

How we can define a network?

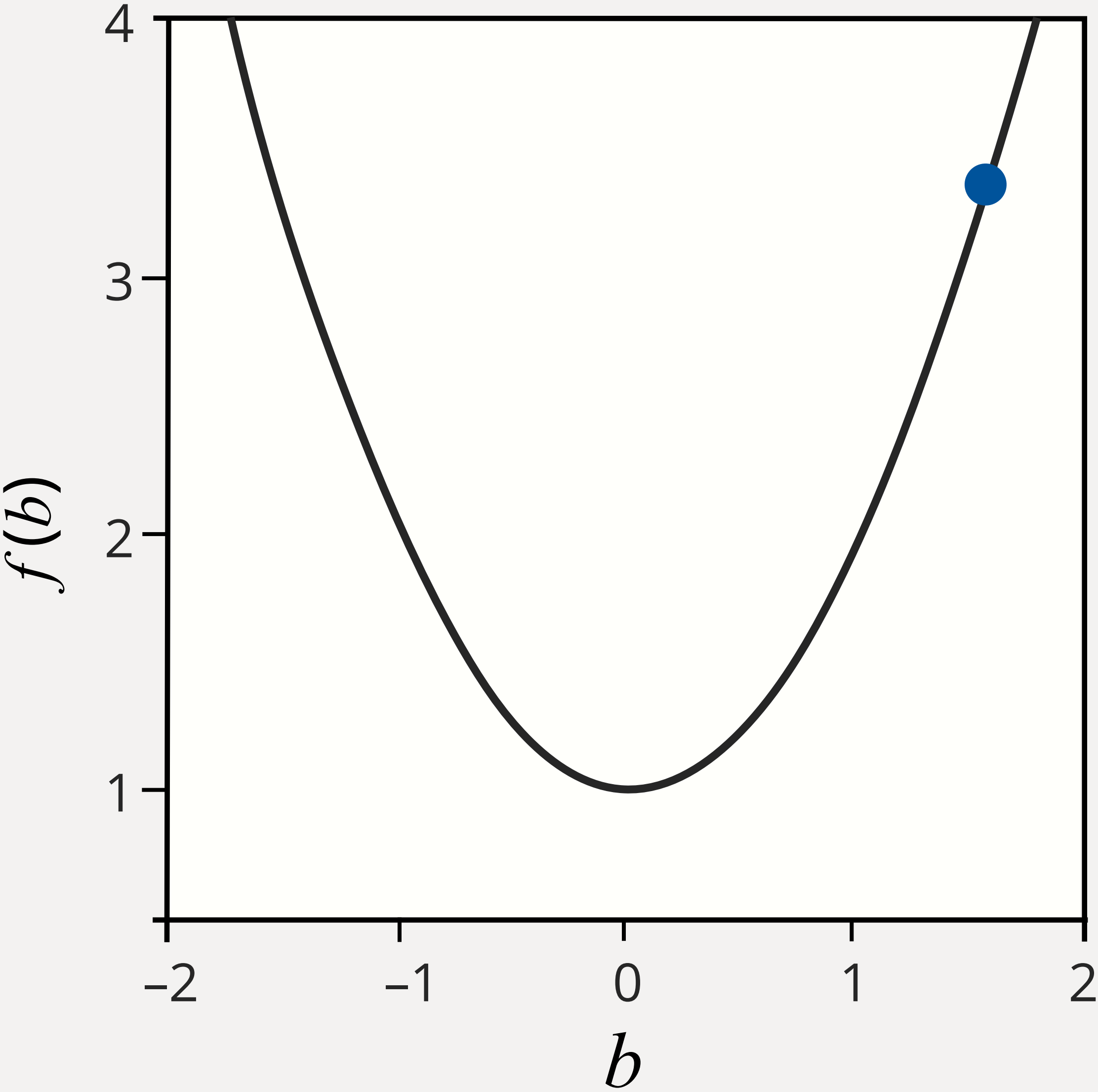
How we can mathematically
set up our learning goal?

How do we learn the network?

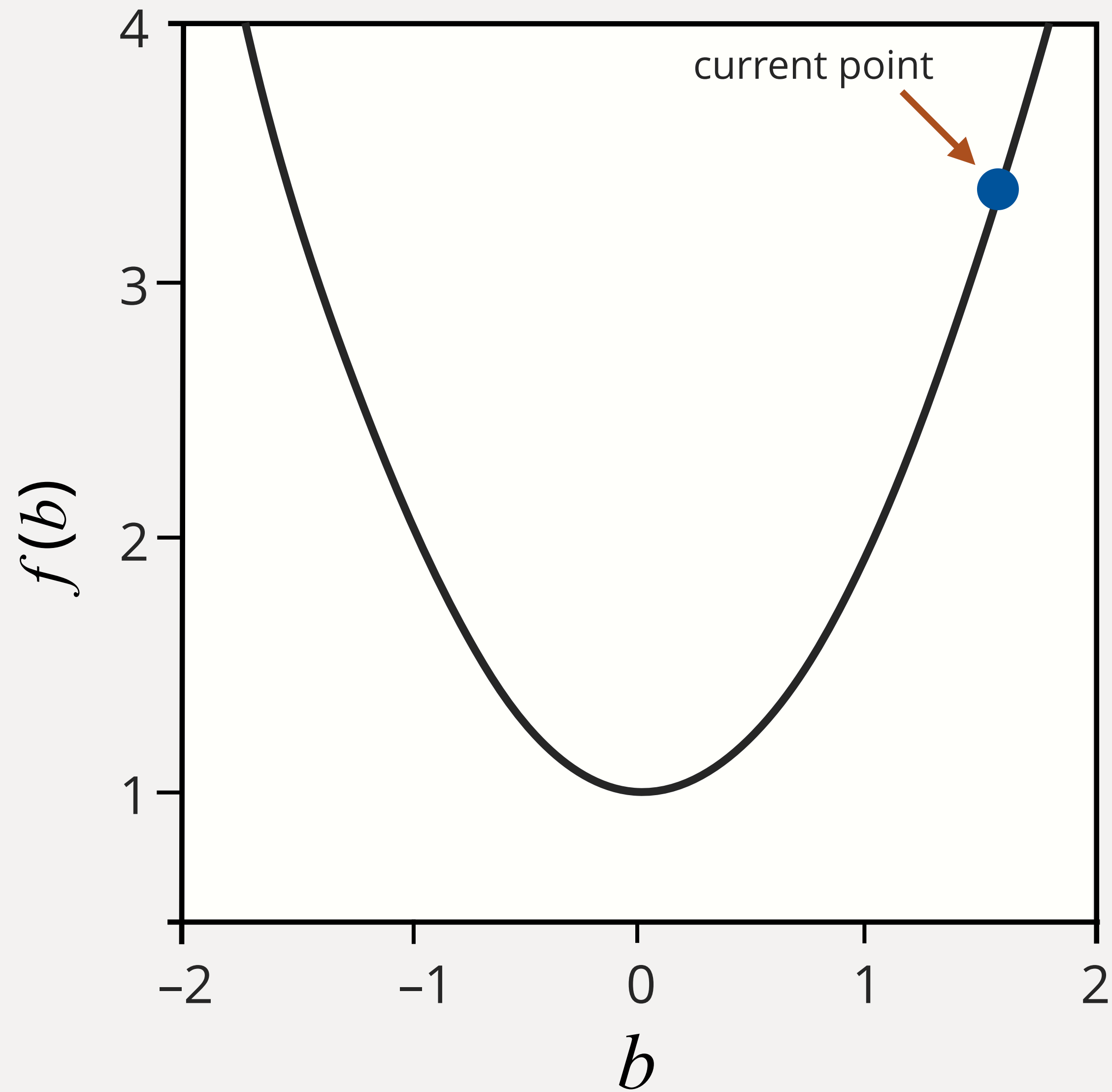
Learning Is Optimization

$$b^* = \arg \min_b \frac{1}{N} \sum_i^N \ell(y_i, \sigma(z_i))$$

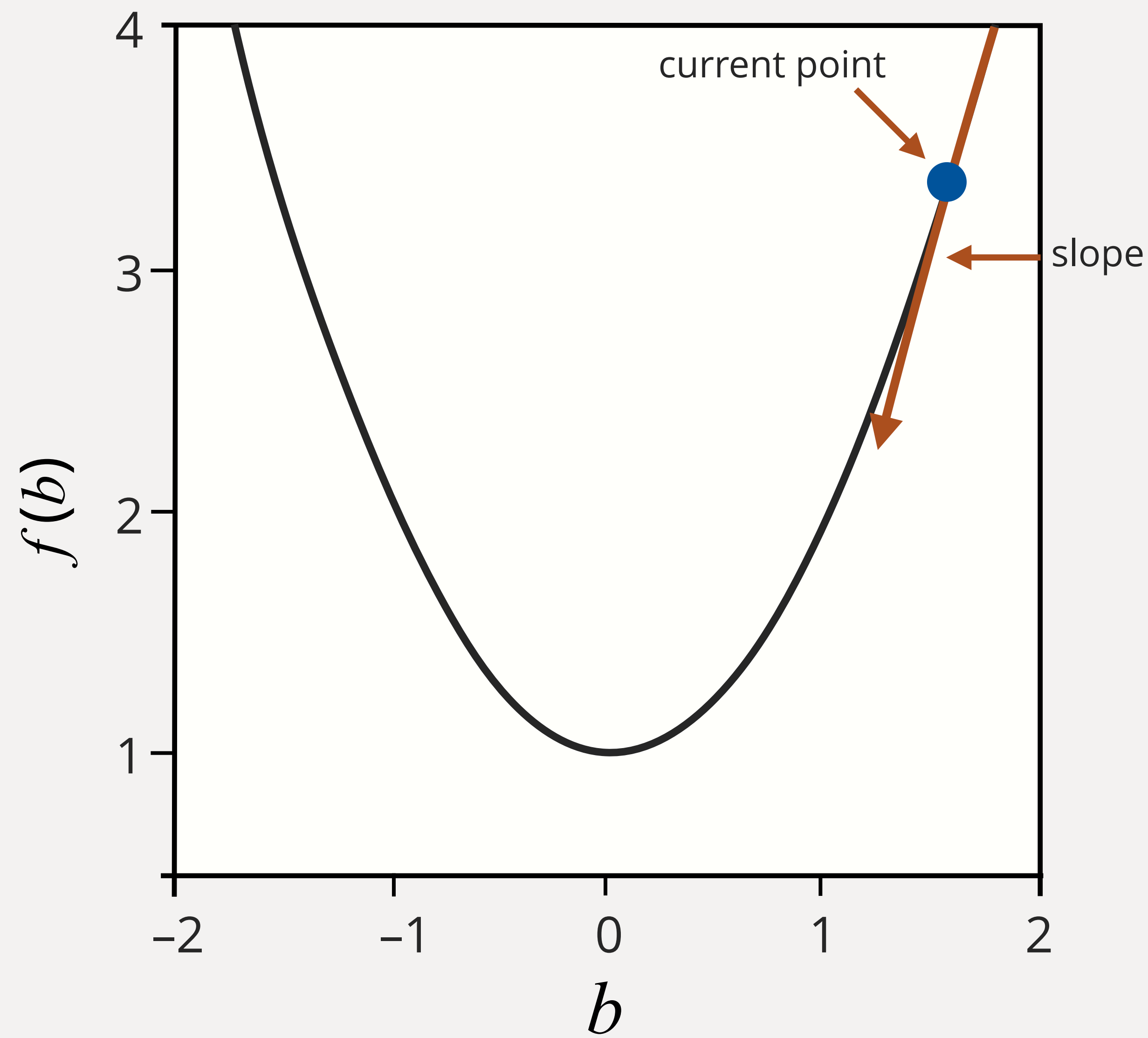
Gradient Descent



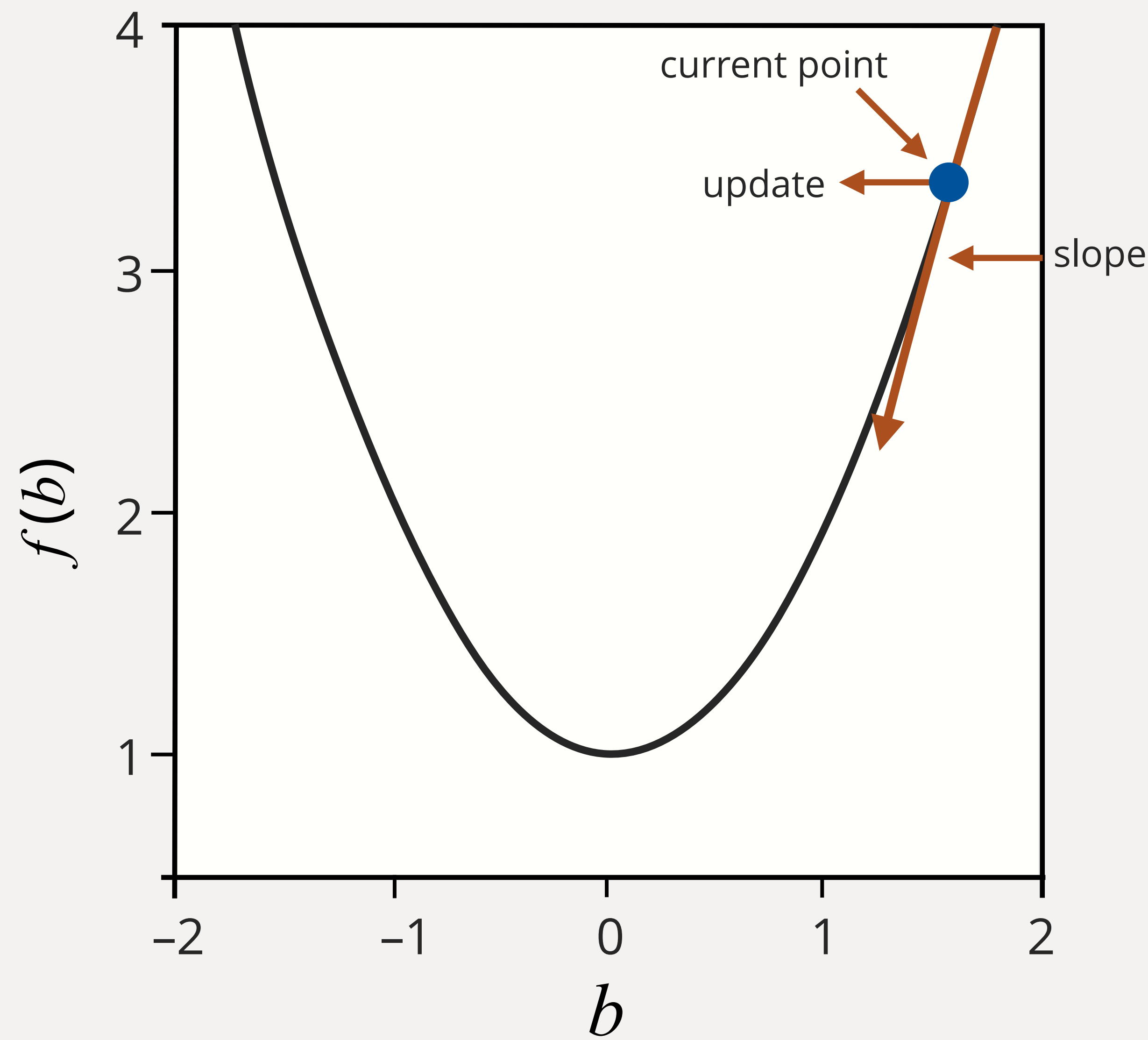
Gradient Descent



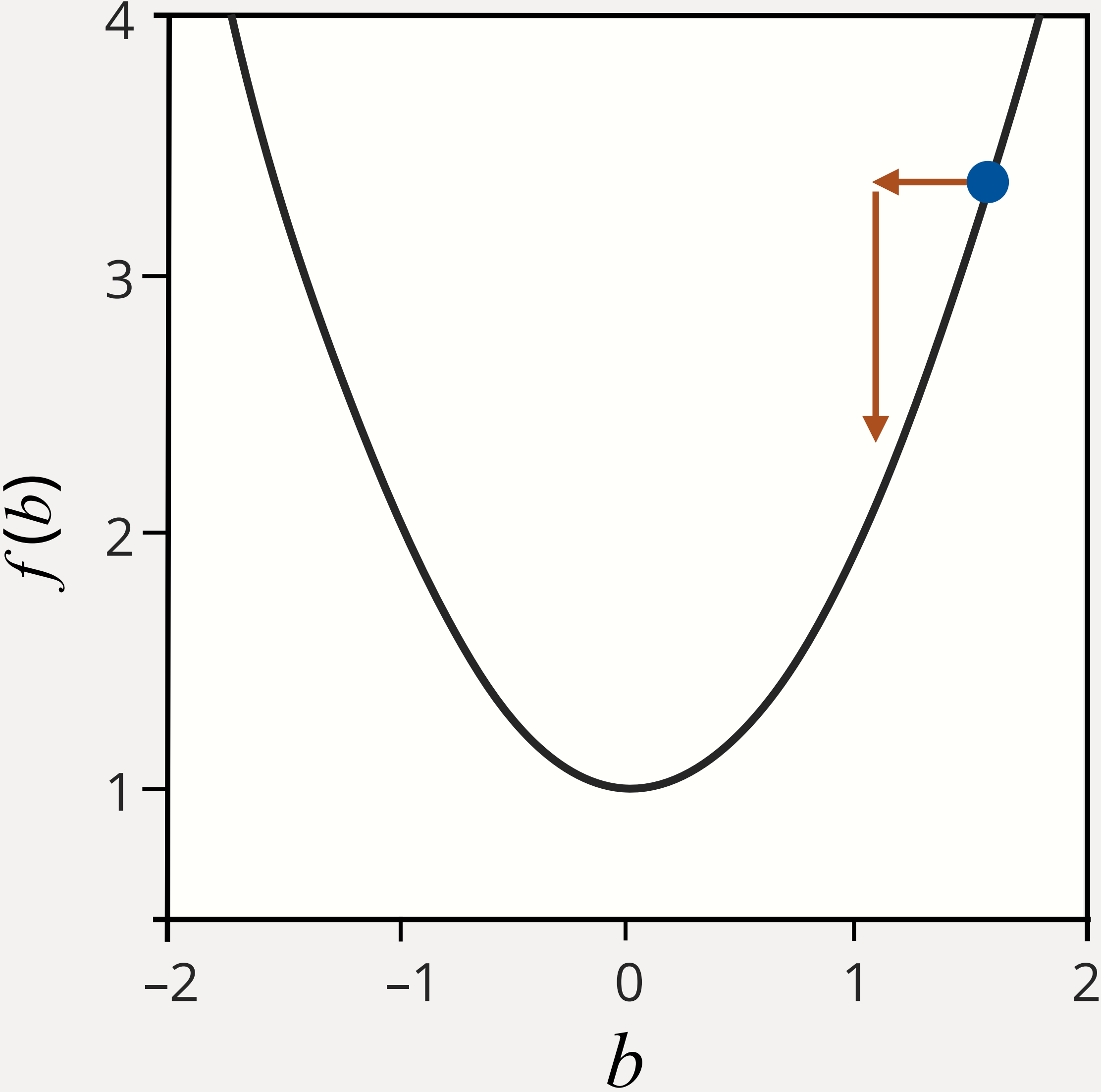
Gradient Descent



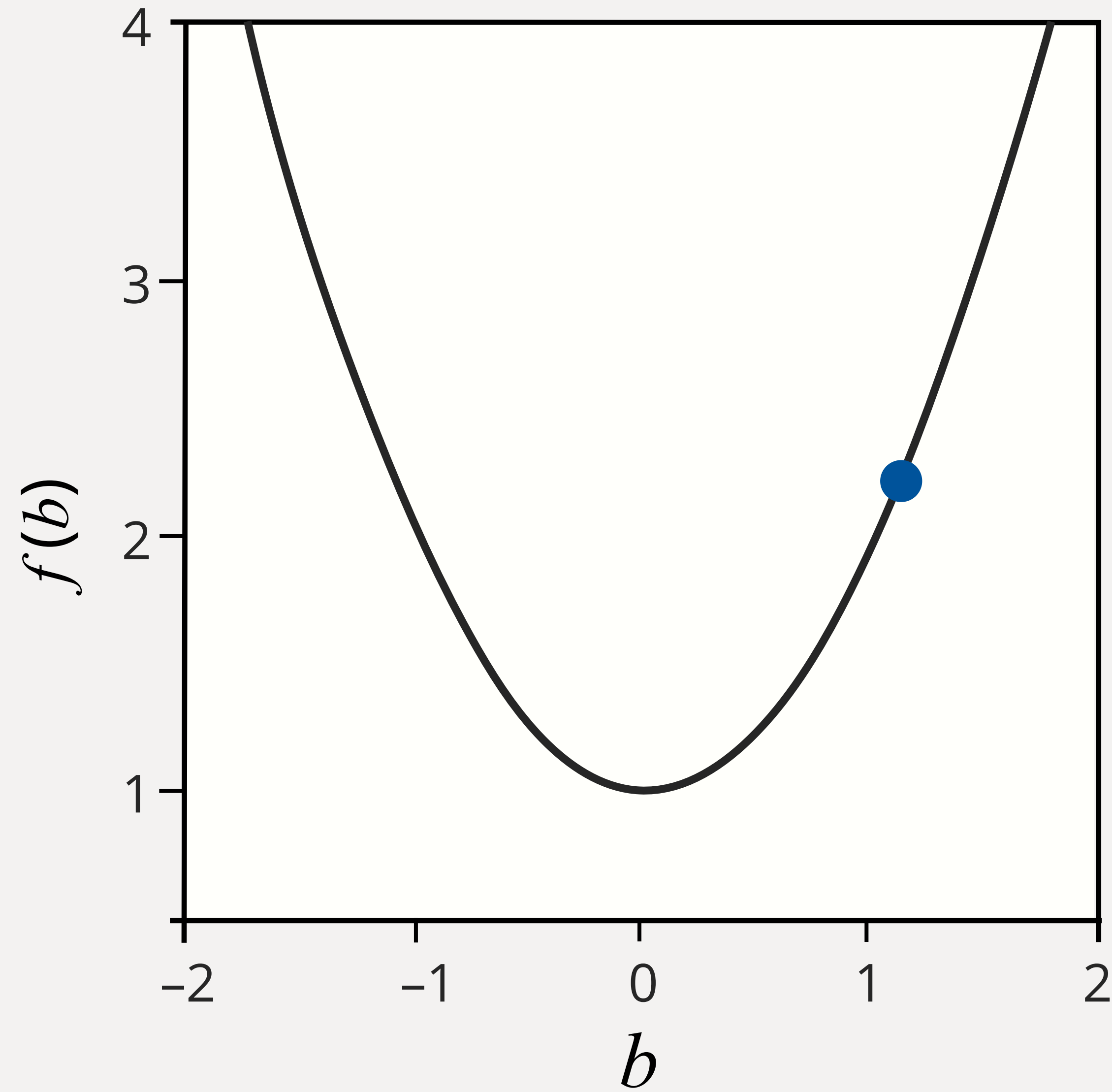
Gradient Descent



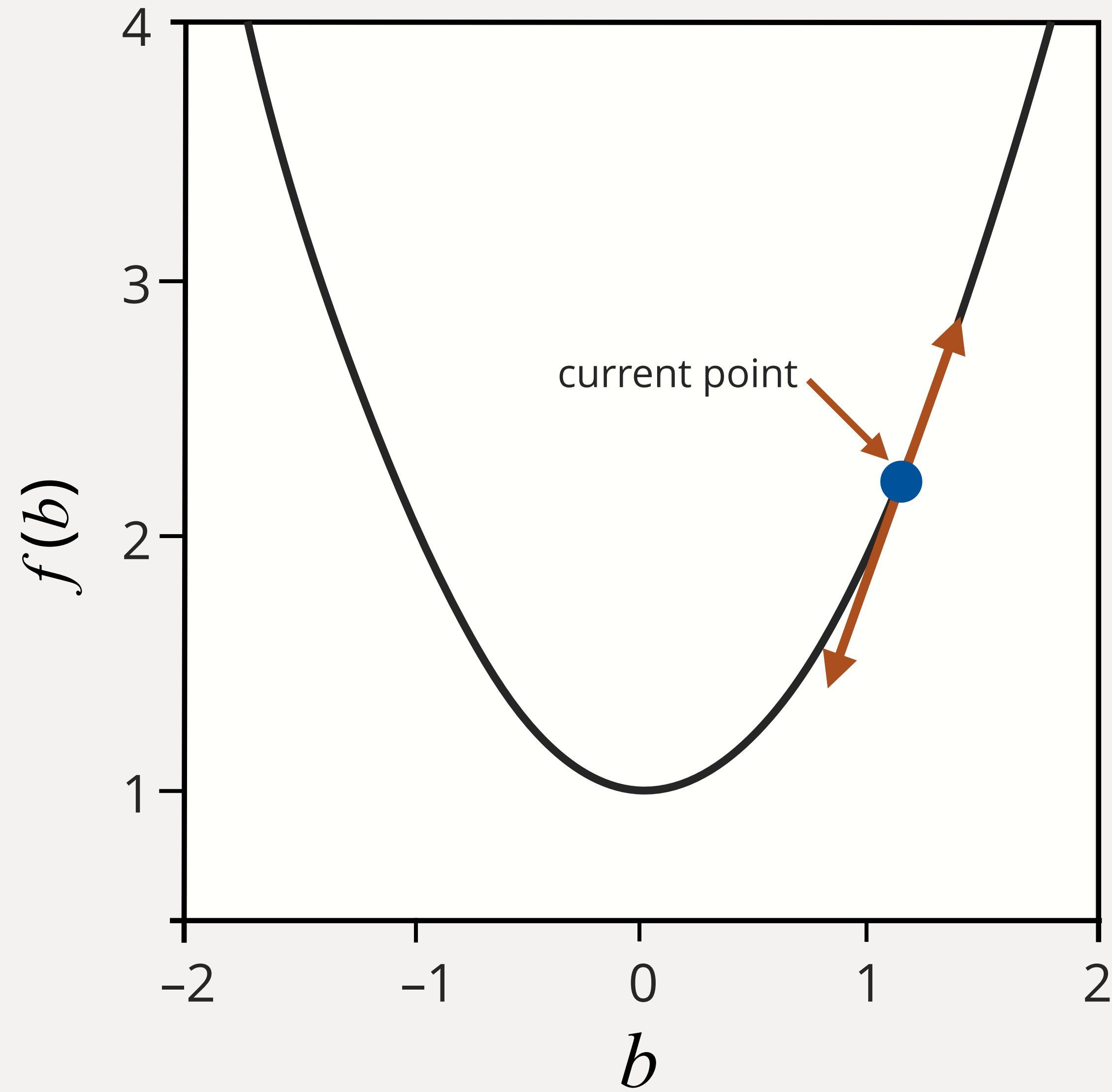
Gradient Descent



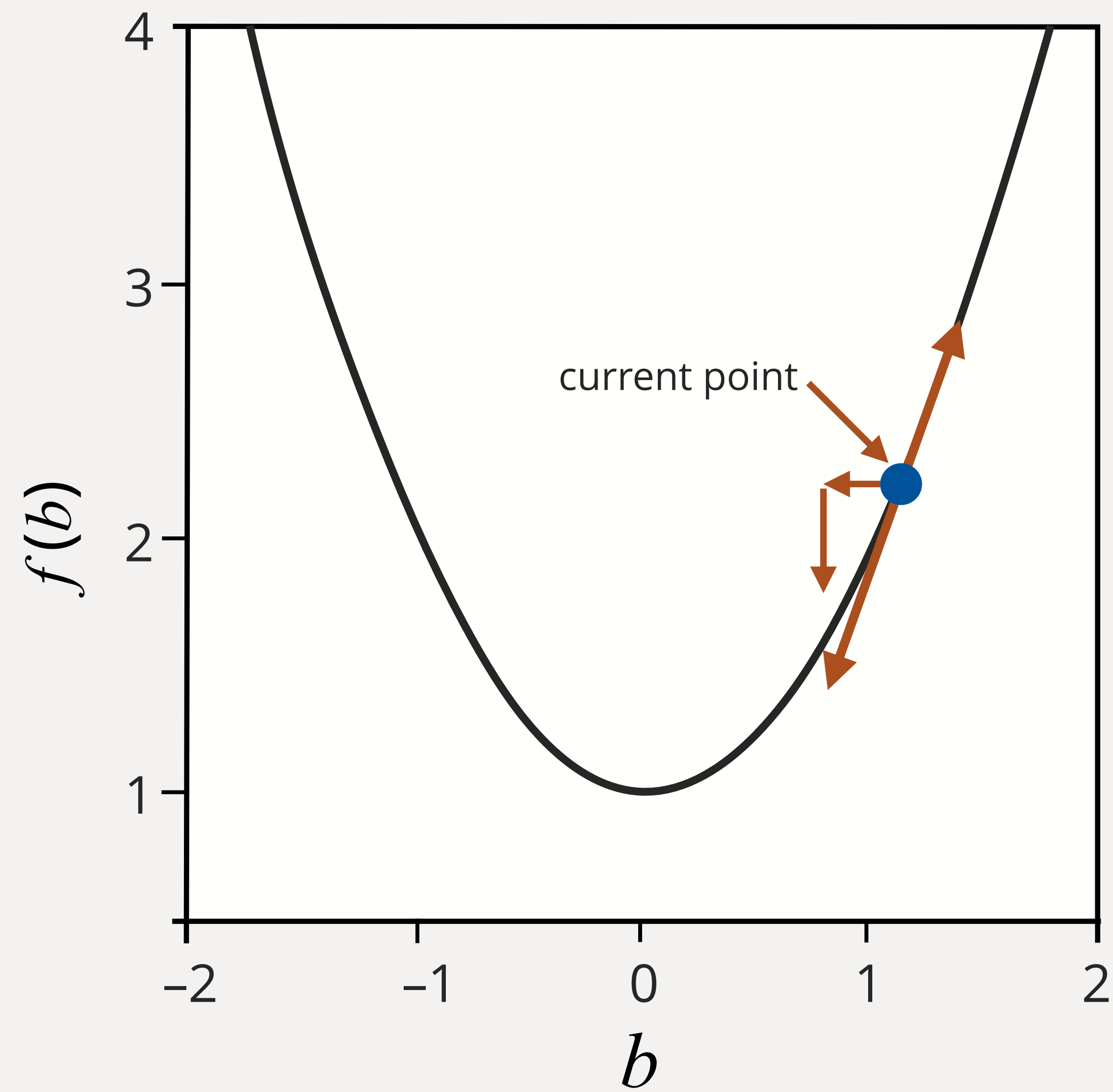
Gradient Descent



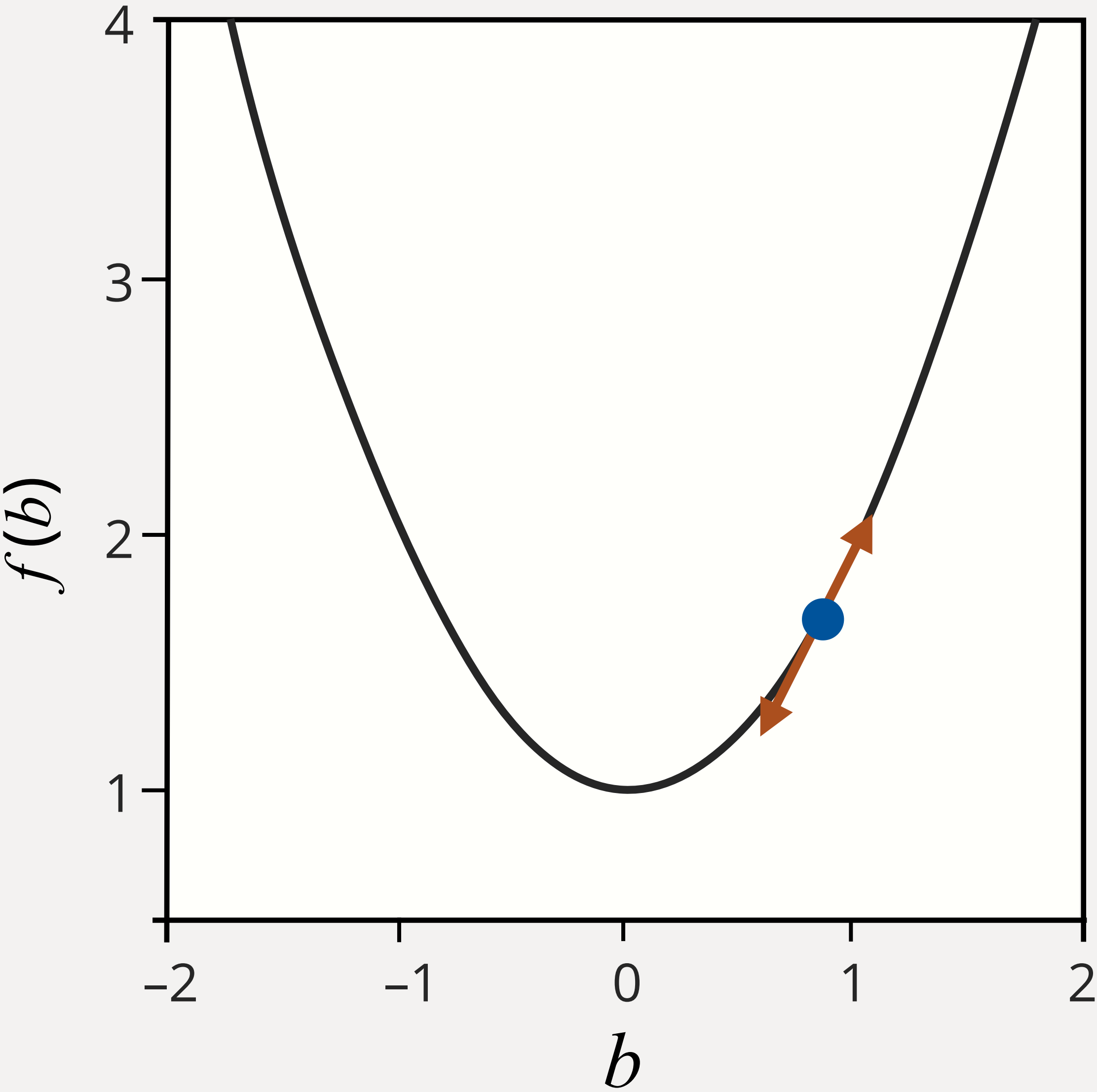
Gradient Descent



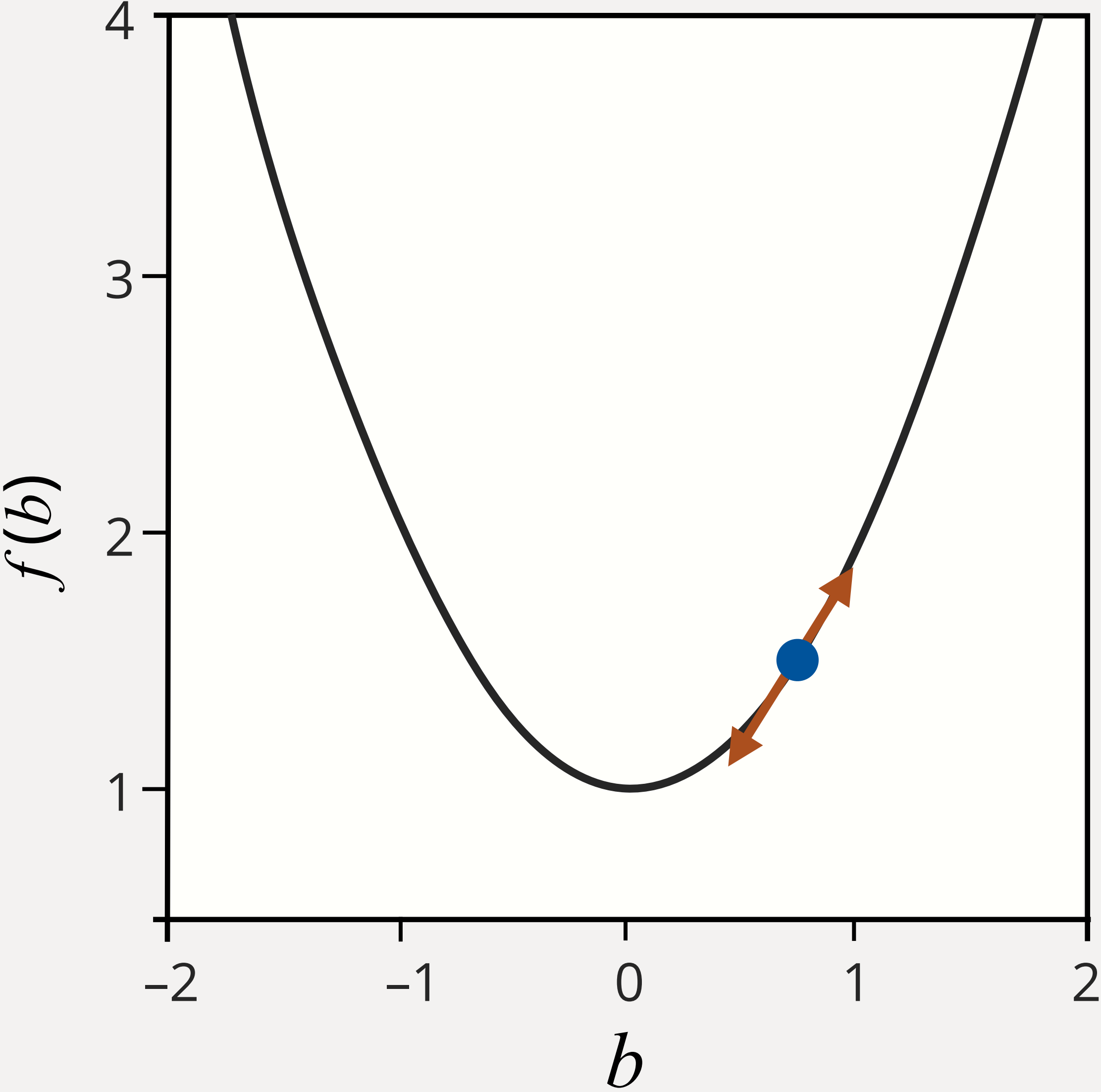
Gradient Descent



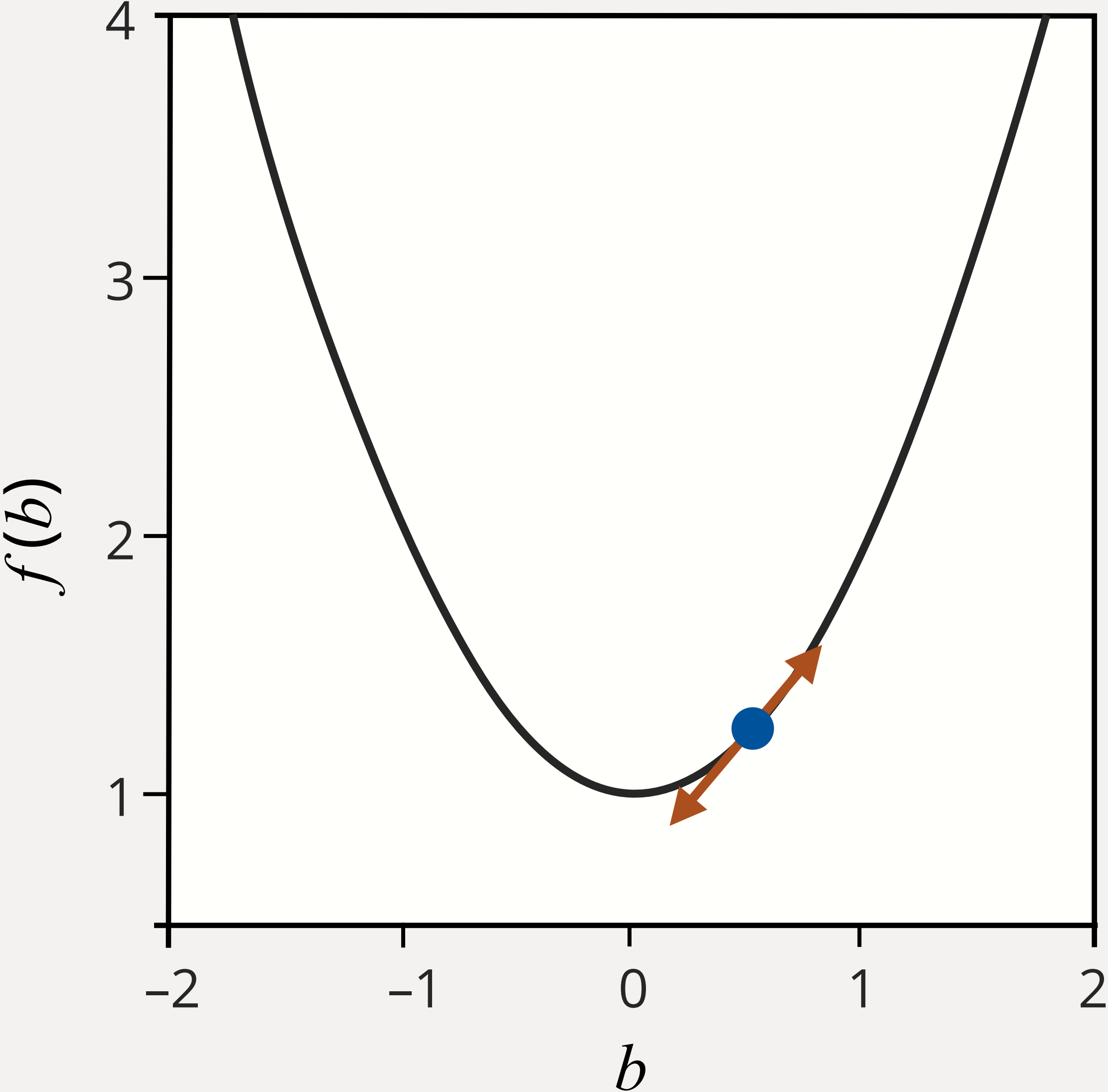
Gradient Descent

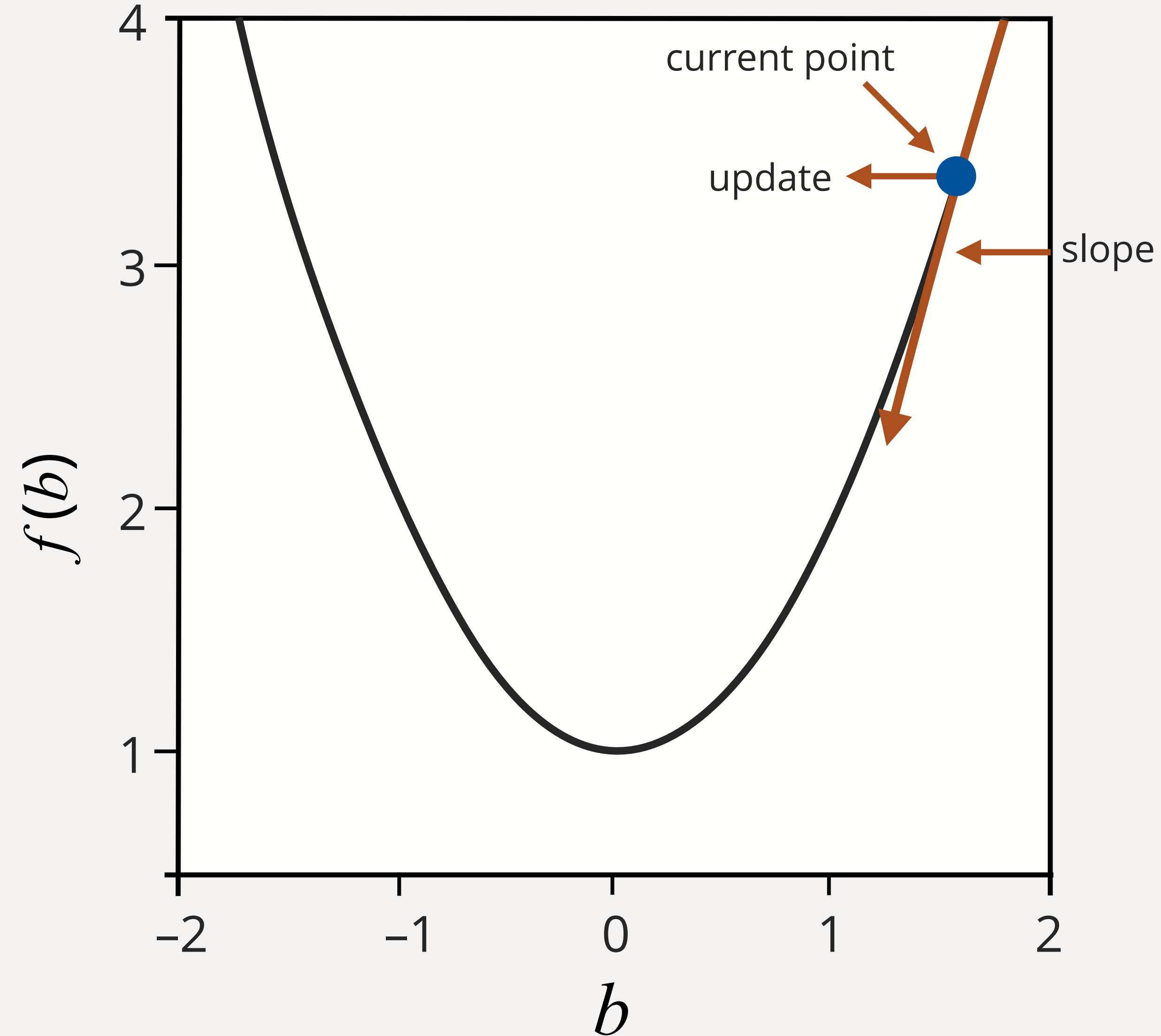


Gradient Descent



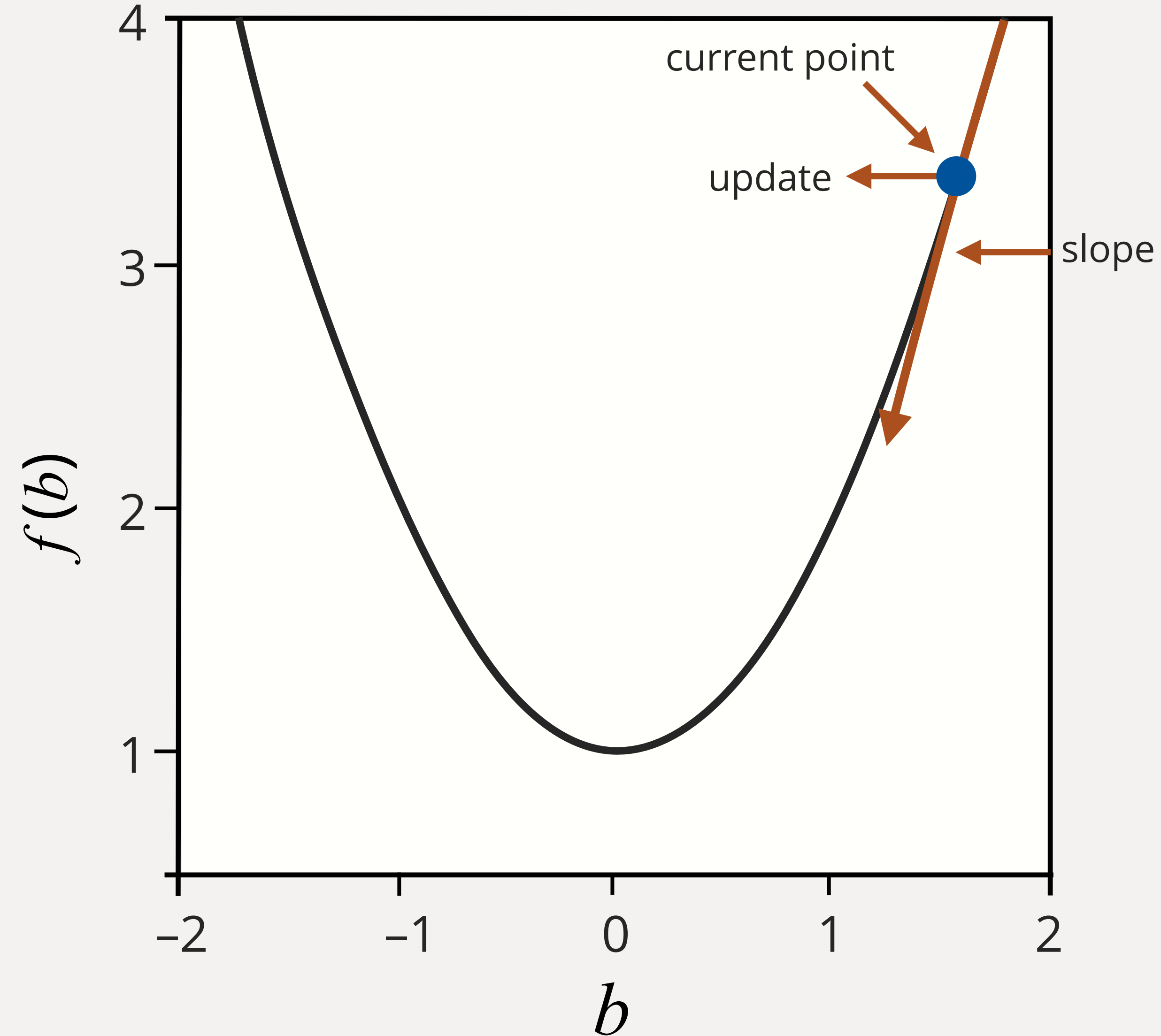
Gradient Descent





Minimize $b^* = \arg \min f(b)$

- Start with initial value b^0
- Run series of updates to move from b^k to b^{k+1}



Run Procedure

- Calculate slope at current point.
One parameter = derivative
Multiple parameters = $\nabla f(\mathbf{b}^k)$
- Move in the direction of negative gradient with step size α^k
- Run update $b^{k+1} \leftarrow b^k - \alpha^k \nabla f(b^k)$
- Repeat until converged