Gradient descent: squared distance to the OLS estimator
$||\theta - \hat{\theta}||^2$
Tolerance 1.0E-07
n=200, d=100
gaussian matrix
OLS time: 4.2E-04

Linear scale

Log scale

- $\gamma = 0.001$
- $\gamma = 0.01$, 84143 iters, 2.8E+00 s
- $\gamma = 0.05$, 16827 iters, 5.7E-01 s
- $\gamma = 0.1$
- Line search, 655 iters, 1.1E-01 s