Exploring Jacobian Inexactness in Second-Order Methods for Variational Inequalities

- Lower Bound
- Optimal Algorithm
- Quasi-Newton Approximation

MOHAMED BIN ZAYED UNIVERSITY OF ARTIFICIAL INTELLIGENCE

MIPT

NEURAL INFORMATION PROCESSING SYSTEMS

Spotlight

Artem Agafonov
Petr Ostroukhov
Roman Mozhaev
Konstantin Yakovlev
Eduard Gorbunov
Martin Takáč
Alexander Gasnikov
Dmitriy Kamzolov

Monotone MVI approximate solution

Find $\hat{x} \in \mathcal{X}$ such that

$$\mathrm{GAP}(\hat{x}) = \sup_{x \in \mathcal{X}} \langle F(x), \hat{x} - x \rangle \leq \varepsilon.$$

Order	Convergence	Lower bound
First	$O\left(\varepsilon^{-1}\right)$	$\Omega\left(\varepsilon^{-1}\right)$
Second	$O\left(\varepsilon^{-2/3}\right)$	$\Omega\left(\varepsilon^{-2/3}\right)$

Inexact Jacobian $\quad \|\nabla F(v) - J(v)\| \leq \delta$

The lower bound for methods with inexact Jacobians

$$\mathrm{GAP}(\hat{x}) = \Omega\left(\frac{L_1 D^3}{T^{3/2}} + \frac{\delta D^2}{T}\right)$$

The Model

Inexact Taylor approximation

$$\Psi_v(x) = F(v) + J(v)[x - v]$$

The model of objective

$$\Omega_v^\eta(x) = \Psi_v(x) + \boxed{\eta\delta(x - v)} + 5L_1\|x - v\|(x - v)$$

The Method

Algorithm 1 VIJI

Input: initial point $x_0 \in \mathcal{X}$, parameters L_1, η, sequence $\{\beta_k\}$, and opt $\in \{0, 1, 2\}$.
Initialization: set $s_0 = 0 \in \mathbb{R}^d$.
for $k = 0, 1, 2, \ldots, T$ **do**
  Compute $v_{k+1} = \mathrm{argmax}_{v \in \mathcal{X}}\{\langle s_k, v - x_0\rangle - \frac{1}{2}\|v - x_0\|^2\}$.
  Compute $x_{k+1} \in \mathcal{X}$ such that

$$\boxed{\sup_{x \in \mathcal{X}} \langle \Omega_{v_{k+1}}(x_{k+1}), x_{k+1} - x\rangle \leq \frac{L_1}{2}\|x_{k+1} - v_{k+1}\|^3 + \delta\|x_{k+1} - v_{k+1}\|^2.}$$

  Compute $\boxed{\lambda_{k+1} \text{ such that } \frac{1}{32} \leq \lambda_{k+1}\left(\frac{L_1}{2}\|x_{k+1} - v_{k+1}\| + \beta_{k+1}\right) \leq \frac{1}{22}.}$
  Compute $s_{k+1} = s_k - \lambda_{k+1} F(x_{k+1})$.

Output: $\hat{x} = \begin{cases} \tilde{x}_T = \frac{1}{\sum_{k=1}^T \lambda_k} \sum_{k=1}^T \lambda_k x_k, & \text{if opt} = 0, \\ x_T, & \text{else if opt} = 1, \\ x_{k_T} \text{ for } k_T = \mathrm{argmin}_{1 \leq k \leq T}\|x_k - v_k\|, & \text{else if opt} = 2. \end{cases}$

Insights

- *Subproblem is monotone VI.* For QN approximation solution is reduced to minimization problem.
- Converging to optimum, β_k starts to have a greater influence on the choice of λ_k, *preventing the method from taking overly aggressive steps*

Convergence

Monotone. L_1-smooth VIs

$$\mathrm{GAP}(\hat{x}) = O\left(\underbrace{\frac{L_1 D^3}{T^{3/2}}}_{\text{optimal second-order rate}} + \underbrace{\frac{\delta D^2}{T}}_{\text{optimal first-order rate}}\right)$$

Tensor generalization. Monotone L_i-smooth VIs

$$\mathrm{GAP}(\hat{x}) = O\left(\frac{L_{p-1} D^{p+1}}{T^{\frac{p+1}{2}}} + \sum_{i=1}^{p-1} \frac{\delta_i D^{i+1}}{T^{\frac{i+1}{2}}}\right)$$

Non-monotone generalization. Minty condition, L_1-smooth VIs

$$\mathrm{RES}(\hat{x}) = \sup_{x \in \mathcal{X}} \langle F(\hat{x}), \hat{x} - x\rangle = O\left(\frac{L_1 D^3}{T} + \frac{\delta D^2}{\sqrt{T}}\right)$$

Quasi-Newton approximation

Damped L-Broyden $\quad J^{i+1} = J^i + \boxed{\frac{1}{m+1}}\frac{(y_i - J^i s_i)s_i^\top}{s_i^\top s_i}$

- $s_i = z_{i+1} - z_i, \quad y_i = F(z_{i+1}) - F(z_i)$
- s_i are sampled, $y_i = \nabla F(x) s_i$

Experiments

$$\min_{x \in \mathbb{R}^d} \max_{y \in \mathbb{R}^d} f(x, y) = y^\top(Ax - b) + \frac{\rho}{6}\|x\|^3$$