Stochastic setting

Here, we consider the stochastic setting, where each of the functions f_c can be written as

$$f_c(\theta) = \mathbb{E}_{Z^c \sim \xi_c}\left[f_c^{Z^c}(\theta)\right] \quad ,$$

where for each $c \in [M]$, Z^c is a random variable with a certain distribution ξ_c. We assume that each client has access to its own function f_c through stochastic sampling of $f_c^{Z^c}$. In this setting, `FedAVG` solves the global optimization problem by performing local stochastic gradient updates on each client. Starting from an initial point θ_0 shared by the central server, the learning procedure is as follows:

- The server sends the current parameter θ_r to all the clients.
- Starting from θ_r, each client performs H local updates:

  $$\theta^c_{r,h+1} = \theta^c_{r,h} - \eta \nabla f_c^{Z^c_{r,h}}(\theta^c_{r,h}) \quad , \quad \text{for } h = 0, \dots, H-1 \quad ,$$

  where $Z^c_{r,h}$ is sampled from the distribution ξ_c.
- Finally, the clients send to the central server their last iterate $\theta^c_{r,H}$, and the server averages all the iterates before broadcasting the updated global model again to all the agents.

For convenience of notation, given $Z := (Z^1, \dots, Z^M)$, we define the following *virtual* unbiased estimator of $\nabla F(\theta)$:

$$\nabla F^Z(\theta) = \frac{1}{M}\sum_{c=1}^{M} \nabla f_c^{Z^c}(\theta) \quad .$$

We make the following assumption on the noise of the gradient:

FL-4. *There exists $\sigma^2 \geq 0$ such that for every agent $c \in [M]$, the gradient estimator satisfies*

$$\mathbb{E}_{Z^c \sim \xi_c}\left[\|\nabla f_c^{Z^c}(\theta) - \nabla f_c(\theta)\|^2\right] \leq \sigma^2 \quad .$$

Under **FL**-4, it holds that

$$\begin{aligned}\mathbb{E}\left[\|\nabla F^Z(\theta) - \nabla F(\theta)\|_2^2\right] &= \frac{1}{M^2}\sum_{c=1}^{M}\sum_{c'=1}^{M}\mathbb{E}\left[\langle \nabla f_c^{Z^c}(\theta) - \nabla f_c(\theta), \nabla f_{c'}^{Z^{c'}}(\theta) - \nabla f_{c'}(\theta)\rangle\right] \\ &= \frac{1}{M^2}\sum_{c=1}^{M}\mathbb{E}\left[\|\nabla f_c^{Z^c}(\theta) - \nabla f_c(\theta)\|_2^2\right] \leq \frac{\sigma^2}{M} \quad ,\end{aligned} \tag{33}$$

where in the last inequality, we used that for $c \neq c'$, $\mathbb{E}\left[\langle \nabla f_c^{Z^c}(\theta) - \nabla f_c(\theta), \nabla f_{c'}^{Z^{c'}}(\theta) - \nabla f_{c'}(\theta)\rangle\right] = 0$ by the independence of Z^c and $Z^{c'}$ and **FL**-4.

F.1. Convergence of `FedAVG` when $\alpha = 1$

Denote by $F^\star = \frac{1}{M}\sum_{c=1}^{M} f_c^\star$ and where $(f_c^\star)_{c\in[M]}$ are defined in **FL**-2. First, we prove the following lemma, which bounds the local drift

Lemma F.1. *Assume **FL**-1, **FL**-2 with $\alpha = 1$, **FL**-3 and **FL**-4. Then, for any $\eta > 0$ that satisfies $\eta \leq \frac{1}{LH}$, the iterates θ_R of Algorithm `FedAVG` satisfies*

$$\frac{1}{MH}\sum_{c=1}^{M}\sum_{h=1}^{H-1}\mathbb{E}\left[\|\theta_r - \theta^c_{r,h}\|_2^2 \middle| \theta_r\right] \leq 4\eta^2(H-1)^2\|\nabla F(\theta_r)\|_2^2 + 4\eta^2(H-1)^2\zeta^2 + 4\eta^2(H-1)\sigma^2 \quad . \tag{34}$$

Proof. Using the expression of $\theta_{r,h}^c$, and decomposing each gradient as $\nabla f_c^{Z_{r,\ell}^c}(\theta_{r,\ell}^c) = \nabla f_c^{Z_{r,\ell}^c}(\theta_{r,\ell}^c) - \nabla f_c(\theta_{r,\ell}^c) + \nabla f_c(\theta_{r,\ell}^c) - \nabla f_c(\theta_r^c) + \nabla f_c(\theta_r)$ we obtain by Young's inequality

$$
\begin{aligned}
&\frac{1}{H}\sum_{h=1}^{H-1}\mathbb{E}\left[\|\theta_r - \theta_{r,h}^c\|_2^2 \middle| \theta_r\right] = \frac{\eta^2}{H}\sum_{h=1}^{H-1}\mathbb{E}\left[\left\|\sum_{\ell=0}^{h-1}\nabla f_c^{Z_{r,\ell}^c}(\theta_{r,\ell}^c)\right\|_2^2 \middle| \theta_r\right] \\
&\leq \frac{\eta^2}{H}\sum_{h=1}^{H-1}2\mathbb{E}\left[\left\|\sum_{\ell=0}^{h-1}\nabla f_c(\theta_r)\right\|_2^2\middle|\theta_r\right] + 4\mathbb{E}\left[\left\|\sum_{\ell=0}^{h-1}\nabla f_c(\theta_r) - \nabla f_c(\theta_{r,\ell}^c)\right\|_2^2\middle|\theta_r\right] + 4\mathbb{E}\left[\left\|\sum_{\ell=0}^{h-1}\nabla f_c(\theta_{r,\ell}^c) - \nabla f_c^{Z_{r,\ell}^c}(\theta_{r,\ell}^c)\right\|_2^2\middle|\theta_r\right] \\
&\leq \frac{2\eta^2}{H}\sum_{h=1}^{H-1}h^2\mathbb{E}\left[\|\nabla f_c(\theta_r)\|_2^2\middle|\theta_r\right] + \frac{4\eta^2}{H}\sum_{h=1}^{H-1}\mathbb{E}\left[\left\|\sum_{\ell=0}^{h-1}\nabla f_c(\theta_r) - \nabla f_c(\theta_{r,\ell}^c)\right\|_2^2\middle|\theta_r\right] + \frac{4\eta^2}{H}\sum_{h=1}^{H-1}h\sigma^2 \quad ,
\end{aligned}
$$

where we used the fact that $\mathbb{E}\left[\nabla f_c(\theta_{r,\ell}^c) - \nabla f_c^{Z_{r,\ell}^c}(\theta_{r,\ell}^c)\middle|\theta_{r,\ell}^c\right] = 0$ in the last inequality. Using the smoothness of the f_c, Jensen's inequality, and the fact that $\sum_{h=1}^{H-1}h^2 \leq \frac{H(H-1)^2}{2}$, we obtain

$$
\begin{aligned}
\frac{1}{H}\sum_{h=1}^{H-1}\mathbb{E}\left[\|\theta_r - \theta_{r,h}^c\|_2^2\middle|\theta_r\right] &\leq \eta^2(H-1)^2\|\nabla f_c(\theta_r)\|_2^2 + \frac{4\eta^2L^2}{H}\sum_{h=1}^{H-1}\sum_{\ell=0}^{h-1}h\mathbb{E}\left[\|\theta_r - \theta_{r,\ell}^c\|_2^2\middle|\theta_r\right] + 4\eta^2(H-1)\sigma^2 \\
&\leq \eta^2(H-1)^2\|\nabla f_c(\theta_r)\|_2^2 + \frac{2\eta^2H(H-1)L^2}{H}\sum_{h=1}^{H-1}\mathbb{E}\left[\|\theta_r - \theta_{r,h}^c\|_2^2\middle|\theta_r\right] + 2\eta^2(H-1)\sigma^2 \quad ,
\end{aligned}
$$

where the second inequality comes from completing the sum from $\ell = 0$ to h until $\ell = H-1$, and the fact that $\sum_{h=1}^{H-1}h = \frac{H(H-1)}{2}$. Using the fact that $\eta HL \leq 1/2$, we have $2\eta^2H(H-1)L^2 \leq 1/2$. Reorganizing the terms and multiplying the previous inequality by 2, we obtain

$$
\frac{1}{H}\sum_{h=1}^{H-1}\mathbb{E}\left[\|\theta_r - \theta_{r,h}^c\|_2^2\middle|\theta_r\right] \leq 2\eta^2(H-1)^2\|\nabla f_c(\theta_r)\|_2^2 + 4\eta^2(H-1)\sigma^2 \quad .
$$

Averaging this inequality for $c = 1$ to M and using Lemma C.1 to bound $\frac{1}{M}\sum_{c=1}^M\|\nabla f_c(\theta_r)\|_2^2 \leq 2\|\nabla F(\theta_r)\|_2^2 + 2\zeta^2$ gives the result. □

Theorem F.2. *Assume* **FL**-*1,* **FL**-*2 with* $\alpha = 1$*,* **FL**-*3 and* **FL**-*4. Then, for any* $\eta > 0$ *that satisfies* $\eta \leq \frac{1}{18LH}$*, the iterates* θ_R *of Algorithm* `FedAVG` *satisfies:*

$$
\mathbb{E}[F(\theta_R)] - F^\star \leq \left(1 - \frac{\eta H\mu}{4}\right)^R(F(\theta_0) - F^\star) + \frac{8\zeta^2}{\mu} + \frac{16\eta L}{M}\sigma^2 + 12\eta^2(H-1)L^2\sigma^2 \quad .
$$

Case $H = 1$ The algorithm in this setting can be rewritten as stochastic gradient descent on the objective F. Using **FL**-1, setting $Z_r := (Z_{r,0}^1, \ldots, Z_{r,0}^M)$ and taking the conditional expectation over θ_r, we have

$$
\begin{aligned}
\mathbb{E}\left[F(\theta_{r+1})|\theta_r\right] &\leq \mathbb{E}\left[F(\theta_r) + \langle\nabla F(\theta_r), \theta_{r+1} - \theta_r\rangle + \frac{L}{2}\|\theta_{r+1} - \theta_r\|_2^2\middle|\theta_r\right] \\
&= F(\theta_r) - \eta\mathbb{E}\left[\|\nabla F(\theta_r)\|_2^2\middle|\theta_r\right] + \frac{L\eta^2}{2}\mathbb{E}\left[\|\nabla F^{Z_r}(\theta_r)\|_2^2\middle|\theta_r\right] \\
&\leq F(\theta_r) - \eta\|\nabla F(\theta_r)\|_2^2 + L\eta^2\mathbb{E}\left[\|\nabla F^{Z_r}(\theta_r) - \nabla F(\theta_r)\|_2^2\middle|\theta_r\right] + L\eta^2\|\nabla F(\theta_r)\|_2^2 \\
&= F(\theta_r) - \eta(1 - L\eta)\|\nabla F(\theta_r)\|_2^2 + \frac{L\eta^2\sigma^2}{M} \quad , \qquad (35)
\end{aligned}
$$

where in the last inequality we used (33). Using (19), and substracting $F^\star$ from both sides of the inequality yields

$$
\mathbb{E}\left[F(\theta_{r+1})|\theta_r\right] - F^\star \leq (1 - \eta\mu(1 - L\eta))(F(\theta_r) - F^\star) + \eta(1 - L\eta)\zeta^2 + \frac{L\eta^2\sigma^2}{M} \quad .
$$

Since $\eta \leq 1/2L$, taking the expectation with respect to all the stochasticity and expanding the recursion gives

$$\mathbb{E}[F(\theta_r)] - F^\star \leq \left(1 - \frac{\eta\mu}{2}\right)^r (F(\theta_0) - F^\star) + \frac{\zeta^2}{\mu} + \frac{2L\eta\sigma^2}{\mu M} \ .$$

which concludes the proof.

General case Using Lemma D.1, we have

$$F(\theta_{r+1}) \leq F(\theta_r) + \langle \nabla F(\theta_r), \theta_{r+1} - \theta_r \rangle + \frac{L}{2}\|\theta_{r+1} - \theta_r\|_2^2 \ .$$

Let $\beta = \frac{1}{\sqrt{\eta H}}$. Using the polarization identity $2\langle a, b\rangle = \|a+b\|_2^2 - \|a\|_2^2 - \|b\|_2^2$, we get

$$\begin{aligned}
&\mathbb{E}\left[F(\theta_{r+1})|\theta_r\right] - F(\theta_r) \leq \mathbb{E}\left[\langle \beta^{-1}\nabla F(\theta_r), \beta(\theta_{r+1} - \theta_r)\rangle + \frac{L}{2}\|\theta_{r+1} - \theta_r\|_2^2 \Big| \theta_r\right] \\
&\leq \langle \beta^{-1}\nabla F(\theta_r), \beta\mathbb{E}\left[\theta_{r+1} - \theta_r|\theta_r\right]\rangle + \frac{L}{2}\mathbb{E}\left[\|\theta_{r+1} - \theta_r\|_2^2 \big|\theta_r\right] \\
&= \frac{1}{2}\left(\|\beta^{-1}\nabla F(\theta_r) - \beta\mathbb{E}\left[\theta_r - \theta_{r+1}|\theta_r\right]\|_2^2 - \|\beta^{-1}\nabla F(\theta_r)\|_2^2 - \|\beta\mathbb{E}\left[\theta_r - \theta_{r+1}|\theta_r\right]\|_2^2\right) + \frac{L}{2}\mathbb{E}\left[\|\theta_{r+1} - \theta_r\|_2^2\big|\theta_r\right] \\
&= \underbrace{\frac{1}{2\beta^2}\|\nabla F(\theta_r) - \beta^2\mathbb{E}\left[\theta_r - \theta_{r+1}|\theta_r\right]\|_2^2}_{(\mathbf{A})} - \frac{1}{2\beta^2}\|\nabla F(\theta_r)\|_2^2 + \underbrace{\frac{L}{2}\mathbb{E}\left[\|\theta_{r+1} - \theta_r\|_2^2\big|\theta_r\right] - \frac{\beta^2}{2}\|\mathbb{E}\left[\theta_{r+1} - \theta_r|\theta_r\right]\|_2^2}_{(\mathbf{B})} \ . \quad (36)
\end{aligned}$$

Bounding $(\mathbf{A})$. Using the fact that $F = \frac{1}{M}\sum_{c=1}^M f_c$, the definition $\beta^2 = 1/\eta H$, the definition of θ_{r+1} and Jensen's inequality, we have

$$\begin{aligned}
\left\|\nabla F(\theta_r) - \beta^2\mathbb{E}\left[\theta_r - \theta_{r+1}|\theta_r\right]\right\|_2^2 &= \left\|\mathbb{E}\left[\frac{1}{M}\sum_{c=1}^M\left(\nabla F(\theta_r) - \frac{1}{H}\sum_{h=0}^{H-1}\nabla f_c^{Z_{r,h}^c}(\theta_{r,h}^c)\right)\Big|\theta_r\right]\right\|_2^2 \\
&\leq \frac{1}{HM}\sum_{c=1}^M\sum_{h=0}^{H-1}\left\|\mathbb{E}\left[\nabla f_c(\theta_r) - \nabla f_c^{Z_{r,h}^c}(\theta_{r,h}^c)\Big|\theta_r\right]\right\|_2^2 \ .
\end{aligned}$$

By independence of $Z_{r,h}^c$ and $\theta_{r,h}^c$, and using Jensen's inequality and the smoothness of the f_c (**FL**-1), we obtain

$$\begin{aligned}
\left\|\nabla F(\theta_r) - \beta^2\mathbb{E}\left[\theta_r - \theta_{r+1}|\theta_r\right]\right\|_2^2 &\leq \frac{1}{HM}\sum_{c=1}^M\sum_{h=0}^{H-1}\left\|\mathbb{E}\left[\nabla f_c(\theta_r) - \nabla f_c(\theta_{r,h}^c)\big|\theta_r\right]\right\|_2^2 \\
&\leq \frac{1}{HM}\sum_{c=1}^M\sum_{h=0}^{H-1}\mathbb{E}\left[\|\nabla f_c(\theta_r) - \nabla f_c(\theta_{r,h}^c)\|_2^2\big|\theta_r\right] \\
&\leq \frac{L^2}{HM}\sum_{c=1}^M\sum_{h=0}^{H-1}\mathbb{E}\left[\|\theta_r - \theta_{r,h}^c\|_2^2\big|\theta_r\right] \ . \quad (37)
\end{aligned}$$

Using Lemma F.1, we obtain

$$\left\|\nabla F(\theta_r) - \beta^2\mathbb{E}\left[\theta_r - \theta_{r+1}|\theta_r\right]\right\|_2^2 \leq 4\eta^2(H-1)^2L^2\|\nabla F(\theta_r)\|_2^2 + 4\eta^2(H-1)^2L^2\zeta^2 + 4\eta^2(H-1)L^2\sigma^2 \ . \quad (38)$$

Multinplying by $1/(2\beta^2) = \eta H/2$, we obtain the following bound on $(\mathbf{A})$

$$(\mathbf{A}) \leq 2\eta^3 H(H-1)^2L^2\|\nabla F(\theta_r)\|_2^2 + 2\eta^3H(H-1)^2L^2\zeta^2 + 2\eta^3H(H-1)L^2\sigma^2 \ .$$

Bounding $(\mathbf{B})$. To bound this second term, we use the following decomposition

$$\begin{aligned}
&\frac{L}{2}\mathbb{E}\left[\|\theta_{r+1}-\theta_r\|_2^2\middle|\theta_r\right]-\frac{\beta^2}{2}\|\mathbb{E}\left[\theta_{r+1}-\theta_r|\theta_r\right]\|_2^2\\
&=\frac{L}{2}\mathbb{E}\left[\|\mathbb{E}\left[\theta_{r+1}|\theta_r\right]-\theta_{r+1}\|^2\middle|\theta_r\right]+\frac{L}{2}\|\mathbb{E}\left[\theta_{r+1}-\theta_r|\theta_r\right]\|_2^2-\frac{\beta^2}{2}\|\mathbb{E}\left[\theta_{r+1}-\theta_r|\theta_r\right]\|_2^2\\
&=\frac{L}{2}\mathbb{E}\left[\|\mathbb{E}\left[\theta_{r+1}|\theta_r\right]-\theta_{r+1}\|^2\middle|\theta_r\right]+\left(\frac{L}{2}-\frac{\beta^2}{2}\right)\|\mathbb{E}\left[\theta_{r+1}-\theta_r|\theta_r\right]\|_2^2\ .
\end{aligned}$$

Since $\eta HL \leq 1$, we have $\frac{L}{2}-\frac{\beta^2}{2}=\frac{L}{2}-\frac{1}{2\eta H}\leq 0$, and the second term is negative. To bound the first term, we write

$$\begin{aligned}
&\mathbb{E}\left[\|\mathbb{E}\left[\theta_{r+1}|\theta_r\right]-\theta_{r+1}\|^2\middle|\theta_r\right]=\mathbb{E}\left[\left\|\frac{\eta}{M}\sum_{c=1}^{M}\sum_{h=1}^{H}\nabla f_c^{Z^c_{r,h}}(\theta^c_{r,h})-\mathbb{E}\left[\nabla f_c(\theta^c_{r,h})\middle|\theta_r\right]\right\|^2\middle|\theta_r\right]\\
&=\mathbb{E}\left[\left\|\frac{\eta}{M}\sum_{c=1}^{M}\sum_{h=1}^{H}\nabla f_c^{Z^c_{r,h}}(\theta^c_{r,h})-\nabla f_c(\theta_r)+\frac{\eta}{M}\sum_{c=1}^{M}\sum_{h=1}^{H}\nabla f_c(\theta_r)-\mathbb{E}\left[\nabla f_c(\theta^c_{r,h})\middle|\theta_r\right]\right\|^2\middle|\theta_r\right]\\
&\leq 2\mathbb{E}\left[\left\|\frac{\eta}{M}\sum_{c=1}^{M}\sum_{h=1}^{H}\nabla f_c^{Z^c_{r,h}}(\theta^c_{r,h})-\nabla f_c(\theta_r)\right\|^2\middle|\theta_r\right]+2\mathbb{E}\left[\left\|\frac{\eta}{M}\sum_{c=1}^{M}\sum_{h=1}^{H}\nabla f_c(\theta_r)-\mathbb{E}\left[\nabla f_c(\theta^c_{r,h})\middle|\theta_r\right]\right\|^2\middle|\theta_r\right]\\
&\leq 4\mathbb{E}\left[\left\|\frac{\eta}{M}\sum_{c=1}^{M}\sum_{h=1}^{H}\nabla f_c^{Z^c_{r,h}}(\theta^c_{r,h})-\nabla f_c(\theta_r)\right\|^2\middle|\theta_r\right]\ ,
\end{aligned}$$

where we used Young's and Jensen's inequalities. Now, using Young's inequality again, we obtain

$$\begin{aligned}
&\mathbb{E}\left[\|\mathbb{E}\left[\theta_{r+1}|\theta_r\right]-\theta_{r+1}\|^2\middle|\theta_r\right]\\
&\leq 4\mathbb{E}\left[\left\|\frac{\eta}{M}\sum_{c=1}^{M}\sum_{h=1}^{H}\nabla f_c^{Z^c_{r,h}}(\theta^c_{r,h})-\nabla f_c(\theta^c_{r,h})+\frac{\eta}{M}\sum_{c=1}^{M}\sum_{h=1}^{H}\nabla f_c(\theta^c_{r,h})-\nabla f_c(\theta_r)\right\|^2\middle|\theta_r\right]\\
&\leq 8\mathbb{E}\left[\left\|\frac{\eta}{M}\sum_{c=1}^{M}\sum_{h=1}^{H}\nabla f_c^{Z^c_{r,h}}(\theta^c_{r,h})-\nabla f_c(\theta^c_{r,h})\right\|^2\middle|\theta_r\right]+8\mathbb{E}\left[\left\|\frac{\eta}{M}\sum_{c=1}^{M}\sum_{h=1}^{H}\nabla f_c(\theta^c_{r,h})-\nabla f_c(\theta_r)\right\|^2\middle|\theta_r\right]\ .
\end{aligned}$$

The first term is a variance term, that we bound using the fact that the $\theta^c_{r,h}$ are independent from the future noise draws $Z^c_{r,h'}$ for $h' \geq h$, and the fact that the $Z^c_{r,h}$ are independent from an agent to another, and bounding each gradient variance using **FL**-4. We bound the second term by decomposing it using Jensen's inequality and the smoothness of the f_c. This gives

$$\mathbb{E}\left[\|\mathbb{E}\left[\theta_{r+1}|\theta_r\right]-\theta_{r+1}\|^2\middle|\theta_r\right]\leq\frac{8\eta^2H}{M}\sigma^2+\frac{8\eta^2L^2H}{M}\sum_{c=1}^{M}\sum_{h=1}^{H}\mathbb{E}\left[\|\theta^c_{r,h}-\theta_r\|^2\middle|\theta_r\right]\ .$$

Using Lemma F.1, we obtain

$$\begin{aligned}
(\mathbf{B})&\leq\frac{4\eta^2HL}{M}\sigma^2+4\eta^2H^2L^3\left(4\eta^2(H-1)^2\|\nabla F(\theta_r)\|_2^2+4\eta^2(H-1)^2\zeta^2+4\eta^2(H-1)\sigma^2\right)\\
&=16\eta^4H^2(H-1)^2L^3\|\nabla F(\theta_r)\|_2^2+16\eta^4H^2(H-1)^2L^3\zeta+\left(\frac{4\eta^2HL}{M}+16\eta^4H^2(H-1)L^3\right)\sigma^2\ .
\end{aligned}$$

Bound on (36). Plugging in the bounds on $(\mathbf{A})$ and $(\mathbf{B})$ in (36) yields

$$\begin{aligned}
&\mathbb{E}\left[F(\theta_{r+1})|\theta_r\right]-F(\theta_r)\\
&=\left(2\eta^3H(H-1)^2L^2+16\eta^4H^2(H-1)^2L^3-\frac{\eta H}{2}\right)\|\nabla F(\theta_r)\|_2^2+\left(2\eta^3H(H-1)^2L^2+16\eta^4H^2(H-1)^2L^3\right)\zeta^2\\
&\quad+\left(\frac{4\eta^2HL}{M}+2\eta^3H(H-1)L^2+16\eta^4H^2(H-1)L^3\right)\sigma^2\ .
\end{aligned}$$

Using that $\eta H L \le 1/18$, it holds that

$$\mathbb{E}\left[F(\theta_{r+1})|\theta_r\right] \le F(\theta_r) - \frac{\eta H}{4}\|\nabla F(\theta_r)\|_2^2 + \eta H \zeta^2 + \frac{4\eta^2 H L}{M}\sigma^2 + 3\eta^3 H(H-1)L^2\sigma^2 \ . \tag{39}$$

Applying (19), we get

$$\mathbb{E}\left[F(\theta_{r+1})|\theta_r\right] - F^\star \le F(\theta_r) - F^\star - \frac{\eta H \mu}{4}(F(\theta_r) - F^\star) + 2\eta H \zeta^2 + \frac{4\eta^2 H L}{M}\sigma^2 + 3\eta^3 H(H-1)L^2\sigma^2 \ .$$

The result follows from taking the expectation and unrolling the recursion.

F.2. Convergence of `FedAVG` for general $1 < \alpha \le 2$

Lemma F.3. *Assume **FL**-1, **FL**-2 with $\alpha > 1$, **FL**-3 and **FL**-4. For any $\eta > 0$ that satisfies $\eta \le 1/L$, the following inequality holds on the last iterate provided by `FedAVG` with $H = 1$*

$$\mathbb{E}[F(\theta_R)] - F^\star \le \frac{F(\theta_0) - F^\star}{(\eta\mu R(\alpha-1)(F(\theta_0) - F^\star)^{\alpha-1}/4 + 1)^{1/(\alpha-1)}} + 2\left(\frac{\zeta^2}{\mu}\right)^{1/\alpha} + 2\left(\frac{(2L\eta\sigma^2}{M\mu}\right)^{1/\alpha} \ ,$$

*where $F^\star = \frac{1}{M}\sum_{c=1}^{M} f_c^\star$ and where $(f_c^\star)_{c\in[M]}$ are defined in **FL**-2.*

Proof. Starting from (35), we have

$$\mathbb{E}\left[F(\theta_{r+1})|\theta_r\right] \le F(\theta_r) - \eta(1 - L\eta)\|\nabla F(\theta_r)\|_2^2 + \frac{L\eta^2\sigma^2}{M} \ .$$

Now, using (19), subtracting $F^\star$ from both sides of the inequality yields, and using that $\eta \le 1/L$, we have,

$$\mathbb{E}\left[F(\theta_{r+1})|\theta_r\right] - F^\star \le F(\theta_r) - F^\star - \frac{\eta\mu}{2}(F(\theta_r) - F^\star)^\alpha + \frac{\eta}{2}\zeta^2 + \frac{L\eta^2\sigma^2}{M} \ . \tag{40}$$

Taking the expectation with respect to all the stochasticity and applying Jensen's inequality gives

$$\mathbb{E}[F(\theta_{r+1})] - F^\star \le \mathbb{E}[F(\theta_r)] - F^\star - \frac{\eta\mu}{2}(\mathbb{E}[F(\theta_r)] - F^\star)^\alpha + \frac{\eta}{2}\zeta^2 + \frac{L\eta^2\sigma^2}{M}$$

Defining $s_r = \mathbb{E}[F(\theta_r)] - F^\star$, the precedent expression can be rewritten as

$$s_{r+1} \le s_r - \frac{\eta\mu}{2}s_r^\alpha + \frac{\eta}{2}\zeta^2 + \frac{L\eta^2\sigma^2}{M} \ .$$

This expression can be interpreted as a difference inequality corresponding to an Euler discretization of Bernoulli's differential equation. To solve it, we first homogenize the recursive relation by introducing the sequence $v_r = s_r - C$, where $C = \left((\zeta^2)/\mu\right)^{1/\alpha} + \left((2L\eta\sigma^2)/(M\mu)\right)^{1/\alpha}$. The sequence v_r then satisfies the following recursive relation:

$$v_{r+1} \le v_r - \frac{\eta\mu}{2}(v_r + C)^\alpha + \eta\zeta^2 \ . \tag{41}$$

We now consider the case where $s_r \ge C$ for all $r \in [R]$ which implies $v_r \ge 0$ for all $r \in [R]$. Since for $\alpha \ge 1$, and $a, b \ge 0$, $(a+b)^\alpha \ge a^\alpha + b^\alpha$, we get

$$v_{r+1} \le v_r - \frac{\eta\mu}{2}v_r^\alpha - \frac{\eta\mu C^\alpha}{2} + \eta\zeta^2 = v_r - \kappa v_r^\alpha \ ,$$

with $\kappa = \frac{\eta\mu}{2}$. Dividing this inequality by v_r^α yields

$$\frac{v_{r+1} - v_r}{v_r^\alpha} \le -\kappa \ . \tag{42}$$

For $x > 0$, define $g(x) = x^{-(\alpha-1)}$. By convexity of g on $\mathbb{R}_+^\star$, we have $g(v_{r+1}) \geq g(v_r) + (v_{r+1} - v_r)g'(v_r)$ which can be rewritten as

$$v_{r+1}^{-(\alpha-1)} \geq v_r^{-(\alpha-1)} + (v_{r+1} - v_r)\,\frac{1-\alpha}{v_r^\alpha} \quad ,$$

and which implies, after dividing by $1 - \alpha < 0$, and using (42)

$$\frac{v_{r+1}^{-(\alpha-1)} - v_r^{-(\alpha-1)}}{1-\alpha} \leq \frac{v_{r+1} - v_r}{v_r^\alpha} \leq -\kappa \ .$$

Summing up both sides over $r = 0 \ldots R-1$ and rearranging the terms yields

$$(s_R - C)^{-(\alpha-1)} \geq \kappa R(\alpha - 1) + s_0^{-(\alpha-1)} \ .$$

Finally, we get

$$s_R \leq C + \left\{\kappa R(\alpha-1) + s_0^{-(\alpha-1)}\right\}^{-1/(\alpha-1)} = C + \frac{s_0}{\left(\kappa R(\alpha-1)s_0^{\alpha-1} + 1\right)^{1/(\alpha-1)}} \ .$$

Now in the case where there exists s_r such that $s_r \leq C$ it is straightforward to see the sequence s_r will stay smaller than $2C$. □

Theorem F.4. *Assume **FL**-1, **FL**-2 with $\alpha > 1$, and **FL**-3. Then, for any $\eta > 0$ that satisfies $\eta \leq \frac{1}{18MLH}$, the iterates θ_R of Algorithm `FedAVG` satisfies:*

$$\mathbb{E}[F(\theta_R)] - F^\star \leq \frac{F(\theta_0) - F^\star}{1 + R^{1/(\alpha-1)} \cdot (F(\theta_0) - F^\star) \cdot (\eta H \mu(\alpha-1)/4)^{1/(\alpha-1)}} + 2\left(\frac{8\zeta^2}{\mu}\right)^{1/\alpha} + 2\left(\frac{16L\eta\sigma^2}{M\mu}\right)^{1/\alpha} \quad ,$$

*where $F^\star = \frac{1}{M}\sum_{c=1}^M f_c^\star$ and where $(f_c^\star)_{c\in[M]}$ are defined in **FL**-2.*

Proof. Let us follow the first steps of the proof of Theorem F.2. Starting from (39) and applying (19) yields

$$\mathbb{E}\left[F(\theta_{r+1})|\theta_r\right] \leq F(\theta_r) - \frac{\eta H}{4}\left(F(\theta_r) - F^\star\right)\|\nabla F(\theta_r)\|_2^2 + 2\eta H\zeta^2 + \frac{4\eta^2 HL}{M}\sigma^2 + 3\eta^3 H(H-1)L^2\sigma^2 \ .$$

We recognise the same type of recursion as in (40) of Lemma F.3. Similarly, setting $C = \left(8\zeta^2/\mu\right)^{1/\alpha} + \left(16L\eta\sigma^2/M\mu\right)^{1/\alpha}$ and $\kappa = \eta H\mu/4$, we obtain

$$\mathbb{E}\left[F(\theta_{r+1})|\theta_r\right] - F^\star \leq C + \frac{F(\theta_0) - F^\star}{\left(\kappa R(\alpha-1)(F(\theta_0) - F^\star)^{\alpha-1} + 1\right)^{1/(\alpha-1)}} \ .$$

Finally using that for $\alpha \geq 1$, and $a, b \geq 0$, $(a+b)^\alpha \geq a^\alpha + b^\alpha$ concludes the proof. □