$$Z_1(S) := \inf_{v \in \mathcal{K}^{LS} \cap \mathcal{S}^{n-1}} \|Sv\|_2^2,$$

$$Z_2(S) := \sup_{v \in \mathcal{K}^{LS} \cap \mathcal{S}^{n-1}} \left| \langle u, (S^T S - I_n) v \rangle \right| \text{ where } \|u\|_2 = 1.$$

Proposition 1 [Pilanci 2016]

For any convex set C and any sketching matrix $S \in \mathbb{R}^{m \times n}$, the Hessian sketch solution $\hat{x}$ satisfies the bound

$$\|\hat{x} - x^{LS}\|_A \leq \frac{Z_2}{Z_1} \|x^{LS}\|_A$$

Guarantees for Hessian Sketch

$$\hat{x} := \arg\min_{x \in C} \left\{ \frac{1}{2} \|SAx\|_2^2 - \langle A^T y, x \rangle \right\} \tag{3}$$

$$\mathcal{K}^{LS} := \left\{ v \in \mathbb{R}^d \middle| v = tA(x - x^{LS}) \text{ for some } t \geq 0 \text{ and } x \in C \right\},$$

Define Gaussian width to measure the size of $\mathcal{K}^{LS}$:

$$\mathcal{W}(\mathcal{K}^{LS}) := \mathbb{E}_g \left[\sup_{v \in \mathcal{K}^{LS} \cap \mathcal{S}^{n-1}} |\langle g, v \rangle| \right].$$

Lemma 1(a) [Pilanci 2016]

For sub-Gaussian sketch matrices, given a sketch size $m \geq \frac{c_0}{\rho^2} \mathcal{W}^2(\mathcal{K}^{LS})$, we have

$$\|\hat{x} - x^{LS}\|_A \leq \rho \|x^{LS}\|_A \quad \text{w.h.p.}$$

Issues of Hessian Sketch

$$\tilde{x} := \arg\min_{x \in C} \|SAx - Sy\|_2^2. \tag{2}$$

$$\hat{x} := \arg\min_{x \in C} \left\{ \frac{1}{2} \|SAx\|_2^2 - \langle A^T y, x \rangle \right\} \tag{3}$$

- Classical sketch: if $m \gtrsim \frac{d}{\epsilon^2}$, then

$$\|\tilde{x} - x^{LS}\|_A \leq \epsilon \|Ax^{LS} - y\|_2, \quad \text{w.h.p.}$$

- Hessian sketch: if $m \geq \frac{c_0}{\rho^2} \mathcal{W}^2(\mathcal{K}^{LS})$, then

$$\|\hat{x} - x^{LS}\|_A \leq \rho \|x^{LS}\|_A \quad \text{w.h.p.}$$

One-step Hessian sketch has the same issue as classical sketch has. But now we can do Hessian Sketch iteratively.

Issues of Hessian Sketch

$$\tilde{x} := \arg\min_{x \in C} \|SAx - Sy\|_2^2. \tag{2}$$

$$\hat{x} := \arg\min_{x \in C} \left\{ \frac{1}{2}\|SAx\|_2^2 - \langle A^T y, x\rangle \right\} \tag{3}$$

- Classical sketch: if $m \gtrsim \frac{d}{\epsilon^2}$, then

$$\|\tilde{x} - x^{LS}\|_A \le \epsilon \|Ax^{LS} - y\|_2, \quad \text{w.h.p.}$$

- Hessian sketch: if $m \ge \frac{c_0}{\rho^2}\mathcal{W}^2(\mathcal{K}^{LS})$, then

$$\|\hat{x} - x^{LS}\|_A \le \rho \|x^{LS}\|_A \quad \text{w.h.p.}$$

One-step Hessian sketch has the same issue as classical sketch has. But now we can do Hessian Sketch iteratively.

Iterative Hessian Sketch

$$x^{LS} := \arg\min \|Ax - y\|_2^2 = \arg\min_{x \in C} \left\{ \frac{1}{2}\|Ax\|_2^2 - \langle y, Ax \rangle \right\} \tag{1}$$

$$x^1 := \arg\min_{x \in C} \left\{ \frac{1}{2}\|SAx\|_2^2 - \langle y, Ax \rangle \right\}$$

$$\|x^1 - x^{LS}\|_A \leq \rho \|x^{LS}\|_A.$$

Iterative Hessian Sketch

$$(1) \qquad x^{LS} := \arg\min \|Ax - y\|_2^2 = \arg\min_{x \in C} \left\{ \frac{1}{2}\|Ax\|_2^2 - \langle y, Ax \rangle \right\}$$

$$x^1 := \arg\min_{x \in C} \left\{ \frac{1}{2}\|SAx\|_2^2 - \langle y, Ax \rangle \right\}$$

$$\|x^1 - x^{LS}\|_A \leq \rho \|x^{LS}\|_A.$$

$$(2) \qquad x^{LS} - x^1 = \arg\min_{x \in C - x^1} \|Ax - (y - Ax^1)\|_2^2$$

$$= \arg\min_{x \in C - x^1} \left\{ \frac{1}{2}\|Ax\|_2^2 - \langle y - Ax^1, Ax \rangle \right\}$$

$$x^{1.5} := \arg\min_{x \in C - x^1} \left\{ \frac{1}{2}\|SAx\|_2^2 - \langle y - Ax^1, Ax \rangle \right\}$$

$$\|x^{1.5} - (x^1 - x^{LS})\|_A \leq \rho \|x^1 - x^{LS}\|_A \leq \rho^2 \|x^{LS}\|_A.$$

Iterative Hessian Sketch

$$x^1 := \arg\min_{x \in C} \left\{ \frac{1}{2} \|SAx\|_2^2 - \langle y, Ax \rangle \right\}$$

$$\|x^1 - x^{LS}\|_A \leq \rho \|x^{LS}\|_A.$$

$$x^{1.5} := \arg \min_{x \in C - x^1} \left\{ \frac{1}{2} \|SAx\|_2^2 - \langle y - Ax^1, Ax \rangle \right\}$$

$$x^2 := \arg\min_{x \in C} \left\{ \frac{1}{2} \|SA(x - x^1)\|_2^2 - \langle y - Ax^1, Ax \rangle \right\} = x^{1.5} + x^1$$

$$\|x^{1.5} - (x^1 - x^{LS})\|_A \leq \rho^2 \|x^{LS}\|_A.$$

$$\|x^2 - x^{LS}\|_A = \|x^{1.5} - (x^{LS} - x^1)\|_A \leq \rho^2 \|x^{LS}\|_A.$$

Iterative Hessian Sketch

Algorithm: Iterative Hessian Sketch [Pilanci 2016]

Given an iteration number $N \geq 1$:

(1) Initialize at $x^0 = 0$.

(2) For iterations $t = 0, 1, 2, \ldots, N-1$, generate an independent sketch matrix $S^{t+1} \in \mathbb{R}^{m \times n}$, and perform the update

$$x^{t+1} := \arg\min_{x \in C} \left\{ \frac{1}{2} \|SA(x - x^t)\|_2^2 - \langle y - Ax^t, Ax \rangle \right\}.$$

(3) Return the estimate $\hat{x} = x^N$.

Iterative Hessian Sketch

Algorithm: Iterative Hessian Sketch [Pilanci 2016]

Given an iteration number $N \geq 1$:
(1) Initialize at $x^0 = 0$.
(2) For iterations $t = 0, 1, 2, \ldots, N-1$, generate an independent sketch matrix $S^{t+1} \in \mathbb{R}^{m \times n}$, and perform the update

$$x^{t+1} := \arg\min_{x \in C} \left\{ \frac{1}{2} \| SA(x - x^t) \|_2^2 - \langle y - Ax^t, Ax \rangle \right\}.$$

(3) Return the estimate $\hat{x} = x^N$.

Iterative Hessian Sketch

Algorithm: Iterative Hessian Sketch [Pilanci 2016]

Given an iteration number $N \geq 1$:

(1) Initialize at $x^0 = 0$.

(2) For iterations $t = 0, 1, 2, \ldots, N-1$, generate an independent sketch matrix $S^{t+1} \in \mathbb{R}^{m \times n}$, and perform the update

$$x^{t+1} := \arg\min_{x \in C} \left\{ \frac{1}{2} \|SA(x - x^t)\|_2^2 - \langle y - Ax^t, Ax \rangle \right\}.$$

(3) Return the estimate $\hat{x} = x^N$.

Guarantees for Iterative Hessian Sketch

Corollary 1 [Pilanci 2016]

Fix some $\rho \in (0, 1/2)$, and choose sub-Gaussian sketches with sketch dimension $m \geq \frac{c_0}{\rho^2}\mathcal{W}^2(\mathcal{K}^{LS})$. If we apply IHS algorithm for $N(\rho, \epsilon) := 1 + \frac{\log(1/\epsilon)}{\log(1/\rho)}$ steps, then the output $\hat{x} = x^N$ satisfies

$$\|\hat{x} - x^{LS}\|_A \leq \epsilon \|x^{LS}\|_A$$

with probability at least $1 - c_1 N(\rho, \epsilon) \exp(-c_2 m \rho^2)$

Geometrical Decrease of the error

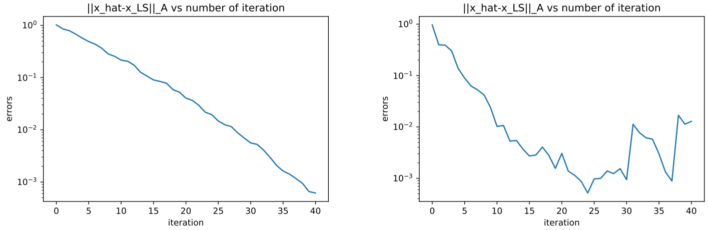


Figure: $\|\hat{x} - x^{LS}\|_A$ vs iteration number: left: unconstrained, right: l_1 constrained.

Classical Sketch vs Iterative Hessian Sketch

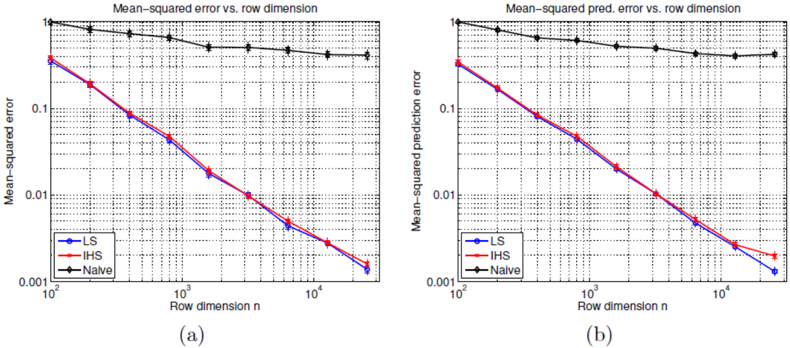
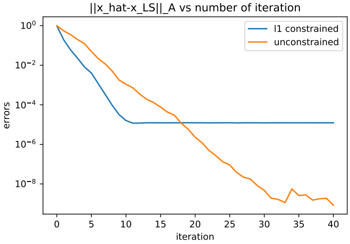


Figure: [Pilanci 2016] Plots of mean-squared error vs the row dimension.

Unconstrained vs l_1 Constrained



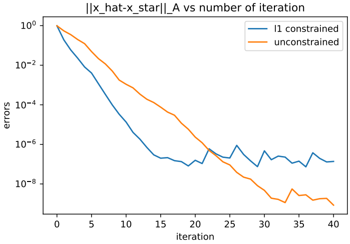


Figure: Plots of mean-squared error vs the row dimension, $m = 5s \ln d \approx 191$.

Unconstrained vs l_1 Constrained

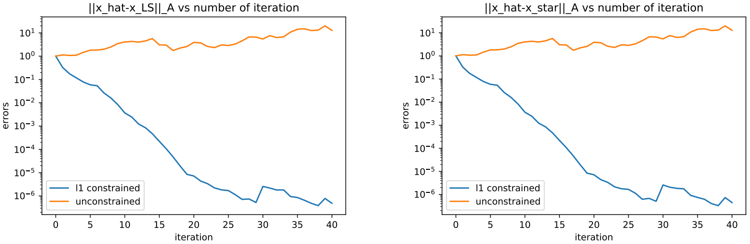


Figure: Plots of mean-squared error vs the row dimension, $m = 4s\ln(ed/s) \approx 90$.

Summary

- Random sketching method can approximately solve least-squares problems.
- Iterative Hessian sketch can solve least-squares problems with better accuracy than classical sketching method.
- The "best" sketch size depends on the width of transformed tangent cone.

References

Pilanci, M., & Wainwright, M. J.
Iterative Hessian sketch: Fast and accurate solution approximation for constrained least-squares.
The Journal of Machine Learning Research, 17(1), 1842-1879. 2016

Wainwright, M.J.
Randomized algorithms for optimization: Statistical and computational guarantees. [Slides]
Retrieved from https://warwick.ac.uk/fac/sci/wdsi/events/yobd/computational/wainwright_warwick15_final.pdf