1-a) Stochastic Gradient Descent with Momentum

$$egin{aligned} v_t &= lpha v_{t-1} + \epsilon g_t \ &\Delta heta_t = -v_t \ &v_{t+1} &= lpha v_t + \epsilon g_{t+1} = lpha (lpha v_{t-1} + \epsilon g_t) + \epsilon g_{t+1} = lpha^2 v_{t-1} + lpha \epsilon g_t + \epsilon g_{t+1} \ &\Delta heta_{t+1} = -v_{t+1} \ &\Delta heta_{t+1} = -lpha^2 v_{t-1} - lpha \epsilon g_t - \epsilon g_{t+1} \ (a) \end{aligned}$$

1-b) Stochastic Gradient Descent with Running Average

$$egin{aligned} v_t &= eta v_{t-1} + (1-eta) g_t \ &\Delta heta_t = -\delta v_t \ &v_{t+1} &= eta v_t + (1-eta) g_{t+1} = eta (eta v_{t-1} + (1-eta) g_t) + (1-eta) g_{t+1} \ &\Delta heta_{t+1} = -\delta v_{t+1} \ &\Delta heta_{t+1} = -\delta eta^2 v_{t-1} - \delta eta (1-eta) g_t - \delta (1-eta) g_{t+1} \ (b) \end{aligned}$$

$$egin{aligned} (a) &= (b) \ &lpha^2 &= \deltaeta^2 \ &lpha\epsilon &= \deltaeta(1-eta) \ &\epsilon &= \delta(1-eta) \end{aligned}$$

2) V In Terms of g