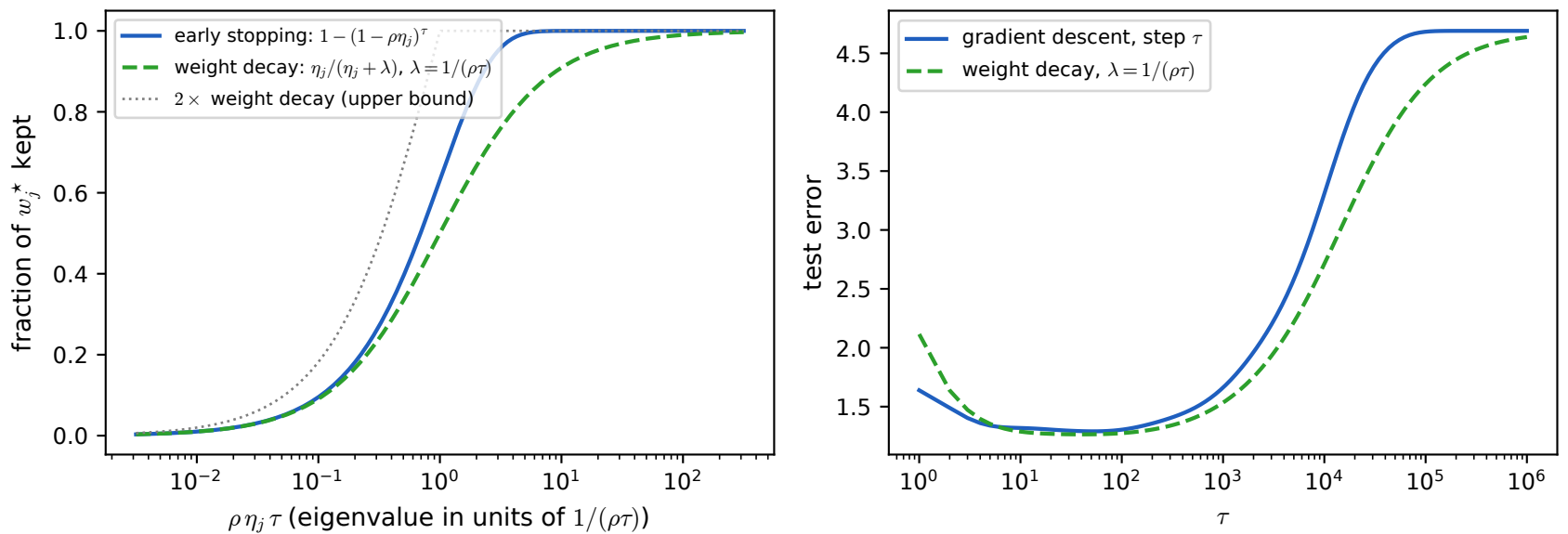

early stopping: $1-(1-\rho\eta_j)^\tau$
weight decay: $\eta_j/(\eta_j+\lambda)$, $\lambda=1/(\rho\tau)$
$2\times$ weight decay (upper bound)
fraction of $w_j^\star$ kept
1.0
0.8
0.6
0.4
0.2
0.0
10^{-2}
10^{-1}
10^{0}
10^{1}
10^{2}
$\rho\,\eta_j\,\tau$ (eigenvalue in units of $1/(\rho\tau)$)
gradient descent, step τ
weight decay, $\lambda=1/(\rho\tau)$
test error
4.5
4.0
3.5
3.0
2.5
2.0
1.5
10^{0}
10^{1}
10^{2}
10^{3}
10^{4}
10^{5}
10^{6}
τ