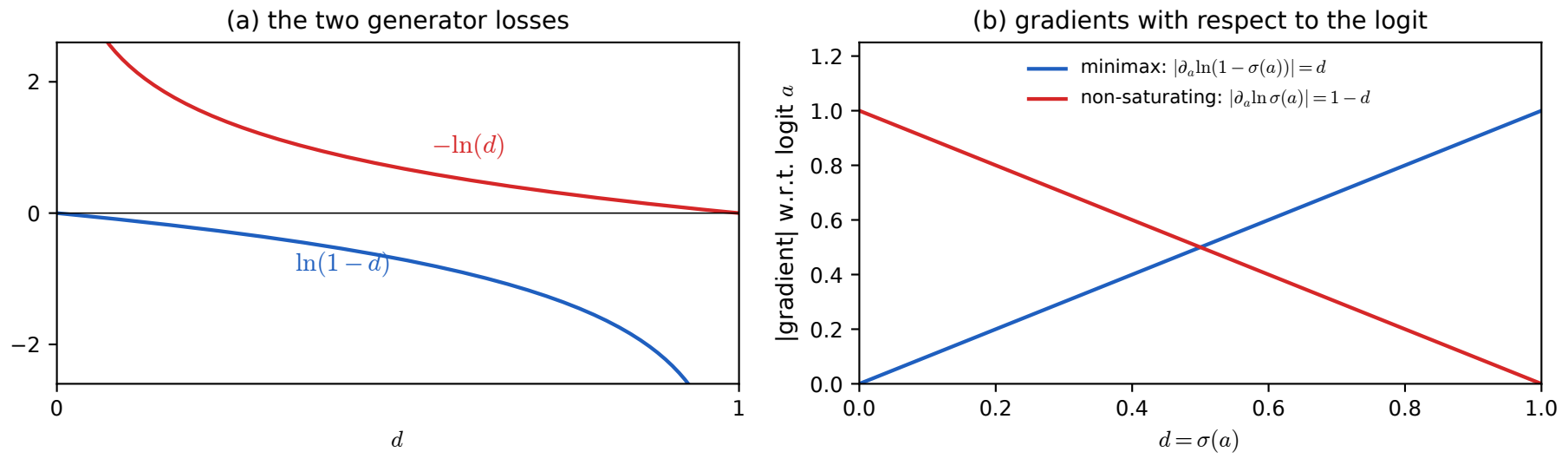

(a) the two generator losses
2
0
−2
0
1
d
$-\ln(d)$
$\ln(1-d)$
(b) gradients with respect to the logit
|gradient| w.r.t. logit a
1.2
1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0
$d = \sigma(a)$
minimax: $|\partial_a \ln(1-\sigma(a))| = d$
non-saturating: $|\partial_a \ln \sigma(a)| = 1-d$