Iterative Hessian Sketch for Constrained Least-Squares

Lydia Hsu, Shuaiwen Wang

December 10, 2018

Paper: *Iterative Hessian Sketch: Fast and Accurate Solution Approximation for Constrained Least-Squares*

Problem description

- Consider the **constrained least-squares**

$$x_{ls} := \arg\min_{x \in C} \frac{1}{2}\|Ax - y\|_2^2$$

where C: constraint; $A \in \mathbb{R}^{n \times p}$: design matrix; y: response.

Problem description

- Consider the **constrained least-squares**

$$x_{ls} := \arg\min_{x \in C} \frac{1}{2}\|Ax - y\|_2^2$$

where C: constraint; $A \in \mathbb{R}^{n \times p}$: design matrix; y: response.

- Issue: when n large, computationally **expensive**;

Problem description

- Consider the **constrained least-squares**

$$x_{ls} := \arg\min_{x \in C} \frac{1}{2}\|Ax - y\|_2^2$$

where C: constraint; $A \in \mathbb{R}^{n \times p}$: design matrix; y: response.

- Issue: when n large, computationally **expensive**;
- Idea: substitute A with its "sketched" version SA;

Problem description

- Consider the **constrained least-squares**

$$x_{ls} := \arg\min_{x \in C} \frac{1}{2}\|Ax - y\|_2^2$$

where C: constraint; $A \in \mathbb{R}^{n \times p}$: design matrix; y: response.

- Issue: when n large, computationally **expensive**;
- Idea: substitute A with its "sketched" version SA;
- $S \in \mathbb{R}^{m \times n}$: sketching matrix; $\mathbb{E} S^\top S = I_n$.

Common sketching matrices

- sub-Gaussian: $S_{ij} \sim_{iid} \frac{1}{\sqrt{m}}\mathcal{N}(0, 1)$;

Common sketching matrices

- sub-Gaussian: $S_{ij} \sim_{iid} \frac{1}{\sqrt{m}}\mathcal{N}(0,1)$;
- random row sampling: randomly select m rows from I_n (rescaled by $\sqrt{\frac{n}{m}}$);

Common sketching matrices

- sub-Gaussian: $S_{ij} \sim_{iid} \frac{1}{\sqrt{m}}\mathcal{N}(0,1)$;
- random row sampling: randomly select m rows from I_n (rescaled by $\sqrt{\frac{n}{m}}$);
- randomized orthogonal system: randomly select m rows from an orthogonal matrix H (rescaled by $\sqrt{\frac{n}{m}}$);

Classical Sketch

- Sketch not only A, but also y:

$$x_{cs} := \arg\min_{x \in C} \frac{1}{2} \|SAx - Sy\|_2^2$$

Classical Sketch

- Sketch not only A, but also y:

$$x_{cs} := \arg\min_{x \in C} \frac{1}{2}\|SAx - Sy\|_2^2$$

- Classical sketch is **sub-optimal**: assume a linear ground truth $y = Ax_* + w$:

$$\|x_{cs} - x_*\| \sim \|x_{ls} - x_*\| \quad \Rightarrow \quad m = \Theta(n)$$

Classical Sketch

- Sketch not only A, but also y:

$$x_{cs} := \arg\min_{x \in C} \frac{1}{2}\|SAx - Sy\|_2^2$$

- Classical sketch is **sub-optimal**: assume a linear ground truth $y = Ax_* + w$:

$$\|x_{cs} - x_*\| \sim \|x_{ls} - x_*\| \quad \Rightarrow \quad m = \Theta(n)$$

- We expect the optimal method satisfies $m \approx O(p)$ Think about this part more carefully?;

Hessian Sketch

- Only sketch A, but not y:

$$x_{hs} := \arg\min_{x \in C} \frac{1}{2}\|SAx\|_2^2 - \langle A^\top y, x \rangle.$$

Hessian Sketch

- Only sketch A, but not y:

$$x_{hs} := \arg\min_{x \in C} \frac{1}{2}\|SAx\|_2^2 - \langle A^\top y, x\rangle.$$

- **Sub-optimal** in the same sense;

Hessian Sketch

- Only sketch A, but not y:

$$x_{hs} := \arg\min_{x \in C} \frac{1}{2}\|SAx\|_2^2 - \langle A^\top y, x\rangle.$$

- **Sub-optimal** in the same sense;
- With $m \sim O(p)$ sample size, we can obtain a weak result:

$$\|x_{hs} - x_{ls}\| \leq \rho\|x_{ls}\|, \quad \text{w.h.p.}$$

where $\rho \in (0, \frac{1}{2})$.

Hessian Sketch

- Only sketch A, but not y:

$$x_{hs} := \arg\min_{x \in C} \frac{1}{2}\|SAx\|_2^2 - \langle A^\top y, x\rangle.$$

- **Sub-optimal** in the same sense;
- With $m \sim O(p)$ sample size, we can obtain a weak result:

$$\|x_{hs} - x_{ls}\| \leq \rho\|x_{ls}\|, \quad \text{w.h.p.}$$

  where $\rho \in (0, \frac{1}{2})$.
- *An iterative idea for improvement*: "$\|x^{t+1} - x_{ls}\| \leq \rho\|x^t - x_{ls}\|$"?

A detailed idea:

- Given x^t, a modified least square with $x_{ls} - x^t$ as minimizer:

$$\min_u \frac{1}{2}\|A(u + x^t)\|_2^2 - \langle A^\top y, (u + x^t)\rangle$$

Iterative Hessian Sketch

A detailed idea:

- Given x^t, a modified least square with $x_{ls} - x^t$ as minimizer:

$$\min_u \frac{1}{2}\|Au\|_2^2 - \langle A^\top (y - Ax^t), u \rangle$$

A detailed idea:

- Given x^t, a modified least square with $x_{ls} - x^t$ as minimizer:

$$\min_u \frac{1}{2}\|Au\|_2^2 - \langle A^\top(y - Ax^t), u\rangle$$

- The corresponding Hessian sketch:

$$u^{t+1} = \arg\min_u \frac{1}{2}\|SAu\|_2^2 - \langle A^\top(y - Ax^t), u\rangle$$

Iterative Hessian Sketch

A detailed idea:

- Given x^t, a modified least square with $x_{ls} - x^t$ as minimizer:

$$\min_u \frac{1}{2}\|Au\|_2^2 - \langle A^\top(y - Ax^t), u\rangle$$

- The corresponding Hessian sketch:

$$u^{t+1} = \arg\min_u \frac{1}{2}\|SAu\|_2^2 - \langle A^\top(y - Ax^t), u\rangle$$

- Property of Hessian sketch: $\|u^{t+1} - (x_{ls} - x^t)\| \le \rho\|x_{ls} - x^t\|$;

A detailed idea:

- Given x^t, a modified least square with $x_{ls} - x^t$ as minimizer:

$$\min_u \frac{1}{2}\|Au\|_2^2 - \langle A^\top(y - Ax^t), u\rangle$$

- The corresponding Hessian sketch:

$$u^{t+1} = \arg\min_u \frac{1}{2}\|SAu\|_2^2 - \langle A^\top(y - Ax^t), u\rangle$$

- Property of Hessian sketch: $\|u^{t+1} - (x_{ls} - x^t)\| \leq \rho\|x_{ls} - x^t\|$;
- Define $x^{t+1} = x^t + u^{t+1}$, then:

$$\|x^{t+1} - x_{ls}\| \leq \rho\|x^t - x_{ls}\|, \quad \text{w.h.p.}$$

Algorithm

1. Initialize $x^0 = 0$.
2. For $t = 1, ...N$, generate independent sketches $S^t \in \mathbb{R}^{m \times n}$ and

$$x^t = \arg\min_{x \in C} \frac{1}{2}\|S^t A(x - x^{t-1})\|_2^2 - \langle A^T(y - Ax^{t-1}), x - x^{t-1}\rangle$$

3. Return $\hat{x} = x^N$.

Iterative Hessian Sketch

Properties of IHS:

- $m = O(p)$ to guarantee $\|x^t - x_{ls}\| \leq \rho \|x^{t-1} - x_{ls}\|$ in each iteration;

Properties of IHS:

- $m = O(p)$ to guarantee $\|x^t - x_{ls}\| \leq \rho \|x^{t-1} - x_{ls}\|$ in each iteration;
- To achieve an accuracy of ϵ, we need $\log(\frac{1}{\epsilon})$ iterations;

Iterative Hessian Sketch

Properties of IHS:

- $m = O(p)$ to guarantee $\|x^t - x_{ls}\| \leq \rho \|x^{t-1} - x_{ls}\|$ in each iteration;
- To achieve an accuracy of ϵ, we need $\log(\frac{1}{\epsilon})$ iterations;
- Optimality:
  - To achieve the same bound as $\|x_{ls} - x_*\| \approx O(\frac{1}{\sqrt{n}})$, we need $N = log(n)$ iterations;

Iterative Hessian Sketch

Properties of IHS:

- $m = O(p)$ to guarantee $\|x^t - x_{ls}\| \leq \rho \|x^{t-1} - x_{ls}\|$ in each iteration;
- To achieve an accuracy of ϵ, we need $\log(\frac{1}{\epsilon})$ iterations;
- Optimality:
  - To achieve the same bound as $\|x_{ls} - x_*\| \approx O(\frac{1}{\sqrt{n}})$, we need $N = log(n)$ iterations;
  - Union bound: $\|x^N - x_{ls}\| \leq \rho^N \|x_{ls}\|$, *w.h.p.*.

Simulations in the paper

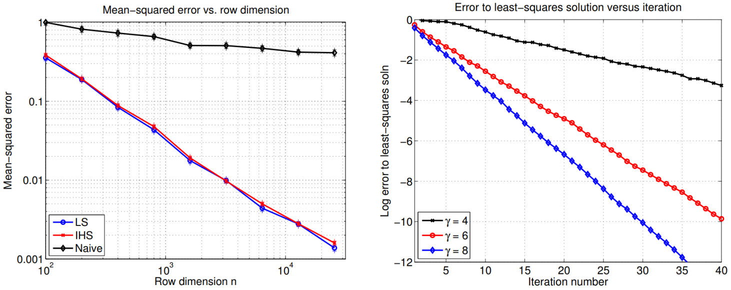

Figure : Left: classical sketch vs. iterative hessian sketch; Right: different choices of $m = \gamma p$.

Simulations - different sketch matrices

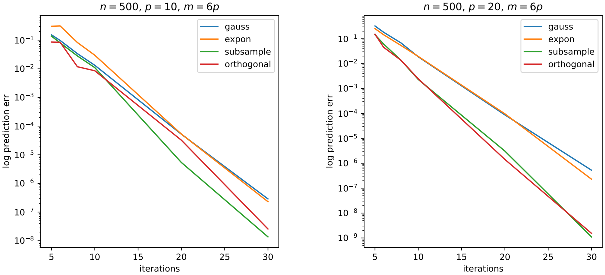

Figure : Using different sketch matrices, under Gaussian design. “expon” means Laplace design.

Simulations - model mis-specifications

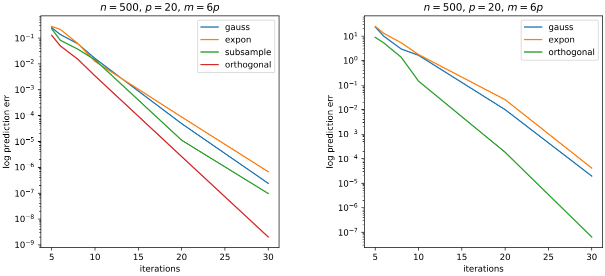

Figure : Left: exponential design; Right: Laplace design.

Simulations - model mis-specifications cont.

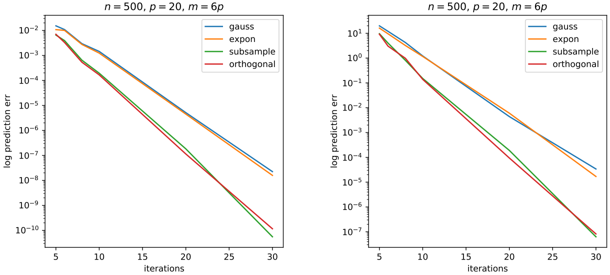

Figure : Left: $y = \sin(Ax) + w$; Right: $y = (Ax)^3 + w$. Both under Gaussian design. Large noise w behabves similarly.

Thank you!

Evaluation for Sketch Performance

Cost Approximation

In terms of f-cost, the approximated $\tilde{x}$ is said to be ϵ-optimal if

$$f(x^{LS}) \leq f(\tilde{x}) \leq (1+\epsilon)^2 f(x^{LS}).$$

Solution Approximation

In terms of f-cost, the prediction norm defined as

$$\|\tilde{x} - x^{LS}\|_A := \frac{1}{\sqrt{n}} \|A(\tilde{x} - x^{LS})\|_2.$$