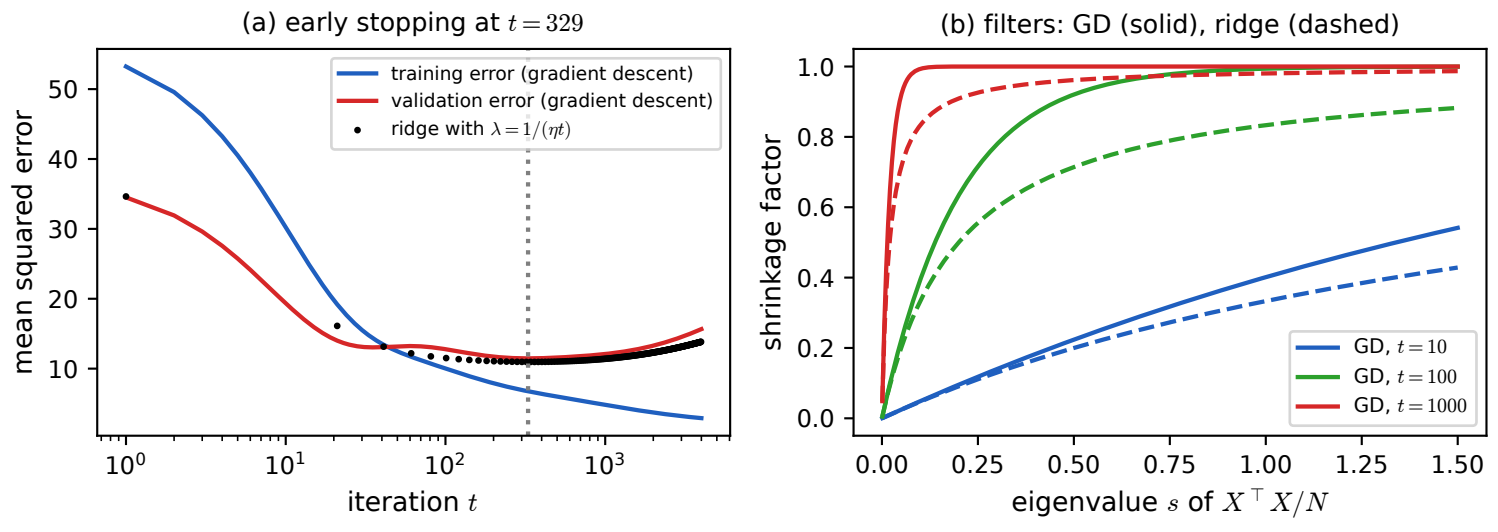

(a) early stopping at $t = 329$
training error (gradient descent)
validation error (gradient descent)
ridge with $\lambda = 1/(\eta t)$
mean squared error
iteration t
(b) filters: GD (solid), ridge (dashed)
shrinkage factor
GD, $t = 10$
GD, $t = 100$
GD, $t = 1000$
eigenvalue s of $X^\top X/N$