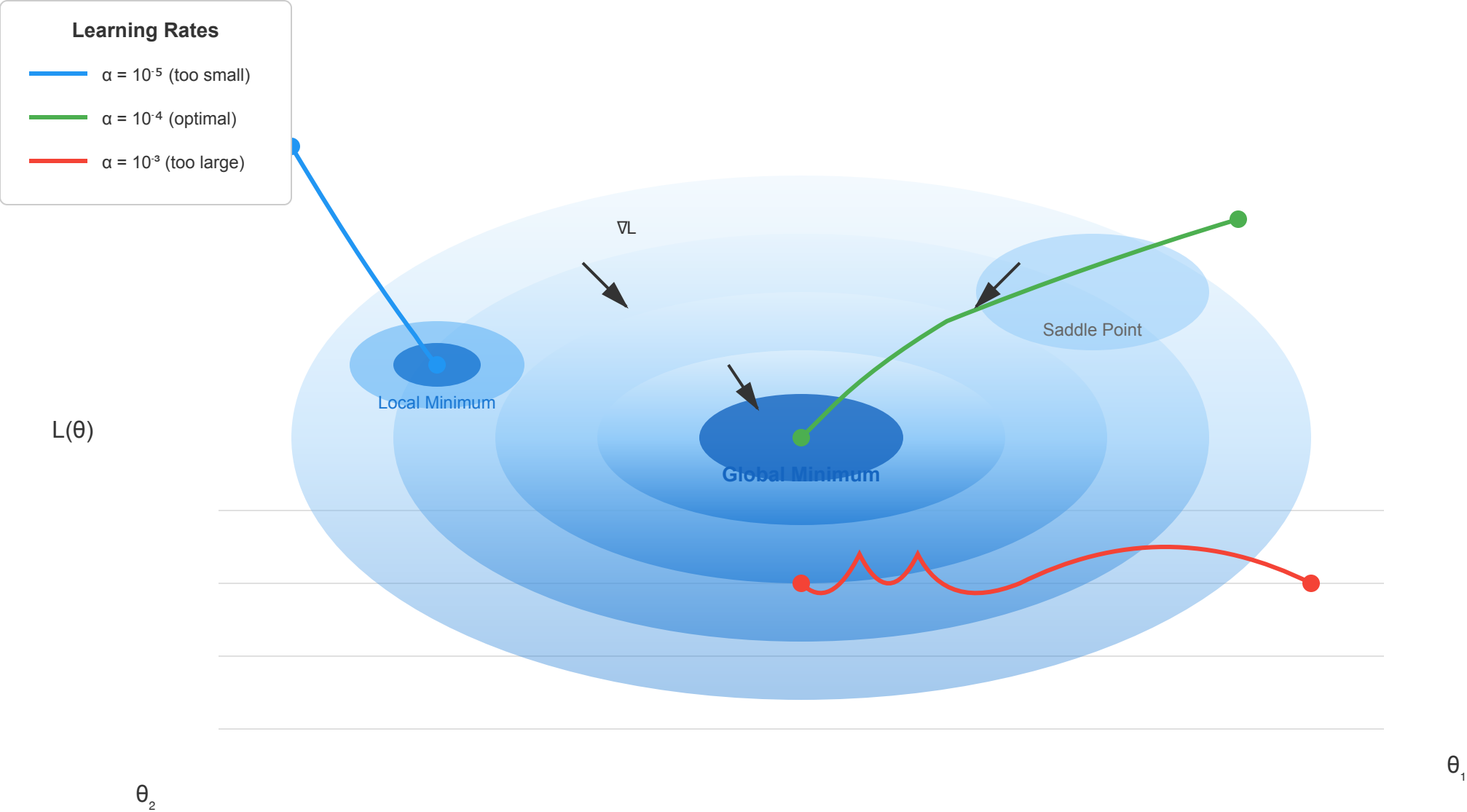


# Loss Landscape with Gradient Descent Trajectories



Gradient descent follows  $\nabla L$  downhill. Learning rate  $\alpha$  controls step size: too small is slow, too large oscillates.