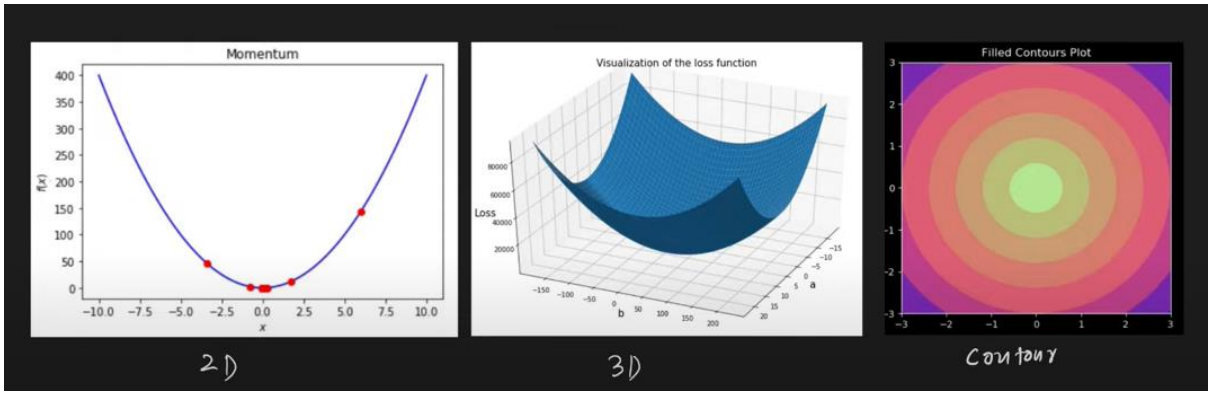
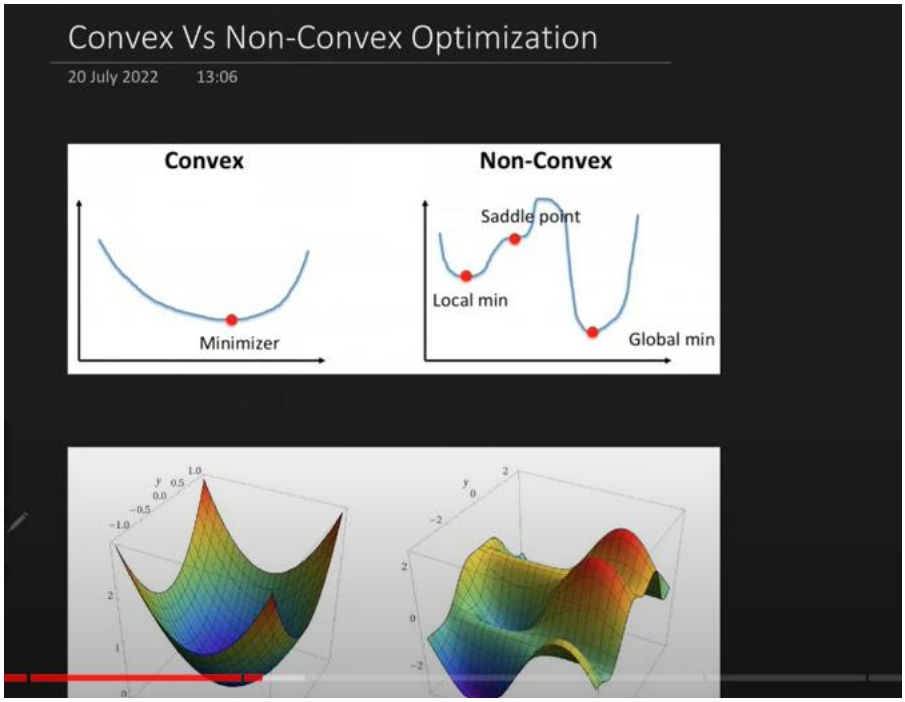
SGD with Momentum



Contour ko vahi 3d graph ko 2d mai smjhneka way

Jo Loss hota hai vo kindoff Function hota hai Weights & Biases ka



JBH Hum Complex Neural network Architecture banarhe hote hai tbhi

Loss Function ka Graph complex hota , vahi show horha

Humko Weights ka esa value nikalna hai jaha Loss sbhse kamm hoo

Convex Optimization:

- A convex optimization problem is one in which the objective function is convex, and the feasible set (constraints) is also convex.
- Convex optimization problems have desirable properties, such as a single global minimum, and any local minimum is also a global minimum.
- The optimization algorithms for convex problems are well-established, and convergence to the global minimum is guaranteed.

Application in Deep Learning:

- Traditional machine learning algorithms often involve convex optimization problems, where the goal is to find the optimal values for model parameters.
- Linear regression and logistic regression are examples of convex optimization problems commonly used in machine learning.

Non-Convex Optimization:

- In deep learning, the optimization problem is typically non-convex due to the highly non-linear and complex nature of neural networks.
- Non-convex optimization problems have multiple local minima, making it challenging to find the global minimum.
- Neural networks have a large number of parameters, and the optimization landscape is highly non-convex, which can lead to difficulties in training.

Challenges in Deep Learning:

- The presence of multiple local minima can make it challenging to find the optimal set of parameters for neural networks.
- Gradient-based optimization algorithms, such as stochastic gradient descent (SGD), may get stuck in suboptimal solutions.
-

Problems of Non Convex minima tak pochne keliye

There are 3 Reasons for that

1. Local Minima
   Vahi Hill climbing vala scene

2. Saddle Point

Saddle Point ek esa point hai 1 Direct mai surface is high & on another direction

It is low.

3. High Curvature

Momentum Optimization – The Why?

Non-Convex Optimization

Problems:

high curvature, consistent gradient, noisy gradient (Local Minimum)

Momentum technique sbhko solve karleta hai

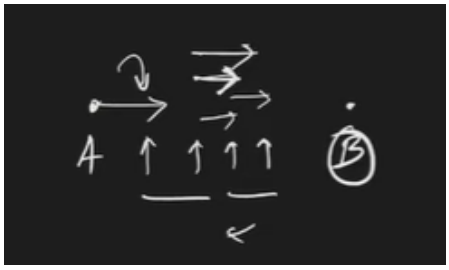
Local Minima se bahar nikalpata hai

High curvature Loss se bahar nikalta hai

Momentum Optimization – The What?

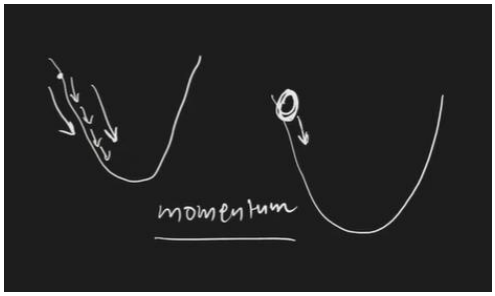
Momentum Optimization Kaam kaise krta hai

Common E.g:



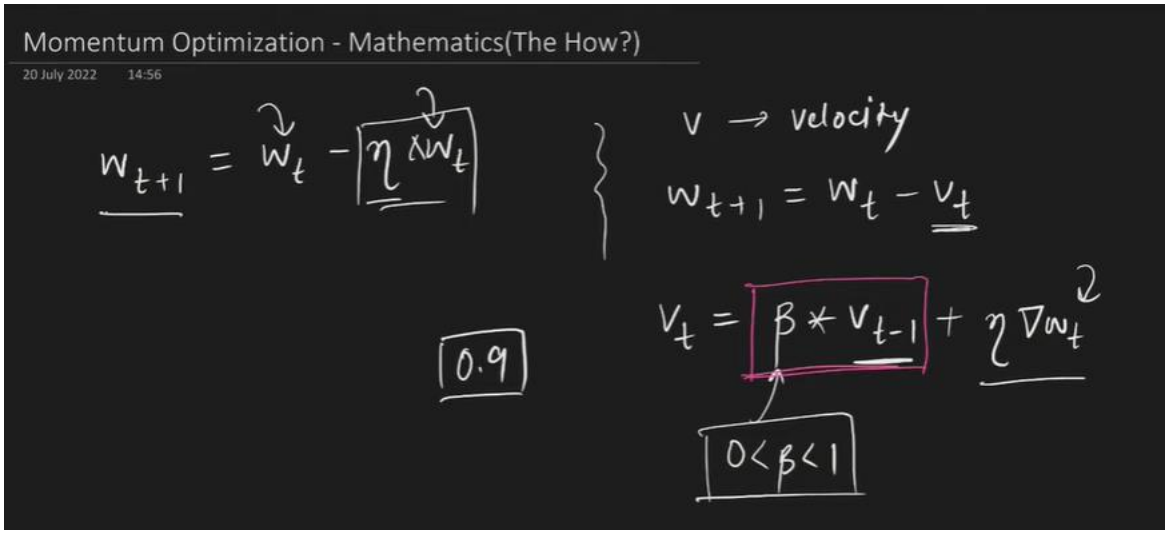
Agar previous gradients aapko 1 direction mai move krarhe then

Speed badake vahi move karenge.

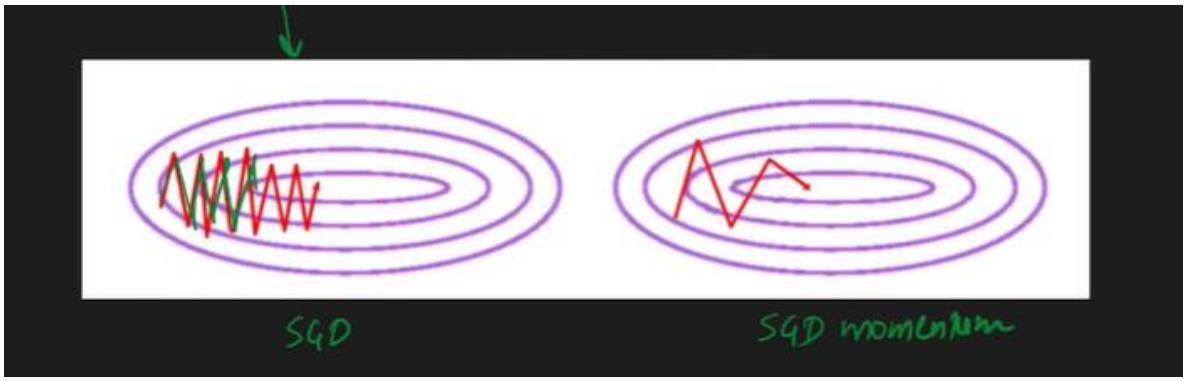


Single most benefit of using Momentum of Optimization

Is Speed.



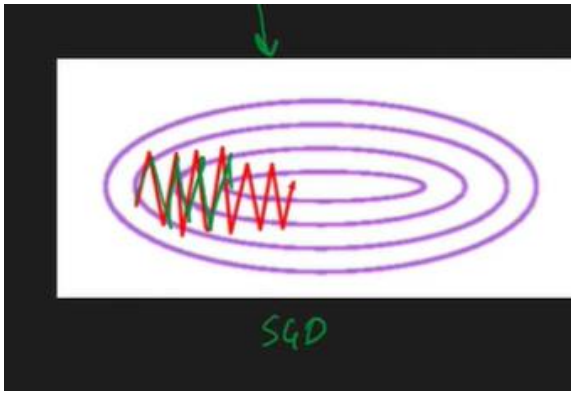
SGD & SGD with Momentum



As we can see ki SGD jo hai vo bht zyaada upar neeche unstability kesath minima achive krrha hai (minimum loss function)

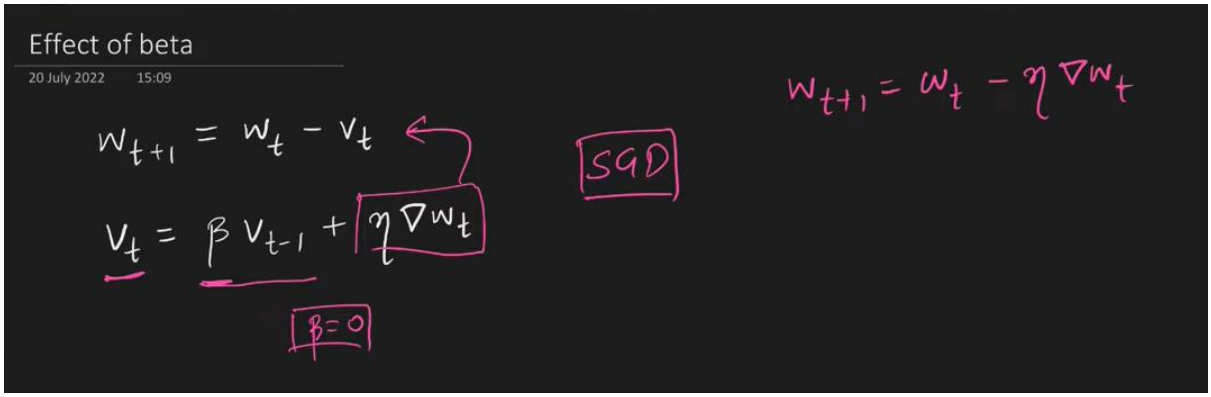
Vertical direction mai zyaada chalraha hai so Convergence mai

Time lagraha hai.



But in SGD Momentum as vo momentum apply krrha hai tou as horizontal direction mai velocity badhjaaraha hai becoz of that

Kamm time mai Convergence hojaraha hai.

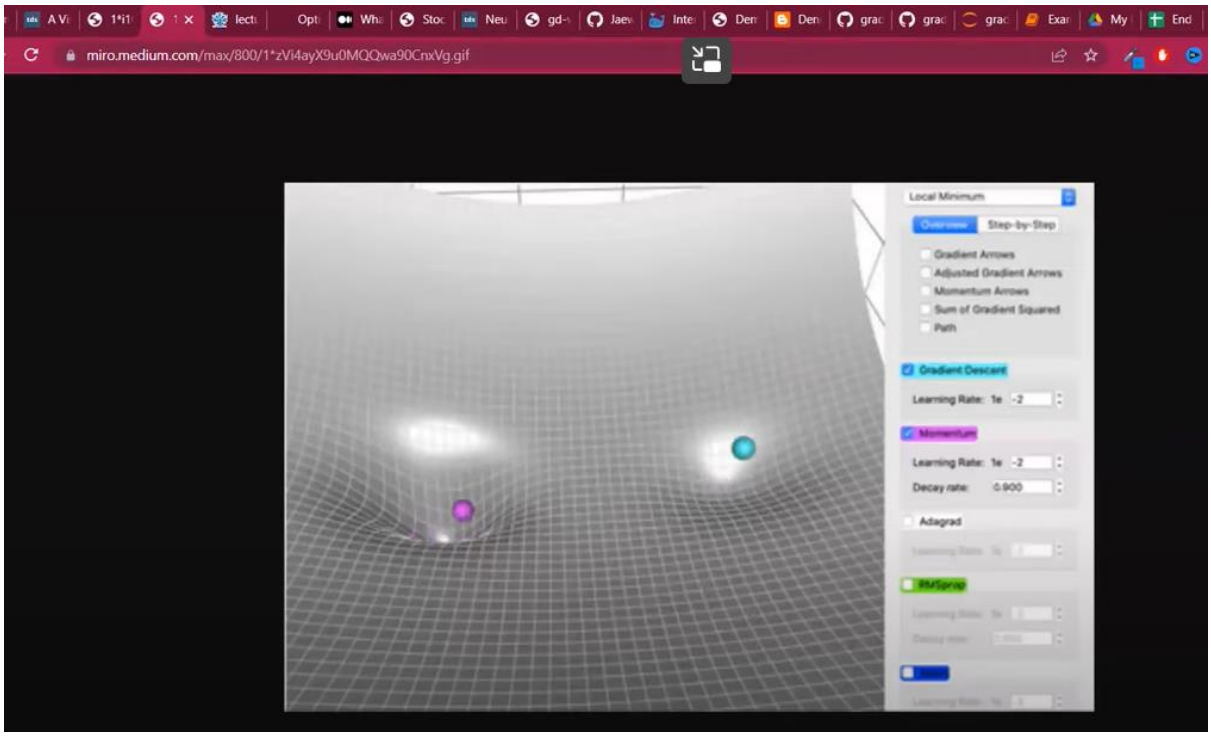


E.g Effect of Beta

If Beta Zero hai then iska mtlb SGD with Momentum Normal SGD hojayega

Problems with Momentum?

Vanilla GD & Momentum here



Benefit

Momentum is Faster

The Blue ball is Coming at constant speed Whereas

Purple Ball ka speed increase horaha hai.

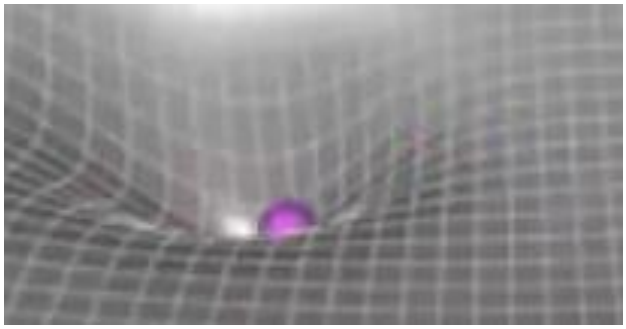
Momentum ka Fayda → Faster Speed

Momentum is able to Escape Local Minima, Global Minima pochjata hai kyuki uskepass Speed hota hai

Whereass GD kepass matlab blue ball kepass Speed nahi hai, that's why Vo atak jaata hai Local Minima ke andar hie.

Momentum is Able to Excape Local Minima

Problems of Momentum?



Benefit hie kbhi kbhi uska problem hojata hai.

Becoz of Speed & momentum , vo Minima pochnekebaad bhi

Speed ke vajese To & Fro krrha thoda oscillate ya unstable hokr Reach horaha.

Jo Oscillation hota hai Vahi thoda time waste hojata hai becoz of that Vo Sbhse Fastest Algo nahi bnnpata hai.

Momentum ki sbhse badi Problem Momentum hie hai BC.