1-a) Stochastic Gradient Descent with Momentum

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

$\Delta\theta_t = -v_t$

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

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

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

1-b) Stochastic Gradient Descent with Running Average

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

$\Delta\theta_t = -\delta v_t$

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

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

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

$(a) = (b)$

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

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

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

2) V In Terms of g

$v_1 = \alpha v_0 + \epsilon g_1$

$v_2 = \alpha(\alpha v_0 + \epsilon g_1) + \epsilon g_2 = \alpha^2 v_0 + \alpha\epsilon g_1 + \epsilon g_2$

$v_3 = \alpha[\alpha^2 v_0 + \alpha\epsilon g_1 + \epsilon g_2] + \epsilon g_3 = \alpha^3 v_0 + \alpha^2\epsilon g_1 + \alpha\epsilon g_2 + \epsilon g_3$

.

.

$v_t = \alpha^t v_0 + \sum_{i=1}^{t} \epsilon\alpha^{t-i} g_i$