Constrained Optimization via Exact Augmented Lagrangian and Randomized Iterative Sketching

Ilgee Hong (Presenter)

Department of Statistics
The University of Chicago

Sen Na

ICSI and Department of Statistics
University of California, Berkeley

Michael Mahoney

ICSI, Lawrence Berkeley National Laboratory and Department of Statistics
University of California, Berkeley

Mladen Kolar

Booth School of Business
The University of Chicago

Motivation

Problem

- **Equality-constrained** optimization

$$\min_{\boldsymbol{x}\in\mathbb{R}^n} f(\boldsymbol{x}) \qquad \text{s.t. } c(\boldsymbol{x}) = \mathbf{0},$$

- $f : \mathbb{R}^n \to \mathbb{R}$, objective function.
- $c : \mathbb{R}^n \to \mathbb{R}^m$, equality constraints.

- **Large scale** setting ($n+m$ is large).

Example

- Constrained Logistic Regression

$$\min_{\boldsymbol{x}\in\mathbb{R}^n} \frac{1}{N}\sum_{i=1}^{N} \log\left(1+\exp\left(-y_i\cdot \boldsymbol{d}_i^T\boldsymbol{x}\right)\right)$$
$$\text{s.t. } A\boldsymbol{x} = \boldsymbol{b},\ \|\boldsymbol{x}\|^2 = 1,$$

- PDE-constrained Problem

$$\min_{x,y} \frac{1}{2}\|x-u\|^2_{L^2(\Omega)} + \frac{\zeta}{2}\|y\|^2_{L^2(\Omega)}$$
$$\text{s.t. } -\Delta x = y \text{ in } \Omega,\ x = \mathbf{0} \text{ on } \partial\Omega,$$

Motivation

Classical Newton method for constrained optimization

- $\mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}) = f(\boldsymbol{x}) + \boldsymbol{\lambda}^T c(\boldsymbol{x})$ is the Lagrangian function.
- At each iteration k, we solve the Lagrangian Newton system to find a search direction.
  $$\Gamma_k \Delta \boldsymbol{z}_k = -\nabla \mathcal{L}_k,$$
  where $\Gamma_k \in \mathbb{R}^{(n+m)\times(n+m)}$ approximates the Lagrangian Hessian $\nabla^2 \mathcal{L}_k$.
- **Problem:** When $n + m$ is large, finding the exact solution $\Delta \boldsymbol{z}_k$ is impractical.

Method

AdaSketch-Newton (Randomized Newton method for constrained optimization)

- We use **exact augmented Lagrangian** as the merit function.
- We use **randomized iterative sketching** for the Lagrangian Newton system.
- We **adaptively control** the accuracy of randomized solver and penalty parameters of exact augmented Lagrangian.

Method

AdaSketch-Newton

- We use smooth exact augmented Lagrangian as the merit function

$$\mathcal{L}_{\boldsymbol{\eta}}(\boldsymbol{x}, \boldsymbol{\lambda}) = \underbrace{\mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda})}_{\text{Lagrangian}} + \overbrace{\frac{\eta_1}{2}}^{\text{penalty}} \underbrace{\|c(\boldsymbol{x})\|^2}_{\text{feasibility error}} + \overbrace{\frac{\eta_2}{2}}^{\text{penalty}} \underbrace{\|\nabla_{\boldsymbol{x}} \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda})\|^2}_{\text{optimality error}}.$$

- **Smoothness:** to overcome the Maratos effect.
- **Exactness:** solution of $\min_{\boldsymbol{x}, \boldsymbol{\lambda}} \mathcal{L}_{\boldsymbol{\eta}}(\boldsymbol{x}, \boldsymbol{\lambda})$ is also the solution of the original constrained problem provided that $\boldsymbol{\eta}$ are suitably specified.

Method

AdaSketch-Newton

- We update inexact Newton direction $\tilde{\Delta}\boldsymbol{z}_k$ by using **sketch-and-project framework**

$$\tilde{\Delta}\boldsymbol{z}_{k,j+1} = \arg\min_{\boldsymbol{u}} \|\boldsymbol{u} - \tilde{\Delta}\boldsymbol{z}_{k,j}\|^2, \quad \text{subject to} \quad \underbrace{S_{k,j}^T \Gamma_k}_{d\times(n+m)} \boldsymbol{u} = -\underbrace{S_{k,j}^T \nabla\mathcal{L}_k}_{d\times 1},$$

where $S_{k,j}$ is a copy of random sketching matrix $S \in \mathbb{R}^{(n+m)\times d} \sim \mathcal{P}$ with d being sketching dimension.

Method

AdaSketch-Newton

- We stop the sketching solver when the **adaptive accuracy condition** hold

$$\left\|\boldsymbol{r}_{k,j}\right\| \leq \theta_k \delta_k C \left\|\nabla \mathcal{L}_k\right\|.$$

- We check if $\tilde{\Delta}\boldsymbol{z}_{k,j}$ satisfies the **descent direction condition**

$$(\nabla \mathcal{L}_{\boldsymbol{\eta}_k}^k)^T \tilde{\Delta}\boldsymbol{z}_{k,j} \leq -\eta_{2,k} \left\|\nabla \mathcal{L}_k\right\|^2 /2.$$

- If $\tilde{\Delta}\boldsymbol{z}_{k,j}$ is a descent direction, we accept it as a search direction and do line search.

- If not, we update $(\eta_{1,k}, \eta_{2,k}, \delta_k)$ and go back to update $\tilde{\Delta}\boldsymbol{z}_{k,j}$.

Main Results

Theoretical Guarantee

Theorem 1 (Global convergence)**.** *Under mild assumptions, with probability one,* $\|\nabla \mathcal{L}_k\| \to 0$ *as* $k \to \infty$.

Theorem 2 (Local linear convergence)**.** *Let* $\boldsymbol{z}^\star$ *be a local solution and* $\theta_k = \theta \in (0, 1]$, $\forall k$. *Under mild assumptions and suppose* $\boldsymbol{z}_k \to \boldsymbol{z}^\star$*, for all sufficiently large* k*, we have* $\alpha_k = 1$ *and (noting that* $\theta\delta_K < 1$*)*

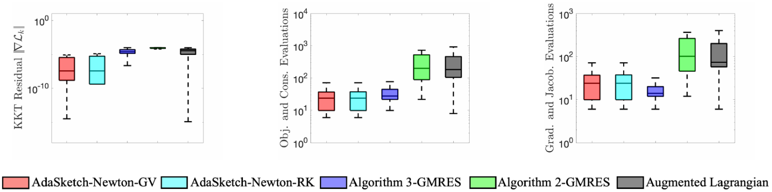
$$\|\boldsymbol{z}_{k+1} - \boldsymbol{z}^\star\| \leq (1 + \varphi)\theta\delta_K \|\boldsymbol{z}_k - \boldsymbol{z}^\star\|, \quad \text{for any } \varphi > 0.$$

Corollary 3 (Local superlinear convergence)**.** *Let* $\boldsymbol{z}^\star$ *be a local solution and* θ_k *be any input sequence such that* $\theta_k \to 0$ *as* $k \to \infty$. *Under mild assumptions and suppose* $\boldsymbol{z}_k \to \boldsymbol{z}^\star$*, for all sufficiently large* k*, we have* $\alpha_k = 1$ *and that*

$$\|\boldsymbol{z}_{k+1} - \boldsymbol{z}^\star\| \leq O(\theta_k + \tau_k) \|\boldsymbol{z}_k - \boldsymbol{z}^\star\| + O\left(\|\boldsymbol{z}_k - \boldsymbol{z}^\star\|^2\right).$$

Experiments

- **Problem:** CUTEst test set
- **Baseline:**
  - Algorithm 2-GMRES: Inexact Newton method with ℓ_1 **penalized merit function** and **GMRES**
  - Algorithm 3-GMRES: Adaptive modification of Algorithm 2
  - Augmented Lagrangian (Nocedal & Wright, 2006, Algorithm 17.3)



Thank you