SGD with momentum:

$$v_t = \alpha v_{t-1} + \epsilon g_t$$

$$\Delta\theta_t = -v_t$$

$$v_{t+1} = \alpha(\alpha v_{t-1} + \epsilon g_t) + \epsilon g_{t+1}$$

$$v_{t+1} = \alpha^2 v_{t-1} + \alpha\epsilon g_t + \epsilon g_{t+1}$$

$$\Delta\theta_{t+1} = -v_{t+1}$$

$$\Delta\theta_{t+1} = -\alpha^2 v_{t-1} - \alpha\epsilon g_t - \epsilon g_{t+1}$$

SGD with running avrage:

$$v_t = \beta v_{t-1} + (1-\beta) g_t$$

$$\Delta\theta_t = -\delta v_t$$

$$v_{t+1} = \beta v_t + (1-\beta) g_{t+1}$$

$$v_{t+1} = \beta(\beta v_{t-1} + (1-\beta) g_t) + (1-\beta) g_{t+1}$$

$$\Delta\theta_{t+1} = -\delta v_{t+1}$$

$$\Delta\theta_{t+1} = -\delta\beta^2 v_{t-1} - \delta\beta(\beta-1) g_t - \delta(1-\beta) g_{t+1}$$

using both results:

We can write:

$$-\alpha^2 v_{t-1} - \alpha\epsilon g_t - \epsilon g_{t-1} = -\delta\beta^2 v_{t-1} - \delta\beta(\beta-1) g_t - \delta(1-\beta) g_{t+1}$$

$\rightarrow \delta\beta^2 = \alpha^2$

$\rightarrow \delta\beta(\beta-1) = \alpha\epsilon$

$\rightarrow \epsilon = \delta(1-\beta)$