Talk 9: Operator Splitting and Federated Learning

WEN Hao

2021-10-28

The Optimization Problem

Let f_j be finite convex, with Lipschitz gradient.

$$\min \quad F(x) := \sum_{j=1}^{m} f_j(x_j)$$

$$\text{s.t.} \quad x_1 = \cdots = x_m \in \mathbb{R}^d, \quad x = (x_1, \cdots, x_m)$$

Main issues of existing FL algorithms (FedSGD, FedProx, etc)

- Convergence
- **Correctness**: fail to preserve the fixed points of the original optimization problem i.e. fixed points produced by the algorithm need not be stationary.

[1] R. Pathak and M. J. Wainwright, "FedSplit: An Algorithmic Framework for Fast Federated Optimization," in *Advances in Neural Information Processing Systems* (H. Larochelle, M. Ranzato, R. Hadsell, M. Balcan, and H. Lin,

More on the Issue of Correctness (FedGD)

Proposition

The sequence $\{x^{(t)}\}_{\infty}^{t=1}$ *generated by* $\texttt{FedGD}(s,e)$ *satisfy*

- *if* $x^{(t)}$ *convergent, then* $x_j^{(t)}$ *share a common limit* x^*
- x^* *satisfy the fixed point relation* $\sum_{i=1}^{e}\sum_{j=1}^{m}\nabla f_j(G_j^{i-1}(x^*)) = 0$

Notations (FedGD)

- $G_j(x_j) := x_j - s\nabla f_j(x_j)$ the gradient mappings
- $G_j^e(x_j) := \underbrace{G_j \circ \cdots \circ G_j}_{e-\text{times}}(x_j)$
- $x_j^{(t+1/2)} := G_j^e(x_j^{(t)}),\ x_j^{(t+1)} = \overline{x}^{(t+1/2)} := \frac{1}{m}\sum_{j=1}^{m} x_j^{(t+1/2)}.$

More on the Issue of Correctness (FedGD)

Note the abuse of the notation x!

Sketch:

Assume $x^{(t)} = (x_1^{(t)}, \cdots, x_m^{(t)}) \to (x_1^*, \cdots, x_m^*)$, then

$$(x_1^*, \cdots) = \mathtt{FedGD}(s, e)(x_1^*, \cdots) = \left(\frac{1}{m} \sum_{j=1}^{m} G_j^e(x_j^*), \cdots \right)$$

Hence $x_1^* = \cdots = x_m^* = x^*$. Write $\frac{1}{m} \sum_{j=1}^m G_j^e(x^*) = x^*$, and substitute G_j^e by its definition, one has

$$\sum_{i=1}^{e} \sum_{j=1}^{m} \nabla f_j(G_j^{i-1}(x^*)) = 0.$$

More on the Issue of Correctness (FedGD)

FedSplit
FedDR
Extension

Indeed, one has

$$
\begin{aligned}
0 &= \frac{1}{m}\sum_{j=1}^{m} G_j^e(x^*) - x^* = \frac{1}{m}\sum_{j=1}^{m} G_j(G_j^{e-1}(x^*)) - x^* \\
&= \frac{1}{m}\sum_{j=1}^{m} \left(G_j^{e-1}(x^*) - s\nabla f_j(G_j^{e-1}(x^*))\right) - x^* \\
&= \frac{1}{m}\sum_{j=1}^{m} G_j^{e-1}(x^*) - x^* - \frac{s}{m}\sum_{j=1}^{m} \nabla f_j(G_j^{e-1}(x^*)) \\
&\ \ \vdots \\
&= \frac{1}{m}\sum_{j=1}^{m} G_j^{0}(x^*) - x^* - \frac{s}{m}\sum_{i=1}^{e}\sum_{j=1}^{m} \nabla f_j(G_j^{i-1}(x^*)) \\
&= -\frac{s}{m}\sum_{i=1}^{e}\sum_{j=1}^{m} \nabla f_j(G_j^{i-1}(x^*))
\end{aligned}
$$

More on the Issue of Correctness (FedProx)

Proposition

The sequence $\{x^{(t)}\}_{\infty}^{t=1}$ generated by `FedProx` *satisfy*

- *if $x^{(t)}$ convergent, then $x_j^{(t)}$ share a common limit x^**
- *x^* satisfy the fixed point relation $\sum_{j=1}^{m} \nabla M_{sf_j}(x^*) = 0$*

Notations (`FedProx`)

- $\text{prox}_{sf_j}(z) := \arg\min_{x_j} \left\{ f_j(x_j) + \frac{1}{2s}\|z - x_j\|^2 \right\}$
- $M_{sf_j} := \inf_{x} \left\{ f_j(x_j) + \frac{1}{2s}\|z - x_j\|^2 \right\}$
- $x_j^{(t+1/2)} := \text{prox}_{sf_j}(x_j^{(t)}), x_j^{(t+1)} = \overline{x}^{(t+1/2)}$.

More on the Issue of Correctness (FedProx)

Sketch:

As f_j are smooth convex, one has

$$\mathrm{prox}_{sf_j}(z) = z - s\nabla M_{sf_j}(z)$$

Hence

$$\begin{aligned}
0 &= x^* - \frac{1}{m}\sum_{j=1}^{m} \mathrm{prox}_{sf_j}(x^*) \\
&= x^* - \frac{1}{m}\sum_{j=1}^{m} \left(x^* - s\nabla M_{sf_j}(x^*)\right) \\
&= x^* - \frac{1}{m}\sum_{j=1}^{m} x^* + \frac{s}{m}\sum_{j=1}^{m} \nabla M_{sf_j}(x^*) \\
&= \frac{s}{m}\sum_{j=1}^{m} \nabla M_{sf_j}(x^*)
\end{aligned}$$

Incorrectness for Least Square Problems

FedSplit
FedDR
Extension

Least Square Problem (LSP)

$f_j(x_j) = \frac{1}{2}\|A_j x_j - b_j\|^2$, and A_j has "full rank" ($= d$).

LSP has unique solution

$$x_{ls}^* = \left(\sum_{j=1}^{m} A_j^T A_j\right)^{-1} \sum_{j=1}^{m} A_j^T b_j$$

Incorrectness for Least Square Problems

Least Square Problem (LSP)

$f_j(x_j) = \frac{1}{2}\|A_j x_j - b_j\|^2$, and A_j has “full rank” ($= d$).

LSP has unique solution

$$x_{ls}^* = \left(\sum_{j=1}^m A_j^T A_j\right)^{-1} \sum_{j=1}^m A_j^T b_j$$

By previous propositions,

- $x_{\mathtt{FedGD}}^* = \left(\sum_{j=1}^m A_j^T A_j G\right)^{-1} \left(\sum_{j=1}^m G A_j^T b_j\right)$, with $G = \sum_{k=0}^{e-1} (I - sA_j^T A_j)^k$
- $x_{\mathtt{FedProx}}^* = \left(\sum_{j=1}^m (I - P_j)\right)^{-1} \left(\frac{1}{s}\sum_{j=1}^m P_j A_j^T b_j\right)$, with $P_j = (I + sA_j^T A_j)^{-1}$

Incorrectness for Least Square Problems

Indeed, for example for `FedGD`, one has

$$\nabla f_j(x_j) = A_j^T A_j x_j - A_j^T b_j$$
$$G_j(x_j) = x_j - s f_j(x_j) = (I - sA_j^T A_j)x_j + sA_j^T b_j$$
$$G_j^{e+1}(x_j) = G_j(G_j^e(x_j)) = (I - sA_j^T A_j)G_j^e(x_j) + sA_j^T b_j$$

Hence

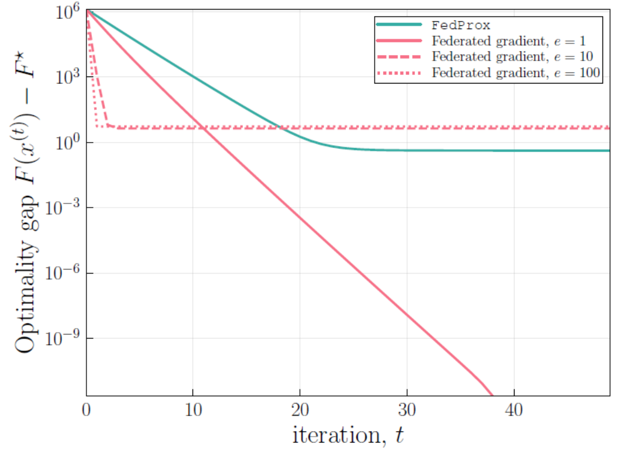
$$\begin{aligned} G_j^e(x_j) &= (I - sA_j^T A_j)^e x_j + (I - (I - sA_j^T A_j)^e)(A_j^T A_j)^{-1} A_j^T b_j \\ &= (I - sA_j^T A_j)^e x_j + (A_j^T A_j)^{-1}(I - (I - sA_j^T A_j)^e) A_j^T b_j \end{aligned}$$

From the fixed point relation $\sum_{i=1}^{e} \sum_{j=1}^{m} \nabla f_j(G_j^{i-1}(x^*)) = 0$, one has

$$0 = \sum_{i=1}^{e} \sum_{j=1}^{m} \left(A_j^T A_j (I - sA_j^T A_j)^{i-1} x^* - (I - sA_j^T A_j)^{i-1} A_j^T b_j \right).$$

Incorrectness for Least Square Problems

Settings: $m = 25$, $d = 100$, $A_j \in \mathrm{Mat}_{500\times 100}$, $(A_j)_{kl} \sim N(0,1)$, $b_j = A_j x_0 + \varepsilon_j$ with $\varepsilon_j \sim N(0, 0.25I)$



Problem Reformulation

The original problem can be reformulated as

$$\min \quad F(x) := \sum_{j=1}^{m} f_j(x_j)$$
$$\text{s.t.} \quad Ax = 0$$

where $x = (x_1, \cdots, x_m)$, $A = \begin{pmatrix} I & -I & & & \\ & I & -I & & \\ & & \ddots & \ddots & \\ & & & \ddots & -I \\ -I & & & & I \end{pmatrix}$

Consider the first-order optimal condition for $L(x, y) = F(x) - \langle y, Ax \rangle$, i.e. $\nabla F(x) - A^T y = 0$, or equiv.

$$\nabla F(x) - \begin{pmatrix} y_1 - y_m \\ \vdots \\ y_m - y_{m-1} \end{pmatrix} = 0$$

Problem Reformulation

Hence is a monotone inclusion problem

$$0 \in \nabla F(x) + \mathcal{N}_E(x)$$

where

$$\mathcal{N}_E(x) = \begin{cases} E^\perp & \text{if } x \in E \\ \emptyset & \text{otherwise} \end{cases} \quad \text{normal cone}$$

$$E = \{x \mid x_1 = \cdots = x_m\}$$

Indeed for $x \in E$,

$$\begin{aligned} \mathcal{N}_E(x) &= \{y \mid \langle y, \tilde{x} - x\rangle \leqslant 0 \ \forall \tilde{x} \in E\} \\ &= \left\{y \,\middle|\, \left\langle \sum\nolimits_{j=1}^m y_j,\ \tilde{x}_1 - x_1 \right\rangle \leqslant 0,\ \forall \tilde{x}_1 \in \mathbb{R}^d \right\} \\ &= \left\{y \,\middle|\, \sum\nolimits_{j=1}^m y_j = 0 \right\} = E^\perp \end{aligned}$$

Problem Reformulation

Another Perspective of Problem Reformulation

Let ι_E be the indicator function of E, then the constrained problem can be reformulated as the following unconstrained one

$$\min\ F(x) + \iota_E(x), \quad x \in \mathbb{R}^{md}$$

The first-order optimal condition gives

$$0 \in \nabla F(x) + \partial \iota_E(x) = \nabla F(x) + \mathcal{N}_E(x)$$

Operator Splitting

Let $\mathcal{F} = A + B$, with A, B maximal monotone. Write

$$R_A = (I + sA)^{-1}, \quad R_B = (I + sB)^{-1}$$
$$C_A = 2R_A - I, \quad C_B = 2R_B - I$$

Then

- $C_A, C_B, C_A C_B$ nonexpansive
- $0 \in A(x) + B(x) \iff C_A C_B(z) = z,\ x = R_B(z)$

i.e. we are reduced to finding fixed points of the nonexpansive operator $C_A C_B$.

Operator Splitting

Now consider $\mathcal{F} = \nabla F + \mathcal{N}_E$, one is reduced to find fixed points of $C_A C_B$ with $A = \nabla F$, $B = \mathcal{N}_E$. One has

$$R_{\nabla F} = \mathrm{prox}_{sF}, \quad R_{\mathcal{N}_E} = \Pi_E$$

and

$$\begin{aligned}
\mathrm{prox}_{sF}(x) &= \arg\min_z \left\{ F(z) + \frac{1}{2s}\|z - x\|^2 \right\} \\
&= \arg\min_z \left\{ \sum_{j=1}^m f_j(z_j) + \frac{1}{2s}\sum_{j=1}^m \|z_j - x_j\|^2 \right\} \\
&= \left(\mathrm{prox}_{sf_j}(x_j)\right)_{j=1}^m
\end{aligned}$$

Operator Splitting

- FedSplit
- FedDR
- Extension

- Peaceman-Rachford $z^{(t+1)} = C_A C_B(z^{(t)})$
- Douglas-Rachford $z^{(t+1)} = \frac{1}{2}(I + C_A C_B)(z^{(t)})$

Peaceman-Rachford

$$
\begin{aligned}
x^{(t+1/2)} &= R_B(z^{(t)}) \\
z^{(t+1/2)} &= 2x^{(t+1/2)} - z^{(t)} \\
x^{(t+1)} &= R_A(z^{(t+1/2)}) \\
z^{(t+1)} &= z^{(t)} + 2x^{(t+1)} \\
&\quad - 2x^{(t+1/2)}
\end{aligned}
$$

Douglas-Rachford

$$
\begin{aligned}
x^{(t+1/2)} &= R_B(z^{(t)}) \\
z^{(t+1/2)} &= 2x^{(t+1/2)} - z^{(t)} \\
x^{(t+1)} &= R_A(z^{(t+1/2)}) \\
z^{(t+1)} &= z^{(t)} + x^{(t+1)} \\
&\quad - x^{(t+1/2)}
\end{aligned}
$$

More generally, $z^{(t+1)} = z^{(t)} + \alpha(x^{(t+1)} - x^{(t+1/2)})$

Operator Splitting

By adjusting ordering and change of variable names

More General and Compressed Form 1

$$
\begin{aligned}
z^{(t+1/2)} &= R_A(2x^{(t)} - z^{(t)}) \\
z^{(t+1)} &= z^{(t)} + \alpha(z^{(t+1/2)} - x^{(t)}) \\
x^{(t+1)} &= R_B(z^{(t+1)})
\end{aligned}
$$

More General and Compressed Form 2

$$
\begin{aligned}
z^{(t+1)} &= z^{(t)} + \alpha(y^{(t)} - x^{(t)}) \\
x^{(t+1)} &= R_B(z^{(t+1)}) \\
y^{(t+1)} &= R_A(2x^{(t+1)} - z^{(t+1)})
\end{aligned}
$$

The FedSplit Algorithm

Algorithm 1: FedSplit

Given initiation $x \in \mathbb{R}^d$, proximal solvers $\texttt{prox\_update}_j : \mathbb{R}^d \to \mathbb{R}^d$

Initialize $x^{(1)} = z_1^{(1)} = \cdots = z_m^{(1)} = x$

for $t = 1, 2, \cdots$ **do**

    for $j = 1, \cdots, m$ **in parallel do**

        Local prox step: $z_j^{(t+1/2)} \leftarrow \texttt{prox\_update}_j(2x^{(t)} - z_j^{(t)})$

        Local centering step: $z_j^{(t+1)} \leftarrow z_j^{(t)} + 2(z_j^{(t+1/2)} - x^{(t)})$

    Compute global average: $x^{(t+1)} \leftarrow \bar{z}^{(t+1)}$

    if *meet convergent criteria* **then**

        $x^* \leftarrow x^{(t+1)}$

        break

return x^*

Note the difference against previous iteration form of Peaceman-Rachford:

first step -> last step; 2, 3 step merges; parameters renamed

Correctness and Convergence

FedSplit
FedDR
Extension

Proposition (Correctness)

If $z^* = (z_1^*, \cdots, z_m^*)$ *is a fixed point of* `FedSplit`*, then* $x^* := \Pi_E(z^*) = \frac{1}{m}\sum_{j=1}^m z_j^*$ *is an optimal solution to the original problem* $\min_x \sum_{j=1}^m f_j(x)$.

Theorem (Convergence)

Let f_j *be* ℓ_j*-strongly convex and* L_j*-smooth,* $\ell_* = \min \ell_j$, $L^* = \max L_j$, $\kappa = L^*/\ell_*$. *Take step size* $s = 1/\sqrt{\ell_* L^*}$*, and assume* $\|\texttt{prox\_update}_j(z) - \mathrm{prox}_{sf_j}(z)\| \leqslant b$*, then*

$$\|x^{(t+1)} - x^*\| \leqslant \left(1 - \frac{2}{\sqrt{\kappa}+1}\right)^t \frac{\|z^{(1)} - z^*\|}{\sqrt{m}} + (\sqrt{\kappa}+1)b$$

Non Strongly Convex Case

Consider a suitable regularization

$$\min F_\lambda(z) = \sum_{j=1}^{m} \left(f_j(z_j) + \frac{\lambda}{2} \|z_j - x^{(1)}\|^2 \right)$$

$$\text{s.t.} z_1 = \cdots z_m$$

Theorem

Let $\lambda \in \left(0, \frac{\varepsilon}{m\|x^* - x^{(1)}\|^2}\right)$, error bound $F(\widehat{x}) - F^* \leqslant \varepsilon$, `FedSplit` with regularized objective F_λ and step size $s = 1/\sqrt{\lambda(L^* + \lambda)}$ converges in at most

$$O\left(\sqrt{\frac{L^* \|x^* - x^{(1)}\|^2}{\varepsilon}}\right)$$

iterations

Motivation and Formulation

Motivation

- Nonconvex Douglas-Rachford splitting
- randomized block-coordinate strategy

Problem Formulation

$$\min_x F(x) = f(x) + g(x) = \frac{1}{n}\sum_{i=1}^{n} f_i(x) + g(x)$$

- f_i nonconvex, L-smooth,
- g closed proper convex

[2] Q. Tran-Dinh, N. Pham, D. T. Phan, and L. M. Nguyen, "FedDR – Randomized Douglas-Rachford Splitting Algorithms for Nonconvex Federated Composite Optimization," in *Advances in Neural Information Processing Systems* (A. Beygelzimer, Y. Dauphin, P. Liang, and J. W. Vaughan, eds.), 2021

Optimal Condition

FedSplit
FedDR
Extension

Necessary Optimal Condition

$$0 \in \nabla f(x) + \partial g(x)$$

This Condition has equivalent forms:

$$\begin{aligned} 0 \in \nabla f(x) + \partial g(x) &\Longleftrightarrow x - \beta \nabla f(x) \in (I + \beta \partial g)(x) \\ &\Longleftrightarrow (I + \beta \partial g)^{-1}(x - \beta \nabla f(x)) = x \\ &\Longleftrightarrow \frac{1}{\beta}(x - \mathrm{prox}_{\beta g}(x - \beta \nabla f(x))) = 0 \end{aligned}$$

Gradient mapping

$$\mathcal{G}_\beta(x) := \frac{1}{\beta}(x - \mathrm{prox}_{\beta g}(x - \beta \nabla f(x)))$$

Problem Reformulation

Block Split Constrained Reformulation

$$\begin{aligned} \min \quad & F(X) = f(X) + g(X) = \sum_{i=1}^{n} f_i(x_i) + g(x_{n+1}) \\ \text{s.t.} \quad & x_1 = \cdots = x_{n+1} \in \mathbb{R}^d, \quad X = (x_1, \cdots, x_{n+1}) \end{aligned}$$

Block Split Unconstrained Reformulation

$$\begin{aligned} \min \quad F(X) &= f(X) + g(X) + \iota_E(X) \\ &= \sum_{i=1}^{n} f_i(x_i) + g(x_{n+1}) + \iota_E(x) \end{aligned}$$

Optimal Condition and Operator Splitting

FedSplit
FedDR
Extension

Necessary Optimal Condition

$$0 \in \nabla f(X) + \partial(g + \iota_E)(X)$$

Douglas-Rachford Splitting

Let $B = \nabla f, A = \partial(g + \iota_E)$, then $R_B = \mathrm{prox}_{nsf}$, $R_A = \mathrm{prox}_{ns(g+\iota_E)}$. Iteration of DR splitting is

$$\begin{aligned} Y^{(t+1)} &= Y^{(t)} + \alpha(\overline{X}^{(t)} - X^{(t)}) \\ X^{(t+1)} &= \mathrm{prox}_{nsf}(Y^{(t+1)}) \\ \overline{X}^{(t+1)} &= \mathrm{prox}_{ns(g+\iota_E)}(2X^{(t+1)} - Y^{(t+1)}) \end{aligned}$$

FedSplit
FedDR
Extension

- $f = \frac{1}{n}\sum_{i=1}^{n} f_i(x_i)$ splits, hence

$$X^{(t+1)} = \mathrm{prox}_{nsf}(Y^{(t+1)}) \Rightarrow \begin{cases} x_i^{(t+1)} = \mathrm{prox}_{sf_i}(y_i^{(t+1)}), i \in [n] \\ x_{n+1}^{(t+1)} = y_{n+1}^{(t+1)} =: y^{(t+1)} \end{cases}$$

- write $\widehat{X}^{(t+1)} = 2X^{(t+1)} - Y^{(t+1)}$, $\widetilde{x}^{(t+1)} = \frac{1}{n}\sum_{i=1}^{n} \widehat{x}_k^{(t+1)}$, then $\overline{X}^{(t+1)} = \mathrm{prox}_{ns(g+\iota_E)}(2X^{(t+1)} - Y^{(t+1)})$ can be simplified (for all $i \in [n+1]$)

$$\begin{aligned} \overline{x}^{(t+1)} := \overline{x}_i^{(t+1)} &= \mathrm{prox}_{\frac{ns}{n+1}g}\left(\frac{1}{n+1}\sum_{i=1}^{n+1}\widehat{x}_i^{(t+1)}\right) \\ &= \mathrm{prox}_{\frac{ns}{n+1}g}\left(\frac{n}{n+1}\widetilde{x}^{(t+1)} + \frac{1}{n+1}y^{(t+1)}\right) \end{aligned}$$

Operator Splitting - Further Analysis

Hence the $(n+1)d$-dim. DRS splits (reduces) to d-dim. parallel DRS

Parallel DRS

$$
\begin{aligned}
y_i^{(t+1)} &= y_i^{(t)} + \alpha(\bar{x}^{(t)} - x_i^{(t)}), \ i \in [n] \\
x_i^{(t+1)} &= \mathrm{prox}_{sf_i}(y_i^{(t+1)}), \ i \in [n] \\
\widehat{x}_i^{(t+1)} &= 2x_i^{(t+1)} - y_i^{(t+1)}, \ i \in [n] \\
\widetilde{x}^{(t+1)} &= \frac{1}{n}\sum\nolimits_{i=1}^{n} \widehat{x}_i^{(t+1)} \\
y^{(t+1)} &= y^{(t)} + \alpha(\bar{x}^{(t)} - y^{(t)}) \\
\bar{x}^{(t+1)} &= \mathrm{prox}_{\frac{ns}{n+1}g}\left(\frac{n}{n+1}\widetilde{x}^{(t+1)} + \frac{1}{n+1}y^{(t+1)}\right)
\end{aligned}
$$

Algorithm 2: FedDR

Initiation: $x^{(0)} \in \mathrm{dom}(F)$, $s > 0$, $\alpha > 0$, $\varepsilon_{i,0} \geqslant 0$

Init server: $\overline{x}^{(0)} = \widetilde{x}^{(0)} = y^{(0)} = x^{(0)}$

Init users: $y_i^{(0)} = x^{(0)}$, $x_i^{(0)} \approx \mathrm{prox}_{sf_i}(y_i^{(0)})$, $\widehat{x}_i^{(0)} = 2x_i^{(0)} - y_i^{(0)}$

for $t = 1, 2, \cdots, T$ **do**

[Active users] Sample $\mathcal{S}_t \subseteq [n]$

[Comm] Each user $i \in \mathcal{S}_t$ receives $\overline{x}^{(t)}$ from server

[Local update] **for** *each user* $i \in \mathcal{S}_t$ **do**

Choose $\varepsilon_{i,t+1} \geqslant 0$, update $y_i^{(t+1)} \leftarrow y_i^{(t)} + \alpha(\overline{x}^{(t)} - x_i^{(t)})$,

$x_i^{(t+1)} \approx \mathrm{prox}_{sf_i}(y_i^{(t+1)})$, $\widehat{x}_i^{(t+1)} \leftarrow 2x_i^{(t+1)} - y_i^{(t+1)}$

[Comm] Each user $i \in \mathcal{S}_t$ sends $\Delta\widehat{x}_i^{(t)} = \widehat{x}_i^{(t+1)} - \widehat{x}_i^{(t)}$ to server

[Server update]: $y^{(t+1)} \leftarrow y^{(t)} + \alpha(\overline{x}^{(t)} - y^{(t)})$,

$\widetilde{x}^{(t+1)} \leftarrow \widetilde{x}^{(t)} + \frac{1}{n}\sum_{i\in\mathcal{S}_t} \Delta\widehat{x}_i^{(t)}$,

$\overline{x}^{(t+1)} \leftarrow \mathrm{prox}_{\frac{ns}{n+1}g}\left(\frac{n}{n+1}\widetilde{x}^{(t+1)} + \frac{1}{n+1}y^{(t+1)}\right)$

The FedDR Algorithm - Convergence

As in `FedSplit`, inexact proximal operator is used

$$x_i^{(t)} \approx \mathrm{prox}_{sf_i}(y_i^{(t)}) =: z_i^{(t)}$$

approximated up to a given accuracy $\varepsilon_{i,t}$, i.e.

$$x_i^{(t)} = z_i^{(t)} + e_i^{(t)}, \qquad \text{with } \|e_i^{(t)}\| \leqslant \varepsilon_{i,t}$$

Convergence Results

Let $\{(\bar{x}^{(t)}, x_i^{(t)})\}$ be the sequence generated by `FedDR`, then

- Global-Local Difference ($\gamma_1 > 0$)

$$\|\bar{x}^{(t)} - x_i^{(t)}\|^2 \leqslant \frac{2(1+s^2L^2)}{\alpha}\Big[(1+\gamma_1)\|x_i^{(t+1)} - x_i^{(t)}\|^2 + \frac{2(1+\gamma_1)}{\gamma_1}(\|e_i^{(t+1)}\|^2 + \|e_i^{(t)}\|^2)\Big]$$

Convergence Results (Continued)

- Bounded Global Gradient Mapping ($\gamma_2 > 0$)

$$\|\mathcal{G}_{\frac{ns}{n+1}}(\bar{x}^{(t)})\|^2 \leqslant \frac{n+1}{n^2 s^2} \left\{ (1+sL)^2 \sum_{i=1}^{n} (1+\gamma_2) \left[\|\bar{x}^{(t)} - x_i^{(t)}\|^2 + \frac{1+\gamma_2}{\gamma_2} \|e_i^{(t)}\|^2 \right] + \|y^{(t)} - \bar{x}^{(t)}\|^2 \right\}$$

- Global Convergence (C_1, C_2, C_3 are constants)

$$\frac{1}{T+1} \sum_{t=1}^{T} \mathbb{E}\left[\|\mathcal{G}_{\frac{ns}{n+1}}(\bar{x}^{(t)})\|^2 \right] \leqslant \frac{C_1 (F(x^{(0)} - F^*)}{T+1} + \frac{1}{n(T+1)} \sum_{t=1}^{T} \sum_{i=1}^{n} (C_2 \varepsilon_{i,t}^2 + C_3 \varepsilon_{i,t+1}^2)$$

Personalization and Operator Splitting

Consider a general form of personalization in FL

$$\sum_{i=1}^{n} \{f_i(x_i) + g(x_i, x)\}$$

or for the most cases

$$\sum_{i=1}^{n} \{f_i(x_i) + g(x_i - x)\}$$

which can be reformulated as a constrained problem

$$\begin{aligned} \min \quad & \sum_{i=1}^{n} \{f_i(x_i) + g(y_i)\} \\ \text{s.t.} \quad & u_i = x_i - y_i, u_1 = \cdots = u_n \end{aligned}$$

How can operator splitting be applied in such problems?

Customized DR Model

$$
\begin{aligned}
\min \quad & F(x) + G(y) \\
\text{s.t.} \quad & Ax + By = b \\
& x \in \mathcal{X}, y \in \mathcal{Y}
\end{aligned}
$$

Let $F(x) = \sum_{i=1}^{n} f_i(x_i)$, $G(y) = \sum_{i=1}^{n} g(y_i)$,

$A = \begin{pmatrix} I & -I & & & \\ & I & -I & & \\ & & \ddots & \ddots & \\ & & & \ddots & -I \\ -I & & & & I \end{pmatrix}$, then the personalization problem can be stated as

$$
\begin{aligned}
\min \quad & F(x) + G(y) \\
\text{s.t.} \quad & Ax - Ay = 0
\end{aligned}
$$

References I

[1] R. Pathak and M. J. Wainwright, "FedSplit: An Algorithmic Framework for Fast Federated Optimization," in *Advances in Neural Information Processing Systems* (H. Larochelle, M. Ranzato, R. Hadsell, M. Balcan, and H. Lin, eds.), vol. 33, pp. 7057–7066, Curran Associates, Inc., 2020.

[2] Q. Tran-Dinh, N. Pham, D. T. Phan, and L. M. Nguyen, "FedDR – Randomized Douglas-Rachford Splitting Algorithms for Nonconvex Federated Composite Optimization," in *Advances in Neural Information Processing Systems* (A. Beygelzimer, Y. Dauphin, P. Liang, and J. W. Vaughan, eds.), 2021.