

# HUST

**ĐẠI HỌC BÁCH KHOA HÀ NỘI**  
HANOI UNIVERSITY OF SCIENCE AND TECHNOLOGY

ONE LOVE. ONE FUTURE.



**ĐẠI HỌC**  
**BÁCH KHOA HÀ NỘI**  
HANOI UNIVERSITY  
OF SCIENCE AND TECHNOLOGY

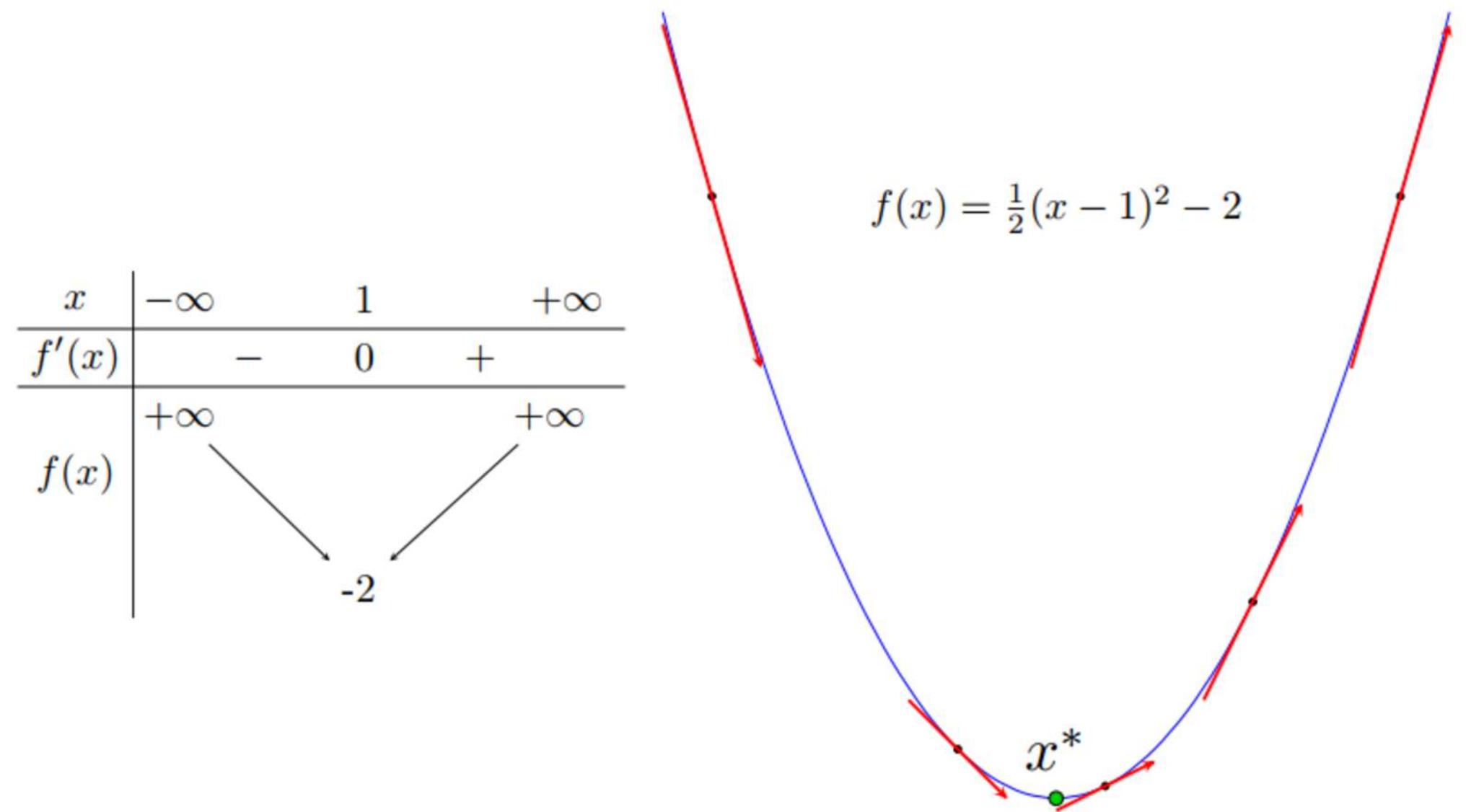
# Gradient Descent

ONE LOVE. ONE FUTURE.

- Gradient Descent
  - Learning rate, Momentum
  - Stochastic Gradient Descent
  - Mini-batch Gradient Descent

# Gradient Descent

- Hướng tiếp cận phổ biến nhất để giải một bài toán tối ưu là xuất phát từ một điểm gần với nghiệm của bài toán, sau đó dùng một phép toán lặp để tiến dần đến điểm cần tìm

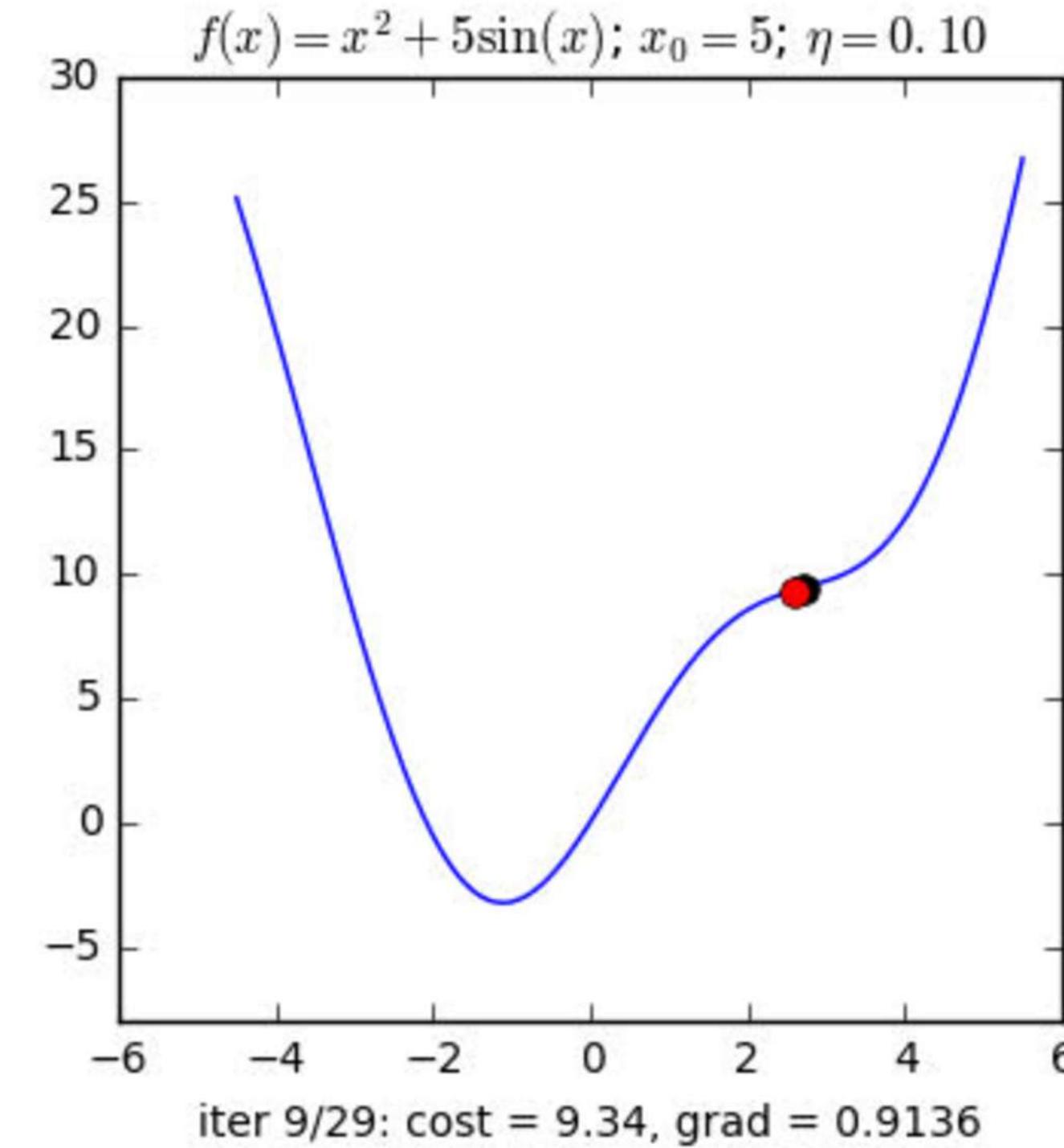
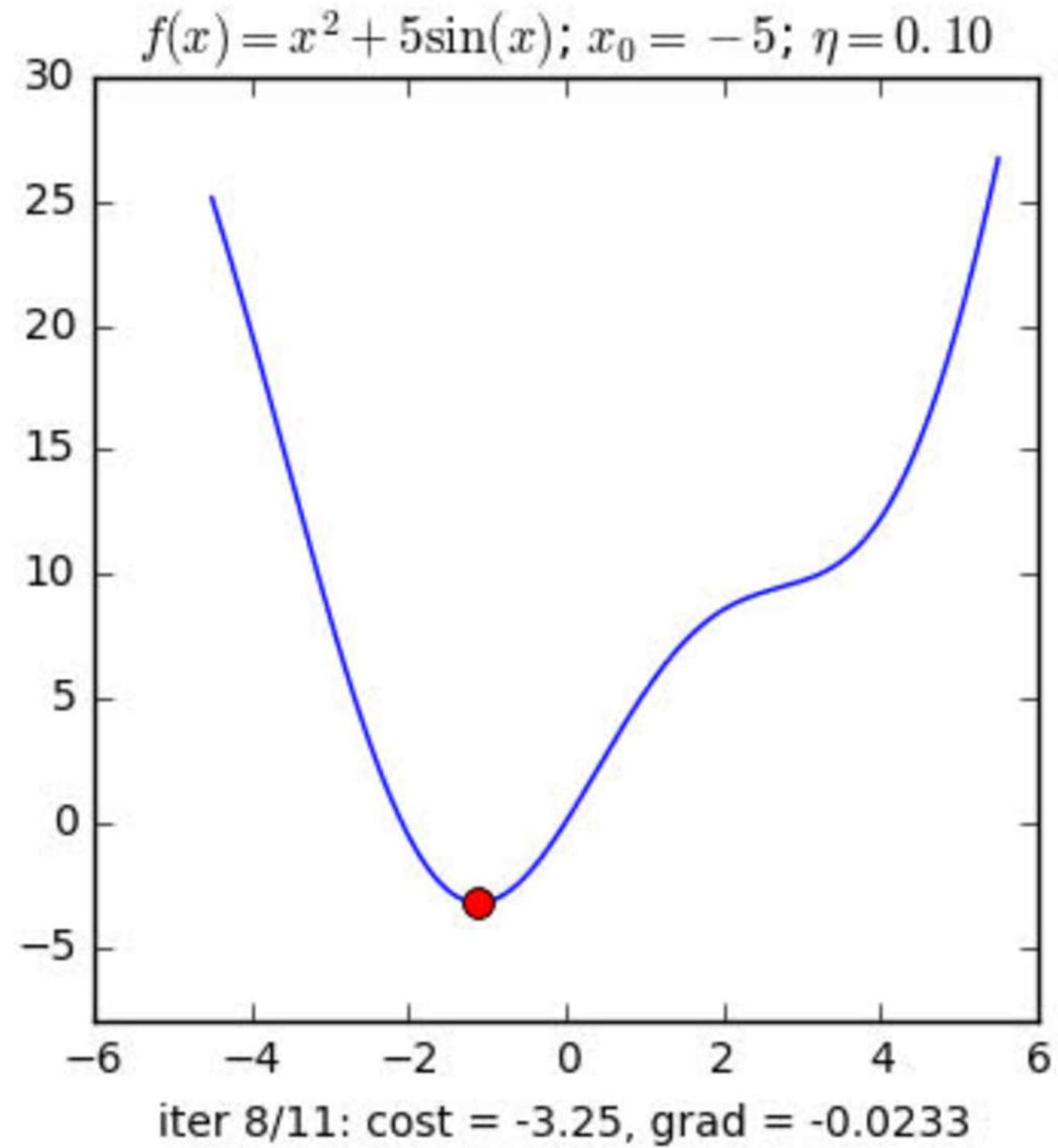


- Ta cần di chuyển ngược dấu với đạo hàm

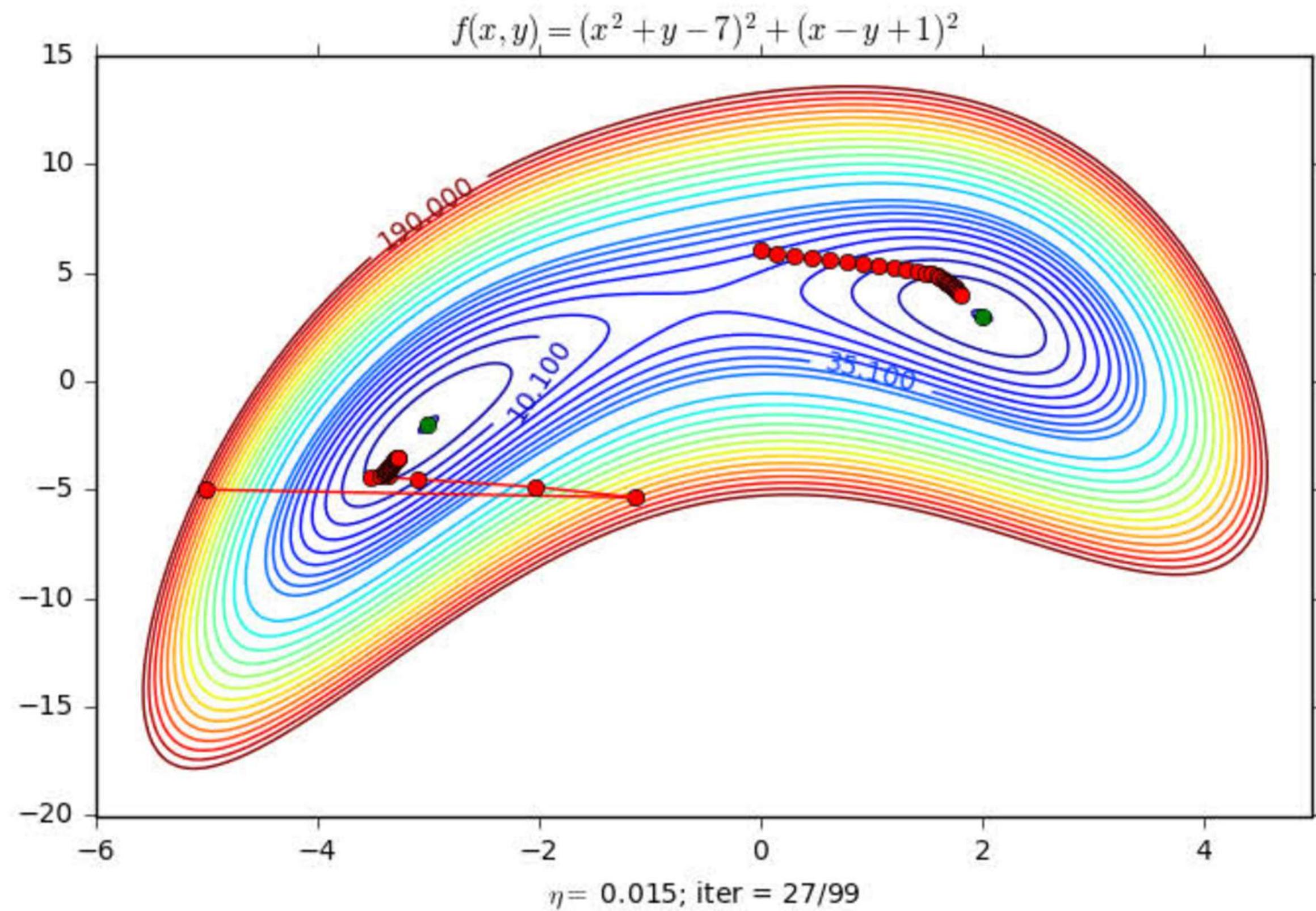
$$\theta_{t+1} = \theta_t - \eta \nabla_{\theta} f(\theta_t)$$

- Tốc độ hội tụ của Gradient Descent phụ thuộc vào điểm khởi tạo ban đầu và tốc độ học (learning rate)

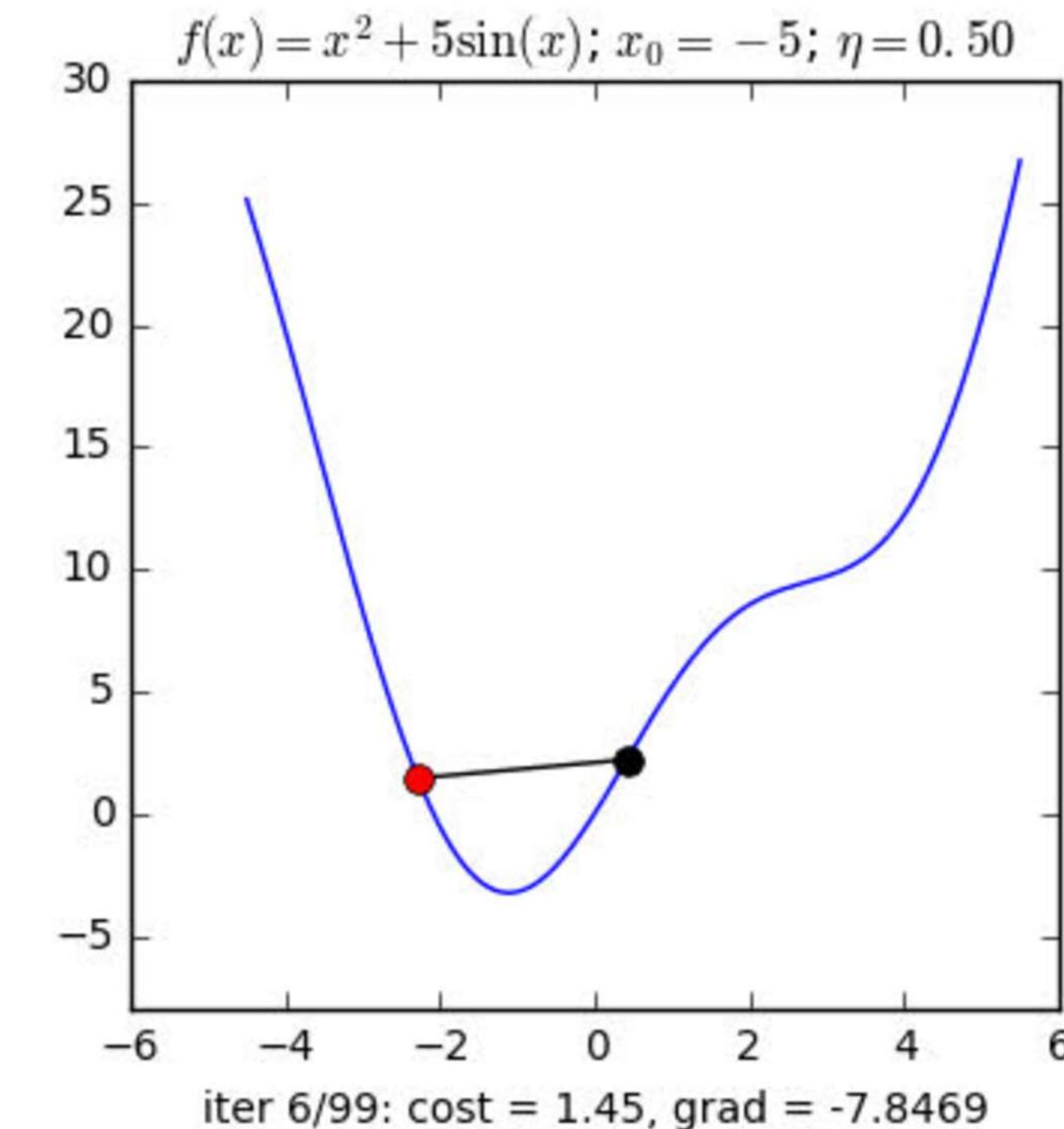
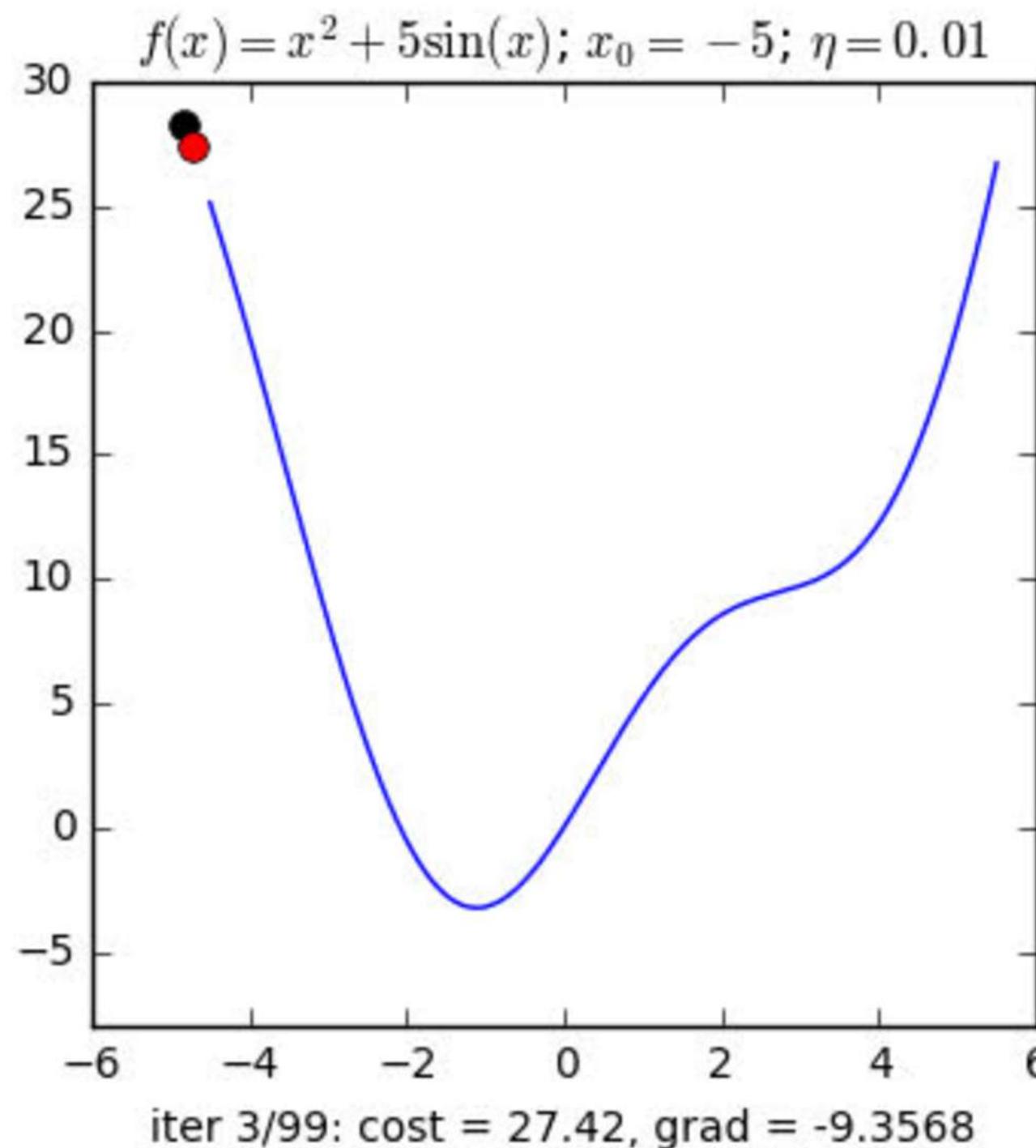
# Gradient Descent



# Gradient Descent

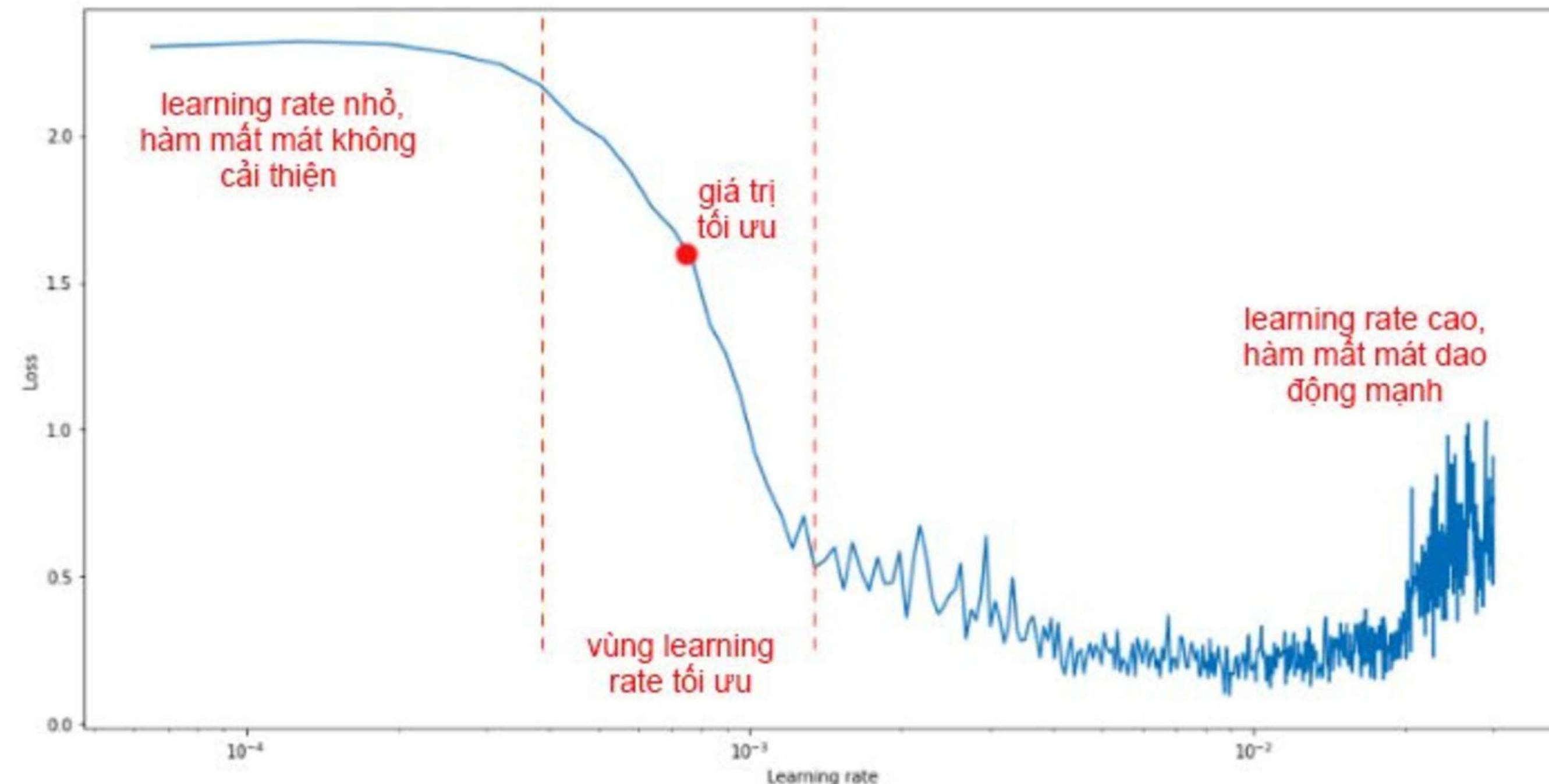


# Learning rate

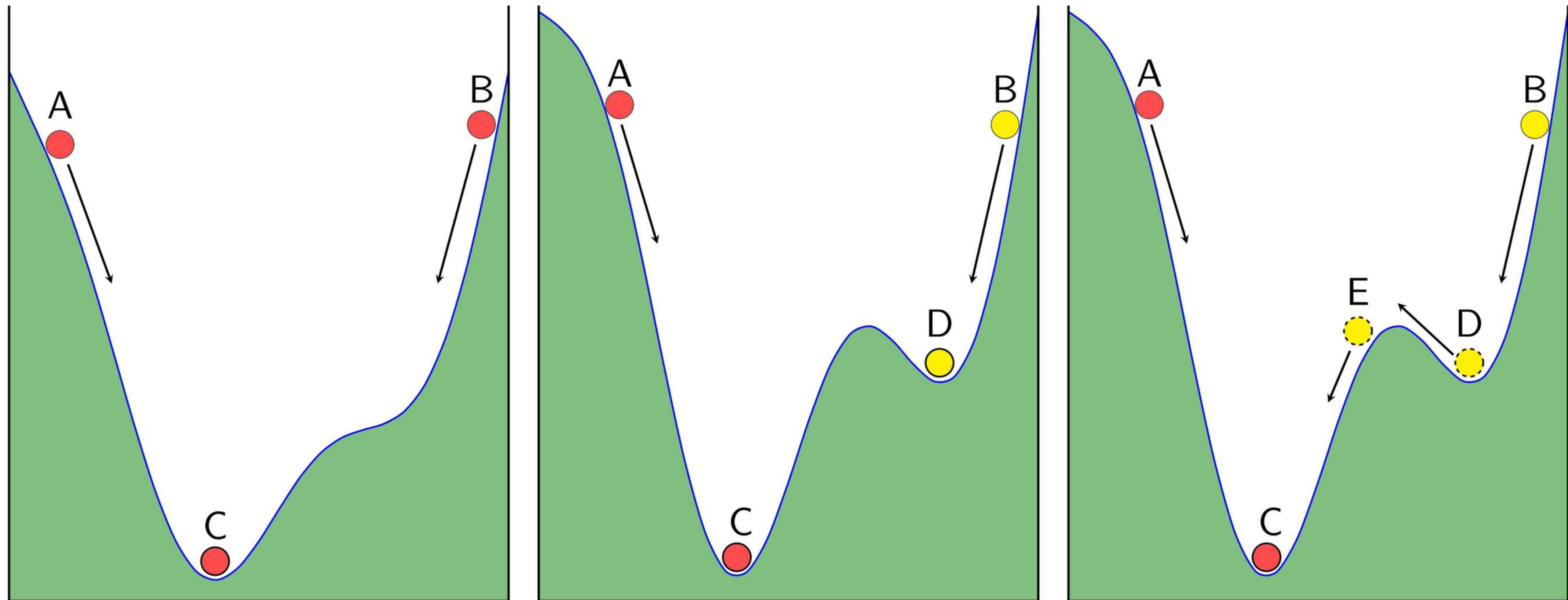


# Learning rate

- Learning rate Finder



# Momentum



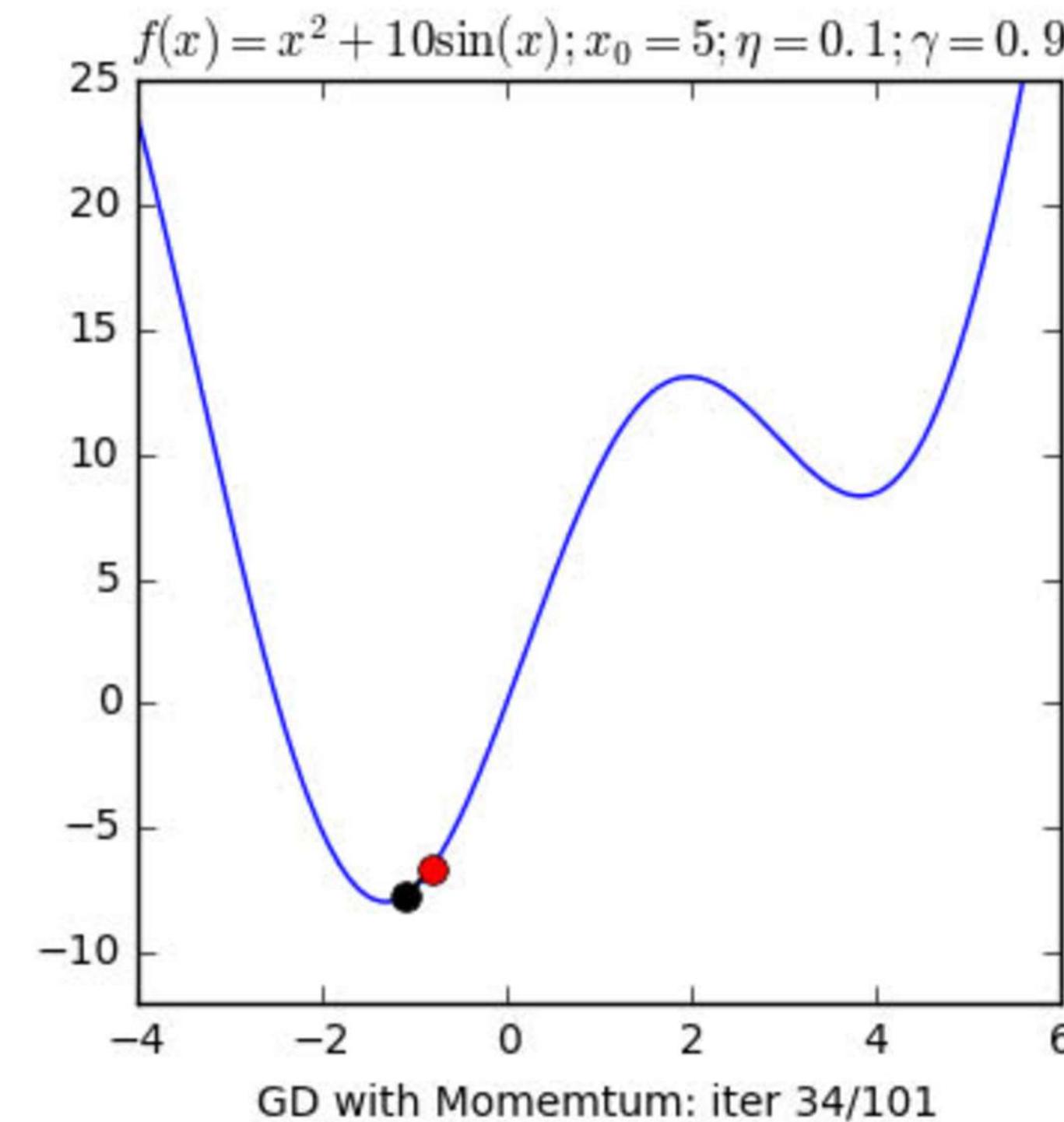
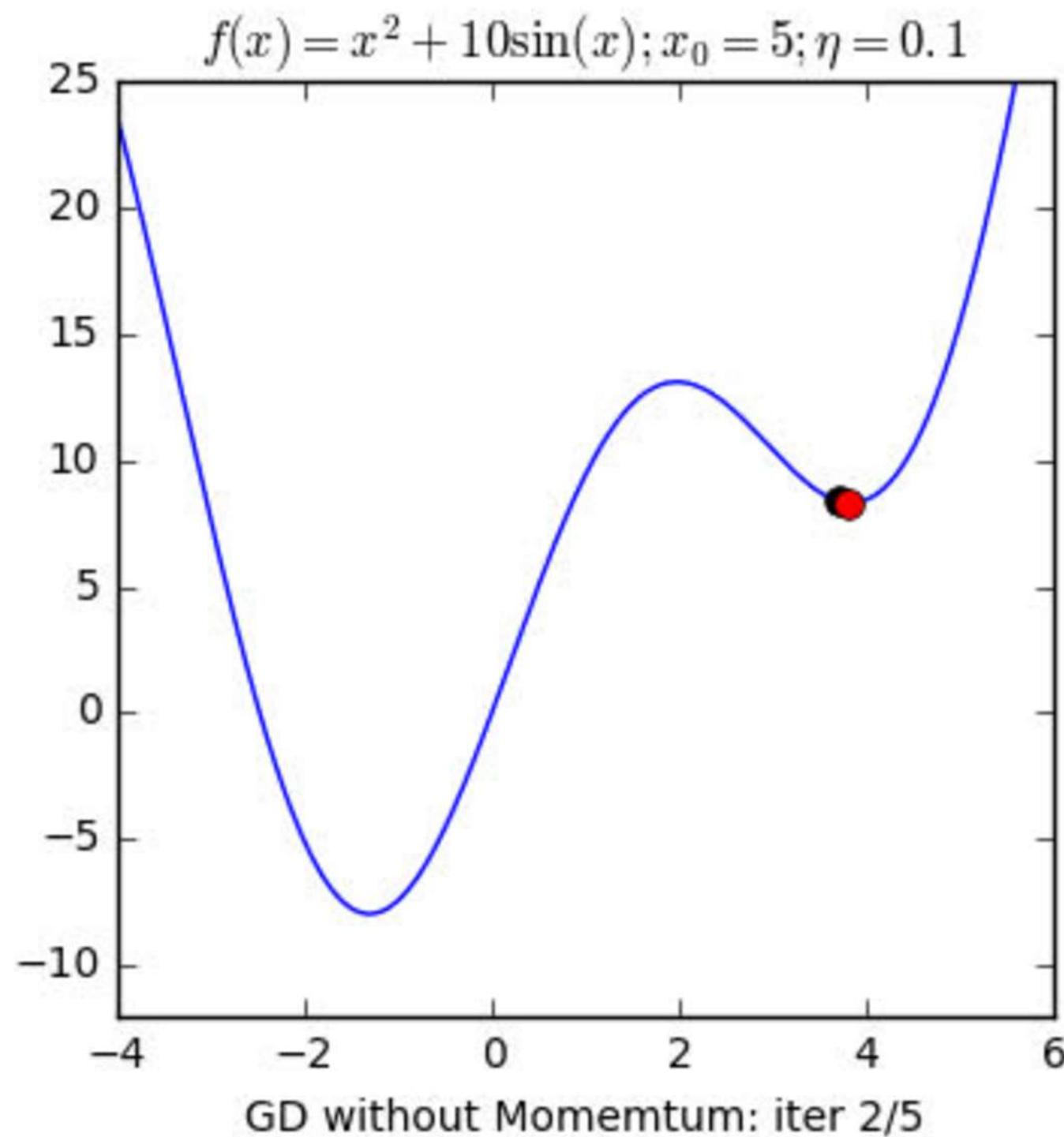
- Vận tốc tại thời điểm t:

$$v_t = \gamma v_{t-1} + \eta \nabla_{\theta} J(\theta)$$

- Vị trí tại thời điểm t:

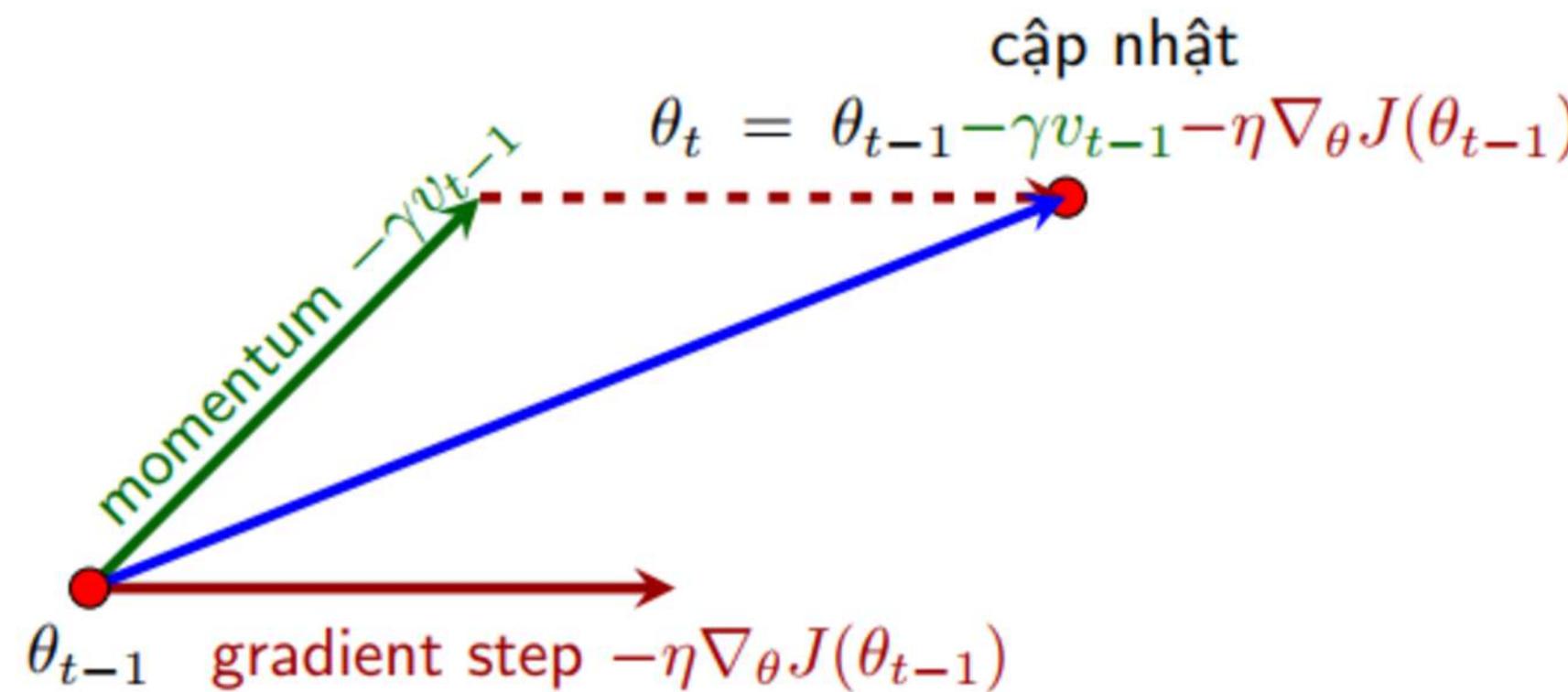
$$\theta \leftarrow \theta - v_t = \theta - \eta \nabla_{\theta} J(\theta) - \gamma v_{t-1}$$

# Momentum

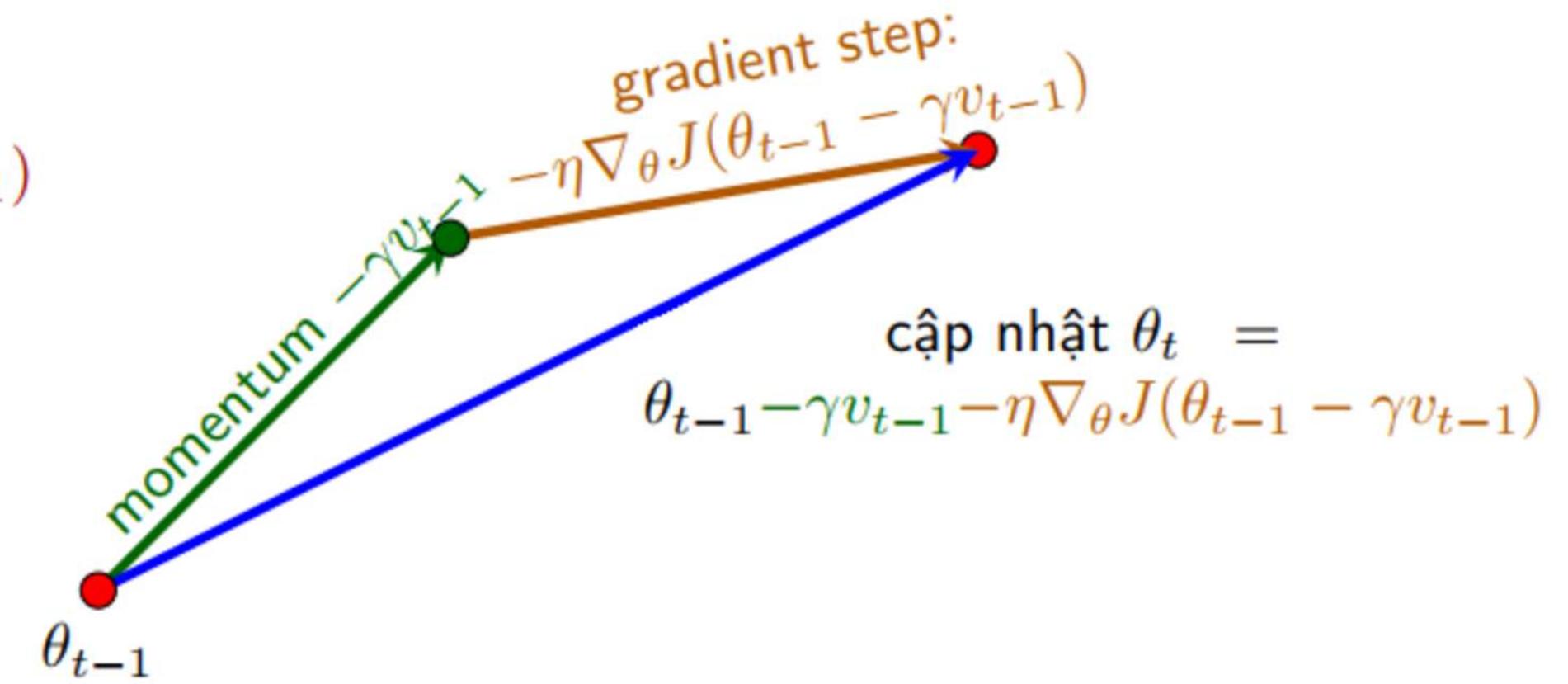


# Momentum

- Phương pháp Nesterov accelerated gradient (NAG):

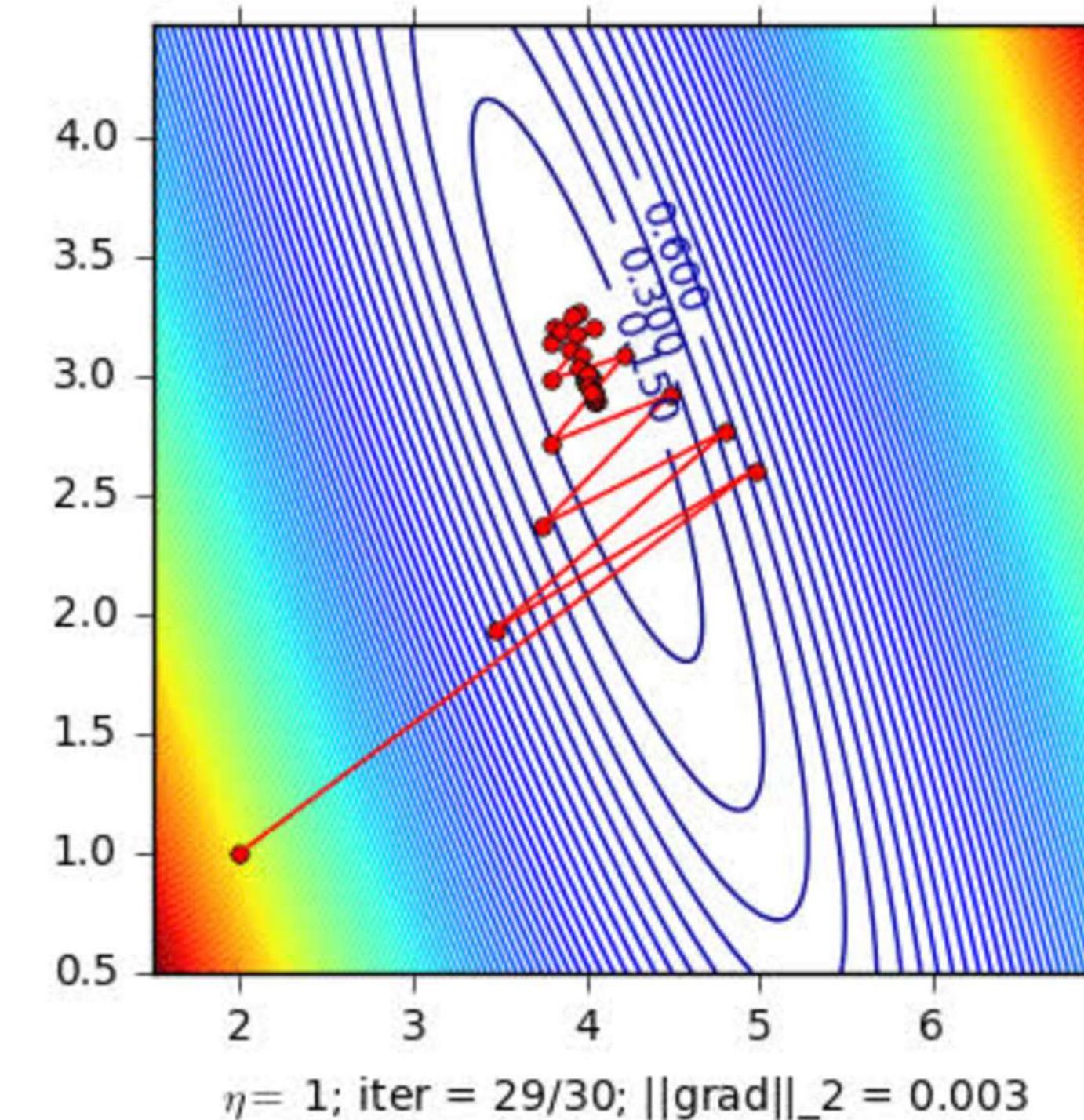
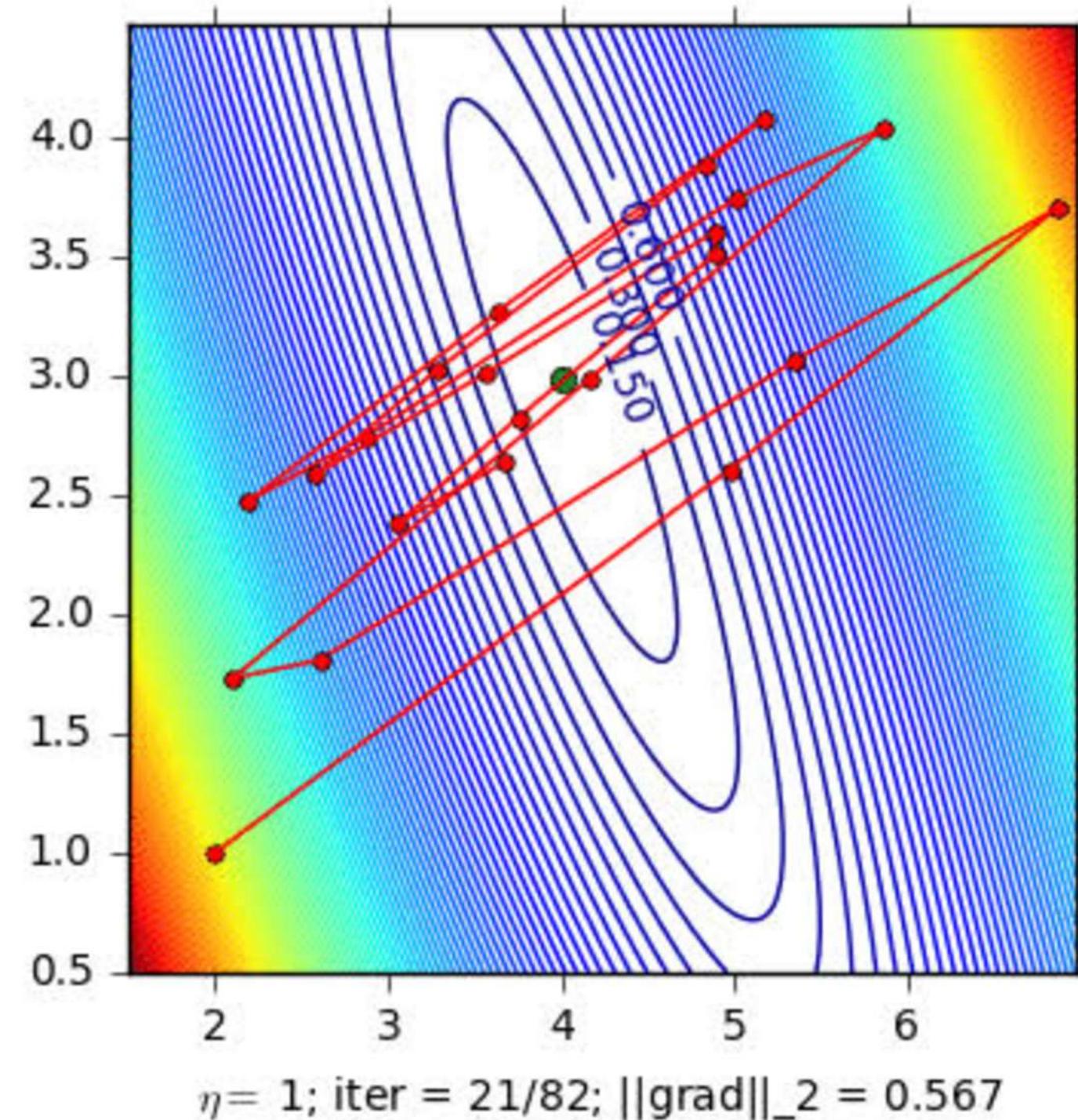


(a) Momentum gradient descent.

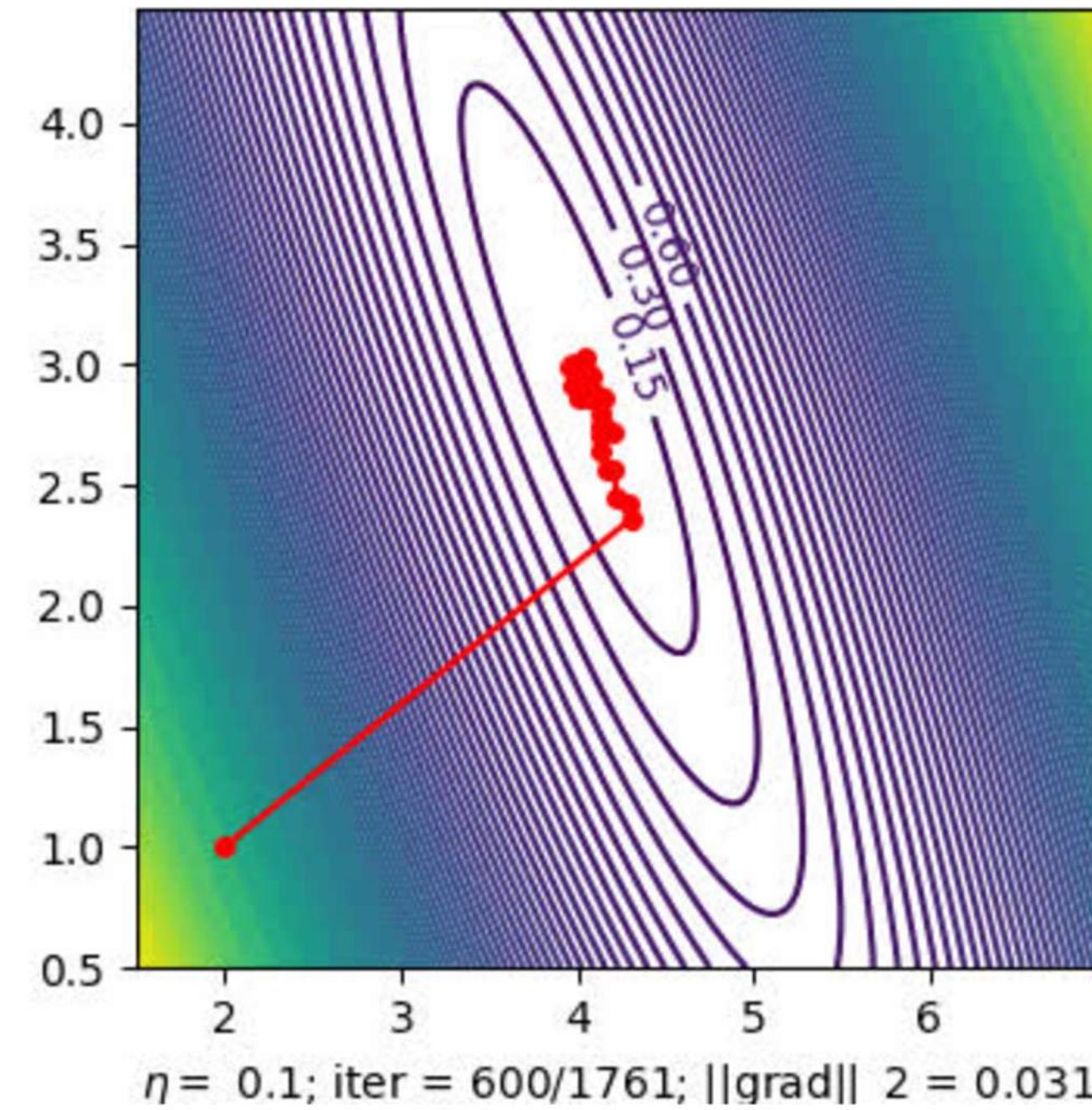
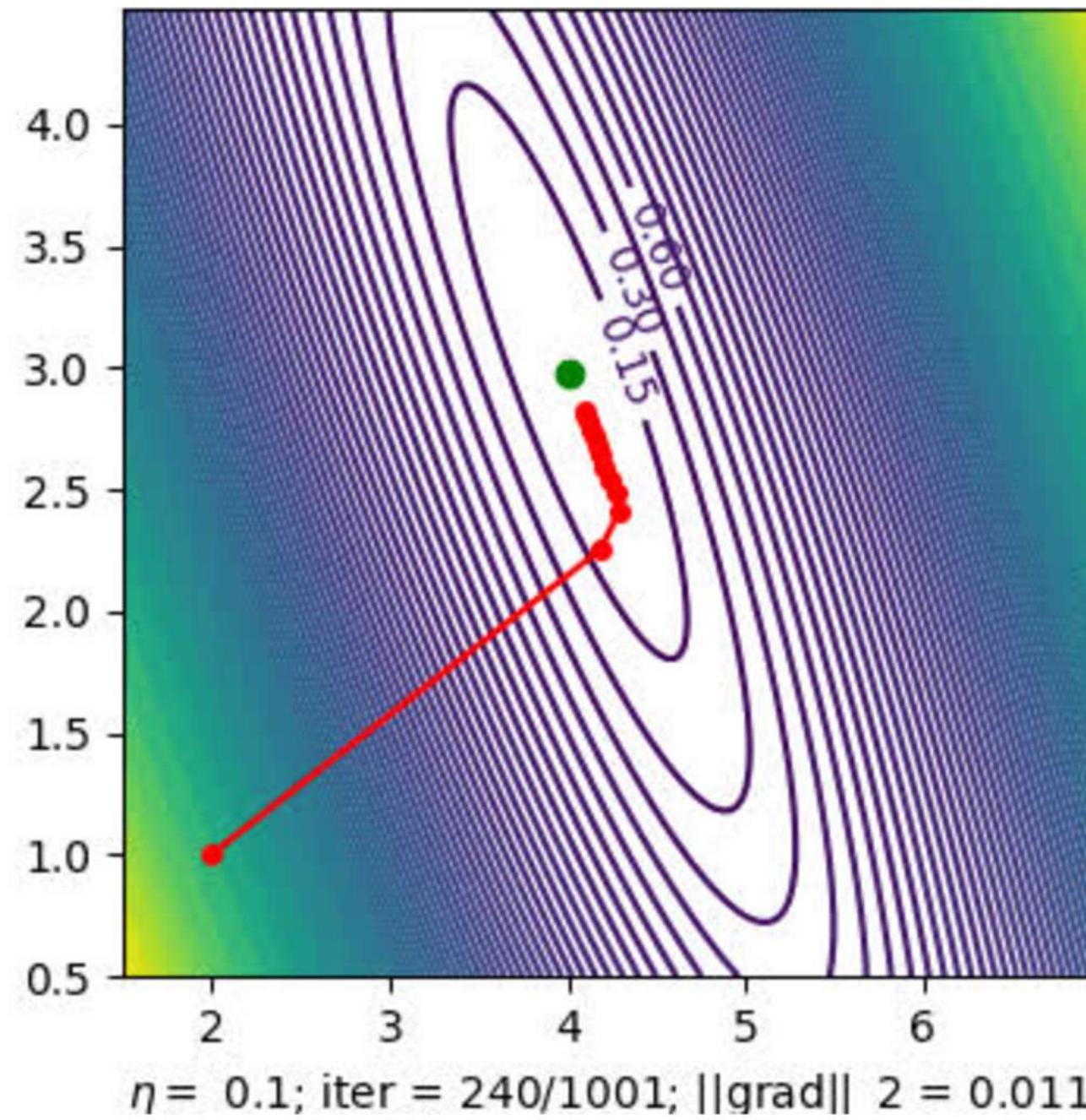


(b) Nesterov accelerated gradient.

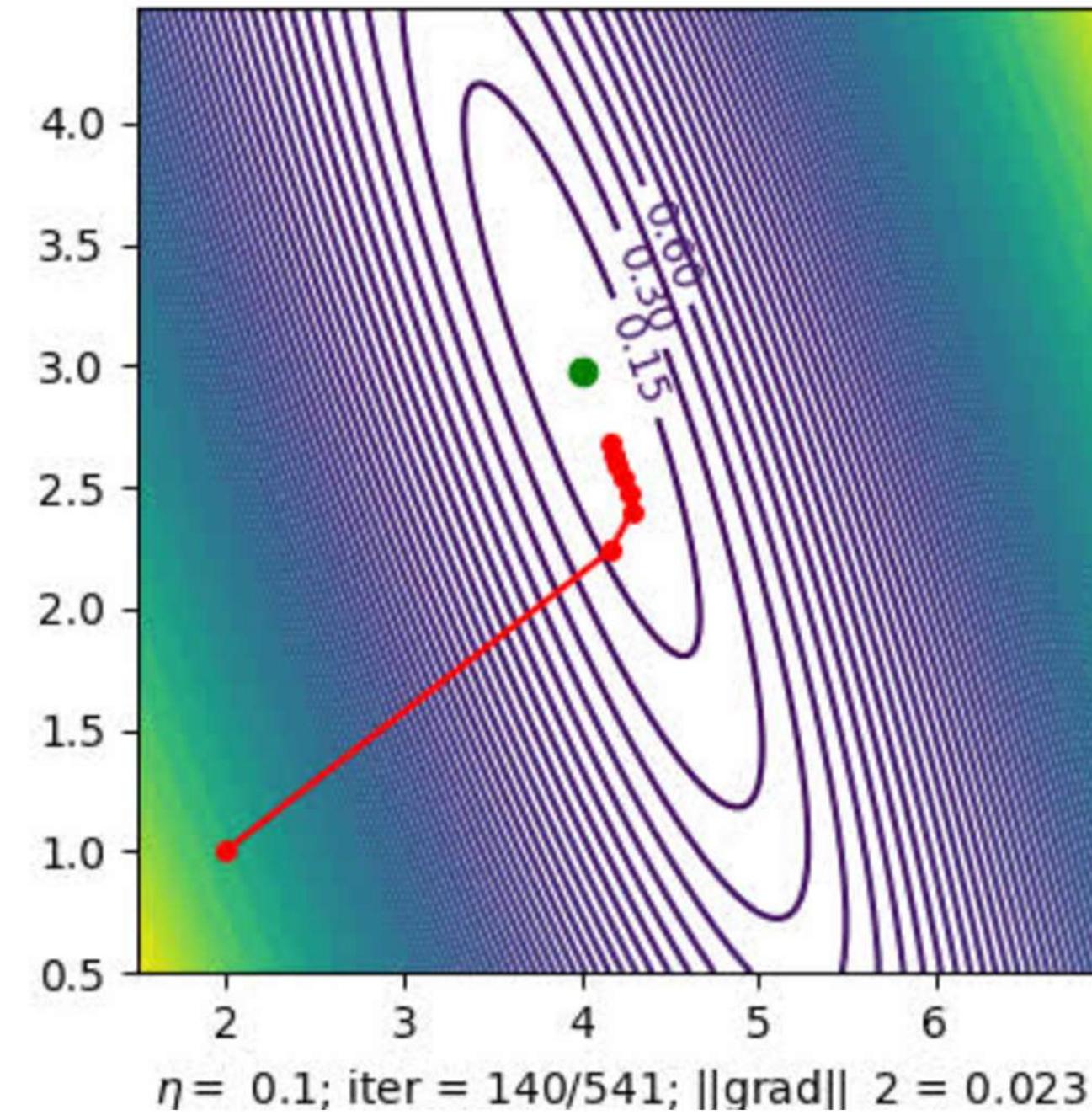
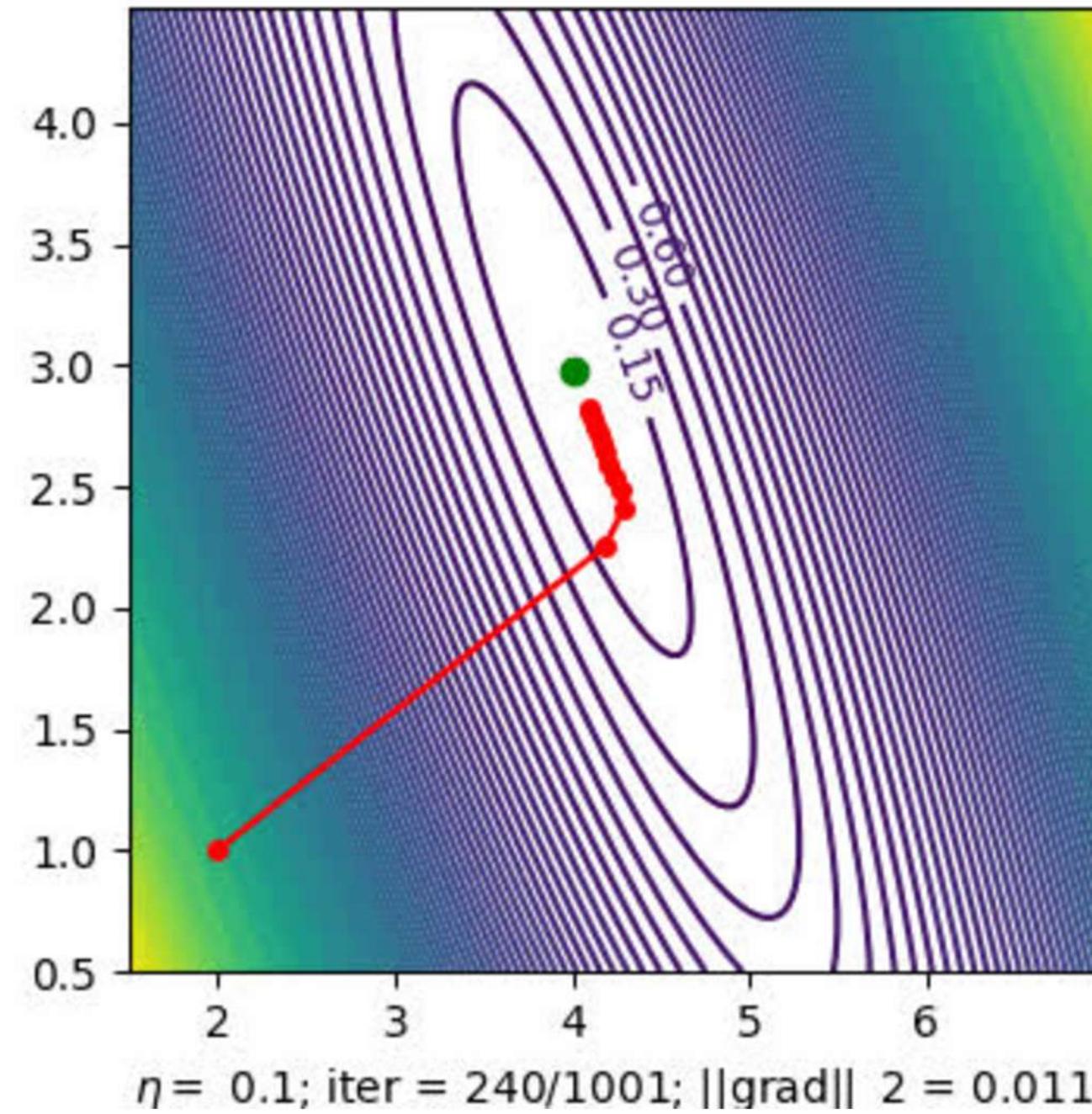
# Momentum



# Stochastic Gradient Descent

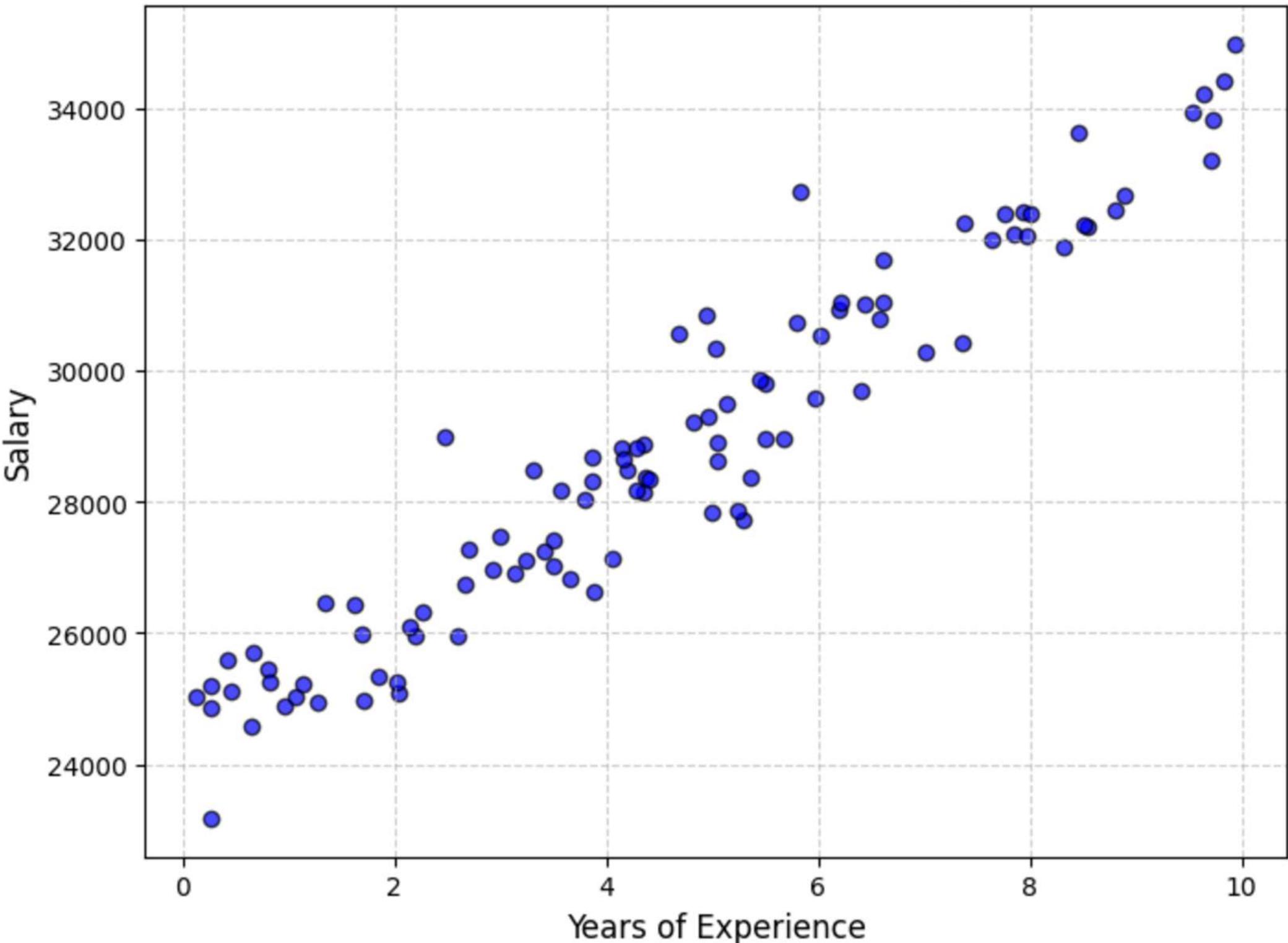


# Mini-batch Gradient Descent



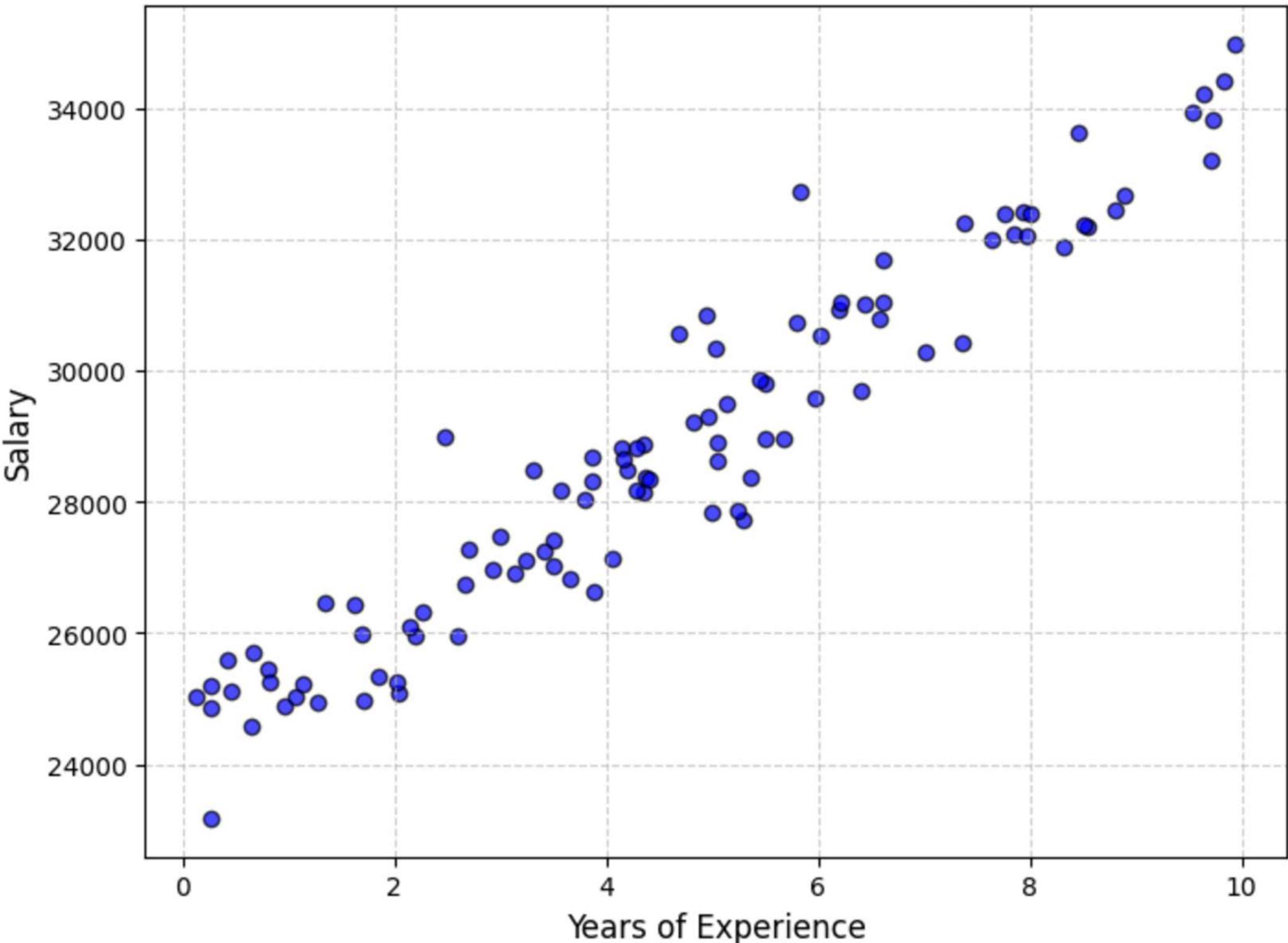
# Ứng dụng bài toán thực tế

- Bộ dữ liệu: Salary Dataset
- Gồm dữ liệu 100 người về số năm kinh nghiệm trong ngành và lương hằng năm



# Ứng dụng bài toán thực tế

- Bộ dữ liệu: Salary Dataset
- Gồm dữ liệu 100 người về số năm kinh nghiệm trong ngành và lương hằng năm



- Hàm mất mát

$$y \approx \hat{y} = f(\mathbf{x}) = w_1x + w_2 = \mathbf{x}^T \mathbf{w}$$

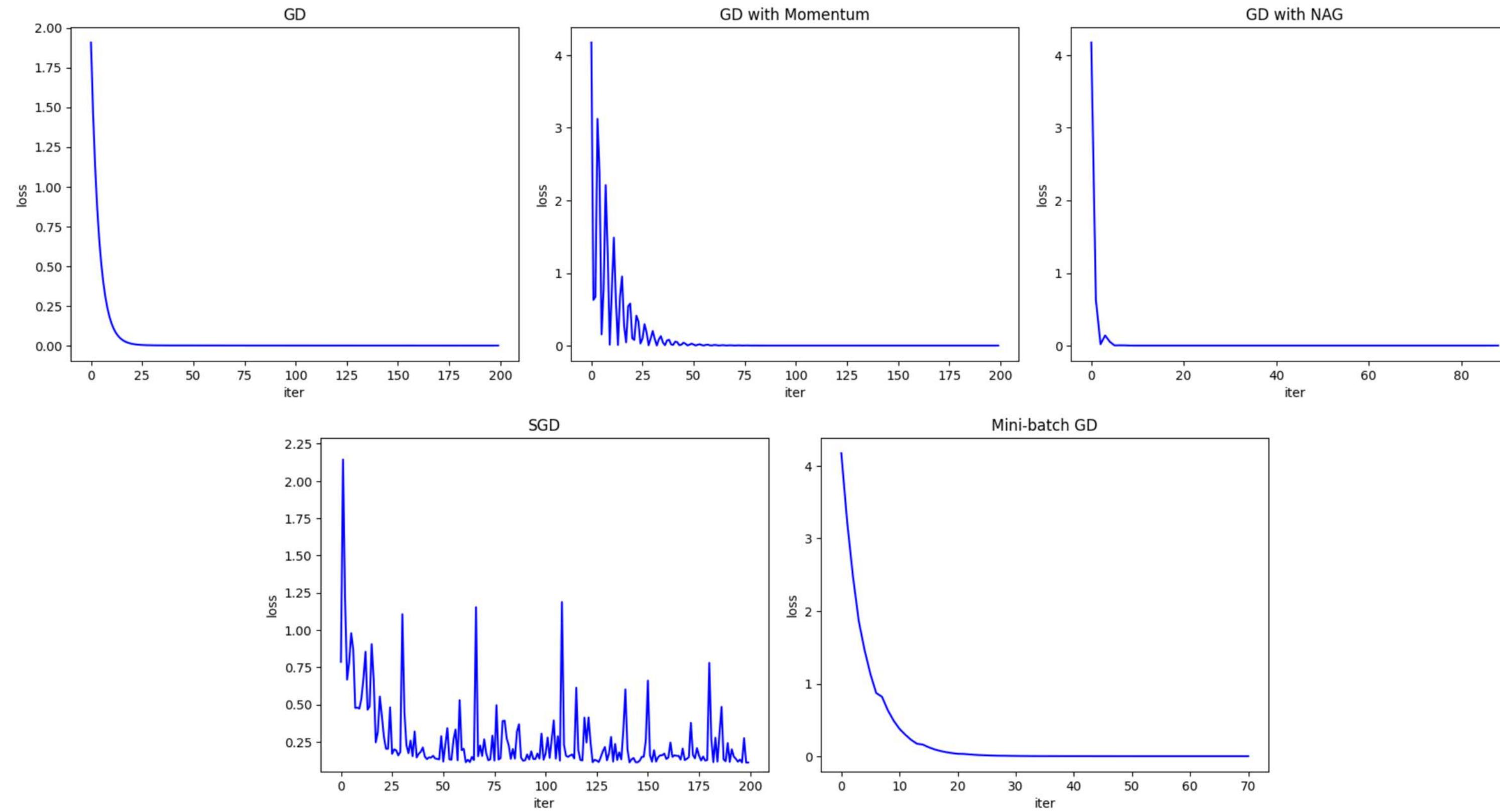
$$\mathcal{L}(\mathbf{w}) = \frac{1}{2N} \sum_{i=1}^N (y_i - \mathbf{x}_i^T \mathbf{w})^2 = \frac{1}{2N} \|\mathbf{y} - \mathbf{X}^T \mathbf{w}\|_2^2$$

$$\frac{\nabla \mathcal{L}(\mathbf{w})}{\nabla \mathbf{w}} = \frac{1}{N} \mathbf{X} (\mathbf{X}^T \mathbf{w} - \mathbf{y})$$

- Công thức cập nhật

$$\mathbf{w}_{t+1} = \mathbf{w}_t - \frac{\eta}{N} \mathbf{X} (\mathbf{X}^T \mathbf{w}_t - \mathbf{y})$$

# Ứng dụng bài toán thực tế



# Ứng dụng bài toán thực tế

|                                  | Number of Epochs | MSE        |
|----------------------------------|------------------|------------|
| Gradient Descent                 | 3801             | 0.02159802 |
| Gradient Descent (with Momentum) | 230              | 0.02191542 |
| Gradient Descent (with NAG)      | 83               | 0.02780732 |
| Stochastic Gradient Descent      | 13               | 0.02356953 |
| Mini-batch Gradient Descent      | 11               | 0.02436395 |



A large, faint watermark of the HUST logo is visible across the entire background of the slide.

**HUST**

**THANK YOU !**