Distributed Optimization Problem

"Decentralized" Optimization Problem

Agent i knows only its own private objective fn

$\{f_i\text{'s}\}$ are private

$$\min_{\{x_i \in \mathbb{R}^n\}} \sum_{i=1}^{N} f_i(x_i)$$

$$\text{s.t.} \quad x_1 = x_2 = \dots = x_N$$

→ Decentralized / Distributed optimization problem

Equivalent

$$\min_{x \in \mathbb{R}^n} \sum_{i=1}^{N} f_i(x)$$

→ Centralized optimization problem

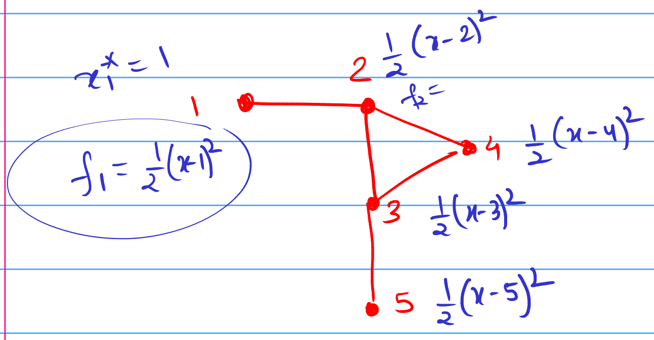
ex: $f_i(x) = \frac{1}{2}(x-i)^2$

Minimize this

$$\sum_{i=1}^{5} f_i(x) = \sum_{i=1}^{5} \frac{1}{2}(x-i)^2 \rightarrow \sum_{i=1}^{5} \nabla f_i(x) = 0$$

$$\boxed{x^* = 3}$$

$$\sum_{i=1}^{5} (x^* - i) = 0$$



Our goal is to develop algorithms for solving DOP.

→ Information exchange b/w neighbors to facilitate consensus on the global optimizer.

What can agents exchange?

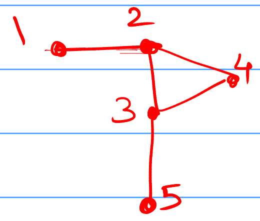
Agent i can share its current estimate

$x_i(k) \rightarrow n$

$\nabla f_i(x_i(k)) \rightarrow n$

$\Big\} \; \mathcal{O}(n)$

* Distributed Gradient Descent (DGD)



$$w_{ij} = \begin{cases} \dfrac{1}{1+\max\{d_i, d_j\}} & \text{if } (i,j)\in\mathcal{E},\ i\neq j \\ 1 - \sum_{j\neq i} w_{ij} & \text{if } i=j \\ 0 & \text{else} \end{cases}$$

Centralized update (Suppose every agent knowns all f_is)

$$x(k+1) = x(k) - \eta_k \nabla F(x(k))$$

$$F = \sum_{i=1}^{n} f_i$$

Distributed setup:

At $k+1$ iteration

(Mixing step/ Consensus step)

$$\bullet\ x_i(k+\tfrac{1}{2}) = \sum_{j=1}^{N} w_{ij}\, x_j(k)$$

Gradient descent step

$$\bullet\ x_i(k+1) = x_i(k+\tfrac{1}{2}) - \eta_k \nabla f_i(x_i(k+\tfrac{1}{2}))$$

Alternative

$$\bullet\ x_i(k+\tfrac{1}{2}) = x_i(k) - \eta \nabla f_i(x_i(k))$$

$$\bullet\ x_i(k+1) = \sum_{j=1}^{N} w_{ij}\, x_j(k+\tfrac{1}{2})$$

$$\bar{x}(k) = \frac{1}{N}\sum_{i=1}^{N} x_i(k)$$

$$\bar{x}(k+1) = \frac{1}{N}\sum_{i=1}^{N} x_i(k+1)$$

$$x_i(k+1) = \sum_{j=1}^{N} w_{ij}\, x_j(k) - \eta_k \nabla f_i(x_i(k+\tfrac{1}{2}))$$

$$\Rightarrow \bar{x}(k+1) = \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N} w_{ij}\, x_j(k) - \frac{\eta_k}{N}\sum_{i=1}^{N}\nabla f_i(x_i(k+\tfrac{1}{2}))$$

$$= \frac{1}{N}\sum_{j=1}^{N}\Big(\underbrace{\sum_{i=1}^{N} w_{ij}}_{=1}\Big)x_j(k) - \frac{\eta_k}{N}\sum_{i=1}^{N}\nabla f_i(x_i(k+\tfrac{1}{2}))$$

$$\boxed{\bar{x}(k+1) = \bar{x}(k) - \frac{\eta_k}{N}\sum_{i=1}^{N}\nabla f_i(x_i(k+1/2))}$$

Average dynamics

$$\bar{x}(k+1) = \bar{x}(k) - \frac{\eta_k}{N}\nabla F(\bar{x}(k)) + e_k$$

where $e_k = \frac{\eta_k}{N}\nabla F(\bar{x}(k)) - \frac{\eta_k}{N}\sum_{i=1}^{N}\nabla f_i(x_i(k+1/2))$

$$= \frac{\eta_k}{N}\sum_{i=1}^{N}\left(\nabla f_i(\bar{x}(k)) - \nabla f_i(x_i(k+1/2))\right)$$

- So if we can guarantee that error e_k is small, then the average dynamics follows the centralized gradient descent.

$$\eta_k = \frac{1}{k} \quad \left\} \begin{array}{l} \sum_{k=1}^{\infty}\eta_k = \infty \\ \sum_{k=1}^{\infty}\eta_k^2 < \infty \end{array}\right.$$

- EXTRA: shows convergence with fixed learning rate.

Distributed Aggregated Gradient Descent.

(2nd-order algorithm)

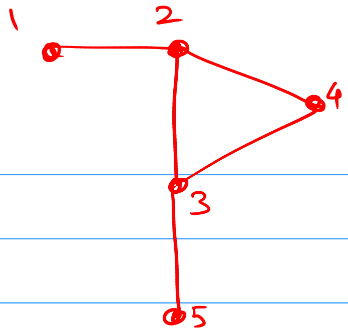
- Initialization step

$\{x_i(0)\}$ $\quad \{y_i(0) := \nabla f_i(x_i(0))\}$

- $x_i(k+1) = \sum_{j=1}^{N} w_{ij}\, x_j(k) - \alpha\, y_i(k)$

- $y_i(k+1) = \sum_{j=1}^{N} w_{ij}\, y_j(k) + \nabla f_i(x_i(k+1)) - \nabla f_i(x_i(k))$

Ex:



$$f_i(x) := \frac{1}{2}(x-i)^2$$

$$\nabla f_i(x) = (x-i)$$

$$\sum_{i=1}^{N=5} f_i \rightarrow \boxed{x^* = 3}$$

Continuous-time distributed optimization algorithm with FxTS guarantees

- N agents in the network
- Each agent has its own private objective function.
- Assume that agents can exchange information about $\{x_i\}$, $\{\nabla f_i(x_i)\}$

$$\min_{\{x_i \in \mathbb{R}^n\}} \sum_{i=1}^{N} f_i(x_i)$$

$$\text{s.t.} \quad x_1 = x_2 = \cdots = x_N$$

- $F(x) = \sum_{i=1}^{N} f_i(x)$ is μ-sc.

Centralized protocol:

* Initialization

$$x_1(0) = x_2(0) = \cdots = x_N(0)$$

Fixed-time consensus on x_i

* Agents have access to global information

$$\left(\sum_{i=1}^{N} \nabla f_i(x_i)\right)$$

We should develop a fixed-time parameter estimation scheme

* $\dot{x}_i = g^*$

$$g^* = -\sum_{i=1}^{N} \nabla f_i(x_i) - \text{sgn}^{\nu_1}\left(\sum_{i=1}^{N} \nabla f_i(x_i)\right) - \text{sgn}^{\nu_2}\left(\sum_{i=1}^{N} \nabla f_i(x_i)\right)$$

$$\dot{x} = -\sum_{i=1}^{N} \nabla f_i(x) - \mathrm{sgn}^{\nu_1}\left(\sum_{i=1}^{N} \nabla f_i(x)\right) - \mathrm{sgn}^{\nu_2}\left(\sum_{i=1}^{N} \nabla f_i(x)\right)$$

Claim: Centralized protocol converges in a fixed-time.

Proof:

$$V = \frac{1}{2}\left(\sum_{i=1}^{N} \nabla f_i(x)\right)^T \left(\sum_{i=1}^{N} \nabla f_i(x)\right)$$

$$\dot{V} = \sum_{i=1}^{N} \nabla f_i^T \nabla^2 f_i \, \dot{x}$$

$$\dot{V} \leq -c_1 V^{\alpha_1} - c_2 V^{\alpha_2}$$ $\Rightarrow$ Fixed-time convergence

- Let θ_i be the agent i's estimate of $\sum_{i=1}^{N} \nabla f_i(x_i)$

$$\theta_c = \frac{1}{N}\sum_{i=1}^{N} \nabla f_i(x_i)$$

$$\tilde{\theta}_i = \theta_i - \theta_c$$

$$\dot{\tilde{\theta}}_i = \dot{\theta}_i - \dot{\theta}_c$$

$$\boxed{\dot{\theta}_i = \omega_i + h_i}$$ $h_i = \frac{d}{dt}\nabla f_i(x_i)$

$$\omega_i = \sum_{j \in N_i}\left[p\,\mathrm{sign}(\theta_j - \theta_i) + \gamma\,\mathrm{sgn}^{\nu_1}(\theta_j - \theta_i) + \delta\,\mathrm{sgn}^{\nu_2}(\theta_j - \theta_i)\right]$$

$$g_i = -\left(N\theta_i + \mathrm{sgn}^{\nu_1}(N\theta_i) + \mathrm{sgn}^{\nu_2}(N\theta_i)\right)$$

$$\boxed{\dot{x}_i = g_i + \tilde{u}_i}$$

$$\tilde{u}_i = -\mathrm{sgn}^{\nu_1}(x_i - x_j) - \mathrm{sgn}^{\nu_2}(x_i - x_j)$$

Theorem: $\theta_i \to \theta_c$ if $\|h_i(t) - h_j(t)\| \leq \rho \quad \forall\, t \geq 0$

where $p > \left(\frac{N-1}{2}\right)\rho$.

—x—x—x—x—x—x—