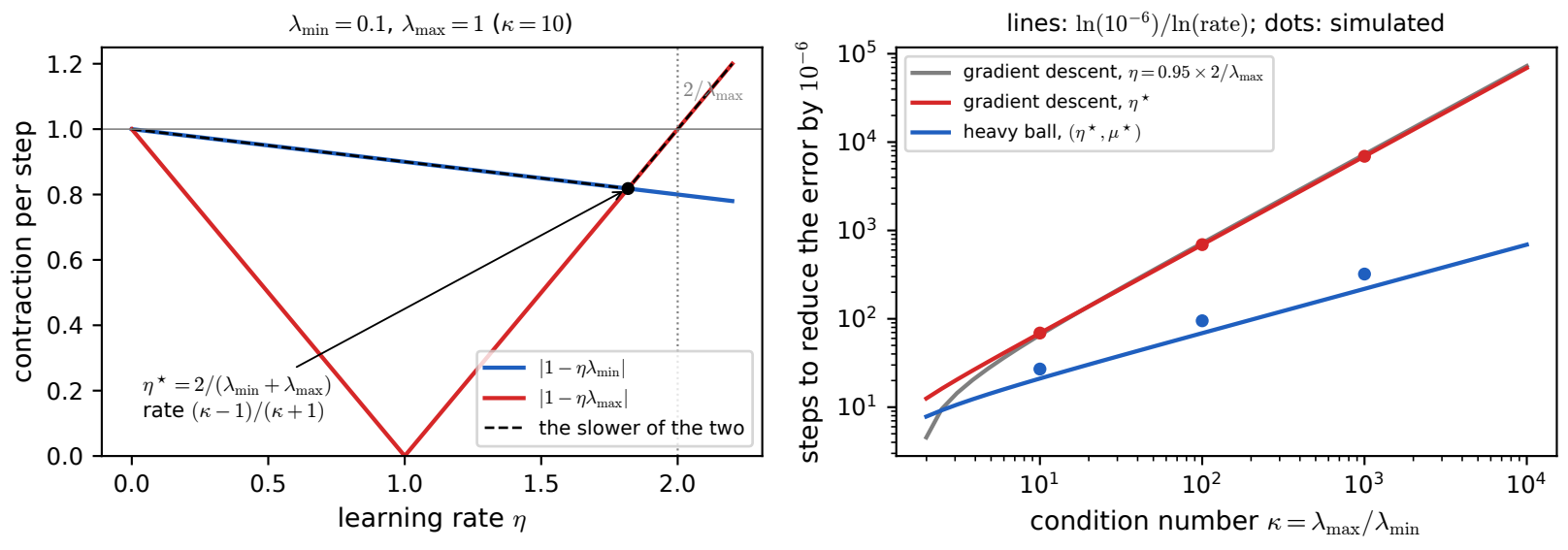

$\lambda_{\min} = 0.1$, $\lambda_{\max} = 1$ ($\kappa = 10$)
contraction per step
learning rate η
$2/\lambda_{\max}$
$\eta^\star = 2/(\lambda_{\min} + \lambda_{\max})$
rate $(\kappa-1)/(\kappa+1)$
$|1 - \eta\lambda_{\min}|$
$|1 - \eta\lambda_{\max}|$
the slower of the two
lines: $\ln(10^{-6})/\ln(\text{rate})$; dots: simulated
steps to reduce the error by 10^{-6}
condition number $\kappa = \lambda_{\max}/\lambda_{\min}$
gradient descent, $\eta = 0.95 \times 2/\lambda_{\max}$
gradient descent, $\eta^\star$
heavy ball, $(\eta^\star, \mu^\star)$