Gradient descent with different learning rates $\alpha = 0.03$: 133 steps $\alpha = 0.028$: 142 steps $\alpha = 0.026$: 153 steps $\alpha = 0.024$: 165 steps $\alpha = 0.022$: 180 steps $\alpha = 0.02$: 197 steps $\alpha = adaptive$: 29 steps

2

 x_1

4

2

0

-2

-4

-6

-2

0

×