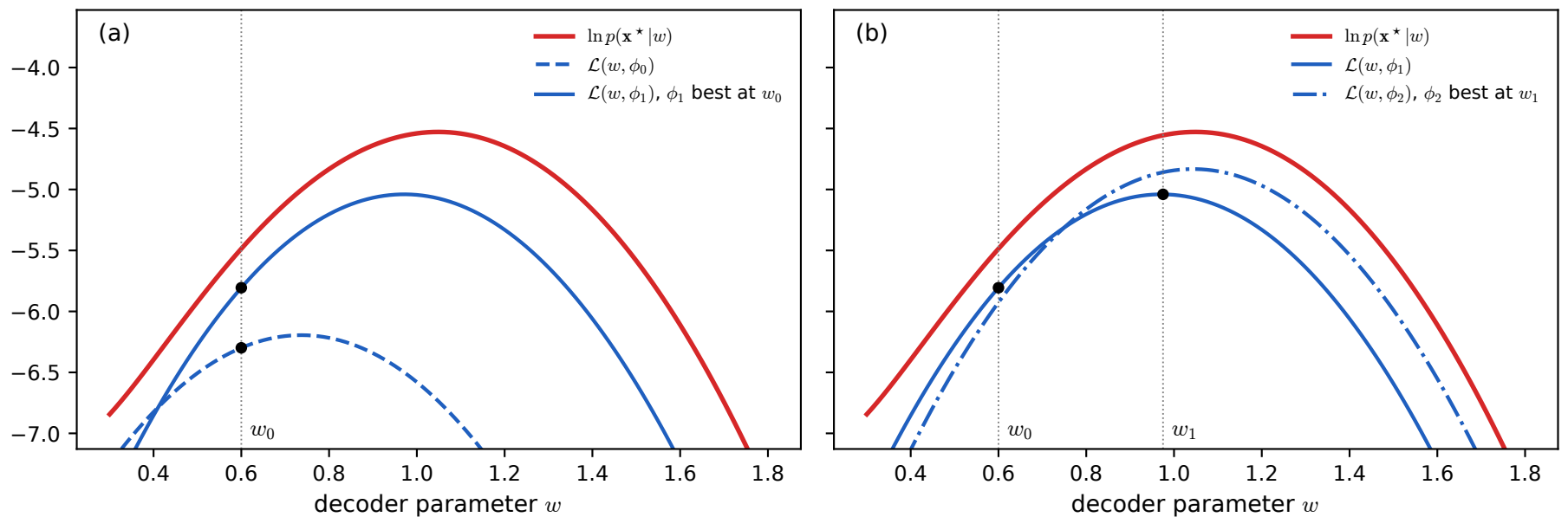

(a)
$\ln p(\mathbf{x}^\star|w)$
$\mathcal{L}(w,\phi_0)$
$\mathcal{L}(w,\phi_1)$, ϕ_1 best at w_0
−4.0
−4.5
−5.0
−5.5
−6.0
−6.5
−7.0
w_0
0.4
0.6
0.8
1.0
1.2
1.4
1.6
1.8
decoder parameter w
(b)
$\ln p(\mathbf{x}^\star|w)$
$\mathcal{L}(w,\phi_1)$
$\mathcal{L}(w,\phi_2)$, ϕ_2 best at w_1
w_0
w_1
0.4
0.6
0.8
1.0
1.2
1.4
1.6
1.8
decoder parameter w