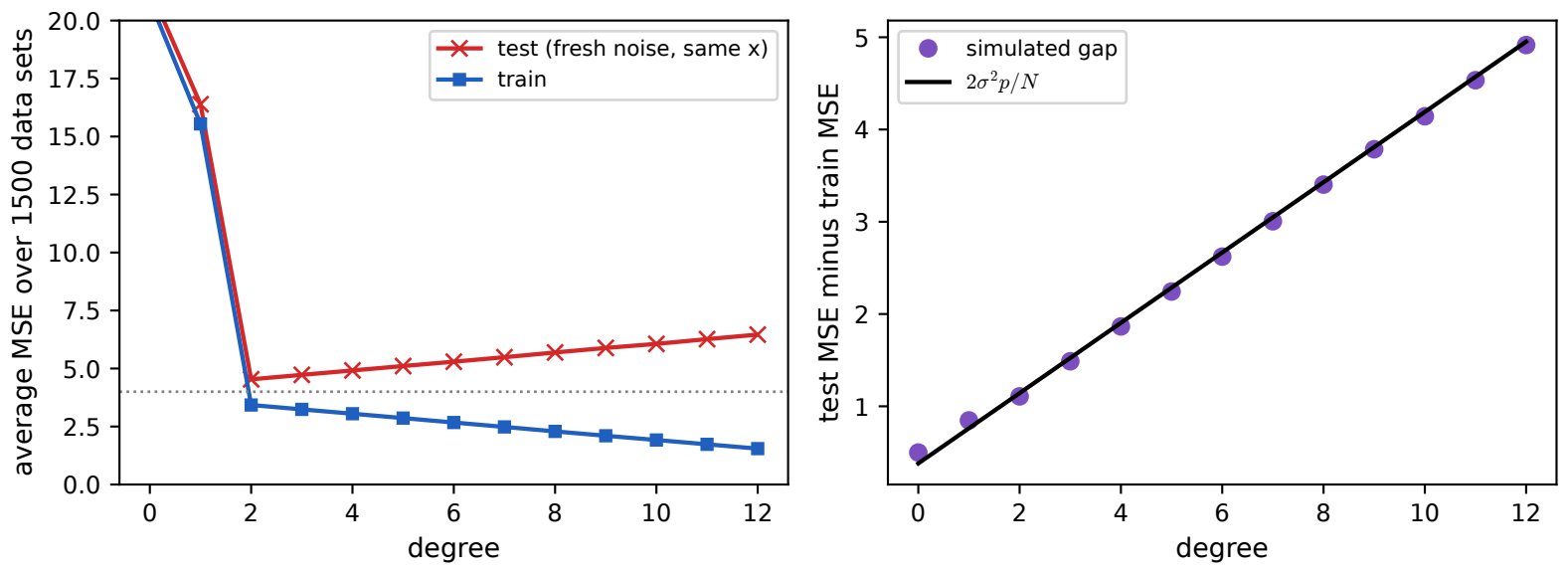

test (fresh noise, same x)
train
average MSE over 1500 data sets
degree
simulated gap
$2\sigma^2 p/N$
test MSE minus train MSE
degree