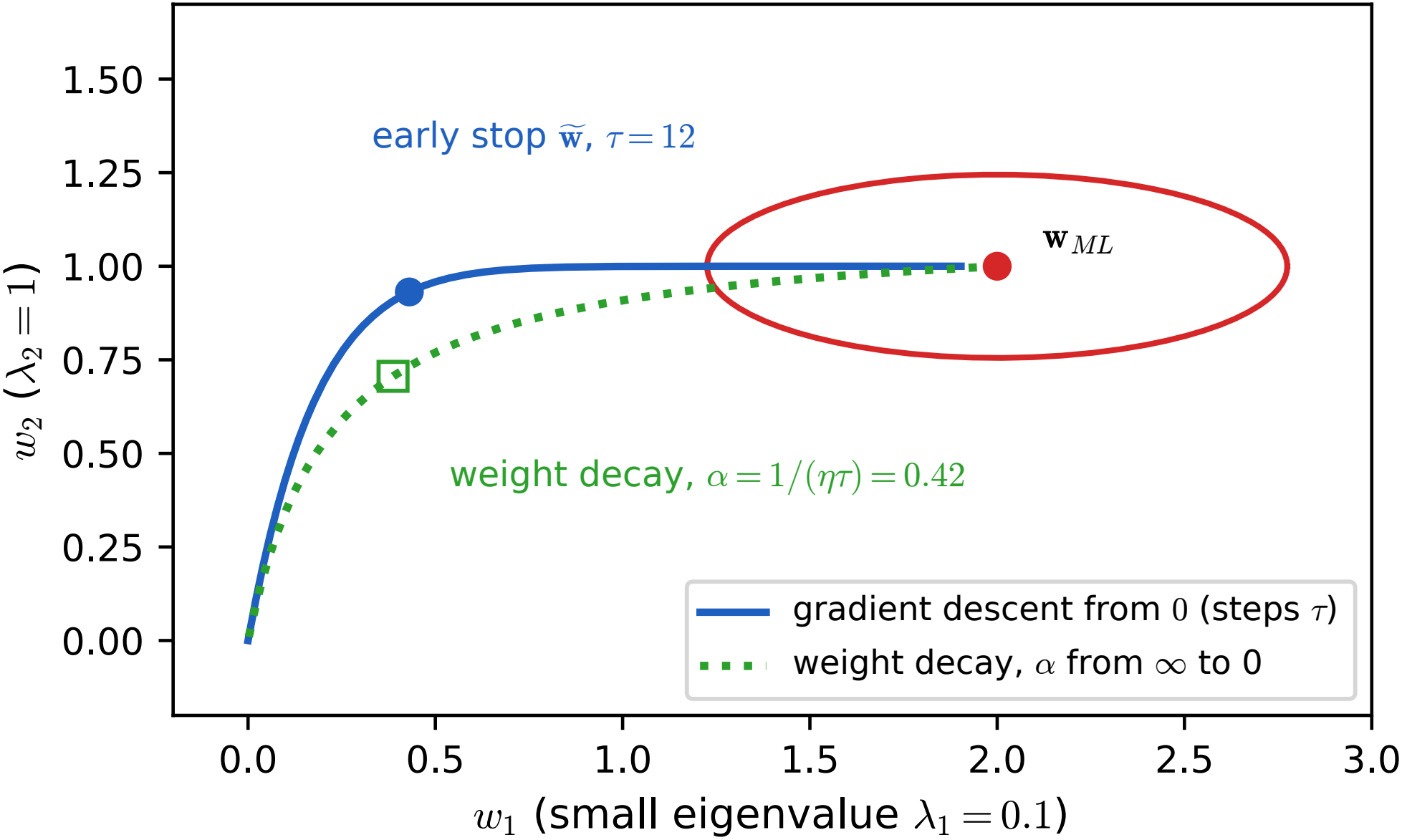

early stop $\widetilde{\mathbf{w}}$, $\tau = 12$
$\mathbf{w}_{ML}$
weight decay, $\alpha = 1/(\eta\tau) = 0.42$
gradient descent from 0 (steps τ)
weight decay, α from ∞ to 0
1.50
1.25
1.00
0.75
0.50
0.25
0.00
w_2 ($\lambda_2 = 1$)
0.0
0.5
1.0
1.5
2.0
2.5
3.0
w_1 (small eigenvalue $\lambda_1 = 0.1$)