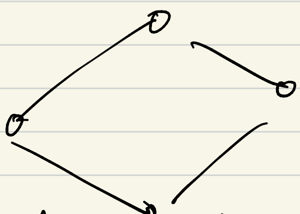
Lecture 5: Convex Algorithms

optimization

Given. $\boxed{\sum_{i=1}^{m} f_i(x)}$ ← 图?

Given a communication Graph. G

Q: How to formulate an optimization problem for decentralized optimization?



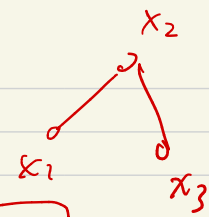
DGD Algorithm does not really address the above question. It starts directly from modifying GD to decentralized setting.

Objective 1: local variable x_i can be optimized using local data.

objective 2 In the end, we want x_i & x_j's to be in consensus.

while optimizing the entire objective.

node



Formulation 1

$$\min \sum_i f_i(x_i)$$

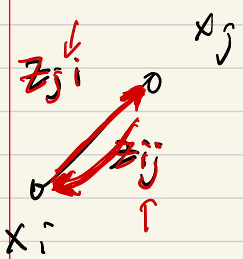
$$\text{s.t. } x_i = x_j, \ \forall (i \sim j)$$

✓ Formulation 2

$$\min \sum_{i=1}^{m} f_i(x_i)$$

$$x_i = z_{ij}$$

$$x_j = z_{ij} \qquad \forall (i \sim j)$$



They are both linearly constrained problems!

$$\min \sum_{i=1}^{n} f_i(x_i)$$

$$Ax + Bz = 0$$

node + edge.

$$\min \sum_{i=1}^{m} f_i(x_i) \quad s.t. \quad \sum_{i=1}^{m} A_i x_i = b$$

$$\boxed{x^{k+1} = \min_{x} \sum_{i=1}^{m} f_i(x_i) + \langle y^k, \sum_{i=1}^{m} A_i x_i - b \rangle}$$

$$y^{k+1} = y^k + \alpha \left(\sum_{i=1}^{m} A_i x_i - b \right)$$

Dual Ascent

m个 independent problem

$$x_i^{k+1} = \min_{x_i} f_i(x_i) + \langle y^k, A_i x_i \rangle \qquad i = 1 \cdots m$$

$$y^{k+1} = \cdots\cdots$$

问题

$$z^{k+1} = \operatorname{argmin} \; g(z) + \langle y^k, \cancel{Ax^{k+1}} + Bz - \cancel{c} \rangle + \frac{\rho}{2} \| Ax^{k+1} + Bz - c \|^2$$

$$\nabla g(z^{k+1}) + B^T y^k + \rho B^T (Bz^{k+1} + Ax^{k+1} - c)^2 = 0$$

$$\nabla g(z^{k+1}) + B^T \big(\underbrace{y^k + \rho(Bz^{k+1} + Ax^{k+1} - c)}_{y^{k+1}} \big) = 0$$

$$\boxed{\nabla g(z^{k+1}) + B^T y^{k+1} = 0}$$

Distributed ADMM

$$\underline{L} = \boxed{\sum f_i(x_i)} + \sum_{i=1}^{m} \left(\langle y_i, x_i - z \rangle + \frac{\rho}{2} \| x_i - z \|^2 \right)$$

$$x^{r+1} = \operatorname{argmin} \sum_i f_i(x_i) + \sum_i \left(\langle y_i^r, x_i - z^r \rangle + \frac{\rho}{2} \| x_i - z^r \|^2 \right)$$

分成 m 个子问题

$$= \operatorname{argmin} \sum_i \left(f_i(x_i) + \frac{\rho}{2} \left\| x_i - z^r + \frac{1}{\rho} y_i^r \right\|^2 \right)$$

$$z^{r+1} = \operatorname{argmin} \sum_{i=1}^{m} \left(\langle y_i^r, x_i^{r+1} - z \rangle + \frac{\rho}{2} \| x_i^{r+1} - z \|^2 \right)$$

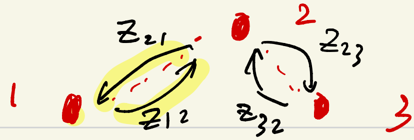
$$= \operatorname{argmin} \sum_{i=1}^{m} \frac{\rho}{2} \left\| x_i^{r+1} + \frac{1}{\rho} y_i^r - z \right\|^2 .$$

$$y_i^{r+1} = y_i^r + \rho \left(x_i^{r+1} - z^{r+1} \right) \quad ☆ \quad i = 1 \cdots m$$

Define $u_i^r = y_i^r / \rho$ Then

$$\begin{cases} x^{r+1} = \operatorname{argmin} \sum_{i=1}^{m} f_i(x_i) + \frac{\rho}{2} \| x_i - z^r + u_i^r \|^2 . \\ z^{r+1} = \operatorname{argmin} \frac{\rho}{2} \sum_{i=1}^{m} \| z - x_i^{r+1} - u_i^r \|^2 = \frac{1}{m} \sum_{i=1}^{m} (x_i^{r+1} + u_i^r) \\ u_i^{r+1} = u_i^r + x_i^{r+1} - z^{r+1} \end{cases}$$

Example



✓ $x_1 = z_{12}$ ✓ $x_2 = z_{12}$ ✓ $x_3 = z_{32}$ $x_2 = z_{32}$

✓ $x_2 = z_{21}$ $x_1 = z_{21}$ ✓ $x_2 = z_{23}$ $x_3 = z_{23}$

$$A_1 = \mathbb{R}^{2E \times m}$$

$$A = \begin{bmatrix} A_1 \\ A_2 \end{bmatrix}^{2E \times m}_{2E \times m}$$

$$B = \begin{bmatrix} -I_{2E} \\ -I_{2E} \end{bmatrix}$$

If $e_{ij} \in E$, and z_{ij} is qth edge.
Then (q, i)th entry of A_1 $= 1$
(q, j)th entry of A_2

otherwise $= 0$.

$$Ax + B = 0$$

$\boxed{A_1}$

$$4\;\begin{array}{c|c} & \overbrace{\begin{matrix}1 & 2 & 3\end{matrix}}^{3} \\ \begin{matrix}12\\21\\23\\32\end{matrix} & \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \end{array} \times \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} z_{12} \\ z_{21} \\ z_{23} \\ z_{32} \end{bmatrix}$$

$\boxed{A_2}$

$$\begin{array}{cc} \begin{matrix}12\\21\\23\\32\end{matrix} & \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix} \end{array} \times \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} z_{12} \\ z_{21} \\ z_{23} \\ z_{32} \end{bmatrix}$$

incidence matrix

$$\boxed{A_1 - A_2 = \text{incidence matrix}}$$

DGD

$$x^{r+1} = W x^{r} - \alpha \nabla f(x^{r}) \qquad \nabla f(x^r) = \begin{bmatrix} \nabla_1 f(x_1^r) \\ \vdots \\ \nabla_m f(x_m^r) \end{bmatrix}$$

$$x_i^{r+1} = \sum_j w_{ij} x_i^{r} - \alpha \nabla_i f_i(x_i^{r})$$

$x^\infty = x_i^\infty = x_j^\infty$

$$x^\infty = x^\infty - \alpha \nabla f_i(x^\infty)$$

$$\Rightarrow \nabla f_i(x^\infty) = 0 \quad \text{impossible for all } i$$

$$x_i^{r+1} = \sum_j w_{ij} x_i^{r} - \alpha \frac{1}{m}\sum_{i=1}^{m} \nabla f_i(x_i^{r})$$

$$x^\infty = x^\infty - \alpha \frac{1}{m}\underbrace{\sum_{i=1}^{m} \nabla f_i(x^\infty)}_{=0}$$

$$x_i^{r+1} = \sum_j w_{ij} x_i^{r} - \alpha g_i^{r}$$

$$g_i^{r+1} = \sum_j w_{ij} g_j^{r} + \nabla f_i(x_i^{r+1}) - \nabla f_i(x_i^{r})$$

Relationship Between Primal Dual / EXTRA.

$$\begin{cases} x^{r+1} = \arg\min \langle \nabla f(x^r), x - x^r\rangle + \langle Ax, \mu^r\rangle - \beta x^{r^T} L_+ x + \beta x^T D x \\ \mu^{r+1} = \mu^r + \beta A x^{r+1} \end{cases}$$

$\equiv$ $\boxed{\text{EXTRA}}$

$$\begin{cases} \nabla f(x^r) + A^T\mu^r - \beta L_+ x^r + 2\beta D x^{r+1} = 0 \\ \nabla f(x^{r-1}) + A^T\mu^{r-1} - \beta L_+ x^{r-1} + 2\beta D x^r = 0 \end{cases}$$

$$0 = \nabla f(x^r) - \nabla f(x^{r-1}) + A^T(\mu^r - \mu^{r-1}) - \beta L_+ (x^r - x^{r-1}) + 2\beta D (x^{r+1} - x^r)$$

$$0 = \nabla f(x^r) - \nabla f(x^{r-1}) + \beta L_- x^r - \beta L_+ (x^r - x^{r-1}) + 2\beta D (x^{r+1} - x^r)$$

$$x^{r+1} = x^r - \frac{1}{2\beta} D^{-1} (\nabla f(x^r) - \nabla f(x^{r-1})) - \frac{1}{2} D^{-1} L_- x^r + \frac{1}{2} D^{-1} L_+ (x^r - x^{r-1})$$

$$= x^r - \frac{1}{2\beta} D^{-1} (\nabla f(x^r) - \nabla f(x^{r-1})) - \frac{1}{2} D^{-1} (L_- - L_+) x^r - \frac{1}{2} D^{-1} L_+ x^{r-1}$$

$$= x^r - \frac{1}{2\beta} D^{-1} (\nabla f^r - \nabla f^{r-1}) + W x^r - \frac{1}{2}(I + W) x^{r-1}$$

$$W = \frac{1}{2} D^{-1} (L^+ - L^-)$$

$$I + W = \frac{1}{2} D^{-1} \underbrace{(L^+ + L^-)}_{2D} + \frac{1}{2} D^{-1} (L^+ - L^-) = D^{-1} L^+$$

$$\underbrace{2D}_{I}$$