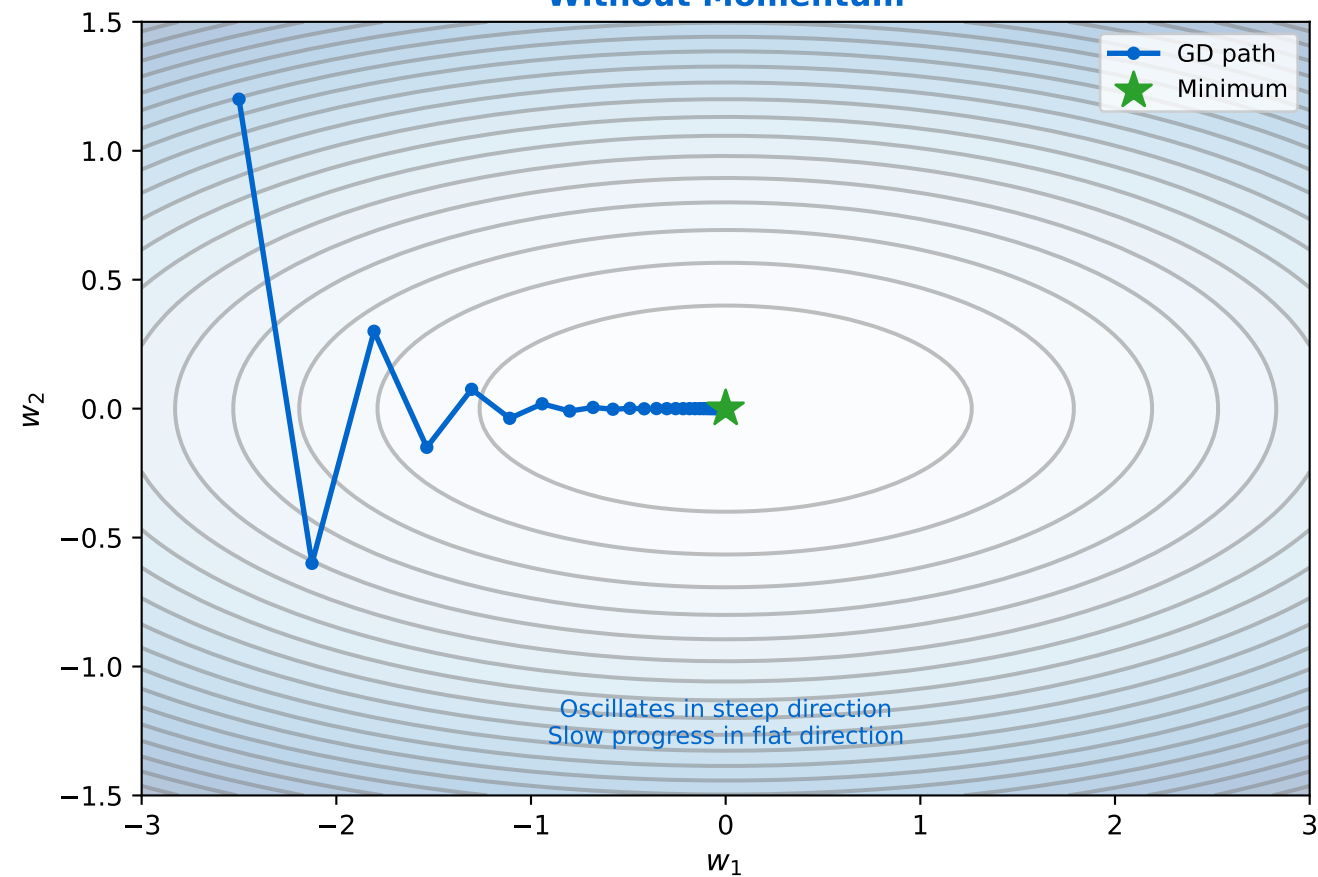
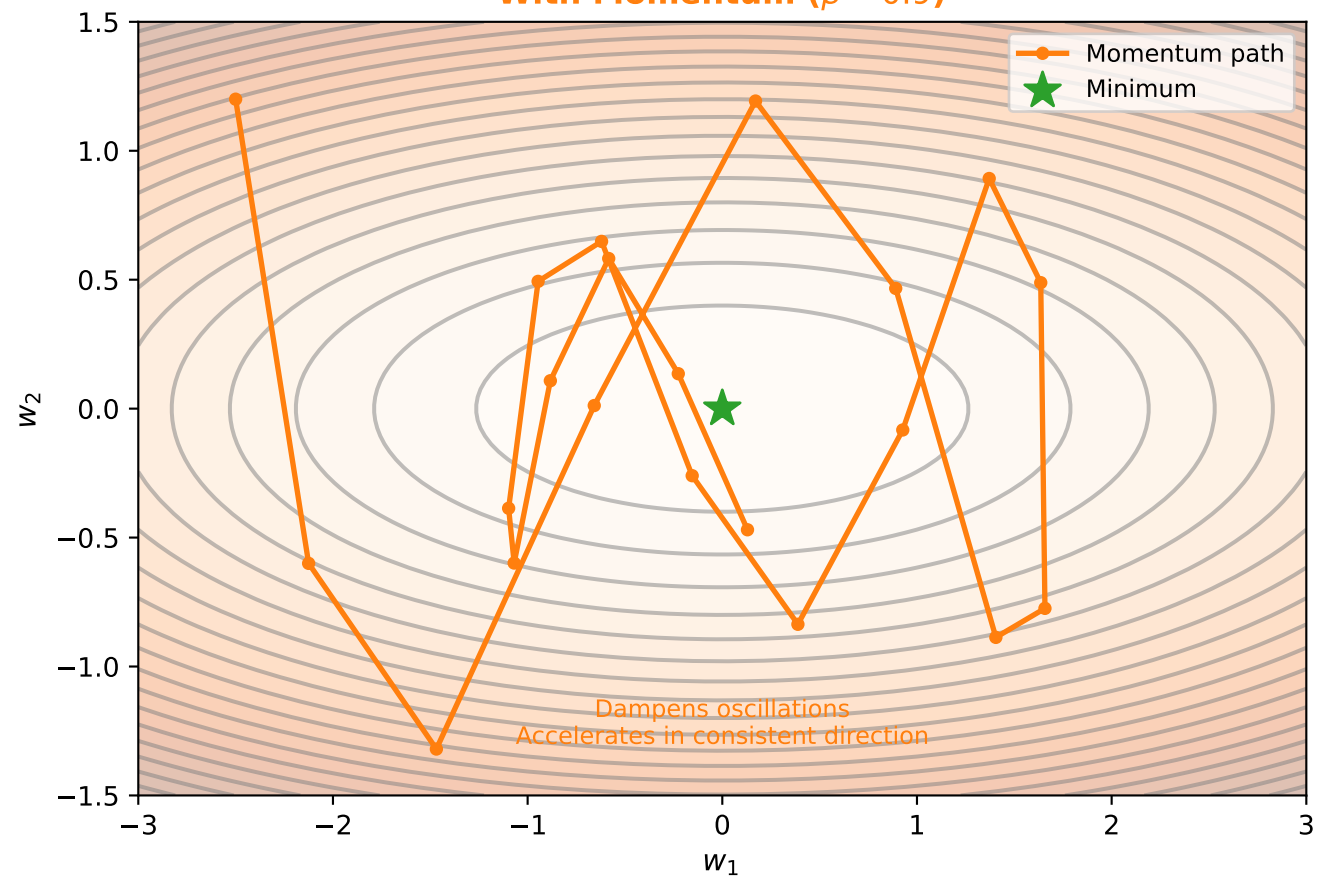


Momentum: Accelerating Gradient Descent

Without Momentum



With Momentum ($\beta = 0.9$)



Momentum Update: $v_t = \beta v_{t-1} + \eta \nabla L$, $w_t = w_{t-1} - v_t$ (Like a ball rolling downhill with inertia)