Technical Notes for Research Project

Summary

Viet-Anh Le[1,2]

[1]University of Delaware
[2]Cornell University

October 14, 2024

Distributed Optimization

- Consider an optimization problem involving N agents with separable objectives and coupling constraints:

$$\begin{aligned} \min_{\boldsymbol{x}_i \in \mathcal{X}_i^{\mathrm{MI}}} \quad & \sum_{i=1}^{N} f_i(\boldsymbol{x}_i), \\ \text{subject to} \quad & \sum_{i=1}^{N} \boldsymbol{A}_i \boldsymbol{x}_i = \boldsymbol{b}, \end{aligned} \tag{1}$$

where $\mathcal{X}_i^{\mathrm{MI}}$ is the mixed-integer-valued set for $\boldsymbol{x}_i$,

$$f_i(\boldsymbol{x}_i) = \boldsymbol{x}_i^\top \boldsymbol{Q}_i \boldsymbol{x}_i + \boldsymbol{q}_i^\top \boldsymbol{x}_i,$$

is the local objective function of each agent–i, $\boldsymbol{Q}_i$, $\boldsymbol{q}_i$, $\boldsymbol{A}_i$, and $\boldsymbol{b}$ are the matrices and vectors of coefficients. Let $\boldsymbol{x}^\top = [\boldsymbol{x}_1^\top, \ldots, \boldsymbol{x}_N^\top]$ be the concatenated vector of optimization variables, and let $\boldsymbol{A} = \left[\boldsymbol{A}_1, \ldots, \boldsymbol{A}_N\right]$.

Distributed Optimization

Overview

- We combine proximal ADMM with sequential convexification of the integrality constraints.
- At each iteration t, we keep a mixed-integer-valued vector $\boldsymbol{x}_i^{(t)} \in \mathcal{X}_i^{\mathrm{MI}}$ and an real-valued solution of the relaxed problem $\tilde{\boldsymbol{x}}_i^{(t)} \in \tilde{\mathcal{X}}_i$ where $\tilde{\mathcal{X}}_i$ is formed from $\mathcal{X}_i$ by relaxing the integrality constraints.

Augmented Lagrangian

$$\mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}) = \sum_{i=1}^{N} f_i(\boldsymbol{x}_i) + \boldsymbol{\lambda}^\top \left(\sum_{i=1}^{N} \boldsymbol{A}_i \boldsymbol{x}_i - \boldsymbol{b} \right) + \frac{\rho}{2} \left\| \sum_{i=1}^{N} \boldsymbol{A}_i \boldsymbol{x}_i - \boldsymbol{b} \right\|_2^2, \tag{2}$$

where $\boldsymbol{\lambda}$ are the dual variables (Lagrangian multipliers), and $\rho > 0$ is a positive constant.

Algorithm

- Agent–i solves the local problem (3).

$$\boldsymbol{x}_i^{(t+1)} = \operatorname*{arg\,min}_{\boldsymbol{x}_i \in \tilde{\mathcal{X}}_i} \mathcal{L}(\boldsymbol{x}_i, \tilde{\boldsymbol{x}}_{-i}^{(t)}, \boldsymbol{\lambda}^{(t)}) + \beta_i \left\| \boldsymbol{x}_i - \tilde{\boldsymbol{x}}_i^{(t)} \right\|_2^2, \tag{3}$$

where $\beta_i \in \mathbb{R}^+$ is a penalty weight.

- Update the dual variables $\boldsymbol{\lambda}$ by

$$\boldsymbol{\lambda}^{(t+1)} = \boldsymbol{\lambda}^{(t)} + \gamma\rho\left(\sum_{i=1}^{N} \boldsymbol{A}_i \boldsymbol{x}_i - \boldsymbol{b}\right). \tag{4}$$

- Compute $\tilde{\boldsymbol{x}}_i^{(t+1)}$ from $\boldsymbol{x}_i^{(t+1)}$ by rounding operator (transform a real-valued solution into an integer one). In other words,

$$\tilde{\boldsymbol{x}}_i^{(t+1)} = \operatorname*{arg\,min}_{\boldsymbol{x}_i \in \mathcal{X}_i^{\mathrm{MI}}} \left\| \boldsymbol{x}_i - \boldsymbol{x}_i^{(t+1)} \right\|_2^2, \tag{5}$$

Overview

- For nonconvex and nonsmooth optimization, to prove convergence, we need to (1) identify a so-called sufficiently decreasing Lyapunov function; and (2) establish the lower boundness property of the Lyapunov function[a].
- Let $(\boldsymbol{x}^*, \boldsymbol{\lambda}^*)$ be a saddle point that satisfies the KKT conditions of the relaxed problem (QP)

$$\begin{aligned} &\boldsymbol{A}_i^\top \boldsymbol{\lambda}^* \in \partial f_i(\boldsymbol{x}_i^*),\ \forall i = 1, \dots, N, \\ &\sum_{i=1}^{N} \boldsymbol{A}_i \boldsymbol{x}_i^* = \boldsymbol{b} \end{aligned} \tag{6}$$

  where $\partial f_i(\boldsymbol{x}_i)$ denotes subdifferential of f_i at $\boldsymbol{x}_i$.
- We consider the following Lyapunov function

$$\Phi^{(t)} = \left\| \boldsymbol{x}^{(t)} - \boldsymbol{x}^* \right\|_{\boldsymbol{P}}^2 + \frac{1}{\gamma\rho} \left\| \boldsymbol{\lambda}^{(t)} - \boldsymbol{\lambda}^* \right\|_2^2 + \eta \left\| \tilde{\boldsymbol{x}}^{(t)} - \boldsymbol{x}^{(t)} \right\|_2^2 \tag{7}$$

  where $\eta > 0$.

[a] Yang et al., "Proximal admm for nonconvex and nonsmooth optimization".

Lemma 1

For $t \geq 1$, we have

$$\left(\left\| \boldsymbol{x}^{(t)} - \boldsymbol{x}^* \right\|_{\boldsymbol{P}}^2 + \frac{1}{\gamma\rho} \left\| \boldsymbol{\lambda}^{(t)} - \boldsymbol{\lambda}^* \right\|_2^2 \right) - \left(\left\| \boldsymbol{x}^{(t+1)} - \boldsymbol{x}^* \right\|_{\boldsymbol{P}}^2 + \frac{1}{\gamma\rho} \left\| \boldsymbol{\lambda}^{(t+1)} - \boldsymbol{\lambda}^* \right\|_2^2 \right)$$
$$\geq \left\| \tilde{\boldsymbol{x}}^{(t)} - \boldsymbol{x}^{(t+1)} \right\|_{\boldsymbol{P}}^2 + \frac{2-\gamma}{\rho\gamma^2} \left\| \boldsymbol{\lambda}^{(t)} - \boldsymbol{\lambda}^{(t+1)} \right\|_2^2 + \frac{2}{\gamma} (\boldsymbol{\lambda}^{(t)} - \boldsymbol{\lambda}^{(t+1)})^\top \boldsymbol{A} (\tilde{\boldsymbol{x}}^{(t)} - \boldsymbol{x}^{(t+1)}). \tag{8}$$

Lemma 2

For $t \geq 1$, we have

$$\begin{aligned}\Phi^{(t)} - \Phi^{(t+1)} \geq \left\|\tilde{\boldsymbol{x}}^{(t)} - \boldsymbol{x}^{(t+1)}\right\|^2_{\boldsymbol{P}-\eta\mathbb{I}} + \frac{2-\gamma}{\rho\gamma^2}\left\|\boldsymbol{\lambda}^{(t)} - \boldsymbol{\lambda}^{(t+1)}\right\|^2_2 \\ + \frac{2}{\gamma}(\boldsymbol{\lambda}^{(t)} - \boldsymbol{\lambda}^{(t+1)})^\top \boldsymbol{A}(\tilde{\boldsymbol{x}}^{(t)} - \boldsymbol{x}^{(t+1)}) + \eta\left\|\tilde{\boldsymbol{x}}^{(t)} - \boldsymbol{x}^{(t)}\right\|^2_2\end{aligned} \tag{9}$$

As a result, if the matrix

$$\boldsymbol{R} = \begin{pmatrix} \boldsymbol{P} - \eta\mathbb{I} & \frac{1}{\rho}\boldsymbol{A}^\top \\ \frac{1}{\rho}\boldsymbol{A} & \frac{2-\gamma}{\rho\gamma^2}\mathbb{I}\end{pmatrix} \tag{10}$$

is positive definite, then $\{\Phi^{(t)}\}$ sufficiently decrease.

Theorem

If the parameters ρ, γ, and β are chosen such that the following conditions are satisfied

$$\begin{aligned} &\beta_i > \eta + \rho\left(1/\epsilon - 1\right) e_i, \\ &2 - \gamma > N\epsilon, \end{aligned} \tag{11}$$

where e_i is the maximum eigenvalue of $\boldsymbol{A}_i^\top \boldsymbol{A}_i$, $\epsilon > 0$ is a positive constant, then $\{\boldsymbol{x}^{(t)} - \tilde{\boldsymbol{x}}^{(t)}\}$ converge to 0, and the sequence $\{\boldsymbol{x}^{(t)}, \boldsymbol{\lambda}^{(t)}\}$ converges.