Non linear optimizers

Combinatorial Decision Making and Optimization Project

Marini Luca

Table of contents

1. Optimizers

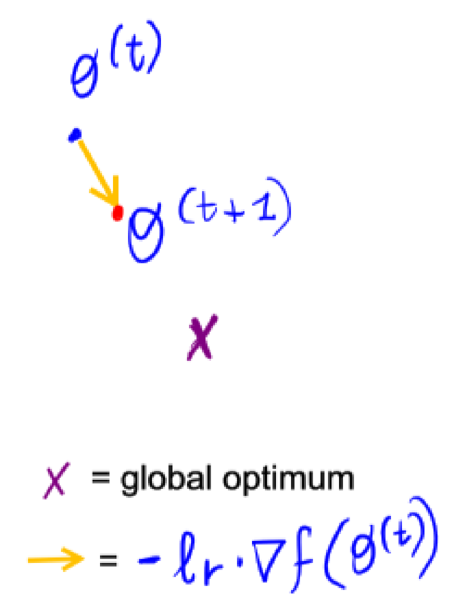
2. Tested functions

Optimizers

Gradient descent

$$\theta^{(t+1)} = \theta^t - lr\nabla f(\theta^{(t)})$$

Gradient descent

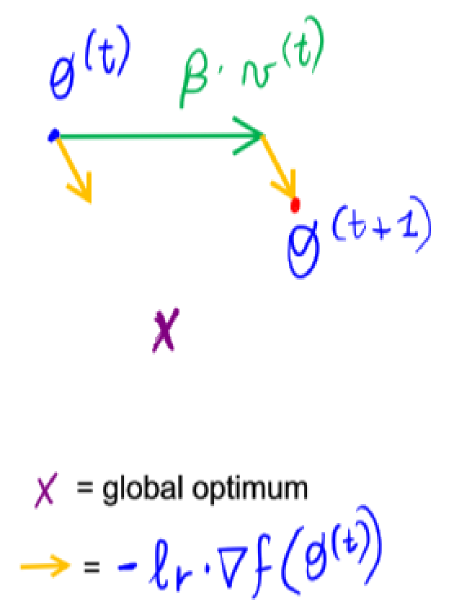


Momentum

$$\text{v}^{(t+1)} = \beta v^{(t)} - lr\nabla f(\theta^{(t)})$$
$$\theta^{(t+1)} = \theta^{(t)} + v^{(t+1)} \quad \text{, with } \beta \in [0,1)$$

if β= 0, then the update step is the same as the one of gradient descent.

Momentum

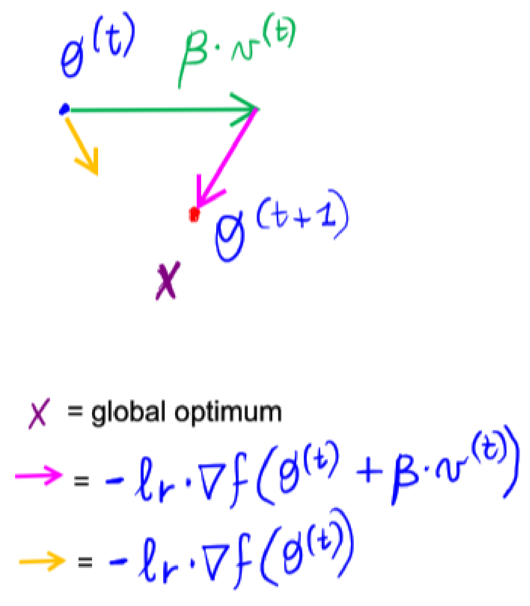


$\theta^{(t)}$
$\beta \cdot v^{(t)}$
$\theta^{(t+1)}$
X = global optimum
$= -l_r \cdot \nabla f(\theta^{(t)})$

Nesterov momentum

$$\text{v}^{(t+1)} = \beta v^{(t)} - lr\nabla f(\theta^{(t)} + \beta v^{(t)})$$
$$\theta^{(t+1)} = \theta^{(t)} + v^{(t+1)} \quad \text{, with } \beta \in [0, 1)$$

Nesterov momentum



Adaptive learning rates method

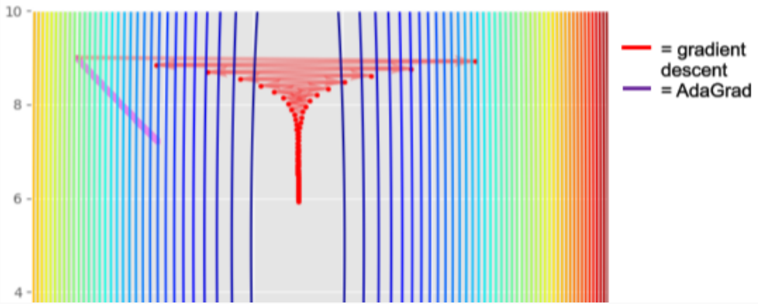
- Methods that adapt the learning rate to the parameters
- they reduce the step of updates for dimensions whose gradient direction is not consistent across iterations
- they increase the step of updates for dimensions whose gradient direction is consistent across iterations

AdaGrad (Adaptive Gradient)

$$
\mathsf{s}^{(t+1)} = s^{(t)} + \nabla f(\theta^{(t)}) \nabla f(\theta^{(t)})
$$
$$
\theta^{(t+1)} = \theta^{(t)} - \frac{lr}{\sqrt{s^{(t+1)}} + \in} \nabla f(\theta^{(t)})
$$

AdaGrad (Adaptive Gradient)

$$f(x1, x2) = 90(x_1 - 3)^2 + (x_2 - 5)^2$$

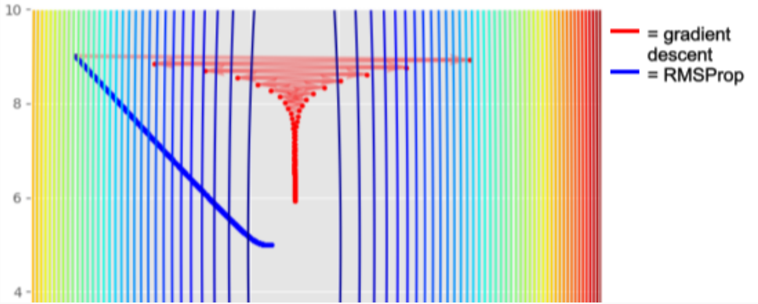


RMSProp

$$\mathsf{s}^{(t+1)} = \beta s^{(t)} + (1-\beta)\nabla f(\theta^{(t)})\nabla f(\theta^{(t)})$$
$$\theta^{(t+1)} = \theta^{(t)} - \frac{lr}{\sqrt{s^{(t+1)}}+\in}\nabla f(\theta^{(t)}) \quad \text{, with } \beta \geq 0.9$$

RMSProp

$$f(x1, x2) = 90(x_1 - 3)^2 + (x_2 - 5)^2$$



Adam

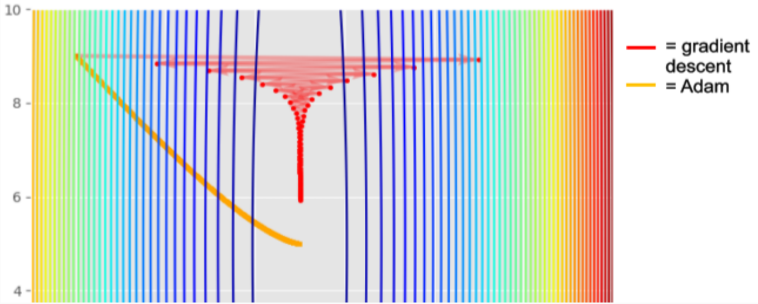
$$\text{g}^{(t+1)} = \beta_1 g^{(t)} + (1-\beta_1)\nabla f(\theta^{(t)})$$
$$\text{s}^{(t+1)} = \beta_2 s^{(t)} + (1-\beta_2)\nabla f(\theta^{(t)})\nabla f(\theta^{(t)})$$
$$\text{g}^{debiased} = \frac{g^{(t+1)}}{1-\beta_1^{t+1}}$$
$$\text{s}^{debiased} = \frac{s^{(t+1)}}{1-\beta_2^{t+1}}$$

$$\theta^{(t+1)} = \theta^{(t)} - \frac{lr}{\sqrt{s^{debiased}}+\in} g^{debiased}$$

β_1= 0.9, β_2= 0.999.

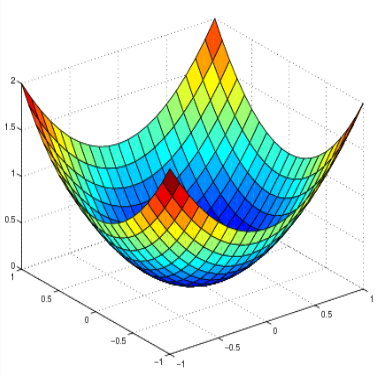
Adam

$$f(x1, x2) = 90(x_1 - 3)^2 + (x_2 - 5)^2$$



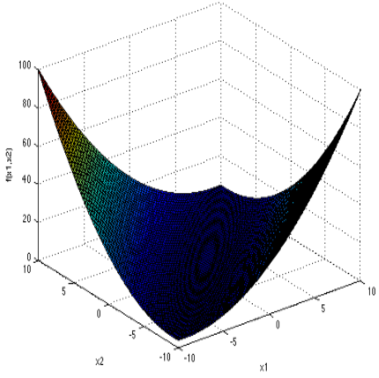
Tested functions

Paraboloid



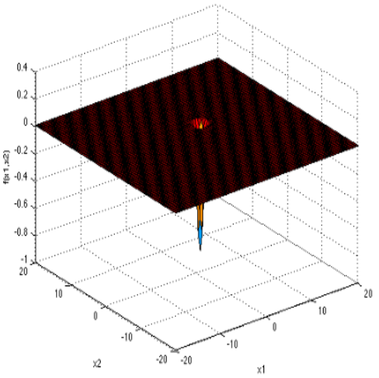
- Function: $x_1^2 + x_2^2$
- Global Minimum: $x = (0, 0)$

Matyas function



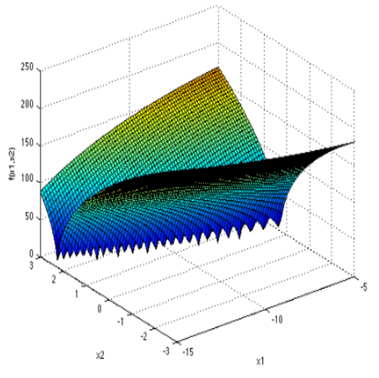
- Function: $0.26(x_1^2 + x_2^2) - 0.48x_1x_2$
- Global Minimum: x = (0, 0)

Easom function



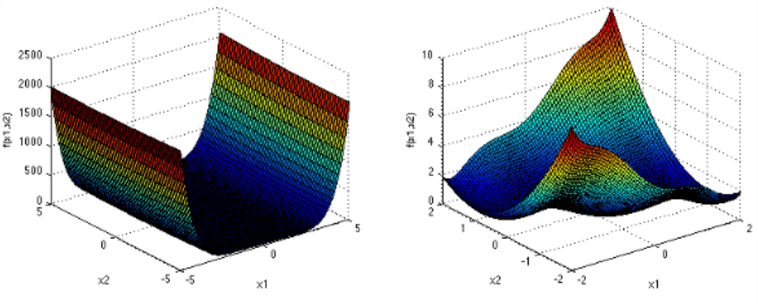
- Function: $-cos(x_1)cos(x_2)exp(-(x_1-\pi)^2-(x_2-\pi)^2)$
- Global Minimum: x = (π, π)

Bukin function



- **Function**: $100\sqrt{|x_2 - 0.01x_1^2|} + 0.01|x_1 + 10|$
- **Global Minimum**: x = (-10, 1)

Three hump camel function



- Function: $2x_1^2 - 1.05x_1^4 + \frac{x_1^6}{6} + x_1x_2 + x_2^2$
- Global Minimum: x = (0, 0)