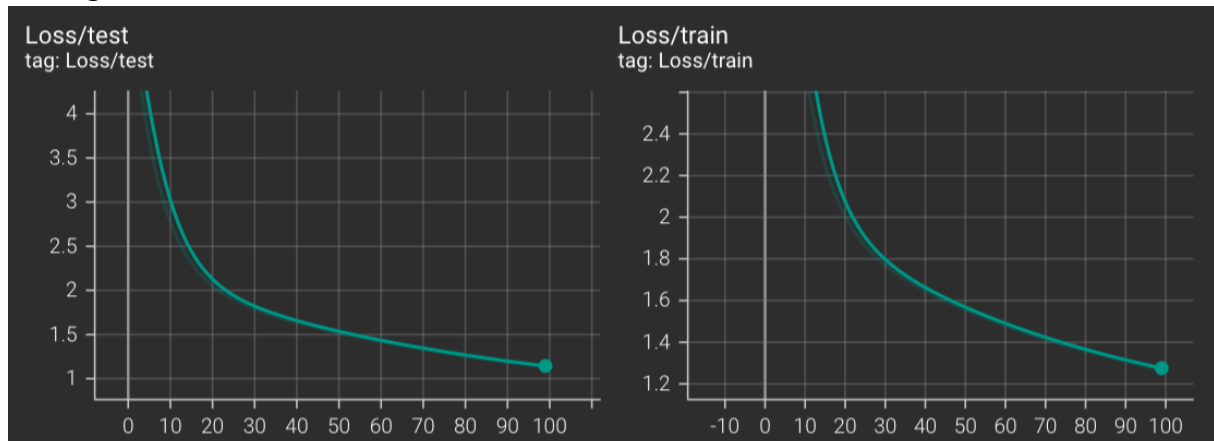


TP2 AMAL

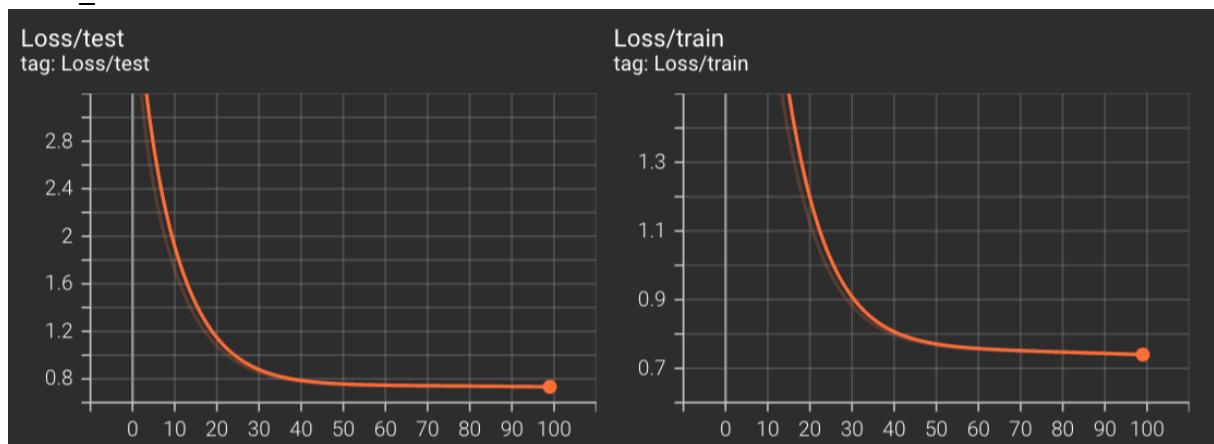
Mathieu Grosso et Thomas Floquet

Question 1 :

Batch gradient descent

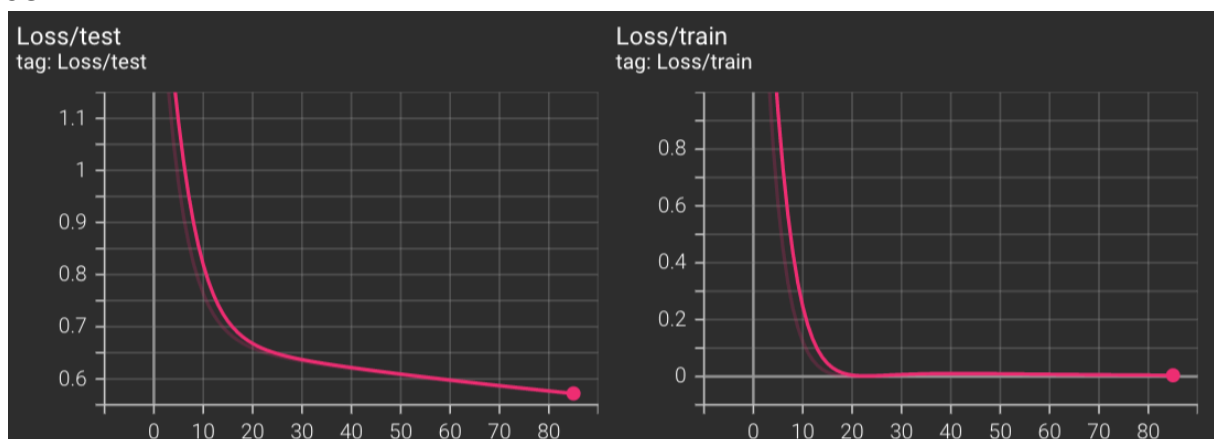


Batch_size = 64 :



On remarque bien que le mini-batch converge plus vite que le batch gradient descent

SGD :

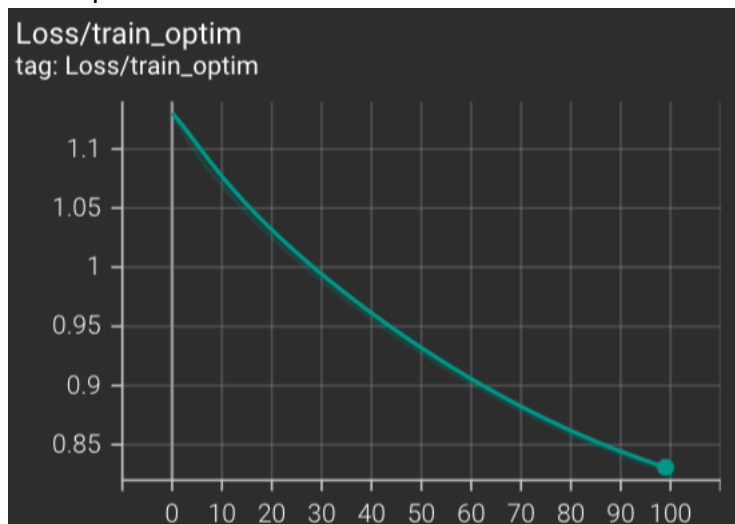


Le SGD converge encore plus vite

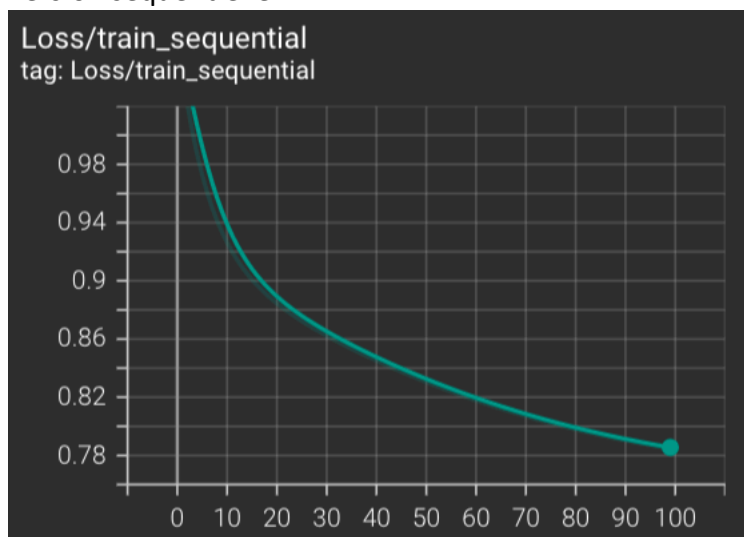
Mais sur cet exemple aussi simple, difficile de montrer que le SGD risque de rester bloquer dans un minimum local ou que l'algo apprend moins bien

Question 2 :

Avec optimiseur :



Version séquentielle :



Les paramètres des modules mis ensemble sont tous mis à jour ensemble automatiquement avec :

```
with torch.no_grad():  
    for param in model.parameters():  
        param -= lr*param.grad
```