Learning curves (logarithmic scale)

ATC gradient tracking
DIGing

$P(w_{-1}^r) - P(w^\star)$

epoch index r