$$f(\theta) = \frac{1}{n}\sum_{i=1}^{n} f_i(\theta) \rightarrow \min_{\theta \in \mathbb{R}^p}$$

$$\theta_{k+1} = \theta_k - \alpha_k \nabla f(\theta_k) \quad \text{GD}$$

$$\theta_{k+1} = \theta_k - \alpha_k \underbrace{\nabla f_i(\theta_k)} \quad \text{SGD}$$

$$\text{ODE:} \quad \underbrace{\frac{\partial \theta}{\partial t} = -\nabla f(\theta)}$$

$$\frac{\partial \theta}{\partial t} = \left[-\nabla f_1(\theta) - \nabla f_2(\theta) \right] \cdot \frac{1}{2}$$

1) $$\frac{\partial \theta}{\partial t} = -\nabla f_1(\theta)$$

2) $$\frac{\partial \theta}{\partial t} = -\nabla f_2(\theta)$$

I order splitting scheme

$$\theta_I(t) - \theta(t) = \frac{t^2}{2}[g_1, g_2]\theta_0 + O(t^3)$$

$$[g_1, g_2] = \frac{\partial g_1}{\partial \theta} g_2 - \frac{\partial g_2}{\partial \theta} g_1$$

1

$$\text{I}: \quad \Theta_{\text{I}}(h) = \Theta_2(h) \circ \Theta_1(h) \circ \Theta_0$$

$$\text{II}: \quad \Theta_{\text{II}}(h) = \Theta_1\left(\frac{h}{2}\right) \circ \Theta_2(h) \circ \Theta_1\left(\frac{h}{2}\right) \circ \Theta_0$$

$$\frac{d\Theta}{dt} = -g_1(\Theta) - g_2(\Theta)$$

$$\Theta_{II}(t) - \Theta(t) = t^3\left(\frac{1}{12}[g_2,[g_2,g_1]] - \frac{1}{24}[g_1,[g_1,g_2]]\right)\Theta_0 + O(t^4)$$

SGD:

$$\theta_{k+1} = \theta_k - \alpha\, g_1(\theta_k)$$

$$\theta_{k+2} = \theta_{k+1} - \alpha\, g_2(\theta_{k+1})$$

T

order

epoch1

epoch2

$$\tilde{\theta}_{k+1} = \theta_k - \alpha g_1(\theta_k)$$
$$\tilde{\theta}_{k+2} = \theta_{k+1} - \alpha g_2(\tilde{\theta}_{k+1})$$
$$\tilde{\theta}_{k+3} = \theta_{k+2} - \alpha g_2(\tilde{\theta}_{k+2})$$
$$\tilde{\theta}_{k+4} = \theta_{k+3} - \alpha g_1(\tilde{\theta}_{k+3})$$

$$\frac{d\theta}{dt} = -g_1(\theta) - g_0(\theta))$$

Θ^* - steady state $\left(\frac{d\Theta^*}{dt} = 0 \right)$

$$\frac{d\theta}{dt} = -g_1(\theta)$$

$\longrightarrow g_1(\Theta^*) + g_2(\theta^*) = 0$

$$\frac{d\theta}{dt} = -g_1 + c - g_2 - c$$

$$c = \frac{1}{2}\left[g_2(\theta^*) - g_1(\theta^*)\right]$$

$$\boxed{\hat{g}_1 = g_1 - c} = $$

$$= g_1 - \frac{1}{2}\left[g_2(\theta^*) - g_1(\theta^*)\right]$$

$$\hat{g}_1(\theta^*) = g_1(\theta^*) + \frac{1}{2}\left[g_2(\theta^*) - g_1(\theta^*)\right] =$$

$$= \frac{g_1(\theta^*) + g_2(\theta^*)}{2} = 0$$

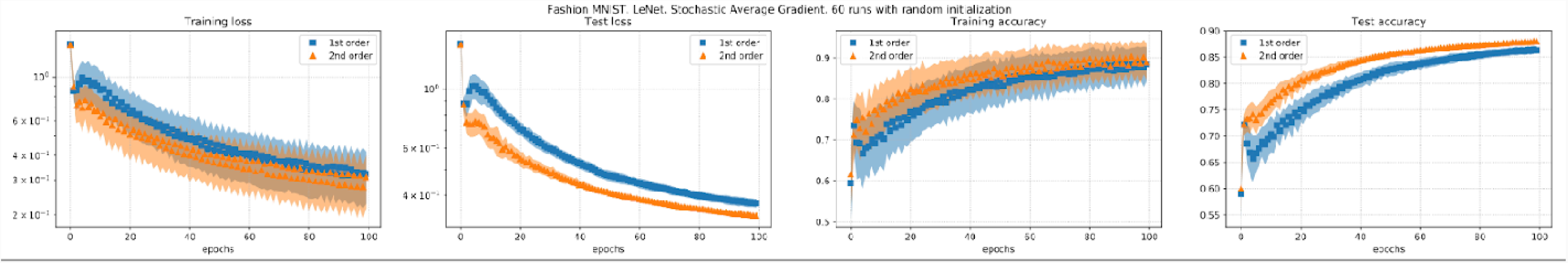
SAG

$$\frac{1}{n}\sum_{i=1}^{n} f_i(\theta) \rightarrow \min$$

SGD	SAG
$\theta_{k+1} = \theta_k - \alpha f_i'(\theta)$	$\theta_{k+1} = \theta_k - \alpha \frac{1}{n}\sum_{i=1}^{n} f_i'(\theta)$

SAG vs SAG2

SAG	SAG2
1 - - - - - n	1 - - - - n
1 - - . . . n	n - - - - . 1


Fashion MNIST. LeNet. Stochastic Average Gradient. 60 runs with random initialization
Training loss
Test loss
Training accuracy
Test accuracy
1st order
2nd order
epochs