1 定义

def 1.

$$\forall i, x_i\left(k+1\right) = x_i\left(k\right) + U_i\left(x^i\left(k\right)\right) \tag{1}$$

T_k 表示第 k 步 *master* 接受更新的 *workers* 的集合。

i 个 *worker* 上面的 *local* 数值

$$x^i\left(k\right) \tag{2}$$

第 k 步的全局 x 值:

$$x\left(k\right) = \left(x_1(k), \cdots, x_n(k)\right) \tag{3}$$

def 2.

$$P_i\left(x^i\left(k\right)\right) = prox_{g_i}^{\eta}\left(x_i^i\left(k\right) - \eta\nabla_i f\left(x^i\left(k\right)\right)\right) \tag{4}$$

$$P_i\left(x\left(k\right)\right) = prox_{g_i}^{\eta}\left(x_i\left(k\right) - \eta\nabla_i f\left(x\left(k\right)\right)\right) \tag{5}$$

$$\begin{aligned} U_i\left(x^i\left(k\right)\right) &= prox_{g_i}^{\eta}\left(x_i^i\left(k\right) - \eta\nabla_i f\left(x^i\left(k\right)\right)\right) - x_i^i\left(k\right) \\ &= P_i\left(x^i\left(k\right)\right) - x_i^i\left(k\right) \end{aligned} \tag{6}$$

def 3.

$$M_F\left(x\left(k+1\right)\right) = x\left(k+1\right) - prox_g^{\eta}\left\{x\left(k+1\right) - \eta\nabla f\left(x\left(k+1\right)\right)\right\} \tag{7}$$

$$M_F\left(x\left(k+1\right)\right)_i = x_i\left(k+1\right) - prox_{g_i}^{\eta}\left\{x_i\left(k+1\right) - \eta\nabla_i f\left(x\left(k+1\right)\right)\right\} \tag{8}$$

2 算法描述

首先按照 workers 的数目 n，把变量 x 分成 n 部分，即 $x = (x_1, x_2, \cdots, x_n)$。master 上面储存有真正的 x 值，也叫做全局 x 值，第 k 步的全局 x 值记作 $x\left(k\right)$；每个 workers 上面储存的是局部 x 值，对于第 i 个 worker，第 k 步的局部 x 值记作 $x^i\left(k\right)$。每一步迭代中，每一个 worker 处理其中的一部分，也就是说 $i-th$worker 处理 x_i，第 k 步时，计算更新程度 $U_i\left(x^i\left(k\right)\right)$。对于 master，每次接收 workers 上的更新程度，并作更新 $x_i\left(k+1\right) = x_i\left(t\right) + \mathrm{I}_{\{i\in T_k\}}U_i\left(x^i\left(t\right)\right)$（$T_k$ 表示第 k 步时，master 会接受的 workers 集合）。因为第 i 个 worker 只负责 x_i 的更新，因而有 $x_i^i\left(k\right) = x_i\left(k\right)$，也就是说，$x_i^i\left(k\right)$ 的值保持最新。

Algorithm 1 Decouped Asyn-SCD, For worker i

repeat
  Obtain the parameter x from the master(shared memory or parameter server).
  Evaluate the gradient of the i^{th} component over parameter x, denoted by $\nabla_i f(x)$.
  Evaluate the proximal operate $[x_i]^{'} = prox_{g_i}^{\eta}(x_i - \eta \nabla_i f(x))$.
  Send the update information $U_i(x) = [x_i]^{'} - x_i$.
until procedure of master ends

Algorithm 2 Decouped Asyn-SCD, For master

for $k = 1, 2, \cdots$ **do**
  Get $U_i(x^i) = [x_i^i]^{'} - x_i^i$.
  Update parameter with $x_i(k+1) = x_i(t) + U_i(x^i(t))$.
  $k = k + 1$
end for

3 定理

Assumption 1. *For all $x \in R^n$ and all $t \in R$, we have*

$$|\nabla_i f(x + te_i) - \nabla_i f(x)| \le L_i |t| \tag{9}$$

L_i is called component Lipshitz constants. $L_{max} = \max L_i, \quad i = 1, 2, \cdots, n$

Assumption 2. *For all $x, y \in R^n$, we have*

$$\|\nabla f(x) - \nabla f(y)\| \le L_f \|x - y\| \tag{10}$$

Assumption 3. *For all $x, y \in R^n$, we have*

$$\|\nabla g(x) - \nabla g(y)\| \le L \|x - y\| \tag{11}$$

Assumption 4. *F is bounded from below F^*.*

Assumption 5. *Delay* 的最大容忍值为 s，也就是说在 s 次迭代中，必定有一个组件（坐标）会得到 *update*。

Th 1. 在 *Assumption1*、*2*、*3*、*4*、*5*满足的情形下，有：

$$\frac{1}{T}\sum_{k=0}^{T}[M_F(x(k))]^2 \le \frac{2ns}{T}(\eta L_{max} + 1)^2 \frac{2}{\frac{1}{\eta} - L_f - 2Ls}[F(x(0)) - F^*] \tag{12}$$

4 证明

因为在 s 个时间之中,第 i 个组件必定能得到 update,$\therefore \exists \delta \in [0,s]$, sit. $x_i(k) = P_i\left(x^i(k-\delta)\right)$

$$\begin{aligned} M_F(x(k))_i &= x_i(k) - prox_{g_i}^{\eta}\left\{x_i(k) - \eta\nabla_i f(x(k))\right\} \\ &= P_i\left(x^i(k-\delta)\right) - P_i(x(k)) \\ &= P_i\left(x^i(k-\delta)\right) - P_i(x(k-\delta)) - \left[P_i(x(k)) - P_i(x(k-\delta))\right] \end{aligned} \tag{13}$$

可得

$$\left\|M_F(x(k))_i\right\|_2 \le \left\|P_i\left(x^i(k-\delta)\right) - P_i(x(k-\delta))\right\|_2 + \left\|P_i(x(k)) - P_i(x(k-\delta))\right\|_2 \tag{14}$$

对于

$$P_i(x) = prox_{g_i}^{\eta}\left\{x_i - \eta\nabla_i f(x)\right\} \tag{15}$$

$$P_i(y) = prox_{g_i}^{\eta}\left\{y_i - \eta\nabla_i f(y)\right\} \tag{16}$$

现在来证明 ∇P_i 的 Lipschitz Continuous 性质。$\because g_i$ 是凸函数，根据凸函数的单调性，

$$\forall\, u \in \partial g_i(P_i(x)), \quad v \in \partial g_i(P_i(y)) \tag{17}$$

$$(u-v)\left[P_i(x) - P_i(y)\right] \ge 0 \tag{18}$$

根据 $P_i(x)$ 及 $P_i(y)$ 的定义，有

$$0 \in \partial g_i(P_i(x)) + \frac{1}{\eta}(P_i(x) - x) + \nabla_i f(x) \tag{19}$$

$$0 \in \partial g_i(P_i(y)) + \frac{1}{\eta}(P_i(y) - y) + \nabla_i f(y) \tag{20}$$

故 $\exists \hat{u}, \hat{v}$，使得

$$\hat{u} + \frac{1}{\eta}(P_i(x) - x) + \nabla_i f(x) = 0 \tag{21}$$

$$\hat{v} + \frac{1}{\eta}(P_i(y) - y) + \nabla_i f(y) = 0 \tag{22}$$

带入18，可得

$$\left[-\frac{1}{\eta}(P_i(x) - x) - \nabla_i f(x) + \frac{1}{\eta}(P_i(y) - y) + \nabla_i f(y)\right]^T \left[P_i(x) - P_i(y)\right] \ge 0 \tag{23}$$

移项可得

$$\begin{aligned} \left\|P_i(x) - P_i(y)\right\|_2^2 &\le \left[(x-y) - \eta\left(\nabla_i f(x) - \nabla_i f(y)\right)\right]^T \left(P_i(x) - P_i(y)\right) \\ \text{Cauchy inequlity} &\le \left\|(x-y) - \eta\left(\nabla_i f(x) - \nabla_i f(y)\right)\right\|_2 \left\|P_i(x) - P_i(y)\right\|_2 \\ &\le \left[\|x-y\|_2 + \eta\left\|\nabla_i f(x) - \nabla_i f(y)\right\|_2\right] \left\|P_i(x) - P_i(y)\right\|_2 \\ &\le (1 + \eta L_f)\|x-y\|_2 \left\|P_i(x) - P_i(y)\right\|_2 \end{aligned} \tag{24}$$

于是可得

$$\|P_i(x) - P_i(y)\|_2 \le (1 + \eta L_f)\|x - y\|_2 \tag{25}$$

将25带入13中，可得

$$\|M_F(x(k))_i\|_2 \le (1 + \eta L_f)\left(\left\|x^i(k-\delta) - x(k-\delta)\right\|_2 + \|x(k) - x(k-\delta)\|_2\right) \tag{26}$$

Difference between local $x^i(k)$ and global $x(k)$

$$\begin{aligned}
\left\|x(k) - x^i(k)\right\|_2 &= \sqrt{\sum_{j=1}^{n} \left\|x_j(k) - x_j^i(k)\right\|_2^2} \\
&\le \sqrt{\sum_{j=1}^{n} \left\|\sum_{t=(k-s)_+}^{k-1} (x_j(t+1) - x_j(t))\right\|_2^2} \\
&= \left\|\sum_{t=(k-s)_+}^{k-1} (x(t+1) - x(t))\right\|_2 \\
\text{(triangle inequality)} &\le \sum_{t=(k-s)_+}^{k-1} \|x(t+1) - x(t)\|_2
\end{aligned} \tag{27}$$

因而

$$\|M_F(x(k))_i\|_2 \le (1 + \eta L_f)\left(\sum_{t=(k-\delta-s)_+}^{k-\delta-1} \|x(t+1) - x(t)\|_2 + \sum_{t=(k-s)_+}^{k-1} \|x(t+1) - x(t)\|_2\right) \tag{28}$$

As $F(x) = f(x) + \sum\limits_{i=1}^{n} g_i(x_i)$, Then

$$F(x(k+1)) - F(x(k)) = f(x(k+1)) - f(x(k)) + \sum_{j=1}^{n} [g_j(x_j(k+1)) - g_j(x_j(k))] \tag{29}$$

对于 $f(x_i(k+1)) - f(x_i(k))$, 根据 ∇f Lipschitz Continuous Gradient

$$f(x(k+1)) - f(x(k)) \le \nabla f(x(k))^T (x(k+1) - x(k)) + \frac{L_f}{2}\|x(k+1) - x(k)\|_2^2 \tag{30}$$

若第 k 步，第 i 坐标得到更新，则

$$x_i(k+1) = prox_{g_i}^{\eta}\left(x_i(k) - \eta \nabla_i f\left(x^i(k)\right)\right) \tag{31}$$

否则

$$x_j(k+1) = x_j(k), \quad \text{Then } g_j(x_j(k+1)) - g_j(x_j(k)) = 0 \tag{32}$$

则对于 $\forall z \in \mathrm{dom}\, g$

$$\begin{aligned}
&g_i(x_i(k+1)) + \nabla_i f\left(x^i(k)\right)^T (x_i(k+1) - x_i(k)) + \frac{1}{2\eta}\|x_i(k+1) - x_i(k)\|_2^2 \\
&\le g_i(z) + \nabla_i f\left(x^i(k)\right)^T (z - x_i(k)) + \frac{1}{2\eta}\|z - x_i(k)\|_2^2
\end{aligned} \tag{33}$$

取 $z = x_i(k)$，则

$$g_i(x_i(k+1)) - g_i(x_i(k)) \le -\left\{\nabla_i f\left(x^j(k)\right)^T (x_i(k+1) - x_i(k)) + \frac{1}{2\eta}\|x_i(k+1) - x_i(k)\|_2^2\right\} \tag{34}$$

把30、34带入29之中，得

$$\begin{aligned}
F(x(k+1)) \le\ & \nabla f(x(k))^T (x(k+1)x(k)) + \frac{L_f}{2}\|x(k+1) - x(k)\|_2^2 \\
& - \sum_{i=1}^{n}\left\{\frac{1}{2\eta}\|(x_i(k+1) - x_i(k)\|_2^2 + \nabla_i f\left(x^i(k)\right)^T (x_i(k+1) - x_i(k))\right\} \\
=\ & \frac{1}{2}\left(L_f - \frac{1}{\eta}\right)\|(x(k+1) - x(k))\|_2^2 + \left(\nabla_i f(x(k)) - \nabla_i f\left(x^i(k)\right)\right)^T (x_i(k+1) - x_i(k)) \\
\le\ & \frac{1}{2}\left(L_f - \frac{1}{\eta}\right)\|(x(k+1) - x(k))\|_2^2 + \left\|\nabla_i f(x(k)) - \nabla_i f\left(x^i(k)\right)\right\|_2 \|x_i(k+1) - x_i(k)\|_2 \\
\le\ & \frac{1}{2}\left(L_f - \frac{1}{\eta}\right)\|(x(k+1) - x(k))\|_2^2 + L_f\left\|x(k) - x^i(k)\right\|_2 \|x_i(k+1) - x_i(k)\|_2 \\
\le\ & \frac{1}{2}\left(L_f - \frac{1}{\eta}\right)\|(x(k+1) - x(k))\|_2^2 + L_f\|x_i(k+1) - x_i(k)\|_2 \sum_{t=(k-s)_+}^{k-1}\|x(t+1) - x(t)\|_2 \\
\le\ & \frac{1}{2}\left(L_f - \frac{1}{\eta}\right)\|(x(k+1) - x(k))\|_2^2 + L_f\|x(k+1) - x(k)\|_2 \sum_{t=(k-s)_+}^{k-1}\|x(t+1) - x(t)\|_2 \\
\le\ & \frac{1}{2}\left(L_f - \frac{1}{\eta}\right)\|(x(k+1) - x(k))\|_2^2 + \frac{L_f}{2}\left(\|x(k+1) - x(k)\|_2 + \sum_{t=(k-s)_+}^{k-1}\|x(t+1) - x(t)\|_2\right) \\
\le\ & \frac{1}{2}\left(L_f - \frac{1}{\eta} + L_f s\right)\|(x(k+1) - x(k))\|_2^2 + \frac{L_f}{2}\sum_{t=(k-s)_+}^{k-1}\|x(t+1) - x(t)\|_2
\end{aligned} \tag{35}$$

改变求和顺序，有不等式:

$$\sum_{k=m_1}^{m_2}\sum_{t=(k-s)_+}^{k-1}\|x(t+1) - x(t)\|_2^2 \le s\sum_{k=m_1}^{m_2}\|x(k+1) - x(k)\|_2^2 \tag{36}$$

以及

$$\begin{aligned}
\sum_{k=m_1}^{m_2}\sum_{t=(k-s-\delta)_+}^{k-\delta-1}\|x(t+1) - x(t)\|_2^2 &\le s\sum_{k=(m_1-\delta)_+}^{m_2-\delta}\|x(k+1) - x(k)\|_2^2 \\
&\le s\sum_{k=(m_1-\delta)_+}^{m_2}\|x(k+1) - x(k)\|_2^2
\end{aligned} \tag{37}$$

对于35，从 $k = m_1$ 加到 $k = m_2$，得

$$\begin{aligned}
F(m_1) - F(m_2) &\le \frac{1}{2}\left(L_f - \frac{1}{\eta} + L_f s\right)\sum_{k=m_1}^{m_2}\|x(k+1) - x(k)\|_2^2 + \frac{L_f}{2}\sum_{k=m_1}^{m_2}\sum_{t=(k-s)_+}^{k-1}\|x(t+1) - x(t)\|_2^2 \\
(36) &\le \frac{1}{2}\left(L_f - \frac{1}{\eta} + L_f s\right)\sum_{k=m_1}^{m_2}\|x(k+1) - x(k)\|_2^2 + \frac{L_f s}{2}\sum_{k=m_1}^{m_2}\|x(k) - x(k-1)\|_2^2 \\
&= \frac{1}{2}\left(L_f - \frac{1}{\eta} + 2L_f s\right)\sum_{k=m_1}^{m_2}\|x(k+1) - x(k)\|_2^2 \qquad (38)
\end{aligned}$$

取 $\eta < \dfrac{1}{L_f + 2L_f s}$，则 $L_f - \dfrac{1}{\eta} + 2L_f s < 0$，则

$$\sum_{k=m_1}^{m_2}\|x(k+1) - x(k)\|_2^2 \le \frac{2}{\frac{1}{\eta} - L_f - 2L_f s}\left(F(x(m_1)) - F(x(m_2))\right) \qquad (39)$$

$$\begin{aligned}
\frac{1}{T}\sum_{k=0}^{T}\left[M_F(x(k))\right]^2 &= \frac{1}{T}\sum_{k=0}^{T}\sum_{i=1}^{n}\left[M_F(x(k))_i\right]^2 \\
&\le \frac{n}{T}(1 + \eta L_f)\left\{\sum_{k=0}^{T}\sum_{t=(k-\delta-s)_+}^{k-\delta-1}\|x(t+1) - x(t)\|_2 + \sum_{k=0}^{T}\sum_{t=(k-s)_+}^{k-1}\|x(t+1) - x(t)\|_2\right\} \\
(36、37) &\le \frac{2ns}{T}(1 + \eta L_f)\sum_{k=0}^{T}\|x(k+1) - x(k)\|_2 \\
&\le \frac{4ns(1 + \eta L_f)}{T\left(\frac{1}{\eta} - L_f - 2L_f s\right)}\left[F(x(0)) - F(x(T))\right] \\
(\text{Assumption 4.}) &\le \frac{4ns(1 + \eta L_f)}{T\left(\frac{1}{\eta} - L_f - 2L_f s\right)}\left[F(x(0)) - F^*\right] \qquad (40)
\end{aligned}$$