# Towards better understanding of decentralized optimization using row stochastic matrix

Liyuan Liang

## 1 Related Works

[?, ?] suggests using only row-stochastic matrix for decentralized optimization. There are some works following this idea:...

## 2 Assumptions and Notations

**Data heterogeneity:** There exists a constant $b$ such that $\|\nabla f_i(x) - \nabla f(x)\| \leq b^2, \forall i = 1, 2, \ldots, n$.

**Notation.** Suppose that we have primitive row-stochastic matrix $A$ satisfying $\pi^\top A = \pi^\top, A\mathbb{1}_n = \mathbb{1}_n$. Here $\pi$ is the Perron vector whose sum is 1. We define $V_0 = I_n, D_k = \mathrm{diag}(V_k) + \underbrace{\epsilon_k I_n}_{\text{New!}}, A_\infty = \lim_{k\to\infty} V_k = \mathbb{1}_n\pi^\top$. We use bold symbol to express parameter matrix. $\mathbf{x}^{(k)} := [{x_1^{(k)}}^T, \ldots, {x_n^{(k)}}^T]^T \in \mathbb{R}^{n\times d}, \nabla f(\mathbf{x}^{(k)}) := \nabla f(\mathbf{x}^{(k)}) = [\nabla f_1(x_1^{(k)})^T; \ldots; \nabla f_n(x_n^{(k)})^T]^T \in \mathbb{R}^{n\times d}, \mathbf{w}^{(k)} = A_\infty \mathbf{x}^{(k)}, w^{(k)} = \pi^T \mathbf{x}^{(k)}$,

## 3 Generalized Pull-Sum Gradient Tracking

To generalize the pull-diging method, we introduce a decay weight for the nonlinear divisor.

**Algorithm.**

$$\mathbf{x}^{(k+1)} = A\mathbf{x}^{(k)} - \alpha\mathbf{y}^{(k)} \tag{1}$$

$$V_{k+1} = AV_k \tag{2}$$

$$\mathbf{y}^{(k+1)} = A\mathbf{y}^{(k)} + D_{k+1}^{-1}\nabla f(\mathbf{x}^{(k+1)}) - D_k^{-1}\nabla f(\mathbf{x}^{(k)}) \tag{3}$$

hereon we set $\mathbf{y}^{(0)} = \nabla f(\mathbf{x}^{(0)})$.

### 3.1 Estimate of Gradient Consensus Error(Lemma 7 in [?])

[LLY: Xinyi, please complete it following the proof in Lemma B.5 in my paper]

**Lemma 1.** *When* $\alpha \leq \dfrac{1-\beta}{12L\kappa_\pi(1+2\beta^2)(\frac{T}{1-\beta}+D_2+1)}$, *we have:*

$$\sum_{k=0}^{T}\|\mathbf{y}^{(k+1)}-A_\infty\mathbf{y}^{(k+1)}\|_\pi^2 \leq \frac{2\kappa_\pi^2(1+\beta)}{(1-\beta)^4}\|\mathbf{y}^{(0)}\|_\pi^2$$
$$+\frac{6L^2\kappa_\pi^2(1+2\beta^2)}{(1-\beta)}(\frac{T}{1-\beta}+D_2+1)\sum_{k=0}^{T}\|\mathbf{x}^{(k)}-\mathbf{w}^{(k)}\|_\pi^2+\frac{6\alpha^2L^2\kappa_\pi^2}{1-\beta}(\frac{T}{1-\beta}+D_2+1)\sum_{k=0}^{T}\|A_\infty\mathbf{y}^{(k)}\|_\pi^2 \tag{4}$$

*Proof.* To bound $\mathbf{y}^{(k-i)}-A_\infty\mathbf{y}^{(k-i)}$, we left multiply $A_\infty$ on each side of (**??**), we have

$$\hat{\mathbf{y}}^{(k+1)}=\hat{\mathbf{y}}^{(k)}+A_\infty D_{k+1}^{-1}\nabla f(\mathbf{x}^{(k+1)})-A_\infty {D_k}^{-1}\nabla f(\mathbf{x}^{(k)}) \tag{5}$$

**Remarks.** Unfold (**??**), we have $\hat{\mathbf{y}}^{(k)}=\hat{\mathbf{y}}_0+A_\infty D_k^{-1}\nabla f(\mathbf{x}^{(k)})-A_\infty {D_0}^{-1}\nabla f^{(0)}=A_\infty {D_k}^{-1}\nabla f_k$. When $k\to\infty$, $\epsilon_k\to 0, \hat{\mathbf{y}}^{(k)}\to \mathbb{1}_n\pi^\top \mathrm{diag}(\pi)^{-1}\nabla f(\mathbf{x}^{(k)})=\mathbb{1}_n\mathbb{1}_n^\top\nabla f(\mathbf{x}^{(k)})$. This is why we use $D_k$ as the divisor.

Define $\Delta^{(k)}=\nabla f(\mathbf{x}^{(k+1)})-\nabla f(\mathbf{x}^{(k)}), \delta_k=D_{k+1}^{-1}-D_k^{-1}$, Subtract (**??**) from (**??**), we have the following equation:

$$\begin{aligned}
&\mathbf{y}^{(k+1)}-A_\infty\mathbf{y}^{(k+1)}\\
=&(A-A_\infty)(\mathbf{y}^{(k)}-A_\infty\mathbf{y}^{(k)})+(D_{k+1}^{-1}-A_\infty D_{k+1}^{-1})\nabla f(\mathbf{x}^{(k+1)})-(D_k^{-1}-A_\infty D_k^{-1})\nabla f(\mathbf{x}^{(k)})\\
=&(A-A_\infty)(\mathbf{y}^{(k)}-A_\infty\mathbf{y}^{(k)})+(D_{k+1}^{-1}-A_\infty D_{k+1}^{-1})\Delta^{(k)}+(I_n-A_\infty)\delta_k\nabla f(\mathbf{x}^{(k)})\\
=&(A-A_\infty)^2(\mathbf{y}^{(k-1)}-A_\infty\mathbf{y}^{(k-1)})+(I_n-A_\infty)D_{k+1}^{-1}\Delta^{(k)}+(I_n-A_\infty)[\delta_k+(A-A_\infty)(D_k^{-1}-A_\infty D_k^{-1})]\Delta^{(k-1)}\\
&\quad+[(I_n-A_\infty)\delta_k+(A-A_\infty)(I_n-A_\infty)\delta_{k-1}]\nabla f(\mathbf{x}^{(k-1)})\\
=&\dots\\
=&(A-A_\infty)^{k+1}(\mathbf{y}^{(0)}-A_\infty\mathbf{y}^{(0)})+\sum_{i=0}^{k}C_i\Delta^{(k-i)}+\sum_{i=0}^{k}(A-A_\infty)^i(I_n-A_\infty)\delta_{k-i}\mathbf{y}^{(0)}\\
=&[(A-A_\infty)^{k+1}+\sum_{i=0}^{k}(A-A_\infty)^i(I_n-A_\infty)\delta_{k-i}]\mathbf{y}^{(0)}+\sum_{i=0}^{k}C_i\Delta^{(k-i)}
\end{aligned} \tag{6}$$

where

$$C_i=(A-A_\infty)^i(I_n-A_\infty)D_{k+1-i}^{-1}+\sum_{j=0}^{i}(A-A_\infty)^j(I_n-A_\infty)\delta_{k-j}$$

Suppose that we have a constant $\kappa_\pi$ satisfying

$$\|D_k^{-1}\|\leq\kappa_\pi, \|\delta_k\|\leq\kappa_\pi\beta^k$$

Let $\beta=\|A-A_\infty\|_\pi$, then

$$\begin{aligned}
&\|A^k-A_\infty\|_F^2=\|(A-A_\infty)^k\|_F^2=\sum_{i=1}^{n}\|(A-A_\infty)^k\mathbf{e}_i\|_F^2\\
=&\sum_{i=1}^{n}\|\Pi^{-\frac{1}{2}}\Pi^{\frac{1}{2}}(A-A_\infty)^k\mathbf{e}_i\|_F^2\leq\sum_{i=1}^{n}\|\Pi^{-\frac{1}{2}}\|_2^2\|(A-A_\infty)^k\mathbf{e}_i\|_\pi^2\leq\underline{\pi}^{-1}\sum_{i=1}^{n}\|(A-A_\infty)^k\|_\pi^2\|\mathbf{e}_i\|_\pi^2\\
\leq&\sum_{i=1}^{n}\frac{1}{\underline{\pi}}\beta^{2k}\pi_i=\frac{1}{\underline{\pi}}\beta^{2k}
\end{aligned} \tag{7}$$

Therefore, for each k and $1 \le i, j \le n$ denote $v_i^{(k)} = (V_k)_{ii}$.We have$|v_i^{(k)} - \pi_i| \le \sqrt{\frac{1}{\underline{\pi}}}\beta^k$.

Let $\epsilon_k = \sqrt{\frac{1}{\underline{\pi}}}\beta^k$ ,then $\|D_k^{-1}\| \le \frac{1}{\underline{\pi}}$ and since

$$|\frac{1}{v_i^{(k)}} - \frac{1}{v_i^{(k+1)}}| = \frac{|v_i^{(k)} - v_i^{(k+1)}|}{v_i^{(k)}v_i^{(k+1)}} \le \frac{\sqrt{\frac{1}{\underline{\pi}}}\beta^k(\beta+1)}{\underline{\pi}^2}$$

we have

$$\|\delta_k\| \le \frac{\sqrt{\frac{1}{\underline{\pi}}}\beta^k(\beta+1)}{\underline{\pi}^2}$$

So we just let $\kappa_\pi = \max\left\{\frac{1}{\pi}, \frac{1+\beta}{\underline{\pi}^{2.5}}\right\} = \frac{1+\beta}{\underline{\pi}^{2.5}}$

It can be easily verified that $\|C_i\|_\pi \le \kappa_\pi(\beta^i + k\beta^k)$. Thus, using Jensen inequality, we have

$$\|\mathbf{y}^{(k+1)} - A_\infty \mathbf{y}^{(k+1)}\|_\pi^2 \le \frac{\kappa_\pi^2 \beta^{k+1}}{1-\beta}(k+2)^2\|\mathbf{y}^{(0)}\|_\pi^2 + \frac{\kappa_\pi^2}{1-\beta}\sum_{i=0}^{k}\beta^i(1+k\beta^{k-i})^2\|\Delta^{(k-i)}\|_\pi^2 \tag{8}$$

Notice that

$$\begin{aligned}
&\|\mathbf{x}^{(k+1)} - \mathbf{w}^{(k+1)}\|^2 = \|(A - \mathbb{1}_n\pi^T)(\mathbf{x}^{(k)} - \mathbf{w}^{(k)}) - \alpha(I_n - \mathbb{1}_n\pi^T)(\mathbf{y}^{(k)} - A_\infty\mathbf{y}^{(k)})\|^2 \\
\le& 2\beta^2\|\mathbf{x}^{(k)} - \mathbf{w}^{(k)}\|^2 + 2\alpha^2\|\mathbf{y}^{(k)} - A_\infty\mathbf{y}^{(k)}\|^2
\end{aligned} \tag{9}$$

here we use the fact that $\|I_n - \mathbb{1}_n\pi^T\|_\pi = 1$. Then we have

$$\begin{aligned}
\|\Delta^{(k)}\|_\pi^2 &\le 3L^2(\|\mathbf{x}^{(k+1)} - \mathbf{w}^{(k+1)}\|_\pi^2 + \|\mathbf{w}^{(k+1)} - \mathbf{w}^{(k)}\|_\pi^2 + \|\mathbf{x}^{(k)} - \mathbf{w}^{(k)}\|_\pi^2) \\
&\le 3L^2(1+2\beta^2)\|\mathbf{x}^{(k)} - \mathbf{w}^{(k)}\|_\pi^2 + 3\alpha^2L^2\|A_\infty\mathbf{y}^{(k)}\|_\pi^2 + 6\alpha^2L^2\|\mathbf{y}^{(k)} - A_\infty\mathbf{y}^{(k)})\|_\pi^2
\end{aligned} \tag{10}$$

Plug (**??**) into (**??**), we should have

$$\begin{aligned}
&\|\mathbf{y}^{(k+1)} - A_\infty\mathbf{y}^{(k+1)}\|_\pi^2 \\
\le& \frac{\kappa_\pi^2\beta^{k+1}}{1-\beta}(k+2)^2\|\mathbf{y}^{(0)}\|_\pi^2 + \frac{3L^2\kappa_\pi^2(1+2\beta^2)}{1-\beta}\sum_{i=0}^{k}\beta^i(1+k\beta^{k-i})^2\|\mathbf{x}^{(i)} - \mathbf{w}^{(i)}\|_\pi^2 \\
&+ \frac{3\alpha^2L^2\kappa_\pi^2}{1-\beta}\sum_{i=0}^{k}\beta^i(1+k\beta^{k-i})^2\|A_\infty\mathbf{y}^{(i)}\|_\pi^2 + \frac{6\alpha^2L^2\kappa_\pi^2(1+2\beta^2)}{1-\beta}\sum_{i=0}^{k}\beta^i(1+k\beta^{k-i})^2\|\mathbf{y}^{(i)} - A_\infty\mathbf{y}^{(i)}\|_\pi^2
\end{aligned} \tag{11}$$

Sum up (**??**) and we obtain:

$$\begin{aligned}
\sum_{k=0}^{T}\|\mathbf{y}^{(k+1)} - A_\infty\mathbf{y}^{(k+1)}\|_\pi^2 \le& \frac{\kappa_\pi^2(1+\beta)}{(1-\beta)^4}\|\mathbf{y}^{(0)}\|_\pi^2 + \frac{3L^2\kappa_\pi^2(1+2\beta^2)}{1-\beta}(\frac{T}{1-\beta} + D_2 + 1)\sum_{k=0}^{T}\beta^k\|\mathbf{x}^{(k)} - \mathbf{w}^{(k)}\|_\pi^2 \\
+\frac{3\alpha^2L^2\kappa_\pi^2}{1-\beta}(\frac{T}{1-\beta} + D_2 + 1)\sum_{k=0}^{T}\beta^k\|A_\infty\mathbf{y}^{(k)}\|_\pi^2 &+ \frac{6\alpha^2L^2\kappa_\pi^2(1+2\beta^2)}{1-\beta}(\frac{T}{1-\beta} + D_2 + 1)\sum_{k=0}^{T}\beta^k\|\mathbf{y}^{(k)} - A_\infty\mathbf{y}^{(k)}\|_\pi^2
\end{aligned} \tag{12}$$

Here is the detailed calculation process.

Since

$$\begin{aligned}\sum_{k=0}^{T} k(k+1)\beta^k =&\beta(\sum_{k=1}^{T}\beta^{k+1})^{(2)}\\ =&\beta\frac{2-(T+1)(T+2)\beta^T+2T(T+2)\beta^{T+1}-(T+1)(T+2)\beta^{T+2}}{(1-\beta)^3}\\ \leq&\beta\frac{2}{(1-\beta)^3}=D_1\end{aligned} \tag{13}$$

And

$$\sum_{k=0}^{T} k\beta^k =\beta(\sum_{k=1}^{T}\beta^k)^{(1)}=\beta\frac{1-(T+1)\beta^T+T\beta^{T+1}}{(1-\beta)^2}\leq\frac{\beta}{(1-\beta)^2}=D_2 \tag{14}$$

Then

$$\begin{aligned}\sum_{k=0}^{T} k^2\beta^k =&\beta\frac{1+\beta-(T+1)^2\beta^T+(2T^2+2T-1)\beta^{(T+1)}-T(T+3)\beta^{T+2}}{(1-\beta)^3}\\ \leq&\beta\frac{1+\beta}{(1-\beta)^3}=D_3\end{aligned} \tag{15}$$

Then

$$\begin{aligned}&\sum_{k=i}^{T}\beta^i(1+k\beta^{k-i})^2=\sum_{k=i}^{T}(\beta^i+2k\beta^k+k^2\beta^{(2k-i)})\\ =&(T-i+1)\beta^i+\beta^i\sum_{s=0}^{T-i}s\beta^s+\beta^i i\sum_{s=0}^{T-i}\beta^s\\ \leq&\beta^i(T+1-i+D_2+\frac{i}{1-\beta})\leq\beta^i(\frac{T}{1-\beta}+D_2+1)\end{aligned} \tag{16}$$

Therefore, we have Choose $\alpha \leq ?$, then we can move the last term to left and we obtain the GCE lemma □

## 3.2 Estimate of Consensus Error

[LLY: Xinyi, please complete it following the proof in Lemma B.6 in my paper]

**Lemma 2.** *When* $\alpha \leq \frac{(1-\beta)^2}{12L\kappa_\pi(1+2\beta^2)(\frac{T}{1-\beta}+D_2+1)\|I_n-\mathbb{1}_n\pi^T\|_\pi}$ *, we have:*

$$\sum_{k=0}^{T+1}\|\mathbf{w}^{(k)}-\mathbf{x}^{(k)}\|^2\leq\frac{4\alpha^2\|I_n-\mathbb{1}_n\pi^T\|_\pi^2\kappa_\pi^2(1+\beta)}{(1-\beta)^6}\|\mathbf{y}^{(0)}\|_\pi^2+\frac{12\alpha^4L^2\kappa_\pi^2\|I_n-\mathbb{1}_n\pi^T\|_\pi^2}{(1-\beta)^3}(\frac{T}{1-\beta}+D_2+1)\sum_{k=0}^{T}\|A_\infty\mathbf{y}^{(k)}\|_\pi^2 \tag{17}$$

*Proof.* $\mathbf{w}^{(k)}$ or $w^{(k)}$ should be the true model parameter. Left multiply $A_\infty$ on each side of (**??**), we have

$$\mathbf{w}^{(k+1)}=\mathbf{w}^{(k)}-\alpha A_\infty\mathbf{y}^{(k)} \tag{18}$$

To estimate consensus error, we unfold the iteration:

$$\begin{aligned}
\mathbf{x}^{(k+1)} - \mathbf{w}^{(k+1)} &= (A - \mathbb{1}_n \pi^T)(\mathbf{x}^{(k)} - \mathbf{w}^{(k)}) - \alpha(I_n - \mathbb{1}_n \pi^T)(\mathbf{y}^{(k)} - A_\infty \mathbf{y}^{(k)}) \\
&= \dots \\
&= -\alpha \sum_{i=0}^{k} (A - \mathbb{1}_n \pi^T)^i (I_n - \mathbb{1}_n \pi^T)(\mathbf{y}^{(k-i)} - A_\infty \mathbf{y}^{(k-i)})
\end{aligned} \tag{19}$$

Then we can use Jensen inequality

$$\|\mathbf{x}^{(k+1)} - \mathbf{w}^{(k+1)}\|^2 \leq \alpha^2 \sum_{i=0}^{k} \frac{\beta^i}{1-\beta} \|I_n - \mathbb{1}_n \pi^T\|_\pi^2 \|\mathbf{y}^{(k-i)} - A_\infty \mathbf{y}^{(k-i)}\|_\pi^2 \tag{20}$$

Sum up, we have

$$\begin{aligned}
&\sum_{k=0}^{T+1} \|\mathbf{w}^{(k)} - \mathbf{x}^{(k)}\|^2 \leq \alpha^2 \|I_n - \mathbb{1}_n \pi^T\|_\pi^2 \sum_{k=0}^{T} \sum_{i=0}^{k} \frac{\beta^{k-i}}{1-\beta} \|\mathbf{y}^{(i)} - A_\infty \mathbf{y}^{(i)}\|_\pi^2 \\
\leq &\frac{\alpha^2 \|I_n - \mathbb{1}_n \pi^T\|_\pi^2}{(1-\beta)^2} \sum_{k=0}^{T} \|\mathbf{y}^{(i)} - A_\infty \mathbf{y}^{(i)}\|_\pi^2 \\
\leq &\frac{\alpha^2 \|I_n - \mathbb{1}_n \pi^T\|_\pi^2}{(1-\beta)^2} \Big( \frac{2\kappa_\pi^2 (1+\beta)}{(1-\beta)^4} \|\mathbf{y}^{(0)}\|_\pi^2 + \frac{6L^2 \kappa_\pi^2 (1+2\beta^2)}{(1-\beta)} \Big( \frac{T}{1-\beta} + D_2 + 1 \Big) \sum_{k=0}^{T} \|\mathbf{x}^{(k)} - \mathbf{w}^{(k)}\|_\pi^2 \\
+ &\frac{6\alpha^2 L^2 \kappa_\pi^2}{1-\beta} \Big( \frac{T}{1-\beta} + D_2 + 1 \Big) \sum_{k=0}^{T} \|A_\infty \mathbf{y}^{(k)}\|_\pi^2 \Big)
\end{aligned} \tag{21}$$

□

### 3.3 Descent Lemma

**Lemma 3.** *When $\alpha \leq ?$, we have:*

$$f(w^{(k+1)}) \leq f(w^{(k)}) - \frac{\alpha}{4c_{k,1}} \|\pi^T \mathbf{y}^{(k)}\|^2 - \frac{\alpha}{2c_{k,1}} \|\mathbb{1}_n^T \nabla f(\mathbf{w}^k)\|^2 + \frac{\alpha L^2 c_{k,2}}{n c_{k,1}} \|\mathbf{x}^{(k)} - \mathbf{w}^{(k)}\|^2 + \frac{\alpha b^2 c_{k,2}}{c_{k,1}} \tag{22}$$

Define $c_{k,1} = \sum_{i=1}^{n} \frac{\pi_i}{D_{k,i}}, c_{k,2} = \sum_{i=1}^{n} \frac{\pi_i^2}{D_{k,i}^2}$, then we have

$$\begin{aligned}
&\|\pi^T \mathbf{y}^{(k)} - \frac{c_{k,1}}{n} \mathbb{1}_n^T \nabla f(\mathbf{w}^{(k)})\|^2 \\
=&\| \sum_{i=1}^{n} \frac{\pi_i}{D_{k,i}} (\nabla f(x_i^{(k)}) - \frac{1}{n} \mathbb{1}_n^T \nabla f(\mathbf{w}^{(k)}))\|^2 \\
\leq &2\| \sum_{i=1}^{n} \frac{\pi_i}{D_{k,i}} (\nabla f_i(x_i^{(k)}) - \nabla f(x_i^{(k)}))\|^2 + 2\| \sum_{i=1}^{n} \frac{\pi_i}{D_{k,i}} \sum_{j=1}^{n} \frac{1}{n} (\nabla f(x_j^{(k)}) - \nabla f(w_j^{(k)}))\|^2 \\
=&2\| \sum_{i=1}^{n} \frac{\pi_i}{D_{k,i}} (\nabla f_i(x_i^{(k)}) - \nabla f(x_i^{(k)}))\|^2 + \frac{2c_{k,2}}{n^2} \|\mathbb{1}_n^T \left( \nabla f(\mathbf{x}^{(k)}) - \nabla f(\mathbf{w}^{(k)}) \right) \|^2 \\
\leq &2b^2 c_{k,2} + \frac{2L^2 c_{k,2}}{n} \|\mathbf{x}^{(k)} - \mathbf{w}^{(k)}\|^2
\end{aligned} \tag{23}$$

Using L-smoothness inequality, we have

$$
\begin{aligned}
f(w^{(k+1)}) &\leq f(w^{(k)}) - \alpha\langle \pi^T \mathbf{y}^{(k)}, \frac{1}{n}\mathbb{1}_n^T \nabla f(\mathbf{w}^k)\rangle + \frac{\alpha^2 L}{2}\|\pi^T \mathbf{y}^{(k)}\|^2 \\
&= f(w^{(k)}) - \frac{\alpha}{c_{k,1}}\langle \pi^T D_k^{-1}\nabla f(\mathbf{x}^{(k)}), \frac{c_{k,1}}{n}\mathbb{1}_n^T \nabla f(\mathbf{w}^{(k)})\rangle + \frac{\alpha^2 L}{2}\|\pi^T D_k^{-1}\nabla f(\mathbf{x}^{(k)})\|^2 \\
&= f(w^{(k)}) - \frac{\alpha}{2c_{k,1}}\|\pi^T D_k^{-1}\nabla f(\mathbf{x}^{(k)})\|^2 - \frac{\alpha c_{k,1}}{2n^2}\|\mathbb{1}_n^T \nabla f(\mathbf{w}^k)\|^2 \\
&\quad + \frac{\alpha}{2c_{k,1}}\|\pi^T \mathbf{y}^{(k)} - \frac{c_{k,1}}{n}\mathbb{1}_n^T \nabla f(\mathbf{w}^{(k)})\|^2 + \frac{\alpha^2 L}{2}\|\pi^T D_k^{-1}\nabla f(\mathbf{x}^{(k)})\|^2 \\
&\overset{(a)}{\leq} f(w^{(k)}) - \frac{\alpha - c_{k,1}\alpha^2 L}{2c_{k,1}}\|\pi^T D_k^{-1}\nabla f(\mathbf{x}^{(k)})\|^2 - \frac{\alpha c_{k,1}}{2n^2}\|\mathbb{1}_n^T \nabla f(\mathbf{w}^k)\|^2 \\
&\quad + \frac{\alpha}{2c_{k,1}}(2b^2 c_{k,2} + \frac{2L^2 c_{k,2}}{n}\|\mathbf{x}^{(k)} - \mathbf{w}^{(k)}\|^2) \\
&\overset{(b)}{\leq} f(w^{(k)}) - \frac{\alpha}{4c_{k,1}}\|\pi^T \mathbf{y}^{(k)}\|^2 - \frac{\alpha c_{k,1}}{2n^2}\|\mathbb{1}_n^T \nabla f(\mathbf{w}^k)\|^2 + \frac{\alpha L^2 c_{k,2}}{nc_{k,1}}\|\mathbf{x}^{(k)} - \mathbf{w}^{(k)}\|^2 \\
&\quad + \frac{\alpha b^2 c_{k,2}}{c_{k,1}}
\end{aligned} \tag{24}
$$

(a) uses (**??**), (b) holds when $\alpha \leq \dfrac{1}{2c_{k,1}L}$ .

## 3.4 Main Theorem

[LLY: Please complete it following the proof of Theorem 4 in my paper.]

**Theorem 1.** *When*

*Since* $|v_i^{(k)} - \pi_i| \leq \sqrt{\frac{1}{\overline{\pi}}}\beta^k$, *we have* $n\dfrac{\overline{\pi}}{\overline{\pi} + \sqrt{\frac{1}{\overline{\pi}}}\beta} \leq c_{k,i} \leq n$. *And notice that* $c_{k,2} \leq c_{k,1}^2 \leq nc_{k,2}$

$$
\begin{aligned}
\|\mathbb{1}_n^T \nabla f(\mathbf{w}^k)\|^2 \leq& \frac{2n^2}{\alpha c_{k,1}}(f(w^{(k)}) - f(w^{(k+1)})) - \frac{n^2}{2c_{k,1}^2}\|\pi^T \mathbf{y}^{(k)}\|^2 + \frac{2nL^2 c_{k,2}}{c_{k,1}^2}\|\mathbf{x}^{(k)} - \mathbf{w}^{(k)}\|^2 + \frac{2b^2 n^2 c_{k,2}}{c_{k,1}^2} \\
\leq& \frac{2n(\overline{\pi} + \sqrt{\frac{1}{\overline{\pi}}}\beta)}{\alpha\overline{\pi}}(f(w^{(k)}) - f(w^{(k+1)})) - \frac{1}{2}\|\pi^T \mathbf{y}^{(k)}\|^2 + 2nL^2\|\mathbf{x}^{(k)} - \mathbf{w}^{(k)}\|^2 + 2b^2 n^2
\end{aligned} \tag{25}
$$

*It then follows*

$$\begin{aligned}
\sum_{k=0}^{T+1} \|\mathbb{1}_n^T \nabla f(\mathbf{w}^k)\|^2 \leq & \frac{2n(\overline{\pi} + \sqrt{\frac{1}{\pi}}\beta)}{\alpha\overline{\pi}} \triangle + 2nL^2 \sum_{k=0}^{T+1} \|\mathbf{x}^{(k)} - \mathbf{w}^{(k)}\|^2 + 2b^2n^2(T+2) \\
\leq & \frac{2n(\overline{\pi} + \sqrt{\frac{1}{\pi}}\beta)}{\alpha\overline{\pi}} \triangle + 2nL^2 (\frac{4\alpha^2 \|I_n - \mathbb{1}_n \pi^T\|_\pi^2 \kappa_\pi^2 (1+\beta)}{(1-\beta)^6} \|\mathbf{y}^{(0)}\|_\pi^2 + \\
& \frac{12\alpha^4 L^2 \kappa_\pi^2 \|I_n - \mathbb{1}_n \pi^T\|_\pi^2}{(1-\beta)^3} (\frac{T}{1-\beta} + D_2 + 1) \sum_{k=0}^{T+1} \|A_\infty \mathbf{y}^{(k)}\|_\pi^2) + 2b^2n^2(T+2) - \frac{1}{2} \sum_{k=0}^{T+1} \|\pi^T \mathbf{y}^{(k)}\|^2 \\
= & \frac{2n(\overline{\pi} + \sqrt{\frac{1}{\pi}}\beta)}{\alpha\overline{\pi}} \triangle + 2nL^2 \frac{4\alpha^2 \|I_n - \mathbb{1}_n \pi^T\|_\pi^2 \kappa_\pi^2 (1+\beta)}{(1-\beta)^6} \|\mathbf{y}^{(0)}\|_\pi^2 \\
& -(\frac{1}{2} - \frac{12\alpha^4 L^2 \kappa_\pi^2 \|I_n - \mathbb{1}_n \pi^T\|_\pi^2}{(1-\beta)^3} (\frac{T}{1-\beta} + D_2 + 1)) \sum_{k=0}^{T+1} \|\pi^T \mathbf{y}^{(k)}\|^2 + 2b^2n^2(T+2) \\
\leq & \frac{2n(\overline{\pi} + \sqrt{\frac{1}{\pi}}\beta)}{\alpha\overline{\pi}} \triangle + 2nL^2 \frac{4\alpha^2 \|I_n - \mathbb{1}_n \pi^T\|_\pi^2 \kappa_\pi^2 (1+\beta)}{(1-\beta)^6} \|\mathbf{y}^{(0)}\|_\pi^2 + 2b^2n^2(T+2) \qquad (26)
\end{aligned}$$

*Let*

*Then*

$$\frac{\sum_{k=0}^{T+1} \|\mathbb{1}_n^T \nabla f(\mathbf{w}^k)\|^2}{T+2} \leq \frac{2n(\overline{\pi} + \sqrt{\frac{1}{\pi}}\beta)}{\alpha\overline{\pi}(T+2)} \triangle + 2nL^2 \frac{4\alpha^2 \|I_n - \mathbb{1}_n \pi^T\|_\pi^2 \kappa_\pi^2 (1+\beta)}{(1-\beta)^6 (T+2)} \|\mathbf{y}^{(0)}\|_\pi^2 + 2b^2n^2 \qquad (27)$$