SGD:

$$v_t = \mu \cdot v_{t-1} - \eta \cdot d_x$$
$$x_t = x_{t-1} + v_t$$
$$x_t = x_{t-1} + \mu \cdot v_{t-1} - \eta \cdot d_x$$

NAG:

$$v_t = \mu \cdot v_{t-1} - \eta \cdot d_{x_a}$$
$$x_a = x_{t-1} + \mu \cdot v_{t-1}$$
$$x_t = x_{t-1} + v_t$$

NAG_v2:

$$v_t = \mu \cdot v_{t-1} - \eta \cdot d_{x_{t-1}}$$
$$x_t = x_{t-1} - \mu \cdot v_{t-1} + v_t + \mu \cdot v_t$$
$$x_t = x_{t-1} + v_t + \mu \cdot (\mu - 1)v_{t-1} - \mu\eta d_{x_{t-1}}$$
$$x_t = x_{t-1} + \mu^2 v_{t-1} - \eta(1 + \mu)d_{x_{t-1}}$$

MXNet Implementation:

$$v_t = \mu \cdot v_{t-1} + d_{x_{t-1}}$$
$$x_t = x_{t-1} - \eta \cdot \mu\, v_t$$
$$= x_{t-1} - \eta\mu^2 v_{t-1} - \eta\mu \cdot d_{x_{t-1}}$$