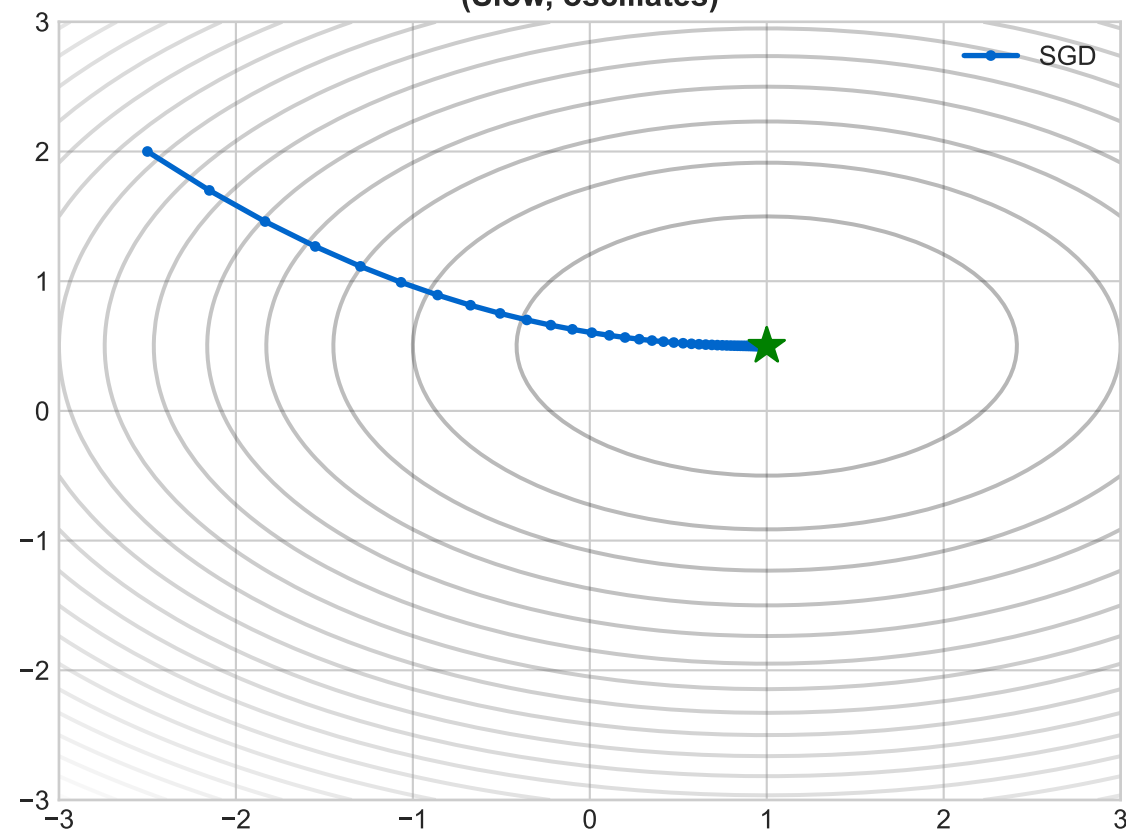
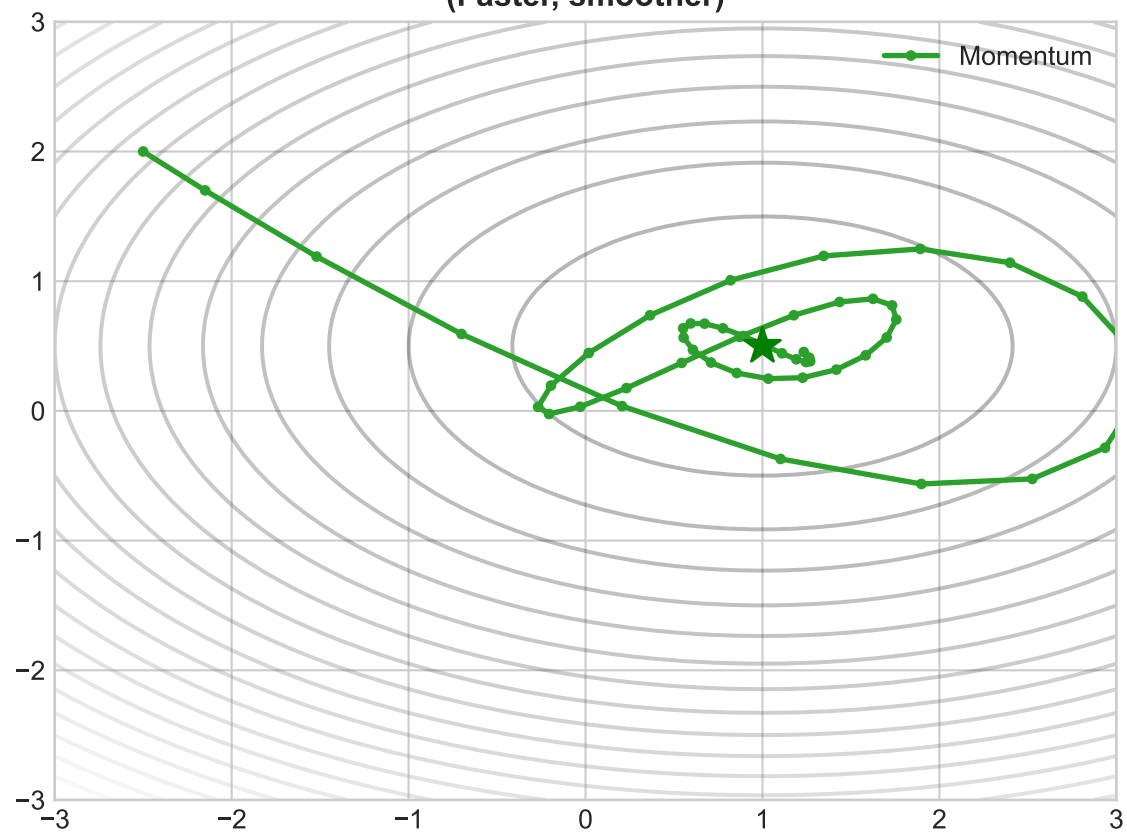


Optimization Algorithm Comparison

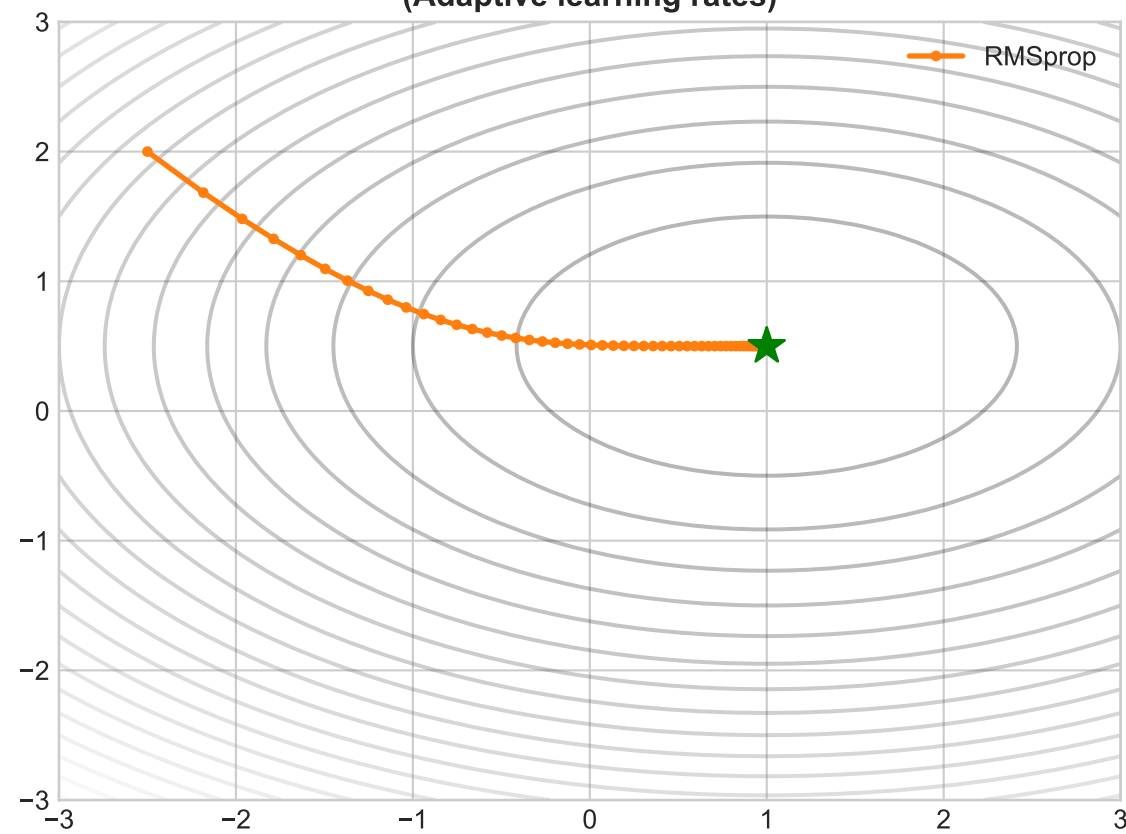
Vanilla SGD
(Slow, oscillates)



SGD + Momentum
(Faster, smoother)



RMSprop
(Adaptive learning rates)



Adam
(Best of both worlds)

