AVERAGING RANDOM PROJECTION: A FAST ONLINE SOLUTION FOR LARGE-SCALE CONSTRAINED STOCHASTIC OPTIMIZATION

Jialin Liu¹, Yuantao Gu², and Mengdi Wang³

¹Department of Automation, ²Department of Electronic Engineering, Tsinghua University, Beijing, China
³Department of Operations Research and Financial Engineering, Princeton University, NJ, USA

Email: danny19921123@126.com, gyt@tsinghua.edu.cn, mengdiw@princeton.edu

Overview

- **Stochastic Optimization (SO) problem:** Finding an optimal solution for a stochastic convex function with intersected convex sets:
  \[
  \arg\min_x \{ F(x) = E[f(x; \xi)] \}
  \]
  s.t. \( x \in X = \bigcap_{i=1}^m X_i \)
  where \( F(x) \) is convex and \( X_i \) are convex sets.

- **Incremental Constraint Projection Method (ICPM):**
  - Choose an arbitrary initial point \( x_0 \in \mathbb{R}^n \).
  - Repeat as \( k = 0, 1, 2, ... \):
    1) Sample a random (sub)gradient \( \nabla f(x; \xi_k) \);
    2) Sample \( \alpha_i \) from the sets of constraints \( X_1, X_2, ..., X_m \);
    3) Compute next iterate \( x_{k+1} \):
      \[
      x_{k+1} = \Pi_{\alpha_k} \left[ x_k - \alpha_k \nabla f(x; \xi_k) \right]
      \]

Algorithm Description

- **Incremental Constraint Averaging Projection Method (ICAPM):**
  - Choose an arbitrary initial point \( x_0 \in \mathbb{R}^n \).
  - Repeat as \( k = 0, 1, 2, ... \):
    1) Sample a random (sub)gradient \( \nabla f(x; \xi_k) \);
    2) Sample \( M_k \) constraints \( \{ \alpha_{i_k} \}_{i_k} \) from the sets of constraints \( X_1, X_2, ..., X_m \);
    3) Compute next iterate \( x_{k+1} \):
      \[
      x_{k+1} = x_k - \alpha_k \nabla f(x; \xi_k)
      \]
      \[
      x_{k+1} = \frac{1}{M_k} \sum_{i_k} \Pi_{\alpha_{i_k}} Y_{i_k}
      \]

Motivation

- Non-controlled Randomness in ICPM:
  - Increasing variance as \( k \) increases.
  - Non-guaranteed convergence rate.
  - More real scenes require sampling multiple constraints:
    - Decentralized computing.
    - Online computing.

Convergence Analysis

- **Theorem 1 (Convergence of ICPM):** Under proper conditions [1], iterates generated by ICPM converge almost surely to a random optimal point.
- **Theorem 2 (Convergence Rate with Controlled Probability):** Assume the objective function is strongly convex, there exist constants \( C_i, C_j > 0 \):
  \[
  \Pr \left[ \| x_k - x^* \| \leq \frac{C_i}{k+1}, \forall 0 \leq k \leq T \right] \geq \prod_{i=1}^{T} \left( 1 - \frac{C_i}{M_k} \right)
  \]
  where \( x^* \) is the optimal point.

Simulation Results

- **Decentralized Network Problem:** A optimization problem is computed on \( m \) decentralized agents:
  \[
  \arg\min_{x_1, x_2, ..., x_m} \left\{ \sum_{i=1}^{m} f_i(x_i) \right\}, \text{s.t.} \ x_i = x_1 = ... = x_m,
  \]
  where \( f_i \) is the local objective function on the \( i \)-th agent (unknown to others), and \( x_i \) is the local variable.
- Use ICAPM to solve this problem, (sub)gradient descent should be computed locally, and communicate their local variable randomly (i.e. random projection). Its performance is largely depended on sample number \( M_k \), as the left figure.
- **Online Linear SVM:**
  \[
  \arg\min_{x, b} \left\{ \| \phi(x) \|_2^2 \right\}, \text{s.t.} \phi(x) \cdot x + b \geq 1, \forall i = 1, 2, ..., m,
  \]
  where \( x_i, y_i \) are data sampled or generated online.

Conclusion

- ICAPM is almost surely convergent under proper conditions (at most the same as ICPM).
- The convergence rate of ICAPM could be controlled by parameters (sample number) every step. Proper parameters could guarantee convergence rate as \( O(1/k) \) with high probability.

References