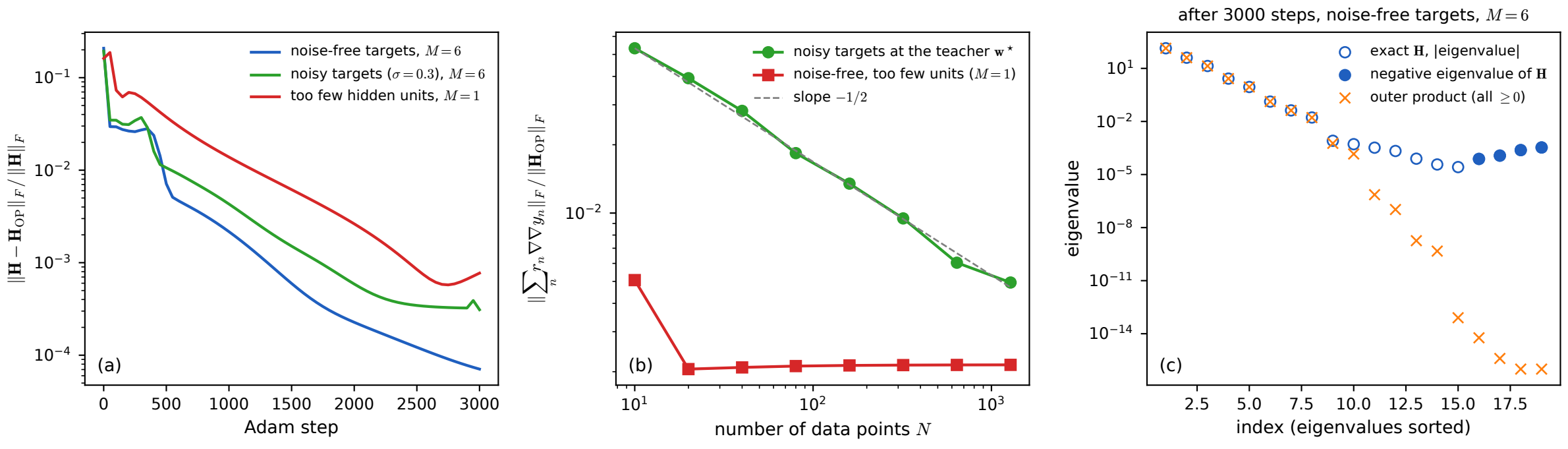

noise-free targets, $M=6$
noisy targets ($\sigma=0.3$), $M=6$
too few hidden units, $M=1$
$\|\mathbf{H}-\mathbf{H}_{\mathrm{OP}}\|_F / \|\mathbf{H}\|_F$
10^{-1}
10^{-2}
10^{-3}
10^{-4}
(a)
0
500
1000
1500
2000
2500
3000
Adam step
noisy targets at the teacher $\mathbf{w}^\star$
noise-free, too few units ($M=1$)
slope $-1/2$
$\|\sum_n r_n \nabla\nabla y_n\|_F / \|\mathbf{H}_{\mathrm{OP}}\|_F$
10^{-2}
(b)
10^1
10^2
10^3
number of data points N
after 3000 steps, noise-free targets, $M=6$
exact $\mathbf{H}$, |eigenvalue|
negative eigenvalue of $\mathbf{H}$
outer product (all ≥ 0)
eigenvalue
10^1
10^{-2}
10^{-5}
10^{-8}
10^{-11}
10^{-14}
(c)
2.5
5.0
7.5
10.0
12.5
15.0
17.5
index (eigenvalues sorted)