Distributed Optimization Methods

Goran Banjac

Large-Scale Convex Optimization
ETH Zurich

May 5, 2020

Global variable consensus optimization

- consider the following convex optimization problem:

$$\text{minimize} \quad \sum_{i=1}^{N} f_i(x)$$

privacy issues, sometimes want to distributed system, also sometimes are just we have n machines

  where f_i are proper closed convex functions

- we want to solve the problem in a distributed way where each term can be handled by its own processing element (subsystem)
- the problem can be reformulated as

you don't have to share same opinions....

$$\begin{array}{ll} \text{minimize} & \sum_{i=1}^{N} f_i(x_i) \\ \text{subject to} & x_i = z, \quad i = 1, \ldots, N \end{array}$$

- this is called the *global consensus problem*

ADMM for consensus problems

- ADMM for the global consensus problem can be derived directly from its augmented Lagrangian

$$
\begin{aligned}
\mathcal{L}_\rho(x_1,\ldots,x_N,z,y_1,\ldots,y_N) &= \textstyle\sum_{i=1}^N \big(f_i(x_i) + y_i^T(x_i - z) + \frac{\rho}{2}\|x_i - z\|_2^2\big) \\
&= \textstyle\sum_{i=1}^N \Big(f_i(x_i) + \frac{\rho}{2}\|x_i - z + \frac{1}{\rho}y_i\|_2^2 - \frac{1}{2\rho}\|y_i\|_2^2\Big)
\end{aligned}
$$

- the resulting ADMM is

$$
\begin{aligned}
x_i^{k+1} &= \operatorname*{argmin}_{x_i} \Big\{ f_i(x_i) + \tfrac{\rho}{2}\|x_i - z^k + \tfrac{1}{\rho}y_i^k\|_2^2 \Big\} \\
\Aboxed{z^{k+1} &= \tfrac{1}{N}\textstyle\sum_{i=1}^N (x_i^{k+1} + \tfrac{1}{\rho}y_i^k)} \\
y_i^{k+1} &= y_i^k + \rho(x_i^{k+1} - z^{k+1})
\end{aligned}
$$

- observe from the z- and y_i-updates that $\sum_{i=1}^N y_i^{k+1} = 0$
- thus, the z-update reduces to

$$
z^{k+1} = \tfrac{1}{N}\textstyle\sum_{i=1}^N x_i^{k+1}
$$

Parallel projections

intersection of sets.

- consider the feasibility problem involving N closed convex sets $\mathcal{C}_i$
- the problem can be formulated as

$$\text{minimize} \quad \sum_{i=1}^{N} \mathcal{I}_{\mathcal{C}_i}(x)$$

- consensus ADMM then reduces to

$$\begin{aligned} x_i^{k+1} &= \Pi_{\mathcal{C}_i}(z^k - \tfrac{1}{\rho} y_i^k) \\ \boxed{z^{k+1} = \tfrac{1}{N} \sum_{i=1}^{N} x_i^{k+1}} \\ y_i^{k+1} &= y_i^k + x_i^{k+1} - z^{k+1} \end{aligned}$$

Global variable consensus with regularization

- consider the regularized global consensus problem:

$$\text{minimize} \quad \sum_{i=1}^N f_i(x) + g(x)$$

where f_i and g are proper closed convex

generalization where we also have g(z)

- the problem can be reformulated as

$$\begin{array}{ll} \text{minimize} & \sum_{i=1}^N f_i(x_i) + g(z) \\ \text{subject to} & x_i = z, \quad i = 1, \ldots, N \end{array}$$

in the previous case, g(z)=0, and the solution is exactly the same

- the ADMM then takes the following form

$$x_i^{k+1} = \underset{x_i}{\operatorname{argmin}} \left\{ f_i(x_i) + \frac{\rho}{2}\|x_i - z^k + \frac{1}{\rho} y_i^k\|_2^2 \right\}$$

$$z^{k+1} = \underset{z}{\operatorname{argmin}} \left\{ g(z) + \frac{N\rho}{2}\|z - \frac{1}{N}\sum_{i=1}^N (x_i^{k+1} + \frac{1}{\rho} y_i^k)\|_2^2 \right\}$$

$$y_i^{k+1} = y_i^k + \rho(x_i^{k+1} - z^{k+1})$$

- when $g \neq 0$, we do not necessarily have $\sum_{i=1}^N y_i^{k+1} = 0$

Distributed model fitting

- a general convex model fitting problem can be written as

$$\text{minimize} \quad l(Ax - b) + r(x)$$

  with parameters $x \in \mathbb{R}^n$, feature matrix $A \in \mathbb{R}^{m \times n}$, and output vector $b \in \mathbb{R}^m$

- the loss function $l : \mathbb{R}^m \to \mathbb{R}$ is often assumed to be additive, *i.e.*,

$$l(Ax - b) = \sum_{i=1}^{m} l_i(a_i^T x - b_i)$$

- the regularization term $r : \mathbb{R}^n \to \overline{\mathbb{R}}$ is often assumed to be separable
  - Tikhonov regularization: $r(x) = \lambda \|x\|_2^2$
  - lasso penalty: $r(x) = \lambda \|x\|_1$
- such problems can be solved efficiently in a distributed fashion

Sharing

- the *sharing problem* has the following form

$$\text{minimize} \quad \sum_{i=1}^N f_i(x_i) + g(\sum_{i=1}^N x_i)$$

- the problem can be reformulated as

$$\begin{array}{ll} \text{minimize} & \sum_{i=1}^N f_i(x_i) + g(\sum_{i=1}^N z_i) \\ \text{subject to} & x_i = z_i, \quad i = 1, \ldots, N \end{array}$$

no global z here, here z is just the compact way to write. for x,it is separable, but for z it is not separable and has higher dimension.

- applying ADMM to the problem above, we obtain

$$\begin{aligned} x_i^{k+1} &= \underset{x}{\operatorname{argmin}} \left\{ f_i(x_i) + \frac{\rho}{2} \| \boxed{x_i - z_i^k +} \frac{1}{\rho} y_i^k \|_2^2 \right\} \\ \boxed{z^{k+1}} &= \underset{z}{\operatorname{argmin}} \left\{ g(\sum_{i=1}^N z_i) + \frac{\rho}{2} \sum_{i=1}^N \| z_i - x_i^{k+1} - \frac{1}{\rho} y_i^k \|_2^2 \right\} \\ y_i^{k+1} &= y_i^k + \rho(x_i^{k+1} - z_i^{k+1}) \end{aligned}$$

where $z = (z_1, \ldots, z_N)$

ADMM for sharing problems

- after some simplifications, the method reduces to

$$
\begin{aligned}
x_i^{k+1} &= \operatorname*{argmin}_{x_i} \left\{ f_i(x_i) + \tfrac{\rho}{2}\|x_i - x_i^k + \bar{x}^k - \bar{z}^k + \tfrac{1}{\rho}y^k\|_2^2 \right\} \\
\bar{x}^{k+1} &= \textstyle\sum_{i=1}^N x_i^{k+1} \\
\bar{z}^{k+1} &= \operatorname*{argmin}_{\bar{z}} \left\{ g(N\bar{z}) + \tfrac{N\rho}{2}\|\bar{z} - \bar{x}^{k+1} - \tfrac{1}{\rho}y^k\|_2^2 \right\} \\
y^{k+1} &= y^k + \rho(\bar{x}^{k+1} - \bar{z}^{k+1})
\end{aligned}
$$

- the x-update can be carried out in parallel
- the $\bar{z}$-update requires gathering x_i^{k+1} to form the averages
- after the y-update, the new value of $\bar{x}^{k+1} - \bar{z}^{k+1} + \frac{1}{\rho}y^{k+1}$ is scattered to the subsystems

Optimal exchange

- the *exchange* or *resource allocation problem* is given by

$$\begin{array}{ll} \text{minimize} & \sum_{i=1}^{N} f_i(x_i) \\ \text{subject to} & \sum_{i=1}^{N} x_i = 0 \end{array}$$

- it can be seen as a special case of the sharing problem with $g = \mathcal{I}_{\{0\}}$
- ADMM then reduces to

$$\begin{aligned} x_i^{k+1} &= \operatorname*{argmin}_{x_i} \left\{ f_i(x_i) + \frac{\rho}{2}\|x_i - x_i^k + \bar{x}^k + \frac{1}{\rho} y^k\|_2^2 \right\} \\ \bar{x}^{k+1} &= \sum_{i=1}^{N} x_i^{k+1} \\ y^{k+1} &= y^k + \rho \bar{x}^{k+1} \end{aligned}$$

Duality between consensus and sharing

- consider again the sharing problem

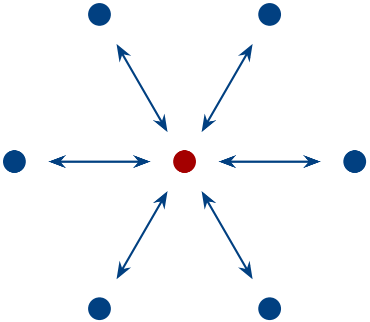
$$\text{minimize} \quad \sum_{i=1}^N f_i(x_i) + g(\sum_{i=1}^N x_i)$$

- its dual has the form

$$\text{minimize} \quad \sum_{i=1}^N f_i^*(-y) + g^*(y)$$

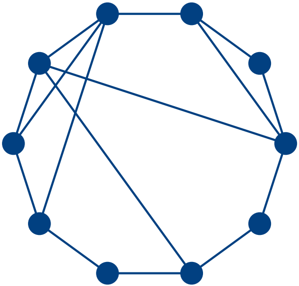
- thus, there is a close dual relationship between consensus and sharing problems
- the sharing problem can be solved by running ADMM on its dual consensus problem, and *vice versa*

Gather & scatter



- the communication structure of the presented algorithms is often referred to as *gather & scatter*
  - the subsystems perform computations in parallel
  - the central collector *gathers* local variables and updates is own variable
  - the updated information is then *scattered* back to the subsystems
- the central collector can evaluate termination criteria

Optimization over graphs



- let $\mathcal{G} = (\mathcal{N}, \mathcal{E})$ be an undirected connected graph
  - $\mathcal{N} = \{1, \ldots, N\}$ is the set of nodes
  - $\mathcal{E} \subseteq \mathcal{N} \times \mathcal{N}$ is the set of edges
- can we solve the consensus problem over $\mathcal{G}$?

$$\text{minimize} \quad \textstyle\sum_{i \in \mathcal{N}} f_i(x)$$

Decentralized ADMM

- we can reformulate the graph optimization problem as

$$\begin{array}{ll} \text{minimize} & \sum_{i\in\mathcal{N}} f_i(x_i) \\ \text{subject to} & x_i = t_{ij}, \quad (i,j)\in\mathcal{E} \\ & x_j = t_{ij}, \quad (i,j)\in\mathcal{E} \end{array}$$

- applying ADMM to the problem formulation above, we obtain

$$x_i^{k+1} = \underset{x_i}{\operatorname{argmin}} \left\{ f_i(x_i) + x_i^T p_i^k + \rho \sum_{j\in\mathcal{N}_i} \|x_i - \tfrac{x_i^k + x_j^k}{2}\|_2^2 \right\}$$

$$p_i^{k+1} = p_i^k + \rho \sum_{j\in\mathcal{N}_i} (x_i^{k+1} - x_j^{k+1})$$

  where $\mathcal{N}_i$ denotes the set of neighbors of node i

- the communication structure is often called the *network gossiping*
- evaluating a termination criterion becomes tricky since no subsystem contains information from all subsystems

because there is no central guy
who knows everything...

Decentralized ADMM for exchange problems

- consider the exchange problem

$$\begin{array}{ll} \text{minimize} & \sum_{i \in \mathcal{N}} f_i(x_i) \\ \text{subject to} & \sum_{i \in \mathcal{N}} x_i = 0 \end{array}$$

- we can apply decentralized ADMM to its dual

$$\text{minimize} \quad \sum_{i \in \mathcal{N}} f_i^*(-y)$$

using Moreau Identity to solve for the proximal operator.

- the resulting algorithm takes the form

show it!

$$\begin{aligned} r_i^{k+1} &= \rho \sum_{j \in \mathcal{N}_i} (y_i^k + y_j^k) - p_i^k \\ \boxed{x_i^{k+1}} &= \operatorname*{argmin}_{x_i} \left\{ f_i(x_i) + \tfrac{1}{4\rho d_i} \|x_i + r_i^{k+1}\|_2^2 \right\} \\ y_i^{k+1} &= \tfrac{1}{2\rho d_i}(x_i^{k+1} + r_i^{k+1}) \\ p_i^{k+1} &= p_i^k + \rho \sum_{j \in \mathcal{N}_i} (y_i^{k+1} - y_j^{k+1}) \end{aligned}$$

where d_i is the number of neighbors of node i (degree of $\mathcal{N}_i$)

Extensions

- decentralized ADMM can be extended to solving a more general exchange problem

$$\begin{array}{ll} \text{minimize} & \sum_{i\in\mathcal{N}} f_i(x_i) \\ \text{subject to} & \sum_{i\in\mathcal{N}}(A_i x_i - b_i) \in \mathcal{K} \end{array}$$

  where $\mathcal{K}$ is a convex cone
- there exist decentralized methods that are not based on ADMM
- some methods do not require that the graph is undirected