Algorithm 1 Modification of Newton's method

Initialize: Choose starting iterate $w^0 \in \mathbb{R}^d$, probability p and $x^0 = w^0$

for k = 0, 1, 2,... **do**

$$x^{k+1} = \left[\frac{1}{n}\sum_{i=1}^{n} \nabla^2 f_i(w^k)\right]^{-1} \left[\frac{1}{n}\sum_{i=1}^{n} \nabla^2 f_i(w^k) w^k - \nabla f_i(w^k)\right]$$

$$w^{k+1} = \begin{cases} x^{k+1}, & \text{with probability } p \\ w^k, & \text{with probability } 1-p \end{cases}$$

end for

Lemma 1. Let $\varphi : \mathbb{R}^d \to \mathbb{R}$ has H-Lipschitz Hessian, i. e. $\|\nabla^2\varphi(x) - \nabla^2\varphi(x)\| \leq H\|x-y\|$. Then φ admits for any $x, y \in \mathbb{R}^d$ following estimate

$$\left\|\nabla\varphi(x) - \nabla\varphi(y) - \nabla^2\varphi(y)(x-y)\right\| \leq \frac{H}{2}\|x-y\|^2. \tag{1}$$

Proof. Indeed,

$$\begin{aligned} \left\|\nabla\varphi(x) - \nabla\varphi(y) - \nabla^2\varphi(y)(x-y)\right\| &= \left\|\int_0^1 \left[\nabla^2\varphi(x+\tau(y-x)) - \nabla^2\varphi(x)\right](y-x)d\tau\right\| \\ &\leq H\|x-y\|^2 \int_0^1 \tau d\tau = \frac{H}{2}\|x-y\|^2. \end{aligned}$$

Lemma 2. Let f_i i be μ-strongly convex and have H-Lipschitz Hessian for all $i = 1, \ldots, n$ and consider the following Lyapunov function:

$$\mathcal{W}^k \stackrel{\text{def}}{=} \|w^k - x^*\|^2. \tag{2}$$

Then the iterates of Algorithm 1 satisfy

$$\|x^{k+1} - x^*\| \leq \frac{H}{2\mu}\mathcal{W}^k. \tag{3}$$

Proof. Letting $H^k = \frac{1}{n}\sum_{i=1}^n \nabla^2 f_i(w^k)$, we have

$$x^{k+1} = \left(H^k\right)^{-1}\left[\frac{1}{n}\sum_{i=1}^{n} \nabla^2 f_i(w^k) w^k - \frac{1}{n}\sum_{i=1}^{n} \nabla f_i(w^k)\right]. \tag{4}$$

Subtracting $x^* = \left(H^k\right)^{-1} H^k x^*$ from both sides, and using the fact that $\sum_{i=1}^n \nabla f_i(x^*) = 0$, this further leads to

$$x^{k+1} - x^* = \left(H^k\right)^{-1}\left[\frac{1}{n}\sum_{i=1}^{n} \nabla^2 f_i(w^k)\left(w^k - x^*\right) - \frac{1}{n}\sum_{i=1}^{n}\left(\nabla f_i(w^k) - \nabla f_i(x^*)\right)\right]. \tag{5}$$

Next, note that since f_i is μ-strongly convex, we have $\nabla^2 f_i(w^k) \succeq \mu I$ for all i. It implies that $H^k \succeq \mu I$, which in turns gives

$$\left\|H^k\right\|^{-1} \leq \frac{1}{\mu}. \tag{6}$$

It gives us following inequality

$$
\begin{aligned}
\left\|x^{k+1}-x^*\right\| &\overset{(5)}{\leq} \left\|\left(H^k\right)^{-1}\right\| \left\|\frac{1}{n}\sum_{i=1}^{n}\left[\nabla^2 f_i(w^k)\left(w^k-x^*\right)-\left(\nabla f_i(w^k)-\nabla f_i(x^*)\right)\right]\right\| \\
&\overset{(6)}{\leq} \frac{1}{\mu}\left\|\frac{1}{n}\sum_{i=1}^{n}\left[\nabla^2 f_i(w^k)\left(w^k-x^*\right)-\left(\nabla f_i(w^k)-\nabla f_i(x^*)\right)\right]\right\| \\
&\leq \frac{1}{n\mu}\sum_{i=1}^{n}\left\|\nabla^2 f_i(w^k)\left(w^k-x^*\right)-\left(\nabla f_i(w^k)-\nabla f_i(x^*)\right)\right\| \\
&\overset{(1)}{\leq} \frac{1}{n\mu}\sum_{i=1}^{n}\frac{H}{2}\left\|w^k-x^*\right\|^2 = \frac{H}{2\mu}\left\|w^k-x^*\right\|^2 \leq \beta\left\|w^k-x^*\right\|^2.
\end{aligned}
$$

Lemma 3. Assume that each f_i is μ-strongly convex and has H-Lipschitz Hessian. If $\|w^0-x^*\| \leq \frac{\mu}{H}$, then for all k

$$
\mathcal{W}^k \leq \frac{\mu^2}{H^2}. \tag{7}
$$

Proof. For $k=0$ it follows from the assumption of the lemma. Assume that the statement of the lemma is true for $k \geq 0$, let's prove for $k+1$. If $w^{k+1}=w^k$ with probability $1-p$, then $\mathcal{W}^{k+1}=\mathcal{W}^k$ and the inequality (7) holds by induction assumption for $\mathcal{W}^k$. If $w^{k+1}=x^{k+1}$ with probability p, then

$$
\mathcal{W}^{k+1} = \left\|x^{k+1}-x^*\right\|^2 \overset{(3)}{\leq} \left(\frac{H}{2\mu}\right)^2\left(\mathcal{W}^k\right)^2 \overset{(7)}{\leq} \left(\frac{H\mu^2}{2\mu H^2}\right)^2 \leq \frac{\mu^2}{H^2}. \tag{8}
$$

Lemma 4. The random iterates of Algorithm 1 satusfy the identity

$$
\mathbb{E}_k\left[\mathcal{W}^{k+1}\right] \overset{\text{def}}{=} \mathbb{E}\left[\mathcal{W}^{k+1} \mid x^k\right] = p\mathbb{E}_k\left[\left\|x^{k+1}-x^*\right\|^2\right] + (1-p)\mathcal{W}^k. \tag{9}
$$

Proof. Obviously.

Theorem 1. Assume that every f_i is μ-strongly convex and has H-Lipschitz Hessian. Then for random iterates of Algorithm 1 we have the recursion

$$
\mathbb{E}_k\left[\mathcal{W}^{k+1}\right] \leq \left(1-p+p\left(\frac{H}{2\mu}\right)^2\mathcal{W}^k\right)\mathcal{W}^k. \tag{10}
$$

Furthemore, if $\|w^0-x^*\| \leq \frac{\mu}{H}$, then

$$
\mathbb{E}_k\left[\mathcal{W}^{k+1}\right] \leq \left(1-\frac{3p}{4}\right)\mathcal{W}^k. \tag{11}
$$

Proof. Using lemmas (2), (3) and (4), we obtain

$$
\begin{aligned}
\mathbb{E}_k\left[\mathcal{W}^{k+1}\right] &\overset{(9)}{=} p\mathbb{E}_k\left[\left\|x^{k+1}-x^*\right\|^2\right] + (1-p)\mathcal{W}^k \overset{(3)}{\leq} p\left(\frac{H}{2\mu}\mathcal{W}^k\right)^2 + (1-p)\mathcal{W}^k \\
&= \left(1-p+p\left(\frac{H}{2\mu}\right)^2\mathcal{W}^k\right)\mathcal{W}^k \overset{(7)}{\leq} \left(1-\frac{3p}{4}\right)\mathcal{W}^k.
\end{aligned}
$$