Iterative Hessian Sketch for Constrained Least-Squares

Lydia Hsu, Shuaiwen Wang

December 9, 2018

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: "sketch" A using 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: "sketch" A using 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:

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

- The above serves as the building block for an iterative method

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;
- The above serves as the building block for an iterative method

Iterative Hessian Sketch

- Idea: construct a new least-squares problem for which the optimal solution is $x^{LS} - x^t$. Then applying HS to the problem will produce a new x^{t+1} whose distance to x^{LS} has been reduced by a factor of ρ.
- $\hat{u} := argmin_{u \in \{C - x^t\}} \{\frac{1}{2} \|Au\|_2^2 - \langle A^T(y - Ax^t), u \rangle\}$.
  By construction, $\hat{u} = x^{LS} - x^t$.
- Advantages: sample size efficiency and computational time saving

Iterative Hessian Sketch Algorithm

Algorithm

1. Initialize $x^0 = 0$.
2. For iterations $t = 0, 1, 2, ...N-1$, generate independent sketches $S^{t+1} \in \mathbb{R}^{mxn}$, and perform the updates $x^{t+1} = argmin_{x \in C}\{\frac{1}{2m}\|S^{t+1}A(x-x^t)\|_2^2 - \langle A^T(y-Ax^t), x\rangle\}$.
3. Return $\hat{x} = x^N$.

References

Sarlos (2006)
Title of the publication
Journal Name 12(3), 45 – 678.

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

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

Thank you!