



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

Optimizers

ONE LOVE. ONE FUTURE.

- Gradient Descent, Learning rate, Momentum
- Các Optimizers phổ biến
 - Adagrad
 - RMSProp
 - Adam
- Các Optimizers hiện đại
- Ví dụ minh họa

- Để giải một bài toán tối ưu, cách tiếp cận phổ biến nhất 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
- Công thức cập nhật của Gradient Descent:

$$\theta_{i,t+1} = \theta_{i,t} - \eta \frac{\partial J}{\partial \theta_i}$$

- Công thức cập nhật của Gradient Descent với Momentum:

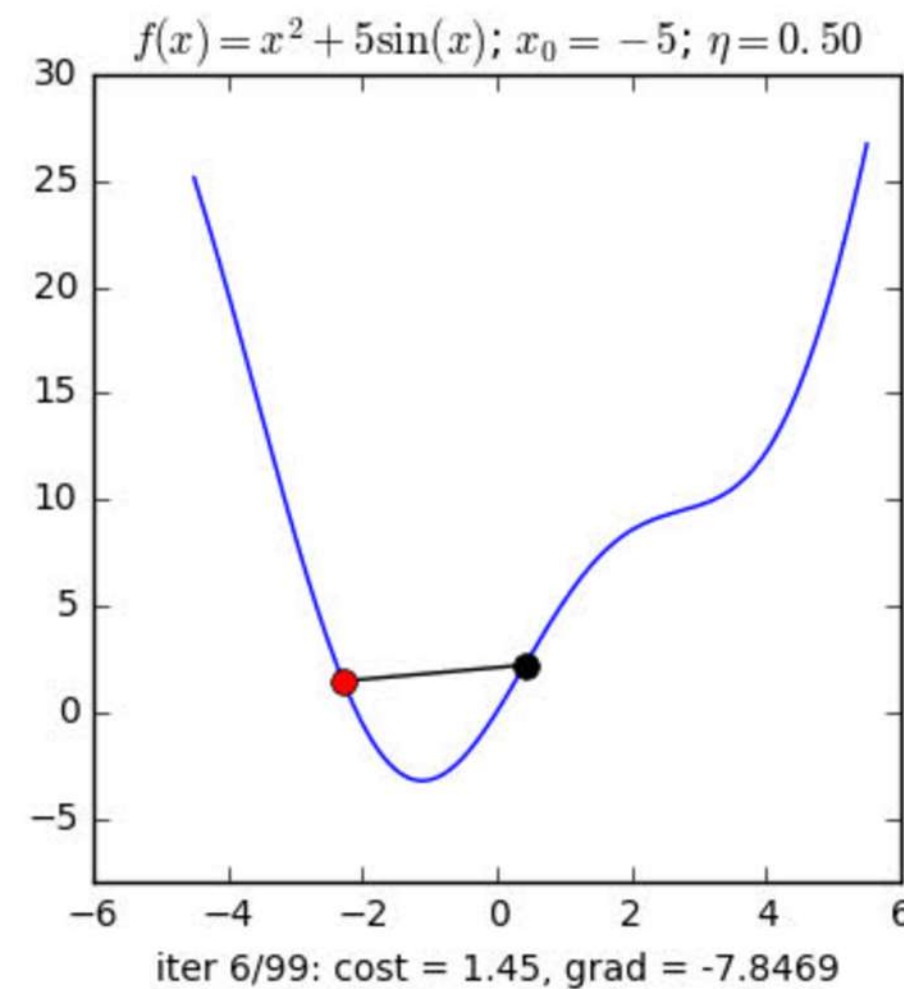
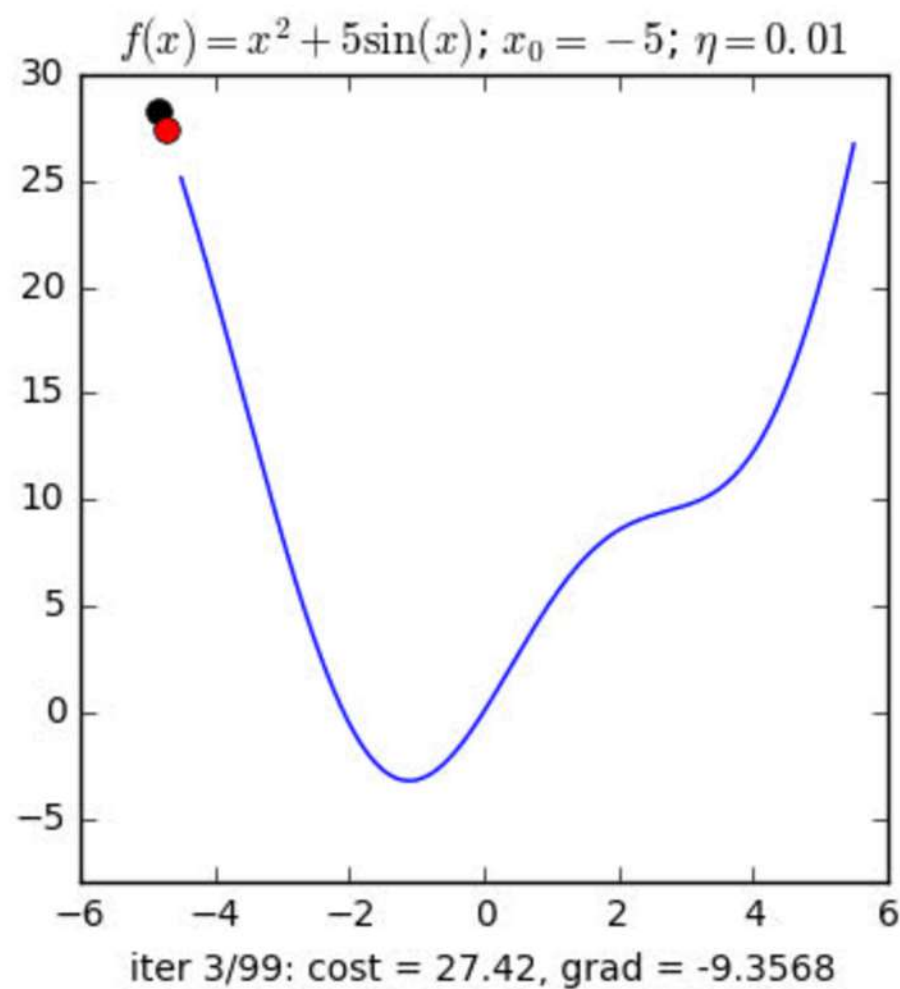
$$\theta_{i,t+1} = \theta_{i,t} - v_t = \theta_{i,t} - \eta \frac{\partial J}{\partial \theta_i} - \gamma v_{t-1}$$

- Công thức cập nhật của Gradient Descent với NAG:

$$\theta_{i,t+1} = \theta_{i,t} - v_t = \theta_{i,t} - \eta \frac{\partial J}{\partial \theta_i} (\theta_{i,t} - \gamma v_{t-1}) - \gamma v_{t-1}$$

Gradient Descent, Learning rate, Momentum

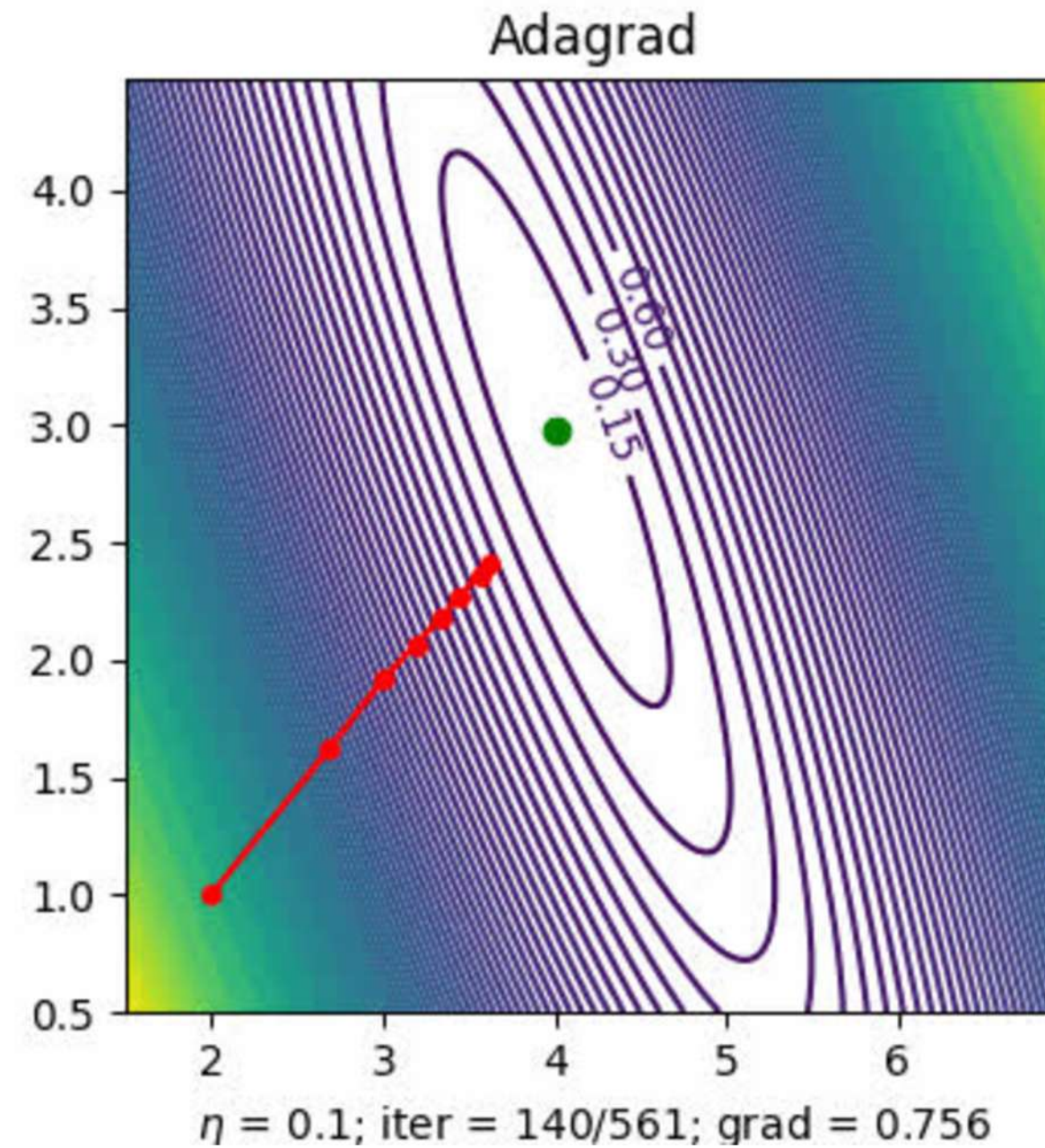
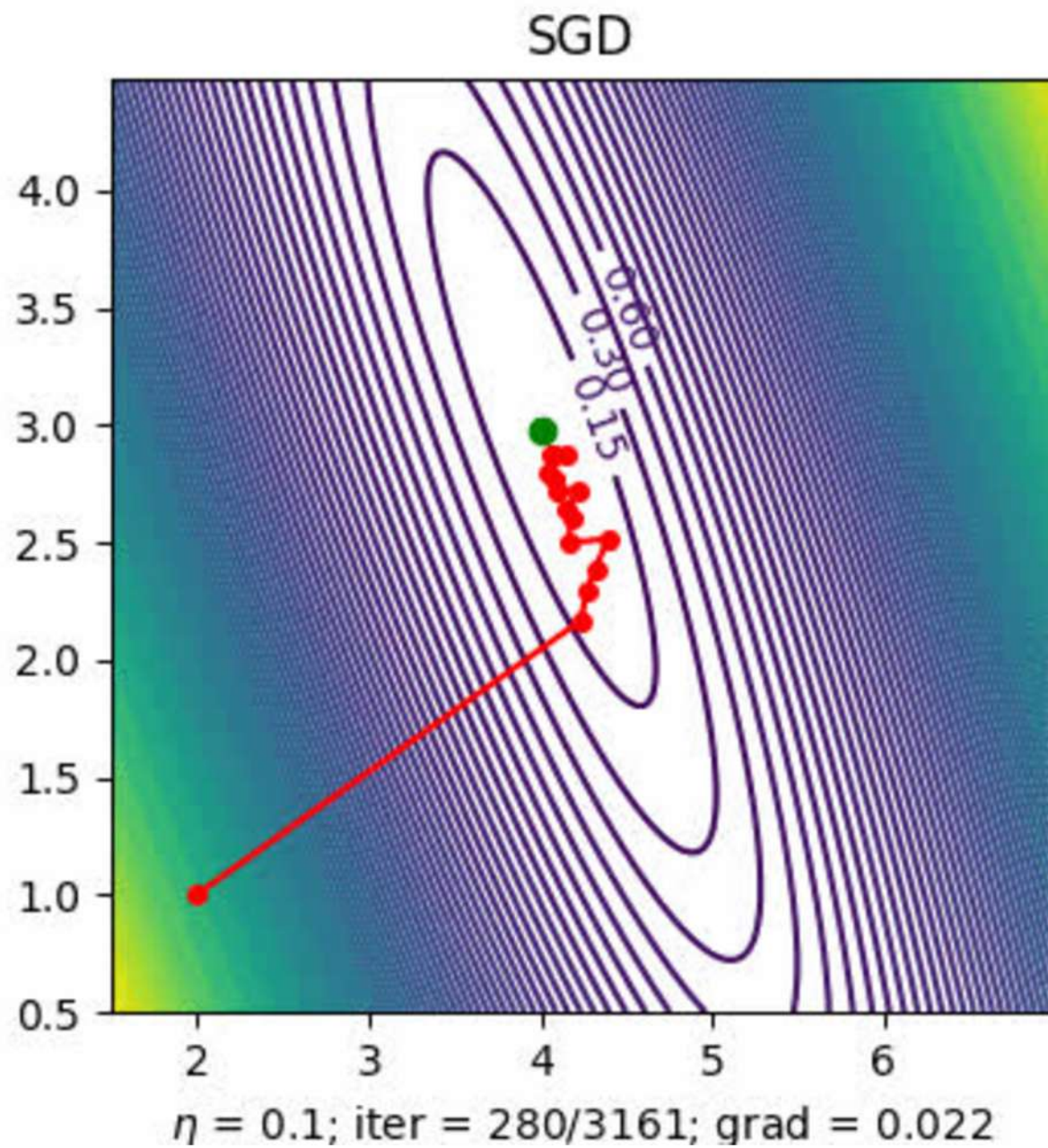
- Hạn chế của Gradient Descent:
 - Learning rate cố định



- Mỗi tham số đều cập nhật với cùng một learning rate

Adagrad

- Adagrad (Adaptive Gradient Algorithm) điều chỉnh learning rate riêng biệt cho từng tham số



- Thuật toán Adagrad:

- Gradient của trọng số thứ i ở thời điểm t :

$$g_{i,t} = \frac{\partial J}{\partial \theta_i}$$

- Tổng bình phương các gradient của trọng số thứ i :

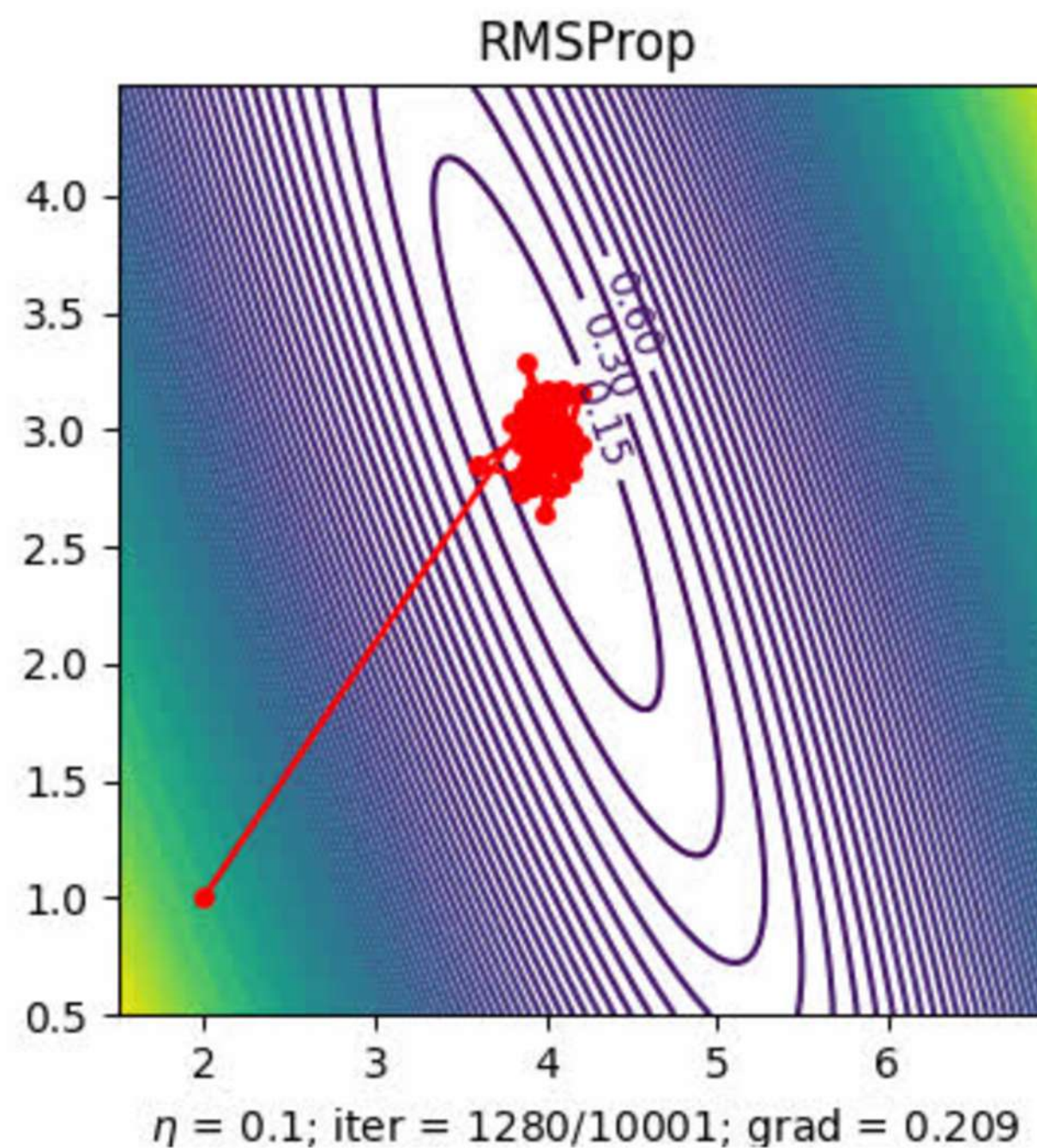
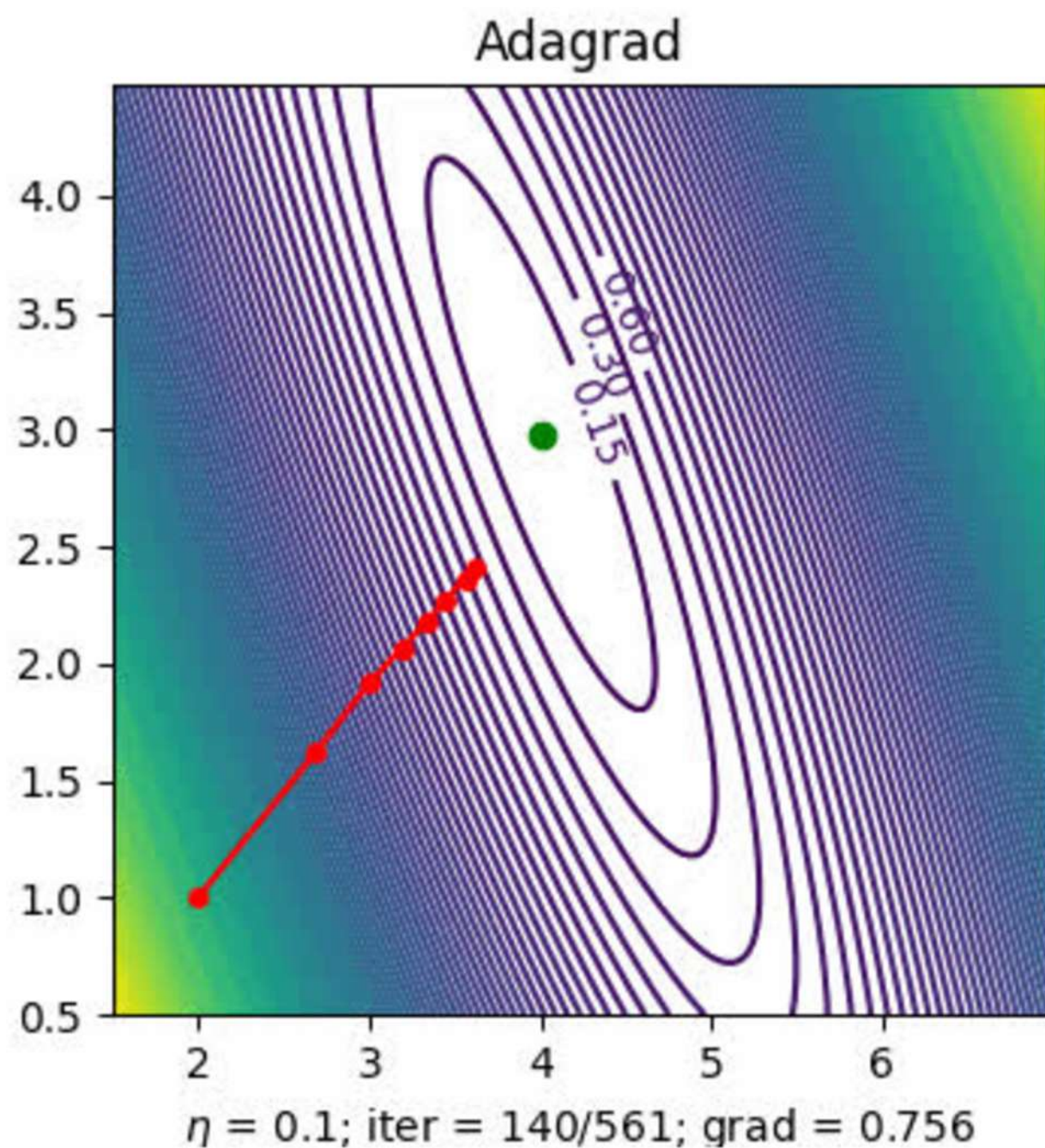
$$G_{i,t} = G_{i,t-1} + [g_{i,t}]^2$$

- Công thức cập nhật:

$$\theta_{i,t+1} = \theta_{i,t} - \frac{\eta}{\sqrt{G_{i,t} + \varepsilon}} g_{i,t}$$

- Hạn chế của Adagrad:
 - Tổng bình phương các gradient cộng dồn vô hạn theo thời gian
 - Learning rate giảm dần về 0, các trọng số gần như ngừng cập nhật
 - Với dữ liệu dày đặc, Adagrad dễ hội tụ sớm và dừng ở nghiệm chưa tối ưu

- RMSProp (Root Mean Square Propagation) thay vì cộng dồn tất cả gradient như Adagrad thì dùng trung bình trượt



- Thuật toán RMSProp:

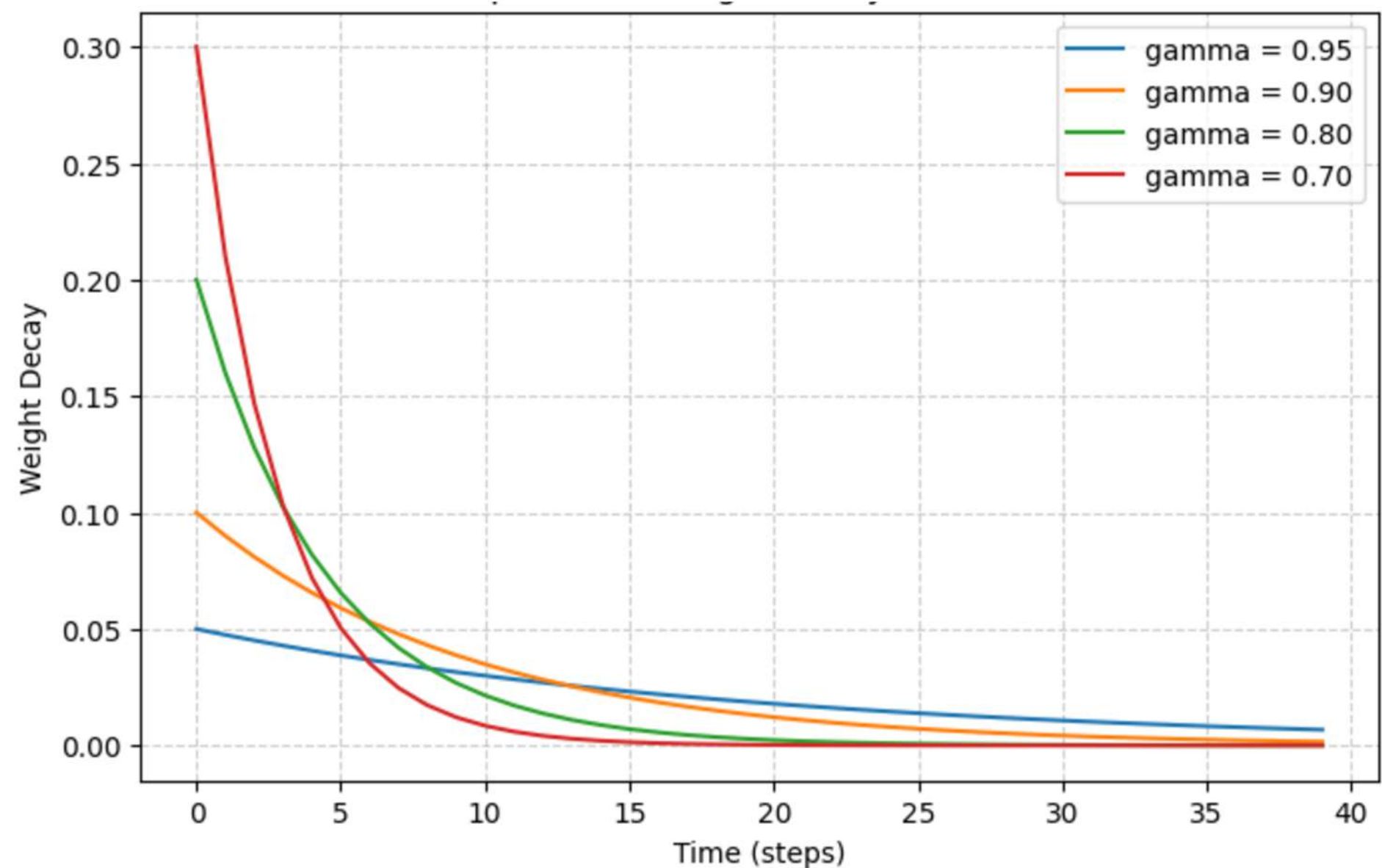
- Trung bình trượt tổng bình phương các gradient của trọng số thứ i :

$$E[g_{i,t}^2] = \gamma E[g_{i,t-1}^2] + (1 - \gamma)[g_{i,t}]^2$$

- Công thức cập nhật:

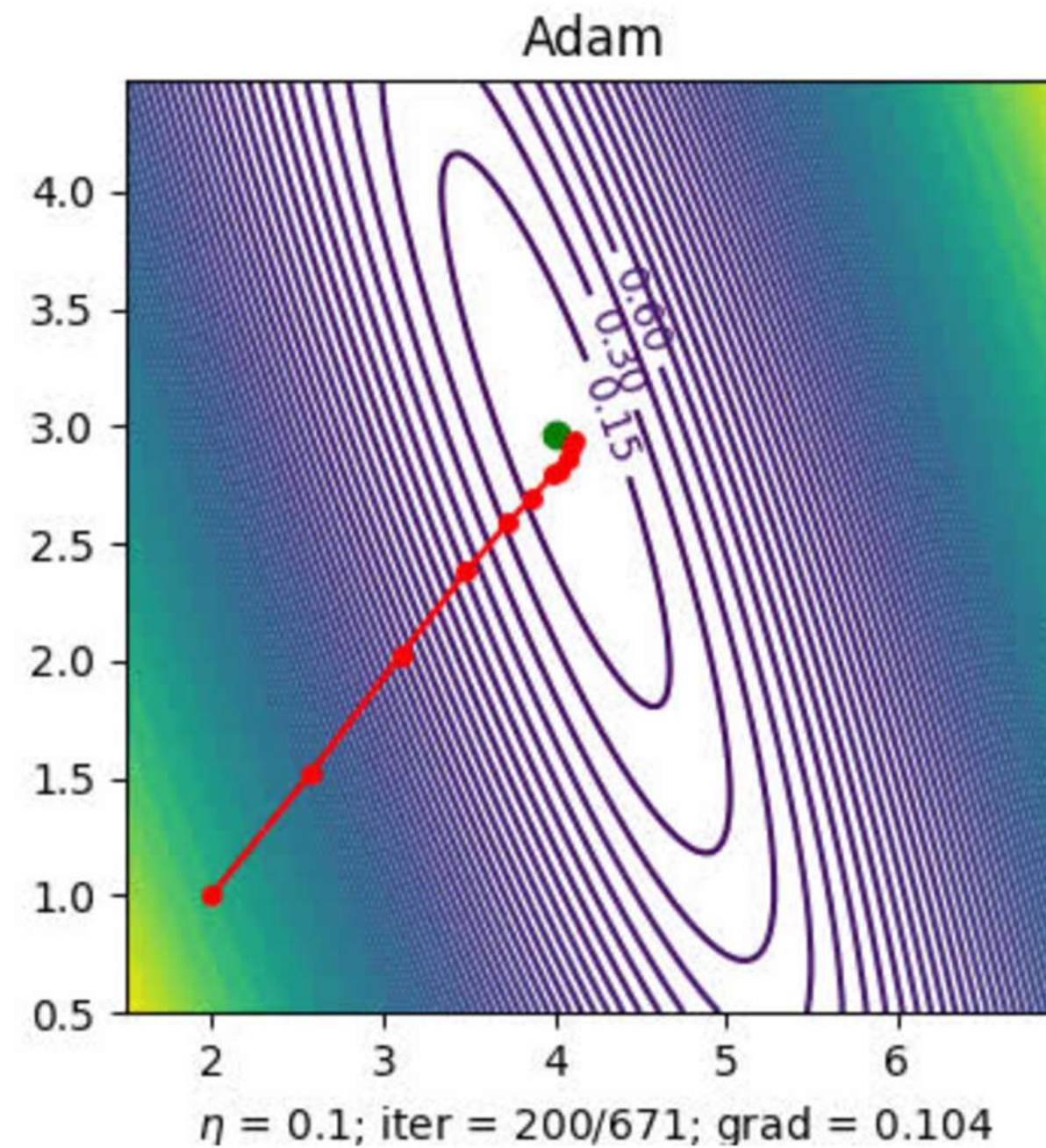
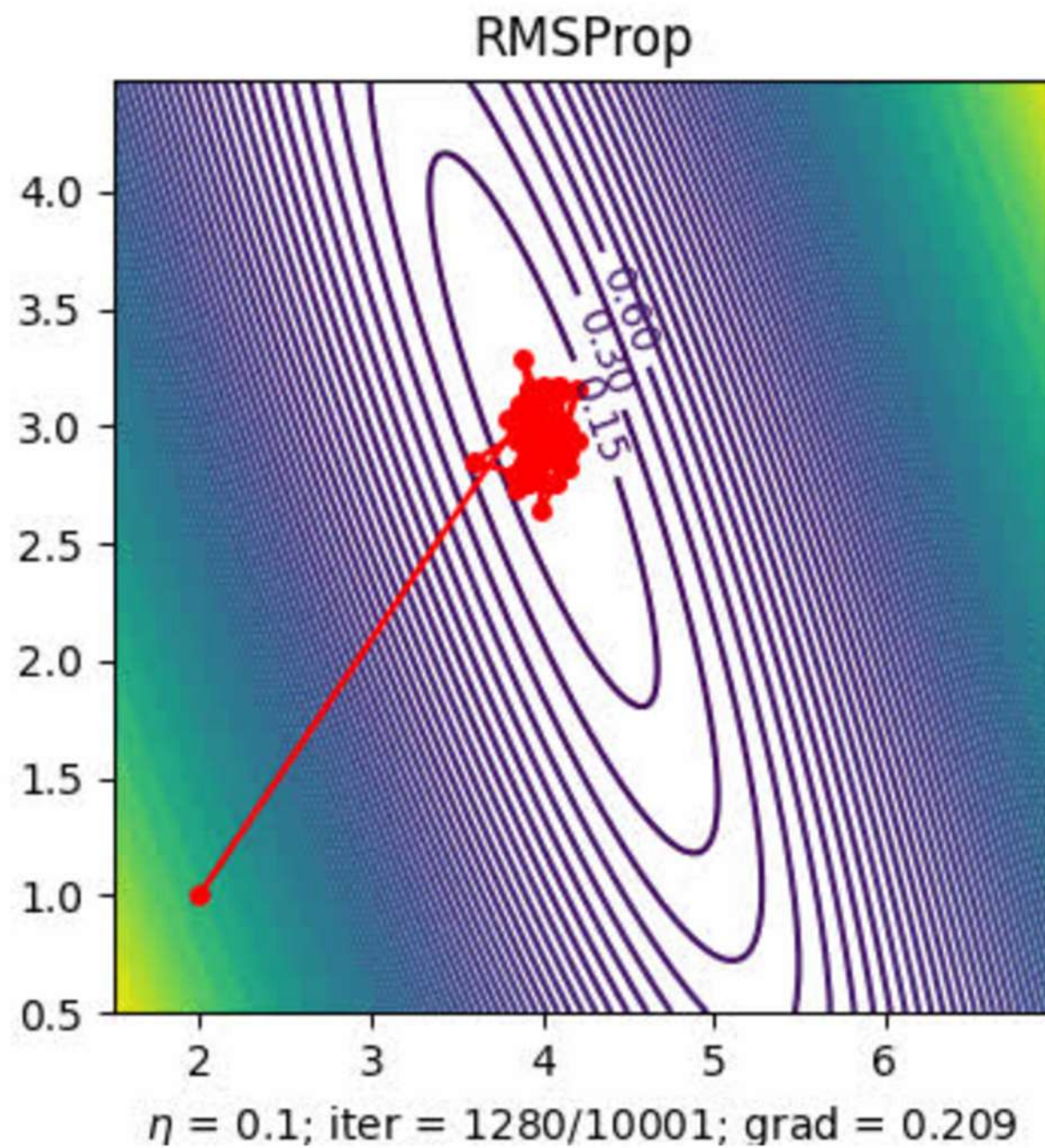
$$\theta_{i,t+1} = \theta_{i,t} - \frac{\eta}{\sqrt{E[g_{i,t}^2] + \varepsilon}} g_{i,t}$$

- Hệ số gamma gọi là “tỉ lệ quên” của bình phương gradient:
 - Gamma lớn: Nhớ lâu hơn, giảm nhiều tốt
 - Gamma nhỏ: Phản ứng tốt với thay đổi, dễ bị nhiễu



- Hạn chế của RMSProp:
 - Không xử lý dữ liệu thưa tốt bằng Adagrad
 - Nhạy với tham số learning rate và gamma, không có quy tắc chọn tối ưu mà phải thông qua thực nghiệm
 - RMSProp hội tụ nhanh hơn Adagrad nhưng vẫn có nguy cơ mắc kẹt ở tối ưu cục bộ

- Adam (Adaptive Moment Estimation) là sự kết hợp của RMSProp và Momentum



- Thuật toán Adam:

- Trung bình động của gradient bậc 1 và 2:

$$m_{i,t} = \beta_1 m_{i,t-1} + (1 - \beta_1) g_{i,t}$$
$$v_{i,t} = \beta_2 v_{i,t-1} + (1 - \beta_2) [g_{i,t}]^2$$

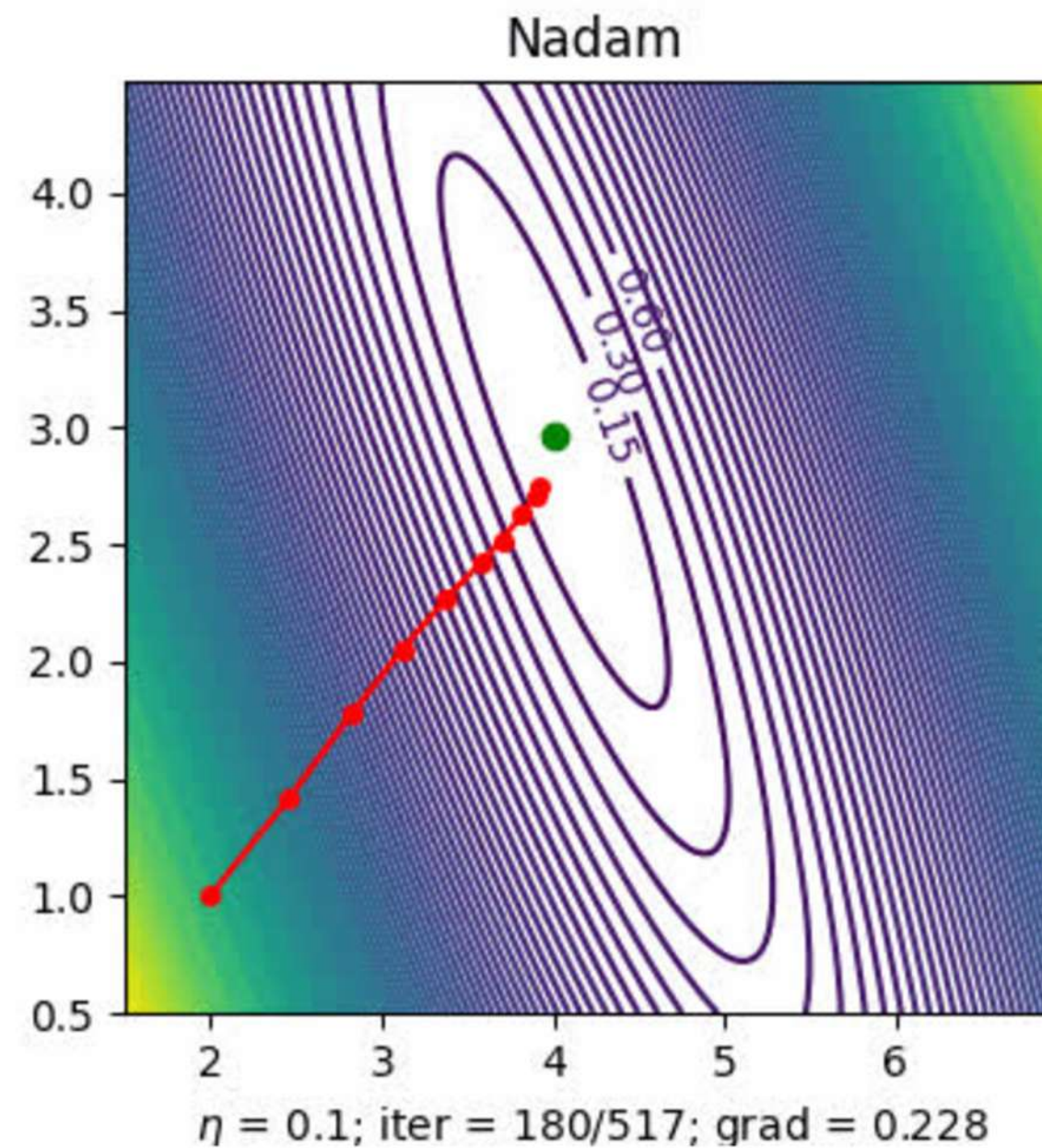
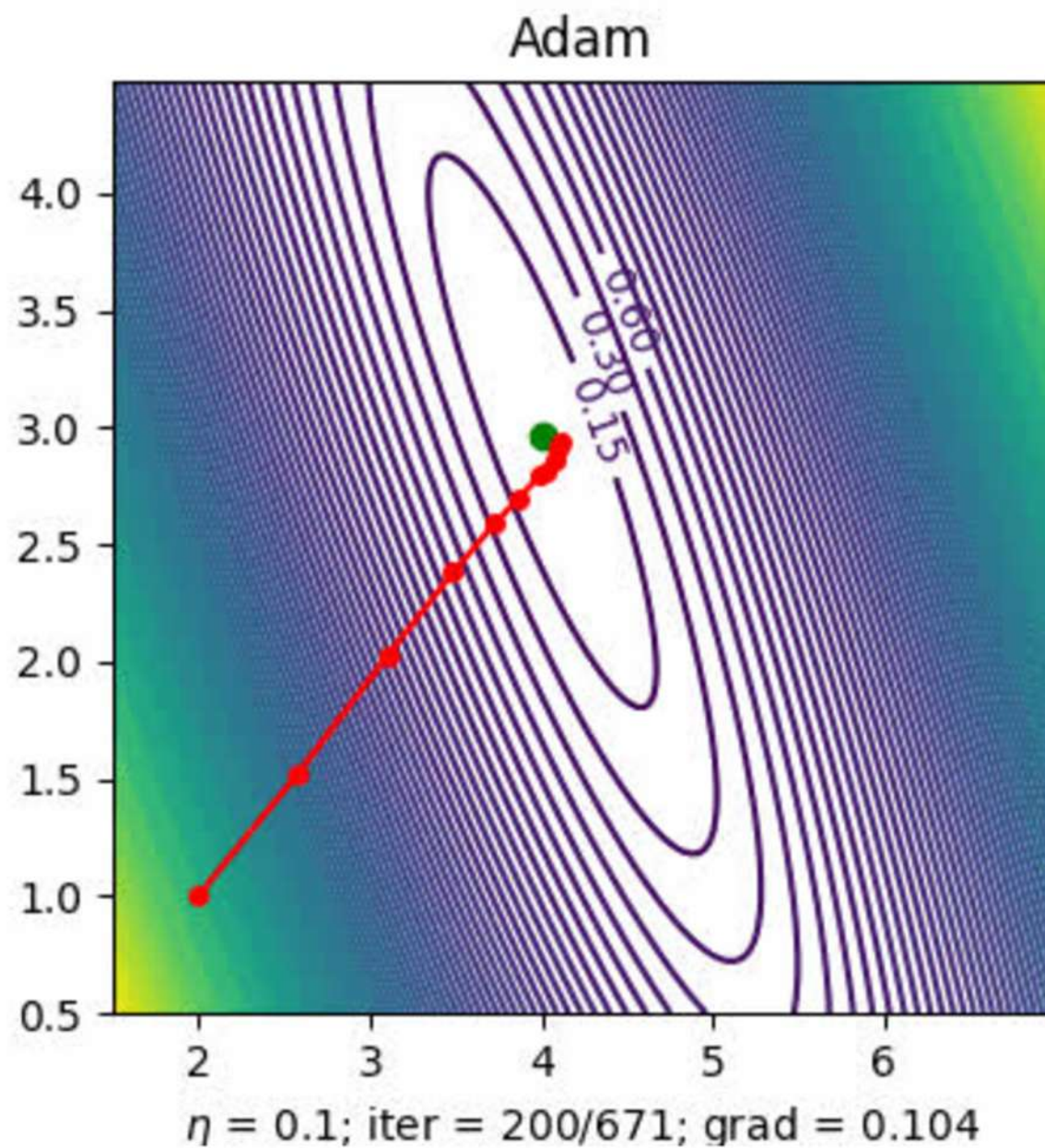
- Hiệu chỉnh tham số:

$$m_{i,t}^o = \frac{m_{i,t}}{1 - \beta_1^t} \quad v_{i,t}^o = \frac{v_{i,t}}{1 - \beta_2^t}$$

- Công thức cập nhật:

$$\theta_{i,t+1} = \theta_{i,t} - \frac{\eta}{\sqrt{v_{i,t-1}^o} + \epsilon} m_{i,t}^o$$

- Adam kết hợp với momentum nên hội tụ nhanh và ổn định, trở thành optimizer phổ biến nhất hiện nay



Các Optimizers hiện đại

Adagrad (2011)
Adadelata (2012)
RMSProp (2012)

AMSGrad (2018)
LAMB (2020)
AdaBelief (2020)

Trước 2000

2010-2014

2014-2017

2018-2020

2020-2025

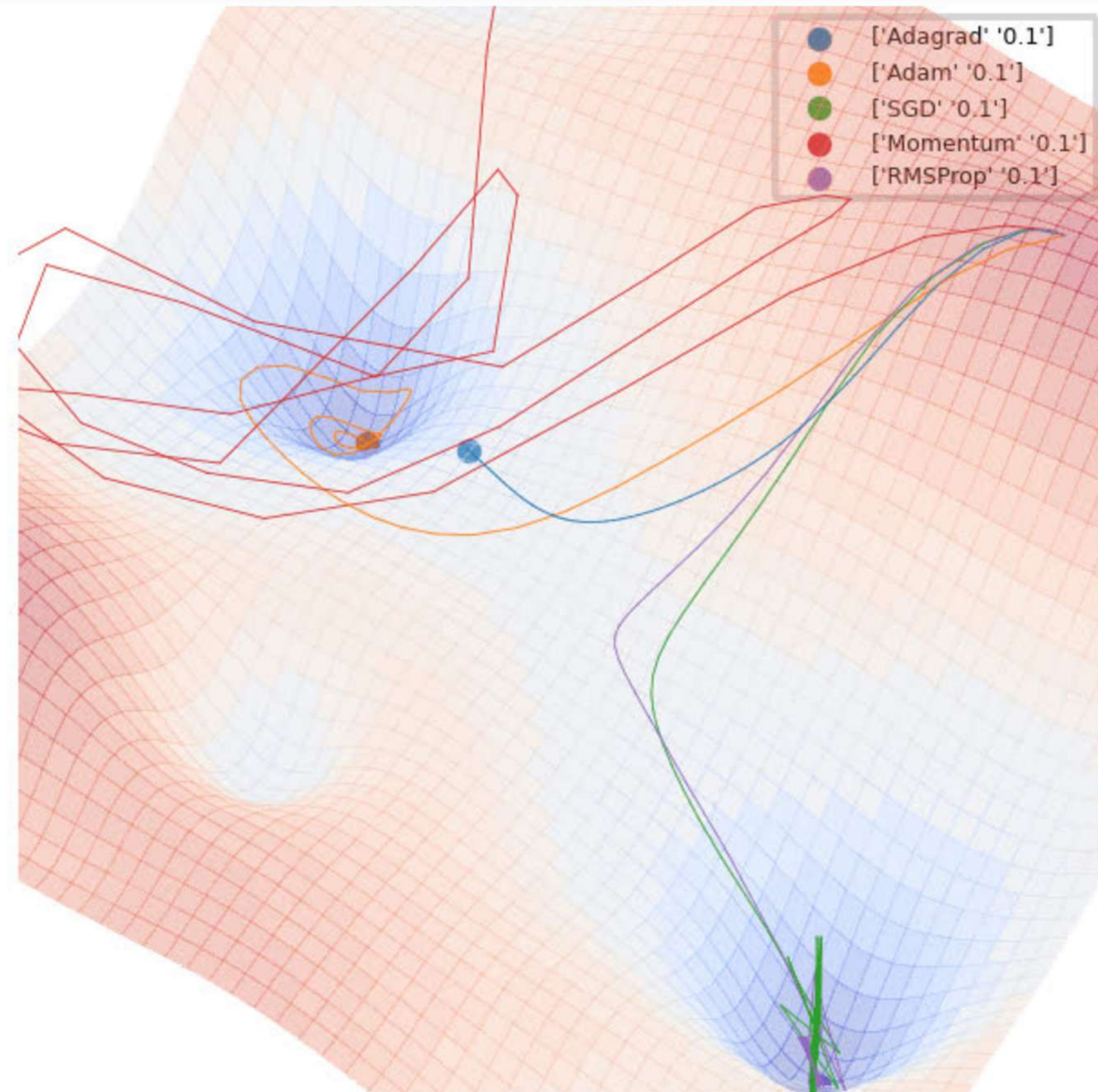
GD, SGD
Momentum
Nag

Adam (2014)
Nadam (2016)
LARS (2017)

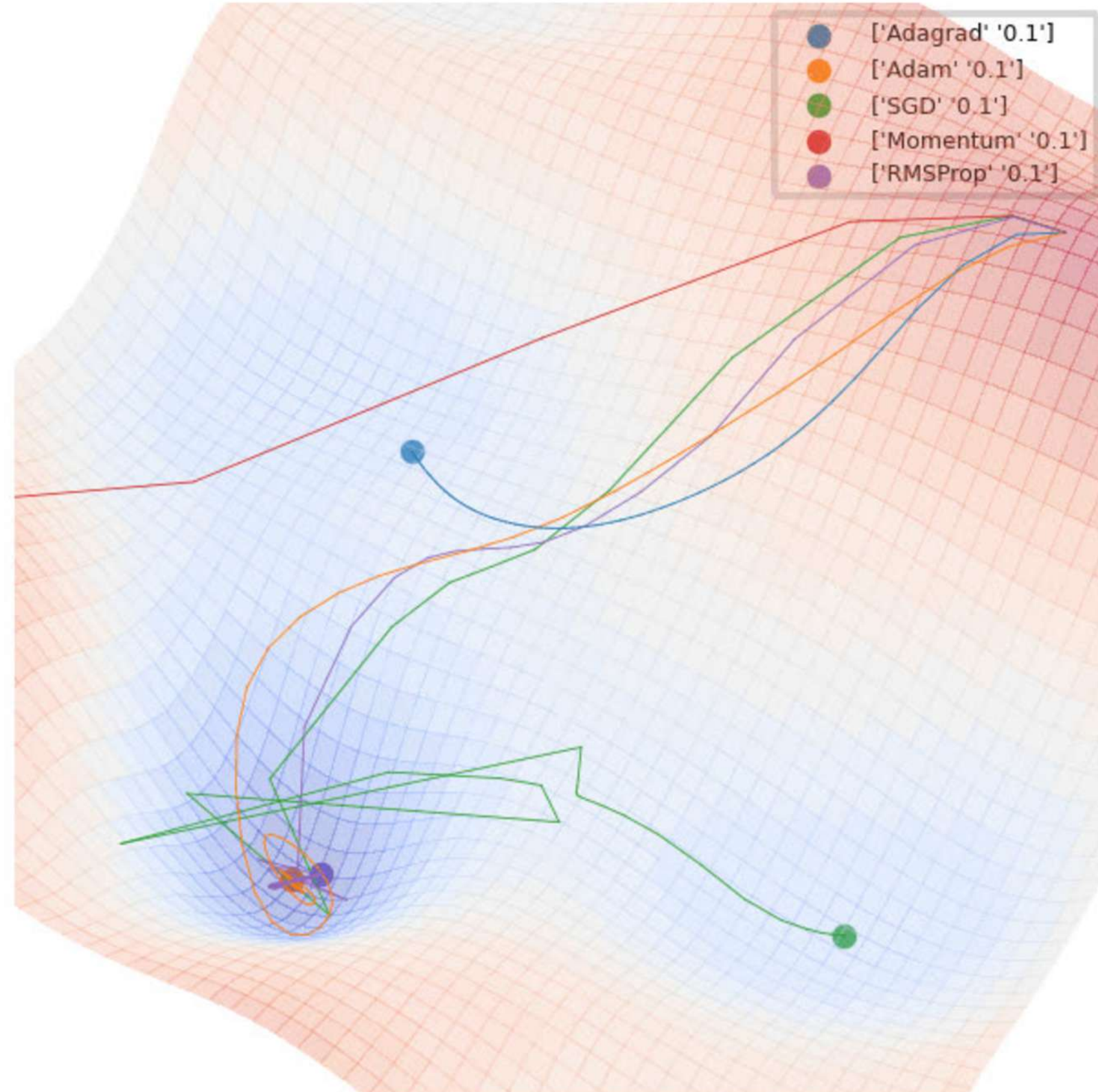
Lion (2023)
Sophia (2023)
Splus (2025)

Ví dụ minh họa

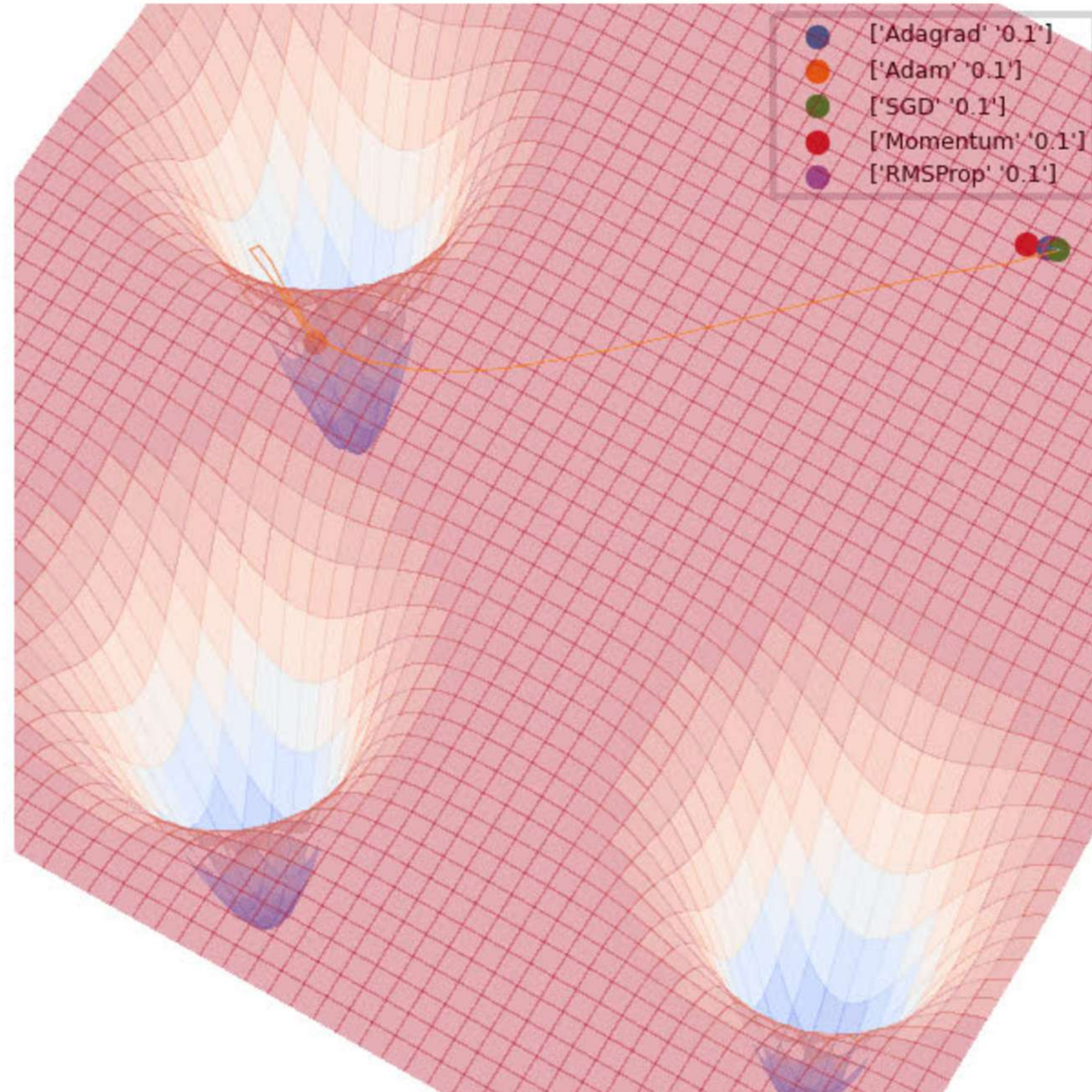
- Ba cực tiểu địa phương
- Hai cực tiểu toàn cục



- Ba cực tiểu địa phương
- Một cực tiểu toàn cục



- Ba cực tiểu địa phương
- Ba cực tiểu toàn cục



A large graphic on the left side of the slide. It features a dark blue background with a pattern of red dots of varying sizes arranged in concentric, slightly irregular circles, creating a sense of depth and movement. The word "HUST" is centered within this graphic.

HUST

THANK YOU !



hust.edu.vn



fb.com/dhbkhn