

1) What effect does the learning rate have on the result?

- A small learning rate causes slow but stable convergence.
- Large learning late (if too large) updates overshoot the valley and the algorithm diverges instead of converging

2) Can you find a learning rate parameter that causes divergence?

- For the paboloid, divergence occurs for learning rates ≥ 1
 - Example: $lr = 1.0$
 - $sdg(0, 0) \rightarrow (0, 0)$