Consider that we have the following problem

$$f(x) = \frac{1}{n}\sum_{i=1}^{n} f_i(x) + \frac{\lambda}{2}\|x\|^2. \tag{1}$$

Each i-th node has its own dataset. That's why each f_i has the same form

$$f_i(x) = \frac{1}{m_i}\sum_{j=1}^{m_i} f_{ij}(x) + \frac{\lambda}{2}\|x\|^2. \tag{2}$$

We assume that all of f_{ij} are convex and f_i is μ-strongly convex. Moreover each f_{ij} has H-Lipshitz Hessian. Let's slightly specify each f_{ij}. We assume that

$$f_{ij}(x) = \varphi_{ij}(a_{ij}^{\mathrm{T}} x) \tag{3}$$

and φ_{ij} is third times continuously differentable. We can easily compute Hessian of f_{ij} at point x

$$H_{ij}(x) = \alpha_{ij}(x) a_{ij} a_{ij}^{\mathrm{T}}, \tag{4}$$

where $\alpha_{ij}(x) = \varphi''_{ij}(a_{ij}^{\mathrm{T}} x)$. We approximate Hessian $H_i(x)$ of f_i by

$$B_i(x) = \frac{\max\limits_{j\in[m_i]} \alpha_{ij}(x)}{m_i} \sum_{j=1}^{m_i} a_{ij} a_{ij}^{\mathrm{T}} \tag{5}$$

Since $\max\limits_{j\in[m_i]} \alpha_{ij}(x) \geq \alpha_{ij}(x) \geq 0$ for all j, the $B_i(x) \succeq H_i(x) \succeq \mu I$. For such problem we want to run the following algorithm.

Algorithm 1 Max Coefficient Newton Method

Initialize: Choose starting iterates $x^0 \in \mathbb{R}^d$
  for $k = 0, 1, 2, \ldots$ **do** in parallel
    broadcast x^k to all workers ← **master node**
    for $i = 0, 1, \ldots n$ **do** ← ***i*-th node**
      compute $\alpha_{ij}^k = \varphi''_{ij}(a_{ij}^{\mathrm{T}} x^k)$
      compute $\beta_i^k = \max\limits_{j\in[m_i]} \alpha_{ij}^k$
      broadcast β_i^k to master node
    end for
    $B_i^k = \frac{\beta_i^k}{m_i} \sum\limits_{j=1}^{m_i} a_{ij} a_{ij}^{\mathrm{T}} + \lambda I$
    $x^{k+1} = x^k - \left[\frac{1}{n}\sum\limits_{i=1}^{n} B_i^k\right]^{-1} \left[\frac{1}{n}\sum\limits_{i=1}^{n} \nabla f_i(x^k)\right]$
  end for

Let's work in one node case: $n=1$. Let's denote $H(x)=\frac{1}{n}\sum\limits_{i=1}^{n}\alpha_i(x)a_ia_i^T$, this is the true Hessian and $B(x)=\frac{\max\limits_i \alpha_i(x)}{n}\sum\limits_{i=1}^{n}a_ia_i^T=\frac{\beta(x)}{n}\sum\limits_{i=1}^{n}a_ia_i^T$, this is the estimator of true Hessian. We know from mathematical analyzes the following Theorem of calculus

$$\nabla f(x)-\nabla f(y)=\int\limits_0^1\nabla^2 f(y+\tau(x-y))(x-y)d\tau. \tag{6}$$

Using this and other properties we obtain

$$\begin{aligned}
\|x^{k+1}-x^*\| &= \|x^k-x^*-(B(x^k)+\lambda I)^{-1}[\nabla f(x^k)+\lambda x^k]\| \qquad (7)\\
&= \|(B(x^k)_\lambda I)^{-1}[[B(x^k)+\lambda I](x^k-x^*)-[\nabla f(x^k)+\lambda x^k]]\|\\
&\overset{\mu-\text{conv.}}{\leq} \frac{1}{\mu}\|[B(x^k)+\lambda I](x^k-x^*)-[\nabla f(x^k)+\lambda x^k]\|\\
&= \frac{1}{\mu}\left\|\frac{\beta(x^k)}{n}\sum_{i=1}^{n}a_ia_i^{\mathrm{T}}(x^k-x^*)+\lambda(x^k-x^*)-\frac{1}{n}\sum_{i=1}^{n}[\nabla f_i(x^k)+\lambda x^k]\right\|\\
&\overset{\nabla f(x^*)+\lambda x^*=0}{=} \frac{1}{\mu n}\left\|\beta(x^k)\sum_{i=1}^{n}a_ia_i^{\mathrm{T}}(x^k-x^*)-\sum_{i=1}^{n}(\nabla f_i(x^k)-\nabla f_i(x^*))\right\|\\
&\overset{(6)}{=} \frac{1}{\mu n}\left\|\beta(x^k)\sum_{i=1}^{n}a_ia_i^{\mathrm{T}}(x^k-x^*)-\sum_{i=1}^{n}\int\limits_0^1 H_i(x^*+\tau(x^k-x^*))(x^k-x^*)d\tau\right\|\\
&= \frac{1}{\mu n}\left\|\beta(x^k)\sum_{i=1}^{n}a_ia_i^{\mathrm{T}}(x^k-x^*)-\sum_{i=1}^{n}\int\limits_0^1 \alpha_i(x^*+\tau(x^k-x^*))a_ia_i^{\mathrm{T}}(x^k-x^*)d\tau\right\|\\
&= \frac{1}{\mu n}\left\|\sum_{i=1}^{n}a_ia_i^T(x^k-x^*)\left[\beta(x^k)-\int\limits_0^1\alpha_i[x^*+\tau(x^k-x^*)]d\tau\right]\right\|\\
&\overset{\text{Jensen's ineq.}}{\leq} \frac{\|x^k-x^*\|}{\mu n}\sum_{i=1}^{n}\|a_i\|^2\left|\beta(x^k)-\int\limits_0^1\alpha_i[x^*+\tau(x^k-x^*)]d\tau\right| \qquad (8)
\end{aligned}$$

We know that all f_i has Lipshitz Hessian

$$\begin{aligned}
\|\alpha_i(x)a_ia_i^{\mathrm{T}}-\alpha_i(y)a_ia_i^{\mathrm{T}}\|=|\alpha_i(x)-\alpha_i(y)|\|a_i\|^2 &\leq \mathrm{H}\|x-y\| \qquad (9)\\
|\alpha_i(x)-\alpha_i(y)| &\leq \frac{\mathrm{H}}{\|a_i\|^2}\|x-y\|.
\end{aligned}$$

That' why we know that $\alpha_i(x)$ is Lipshitz function. Let's fix some $\tau_0\in[0,1]$, then we obtain

$$\begin{aligned}
\left|\alpha_i[x^*+\tau_0\cdot(x-x^*)]-\alpha_i[x^*+\tau(x-x^*)]\right| &\leq \frac{\mathrm{H}}{\|a\|^2}\left\|\{x^*+\tau_0(x-x^*)\}-\{x^*+\tau(x-x^*)\}\right\| \qquad (10)\\
&= \frac{\mathrm{H}}{\|a_i\|^2}\|x-x^*\||\tau_0-\tau|
\end{aligned}$$

Now let's use Lagrangian theorem for the difference between α_i as a function of τ and obtain

$$\left|\alpha_i[x^*+\tau_0\cdot(x-x^*)]-\alpha_i[x^*+\tau(x-x^*)]\right| = \left|\alpha_i'[x^*+\tau(x-x^*)]\Big|_{\tau=\tau^*}\right||\tau_0-\tau|, \qquad \tau^*\in[\tau_0,\tau]. \tag{11}$$

Let's use (11) in (10) and obtain

$$\begin{aligned}
\left|\alpha_i'[x^*+\tau(x-x^*)]\Big|_{\tau=\tau^*}\right||\tau_0-\tau| &\leq \frac{\mathrm{H}}{\|a_i\|^2}\|x-x^*\||\tau_0-\tau| \qquad (12)\\
\left|\alpha_i'[x^*+\tau(x-x^*)]\Big|_{\tau=\tau^*}\right| &\leq \frac{\mathrm{H}}{\|a_i\|^2}\|x-x^*\| \qquad (13)
\end{aligned}$$

We see that the right in (13) doesn't depend on τ. Now if $\tau \to \tau_0$ then from continuity of α' we obtain that $\tau^* \to \tau_0$ and

$$\left| \alpha_i'[x^* + \tau(x - x^*)]_{\tau=\tau_0} \right| \leq \frac{\mathrm{H}}{\|a_i\|^2} \|x - x^*\| \tag{14}$$

One can show that for $x = x^*$ this inequality still true because

$$\begin{aligned}
\alpha_i\Big(x^* + \tau(x - x^*)\Big) &= \varphi_i''\Big(a_i^{\mathrm{T}}(x^* + \tau(x - x^*))\Big) &(15)\\
\left[\alpha_i\Big(x^* + \tau(x - x^*)\Big)\right]'\Bigg|_\tau &= \left[\varphi_i''\Big(a_i^{\mathrm{T}}(x^* + \tau(x - x^*))\Big)\right]'\Bigg|_\tau &(16)\\
\left[\alpha_i\Big(x^* + \tau(x - x^*)\Big)\right]'\Bigg|_\tau &= \varphi_i'''\Big(a_i^{\mathrm{T}}(x^* + \tau(x - x^*))\Big) \times \left[a_i^{\mathrm{T}}(x - x^*)\right] &(17)
\end{aligned}$$

That's why if $x = x^*$ inequality (14) is still true. We have that

$$\begin{aligned}
\left|\beta(x^k) - \int_0^1 \alpha_i[x^* + \tau(x^k - x^*)]d\tau\right| &= \left|\beta(x^k) - \alpha_i(x^k) + \alpha_i(x^k) - \int_0^1 \alpha_i[x^* + \tau(x^k - x^*)]d\tau\right| &(18)\\
&\leq \left|\beta(x^k) - \alpha_i(x^k)\right| + \left|\alpha_i(x^k) - \int_0^1 \alpha_i[x^* + \tau(x^k - x^*)]d\tau\right|\\
&= \left|\beta(x^k) - \alpha_i(x^k)\right| + \left|\int_0^1 \Big\{\alpha_i(x^k) - \alpha_i[x^* + \tau(x^k - x^*)]\Big\}\, d\tau\right|\\
&= \left|\beta(x^k) - \alpha_i(x^k)\right| + \left|\int_0^1 \frac{H}{\|a_i\|^2}\|x^k - (x^* + \tau(x^k - x^*))d\tau\right|\\
&= \left|\beta(x^k) - \alpha_i(x^k)\right| + \int_0^1 \frac{H}{\|a_i\|^2}\|x^k - (x^* + \tau(x^k - x^*))\|d\tau\\
&= \left|\beta(x^k) - \alpha_i(x^k)\right| + \frac{H}{2\|a_i\|^2}\|x^k - x^*\|
\end{aligned}$$

It looks like that we can't prove that $\|B^k - H^k\|$ is upper bounded enough. Other way how we can prove is

$$\begin{aligned}
x^{k+1} - x^* &= x^k - x^* - (B^k + \lambda I)^{-1}(\nabla f(x^k) + \lambda x^k) &(19)\\
&= x^k - x^* - (B^k + \lambda I)^{-1} \int_0^1 \Big[\nabla^2 f(x^* + \tau(x^k - x^*)) + \lambda I\Big](x^k - x^*)d\tau\\
&= x^k - x^* - (B^k + \lambda I)^{-1} \underbrace{\int_0^1 \Big[\nabla^2 f(x^* + \tau(x^k - x^*)) + \lambda I\Big]d\tau}_{G^k}(x^k - x^*)\\
&= \Big(I - (B^k + \lambda I)^{-1}G^k\Big)(x^k - x^*) &(20)\\
&= (B^k + \lambda I)^{-1}\left(B^k - \int_0^1 \nabla^2 f(x^* + \tau(x^k - x^*))d\tau\right)(x^k - x^*). &(21)
\end{aligned}$$

0.1 New notes

Let's work with scaled version of Max. Coefficient Method. Hessian estimator is

$$B^k = \frac{1}{n}\max_i\left\{\frac{\alpha_i(x^k)}{\alpha_i(x^*)}\right\}\sum_{i=1}^n \alpha_i(x^*)a_ia_i^{\mathrm{T}} = \frac{\beta(x^k)}{n}\sum_{i=1}^n \alpha_i(x^*)a_ia_i^{\mathrm{T}}. \tag{22}$$

Using this and other properties we obtain

$$\begin{aligned}
\|x^{k+1}-x^*\| &= \|x^k - x^* - (B(x^k)+\lambda I)^{-1}[\nabla f(x^k)+\lambda x^k]\| \\
&= \|(B(x^k)+\lambda I)^{-1}[[B(x^k)+\lambda I](x^k-x^*) - [\nabla f(x^k)+\lambda x^k]]\| \\
&\overset{\lambda-\text{conv.}}{\leq} \frac{1}{\lambda}\|[B(x^k)+\lambda I](x^k-x^*) - [\nabla f(x^k)+\lambda x^k]\| \\
&= \frac{1}{\lambda}\left\|\frac{\beta(x^k)}{n}\sum_{i=1}^n \alpha_i(x^*)a_ia_i^{\mathrm{T}}(x^k-x^*) + \lambda(x^k-x^*) - \frac{1}{n}\sum_{i=1}^n[\nabla f_i(x^k)+\lambda x^k]\right\| \\
&= \left|\nabla f(x^*)+\lambda x^* = 0\right| \\
&= \frac{1}{\lambda n}\left\|\beta(x^k)\sum_{i=1}^n \alpha_i(x^*)a_ia_i^{\mathrm{T}}(x^k-x^*) - \sum_{i=1}^n(\nabla f_i(x^k)-\nabla f_i(x^*))\right\| \\
&\overset{(6)}{=} \frac{1}{\lambda n}\left\|\beta(x^k)\sum_{i=1}^n \alpha_i(x^*)a_ia_i^{\mathrm{T}}(x^k-x^*) - \sum_{i=1}^n\int_0^1 H_i(x^*+\tau(x^k-x^*))(x^k-x^*)d\tau\right\| \\
&= \frac{1}{\lambda n}\left\|\beta(x^k)\sum_{i=1}^n \alpha_i(x^*)a_ia_i^{\mathrm{T}}(x^k-x^*) - \sum_{i=1}^n\alpha_i(x^*)\int_0^1 \frac{\alpha_i(x^*+\tau(x^k-x^*))}{\alpha_i(x^*)}a_ia_i^{\mathrm{T}}(x^k-x^*)d\tau\right\| \\
&= \frac{1}{\lambda n}\left\|\sum_{i=1}^n \alpha_i(x^*)a_ia_i^T(x^k-x^*)\left[\beta(x^k) - \int_0^1\frac{\alpha_i[x^*+\tau(x^k-x^*)]}{\alpha_i(x^*)}d\tau\right]\right\| \\
&\overset{\text{Jensen}}{\leq} \frac{\|x^k-x^*\|}{\lambda n}\sum_{i=1}^n \alpha_i(x^*)\|a_i\|^2\left|\beta(x^k) - \int_0^1\frac{\alpha_i[x^*+\tau(x^k-x^*)]}{\alpha_i(x^*)}d\tau\right|.
\end{aligned} \tag{23}$$

We know that $\alpha_i(x)$ is Lipshitz conituous with constant $\frac{\mathrm{H}}{\|a_i\|^2}$ from (9). That's why we have can use it for $x = x^*+\tau(x^k-x^*)$ and $y = x^*$

$$\alpha_i(x^*) - \frac{\mathrm{H}}{\|a_i\|^2}\tau\|x^k-x^*\| \leq \alpha_i[x^*+\tau(x^k-x^*)] \leq \alpha_i(x^*) + \frac{\mathrm{H}}{\|a_i\|^2}\tau\|x^k-x^*\| \tag{24}$$

$$1 - \frac{\mathrm{H}}{\alpha_i(x^*)\|a_i\|^2}\tau\|x^k-x^*\| \leq \frac{\alpha_i[x^*+\tau(x^k-x^*)]}{\alpha_i(x^*)} \leq 1 + \frac{\mathrm{H}}{\alpha_i(x^*)\|a_i\|^2}\tau\|x^k-x^*\|$$

$$1 - \frac{\mathrm{H}}{2\alpha_i(x^*)\|a_i\|^2}\|x^k-x^*\| \leq \int_0^1\frac{\alpha_i[x^*+\tau(x^k-x^*)]}{\alpha_i(x^*)}d\tau \leq 1 + \frac{\mathrm{H}}{2\alpha_i(x^*)\|a_i\|^2}\|x^k-x^*\|. \tag{25}$$

Let's $j_k = \arg\max_i\left\{\frac{\alpha_i(x^k)}{\alpha_i(x^*)}\right\}$, then we have $\beta(x^k) = \frac{\alpha_{j_k}(x^k)}{\alpha_{j_k}(x^*)}$. We can write for $\beta(x^k)$ the consequences of Lipshitzness

$$\alpha_{j_k}(x^*) - \frac{\mathrm{H}}{\|a_{j_k}\|^2}\|x^k-x^*\| \leq \alpha_{j_k}(x^k) \leq \alpha_{j_k}(x^*) + \frac{\mathrm{H}}{\|a_{j_k}\|^2}\|x^k-x^*\| \tag{26}$$

$$1 - \frac{\mathrm{H}}{\alpha_{j_k}(x^*)\|a_{j_k}\|^2}\|x^k-x^*\| \leq \frac{\alpha_{j_k}(x^k)}{\alpha_{j_k}(x^*)} \leq 1 + \frac{\mathrm{H}}{\alpha_{j_k}(x^*)\|a_{j_k}\|^2}\|x^k-x^*\|$$

$$1 - \frac{\mathrm{H}}{\min_i\{\alpha_i(x^*)\|a_i\|^2\}}\|x^k-x^*\| \leq \beta(x^k) \leq 1 + \frac{\mathrm{H}}{\min_i\{\alpha_i(x^*)\|a_i\|^2\}}\|x^k-x^*\|. \tag{27}$$

Using (25) and (27) we obtain

$$\begin{aligned}
\left(1-\frac{\mathrm{H}}{\min\limits_i\{\alpha_i(x^*)\|a_i\|^2\}}\|x^k-x^*\|\right) &- \left(1+\frac{\mathrm{H}}{2\alpha_i(x^*)\|a_i\|^2}\|x^k-x^*\|\right) \le \beta(x^k) - \int\limits_0^1 \frac{\alpha_i[x^*+\tau(x^k-x^*)]}{\alpha_i(x^*)}d\tau \\
&\le \left(1+\frac{\mathrm{H}}{\min\limits_i\{\alpha_i(x^*)\|a_i\|^2\}}\|x^k-x^*\|\right) - \left(1-\frac{\mathrm{H}}{2\alpha_i(x^*)\|a_i\|^2}\|x^k-x^*\|\right)
\end{aligned}$$

$$\begin{aligned}
-\mathrm{H}\|x^k-x^*\|\left(\frac{1}{\min\limits_i\{\alpha_i(x^*)\|a_i\|^2\}}+\frac{1}{2\alpha_i(x^*)\|a_i\|^2}\right) &\le \beta(x^k) - \int\limits_0^1 \frac{\alpha_i[x^*+\tau(x^k-x^*)]}{\alpha_i(x^*)}d\tau \\
&\le \mathrm{H}\|x^k-x^*\|\left(\frac{1}{\min\limits_i\{\alpha_i(x^*)\|a_i\|^2\}}+\frac{1}{2\alpha_i(x^*)\|a_i\|^2}\right)
\end{aligned}$$

That's why we have that

$$\left|\beta(x^k) - \int\limits_0^1 \frac{\alpha_i[x^*+\tau(x^k-x^*)]}{\alpha_i(x^*)}d\tau\right| \le \mathrm{H}\|x^k-x^*\|\left(\frac{1}{\min\limits_i\{\alpha_i(x^*)\|a_i\|^2\}}+\frac{1}{2\alpha_i(x^*)\|a_i\|^2}\right). \tag{28}$$

0.2 New notes (version 2)

We need to assume the following properties of $B(x)$ and $\nabla^2 f(x)$

1. f has Lipshitz Hessian:

$$
\begin{aligned}
\nabla^2 f(x^*) - \mathrm{H}\tau\|x^k - x^*\| &\leq \nabla^2 f\left[x^* + \tau(x^k - x^*)\right] \leq \nabla^2 f(x^*) + \mathrm{H}\tau\|x^k - x^*\| \qquad (29)\\
\nabla^2 f(x^*) - \frac{\mathrm{H}}{2}\|x^k - x^*\| &\leq \int\limits_0^1 \nabla^2 f\left[x^* + \tau(x^k - x^*)\right] d\tau \leq \nabla^2 f(x^*) + \frac{\mathrm{H}}{2}\|x^k - x^*\|.
\end{aligned}
$$

2. $B(x^*) = \nabla^2 f(x^*)$ and $B(x^k)$ is also M-Lipshitz

$$
\begin{aligned}
B(x^*) - \mathrm{M}\|x^k - x^*\| &\leq B(x^k) \leq B(x^*) + \mathrm{M}\|x^k - x^*\| \qquad (30)\\
\nabla^2 f(x^*) - \mathrm{M}\|x^k - x^*\| &\leq B(x^k) \leq \nabla^2 f(x^*) + \mathrm{M}\|x^k - x^*\|.
\end{aligned}
$$

Combining the above we get

$$
-\left(M + \frac{H}{2}\right)\|x^k - x^*\| \leq B(x^k) - \int\limits_0^1 \nabla^2 f\left[x^* + \tau(x^k - x^*)\right] d\tau \leq \left(M + \frac{H}{2}\right)\|x^k - x^*\|. \qquad (31)
$$

Now we use it in the following inequality

$$
\|x^{k+1} - x^*\| \leq \frac{1}{\lambda}\|x^k - x^*\| \left\| B(x^k) - \int\limits_0^1 \nabla^2 f\left[x^* + \tau(x^k - x^*)\right] d\tau \right\| \qquad (32)
$$

we get

$$
\|x^{k+1} - x^*\| \leq \frac{\mathrm{M} + \mathrm{H}/2}{\lambda}\|x^k - x^*\|^2. \qquad (33)
$$