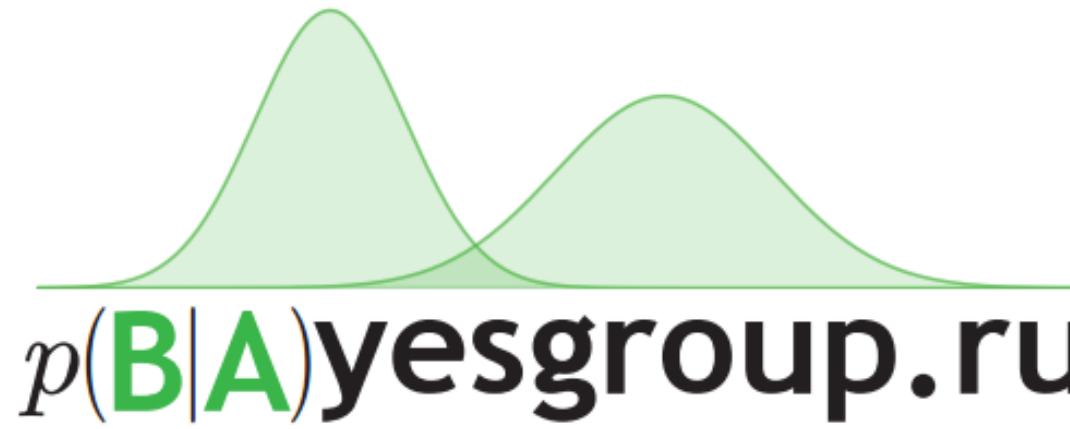
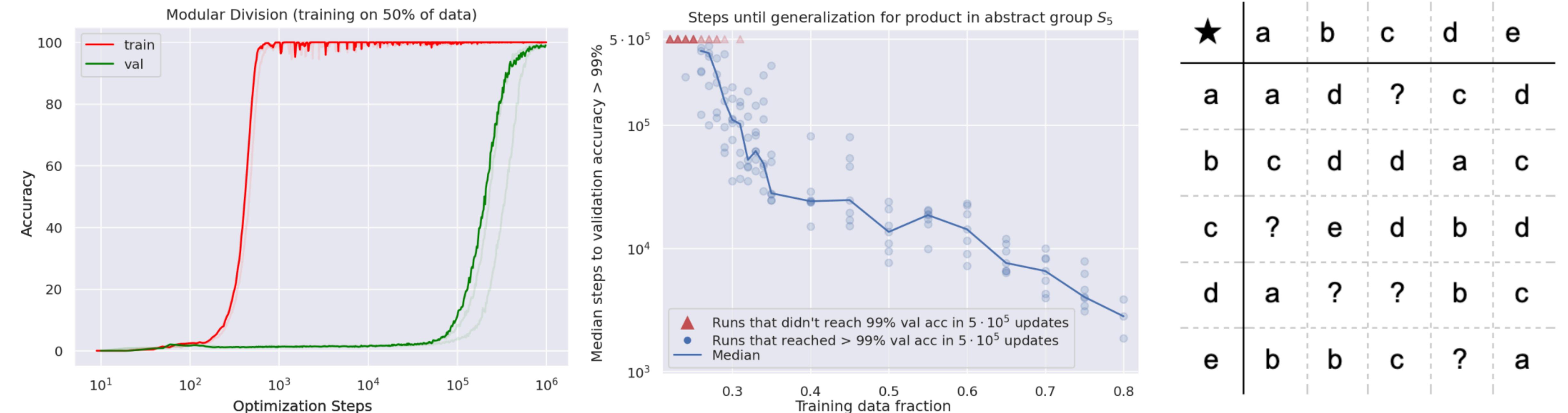


GROKKING: GENERALIZATION BEYOND OVERFITTING ON SMALL ALGORITHMIC DATASETS

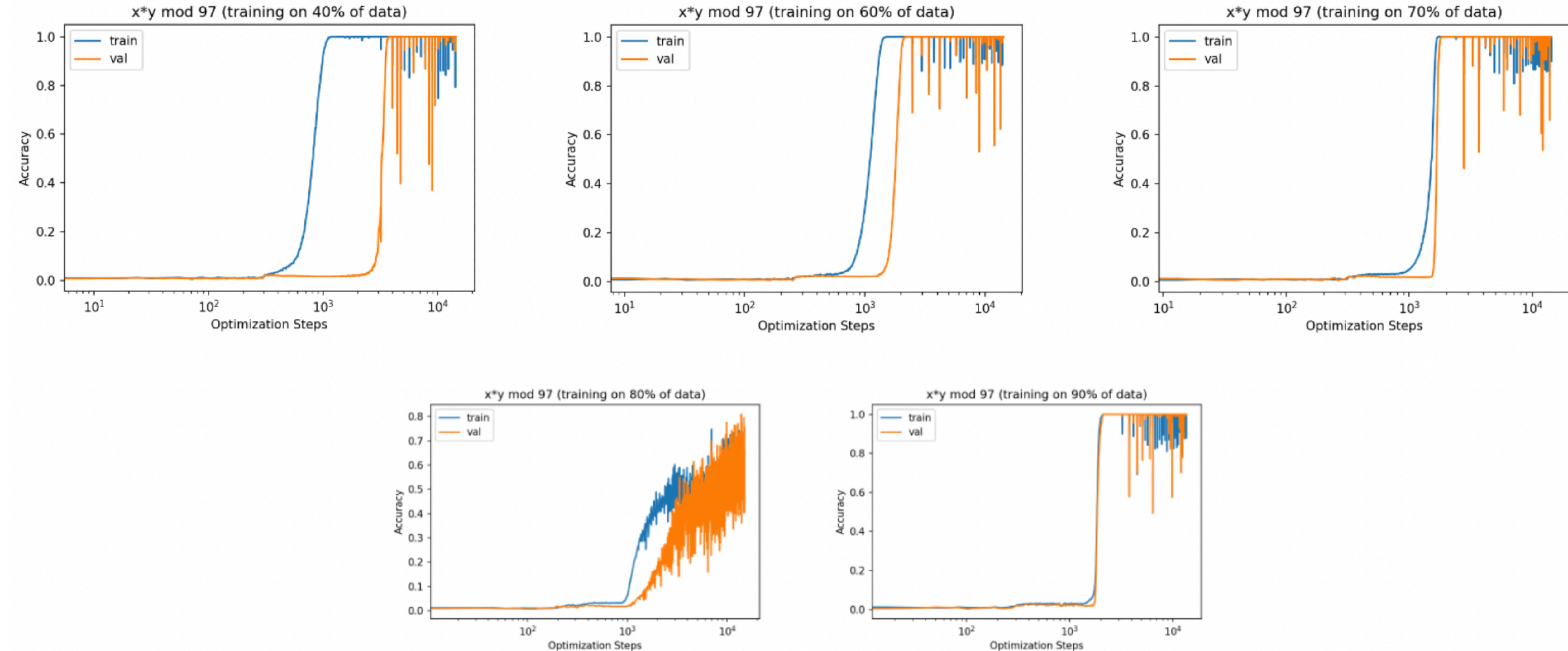


Постановка задачи

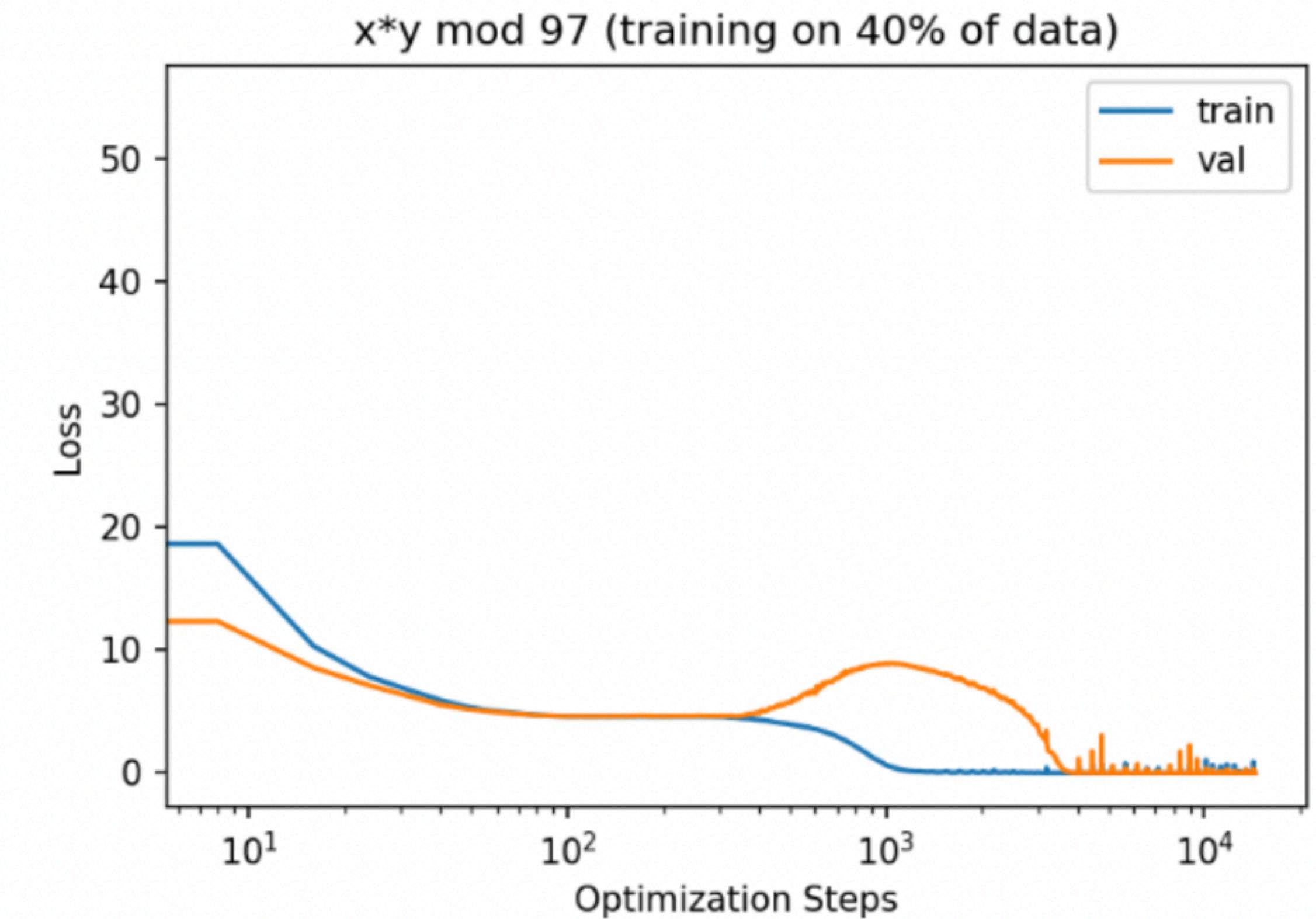
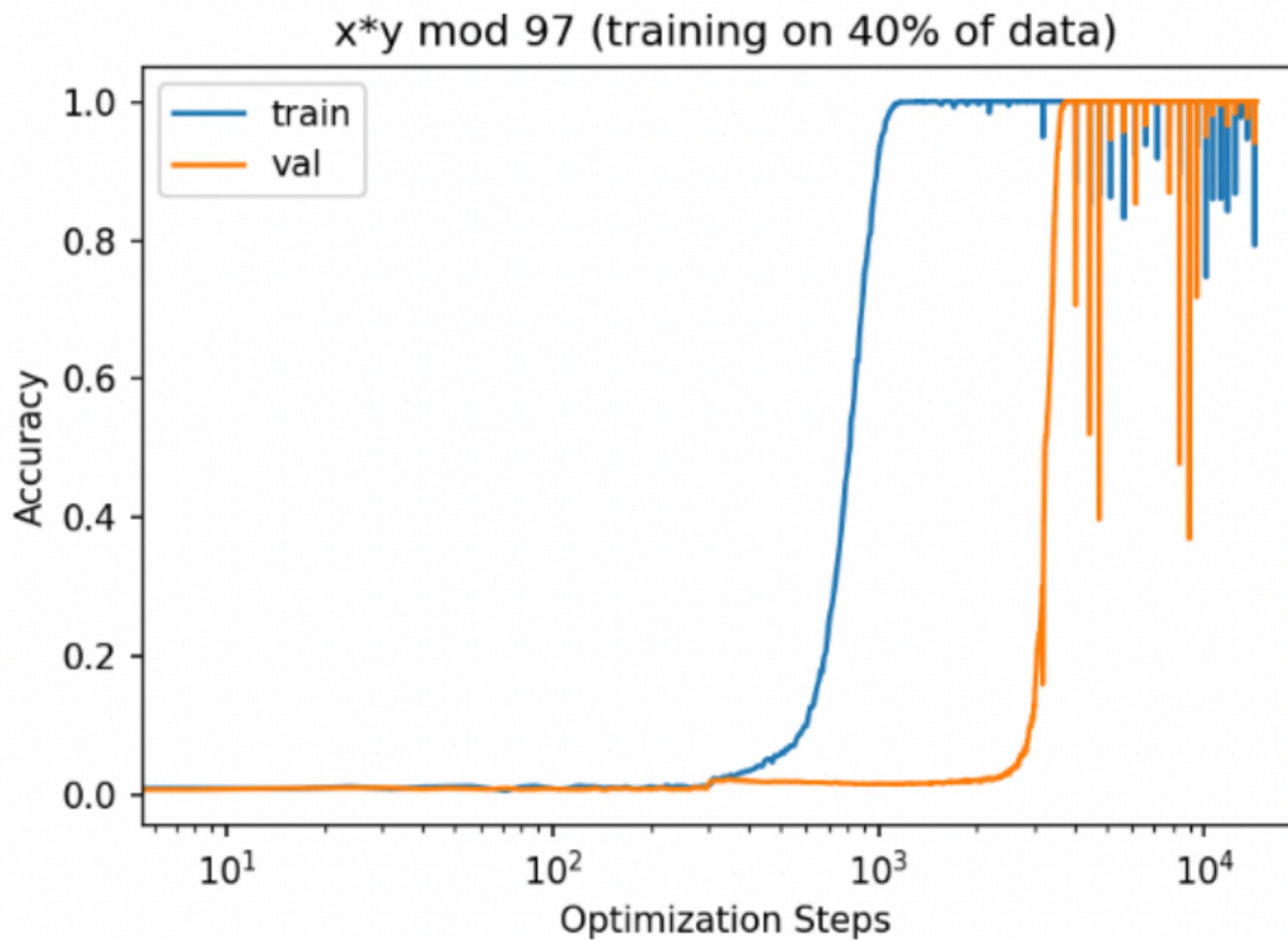


<https://arxiv.org/pdf/2201.02177.pdf>

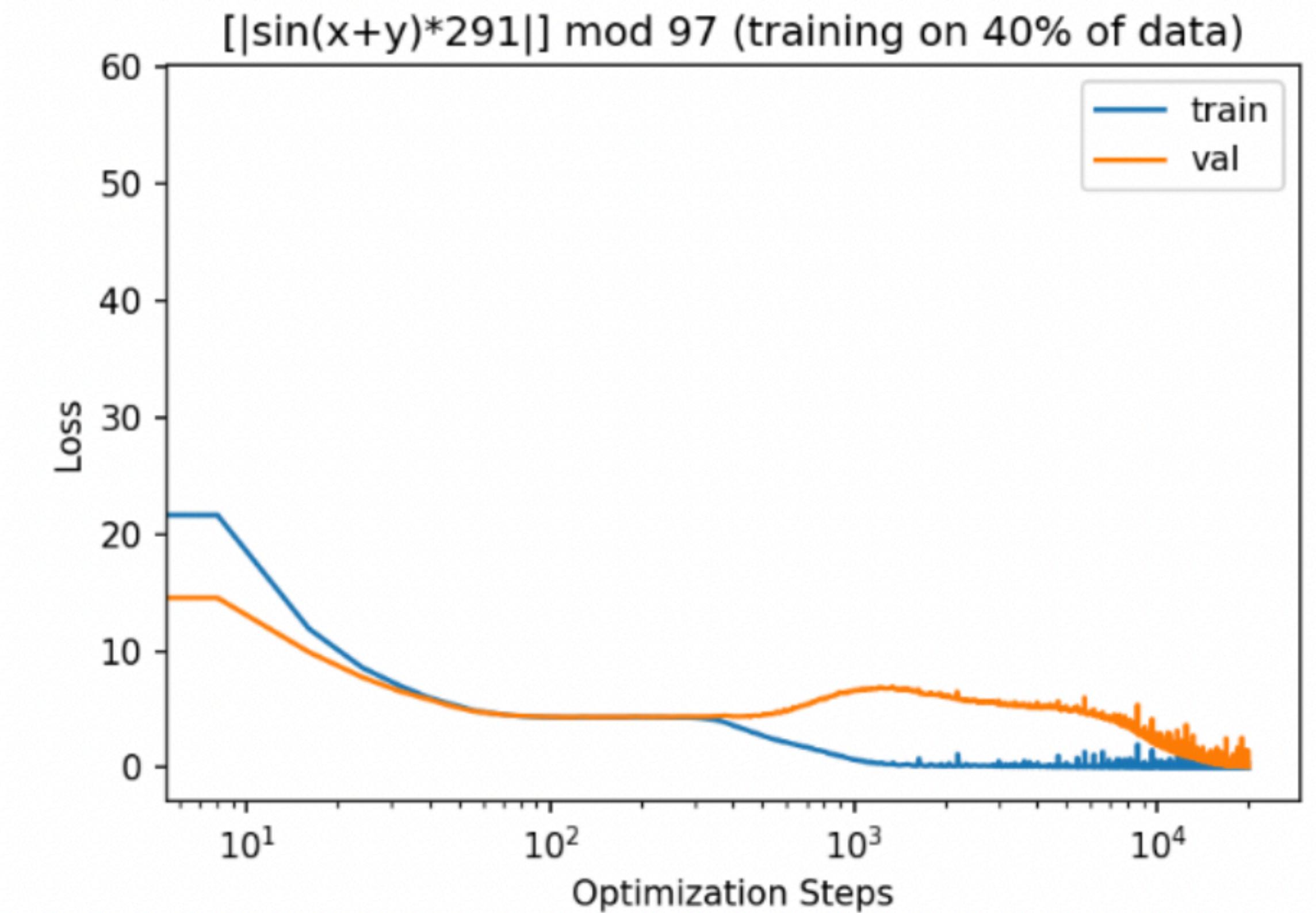
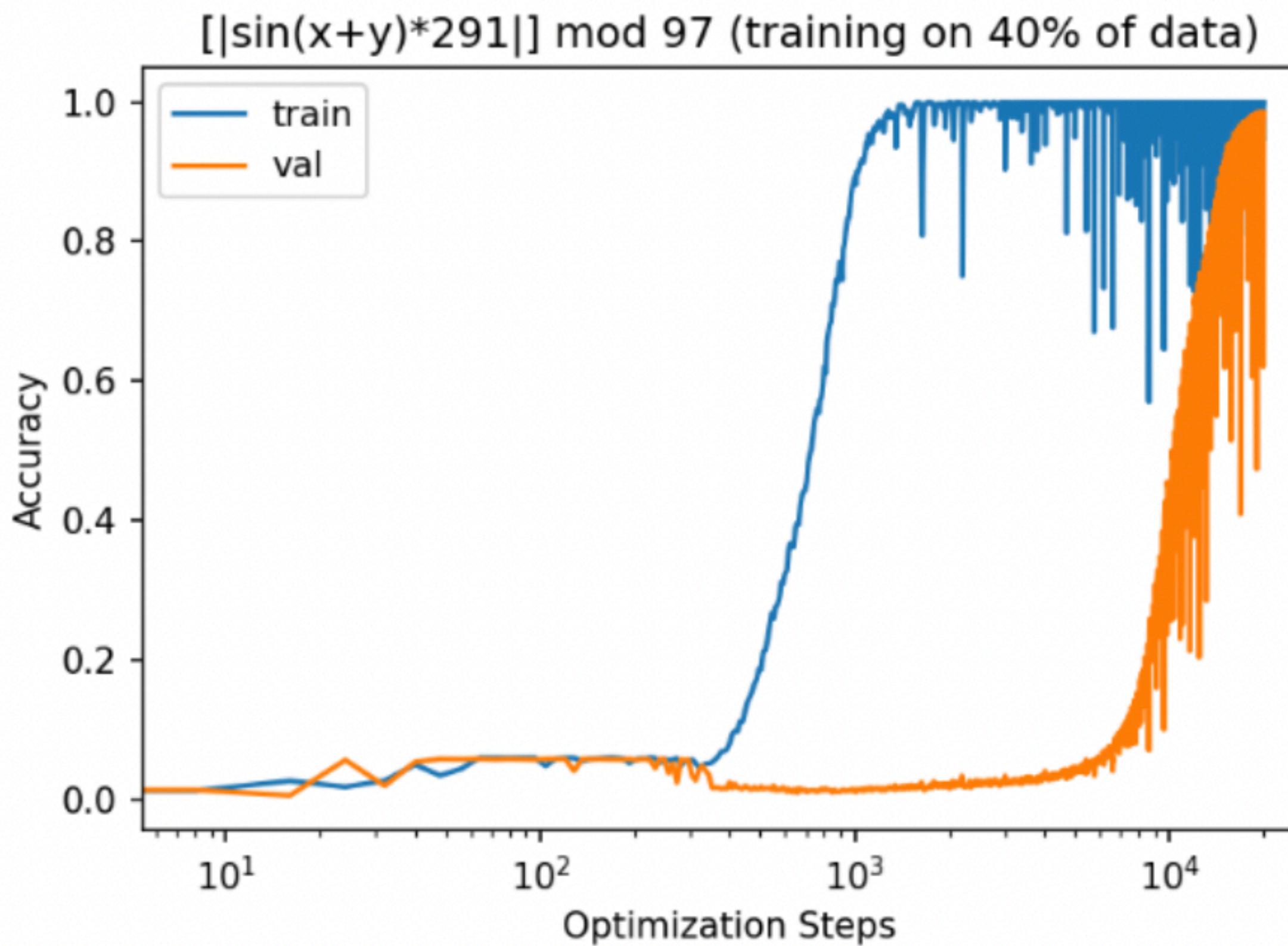
Размер обучающей выборки



Подтверждение результатов

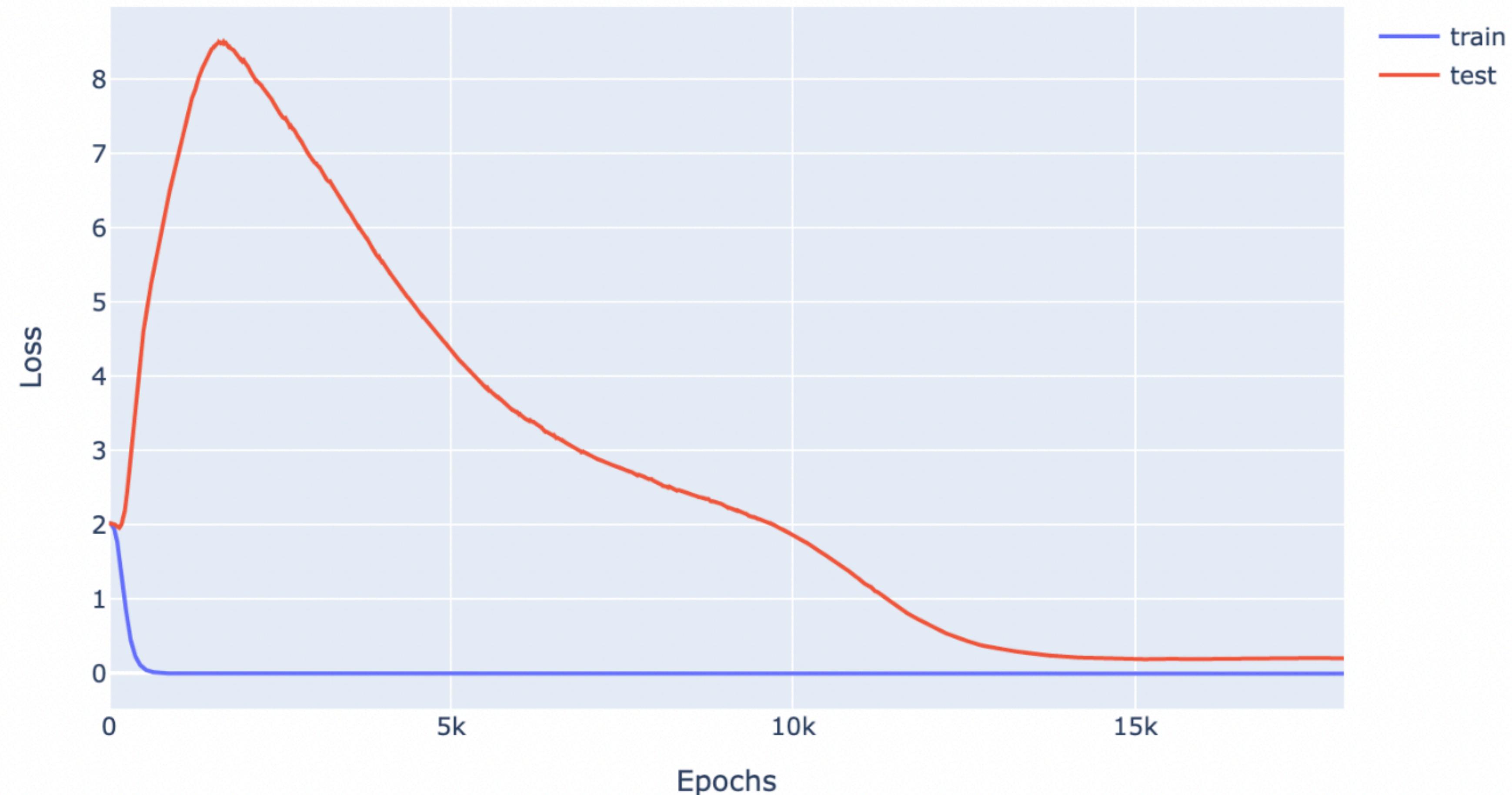


Несимметричные функции

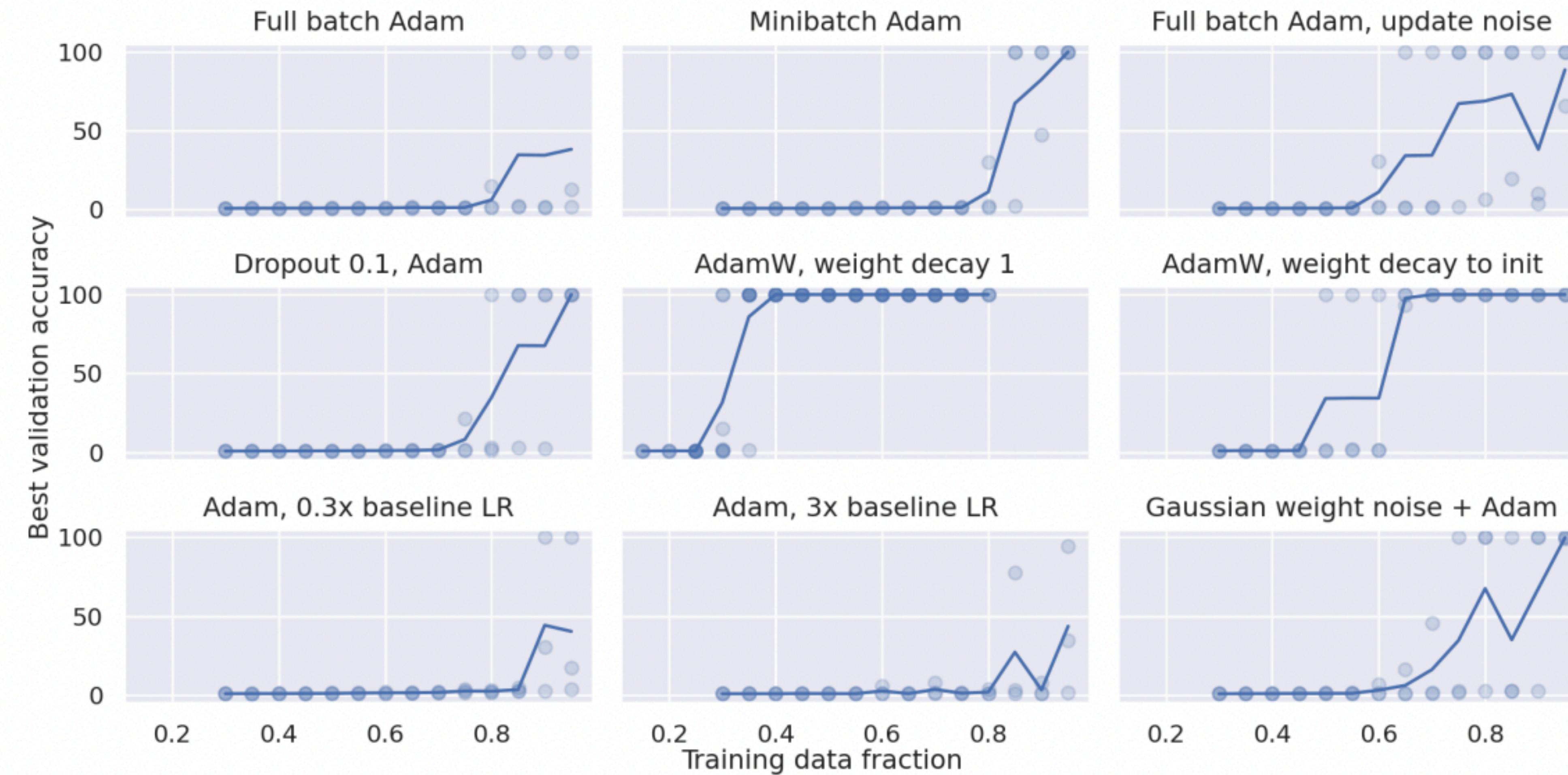


Другие игрушечные задачи

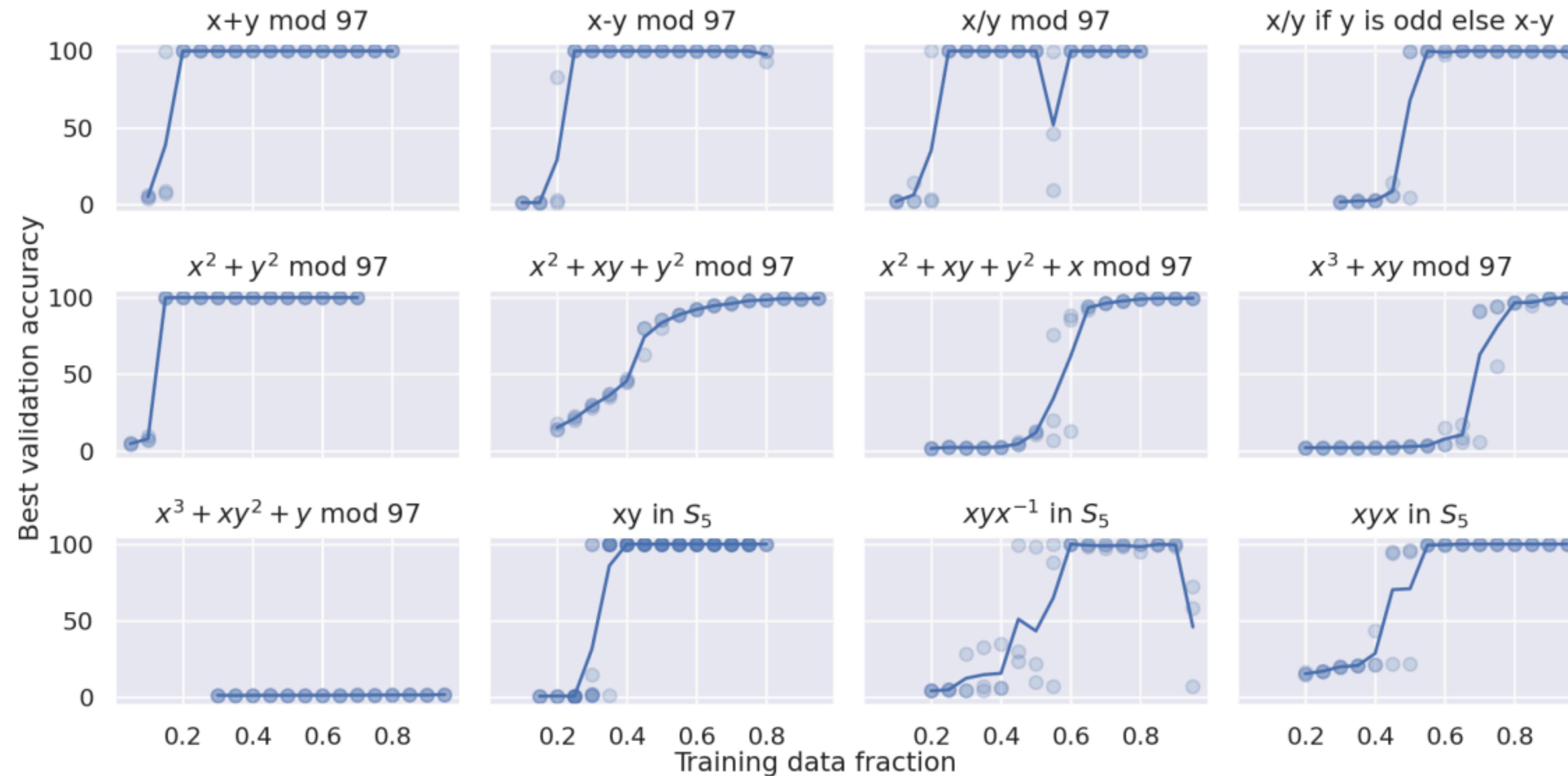
Phase Change in 5 Digit Addition Finite Data Training Curve (Linear Scale)



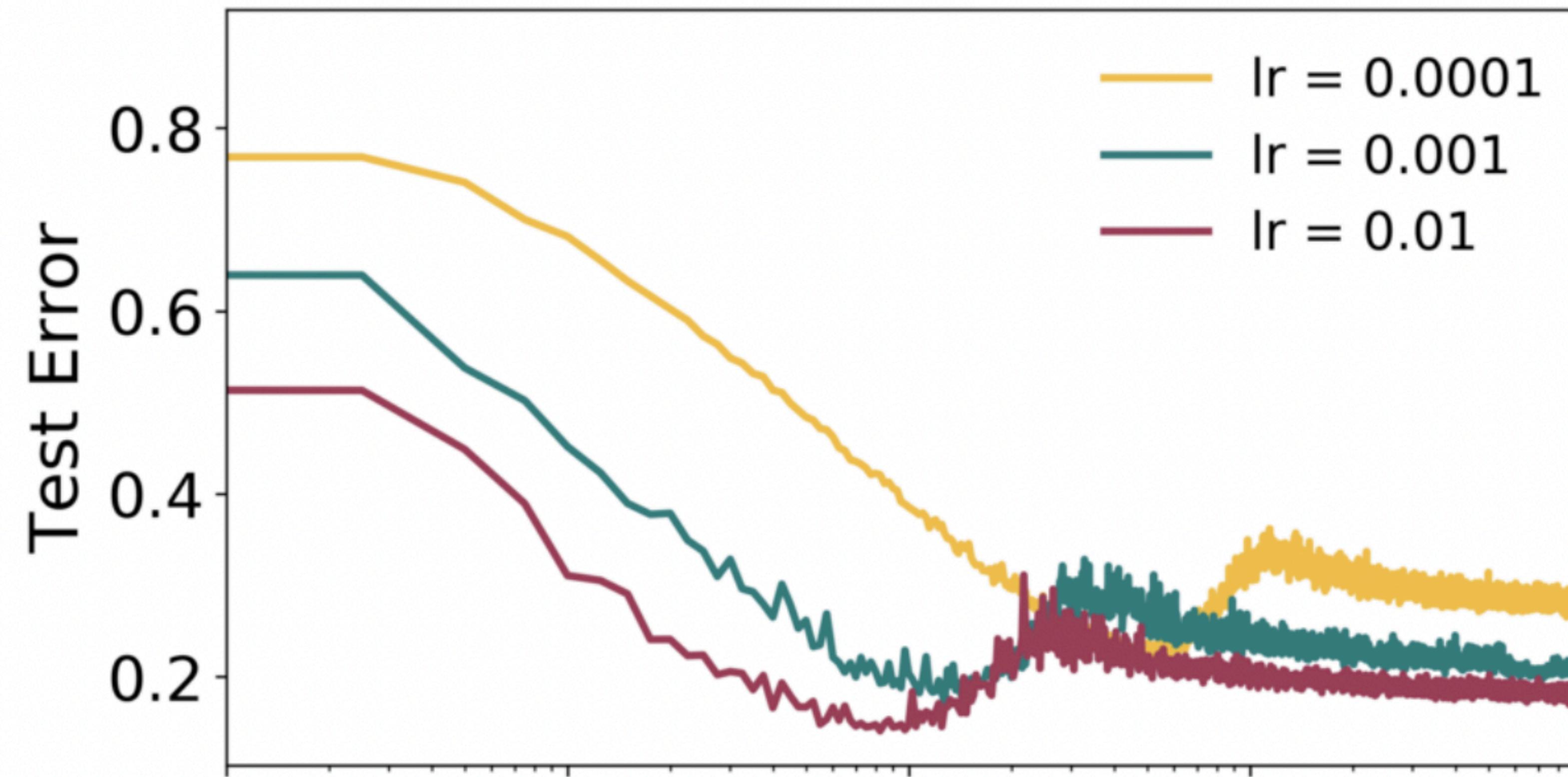
Разные процедуры обучения



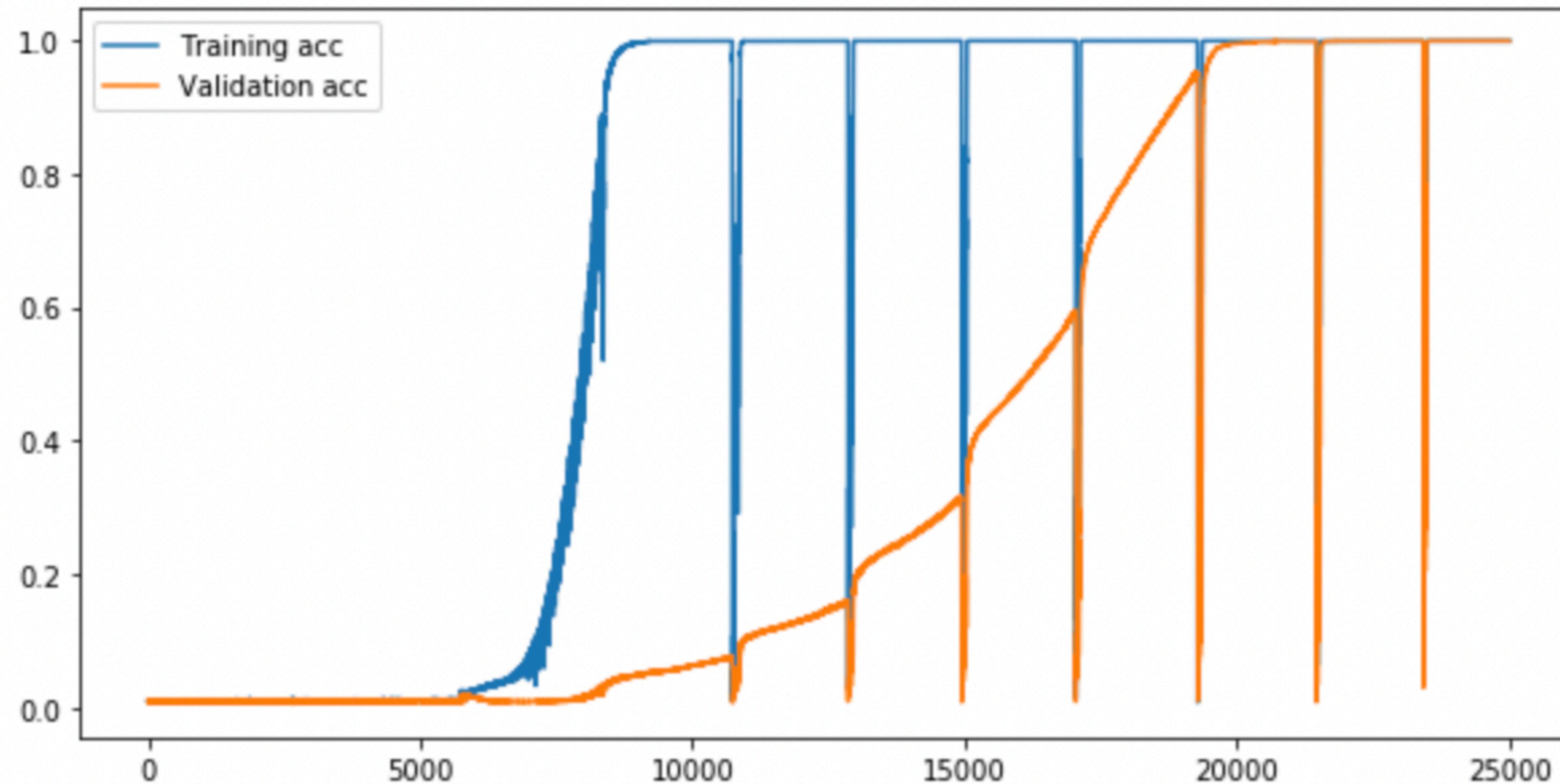
Оптимизация разных функций



Двойной спуск по итерациям

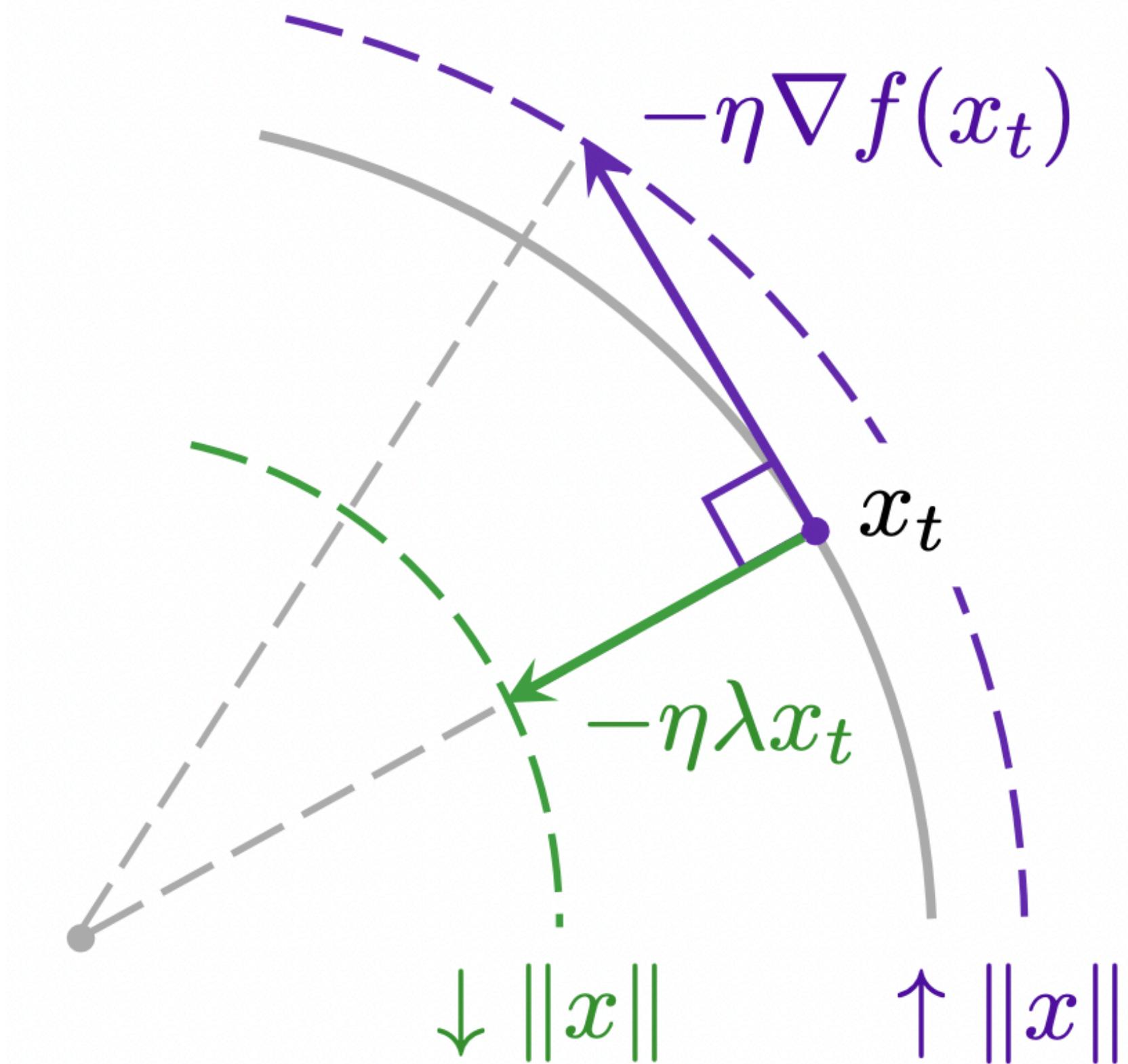
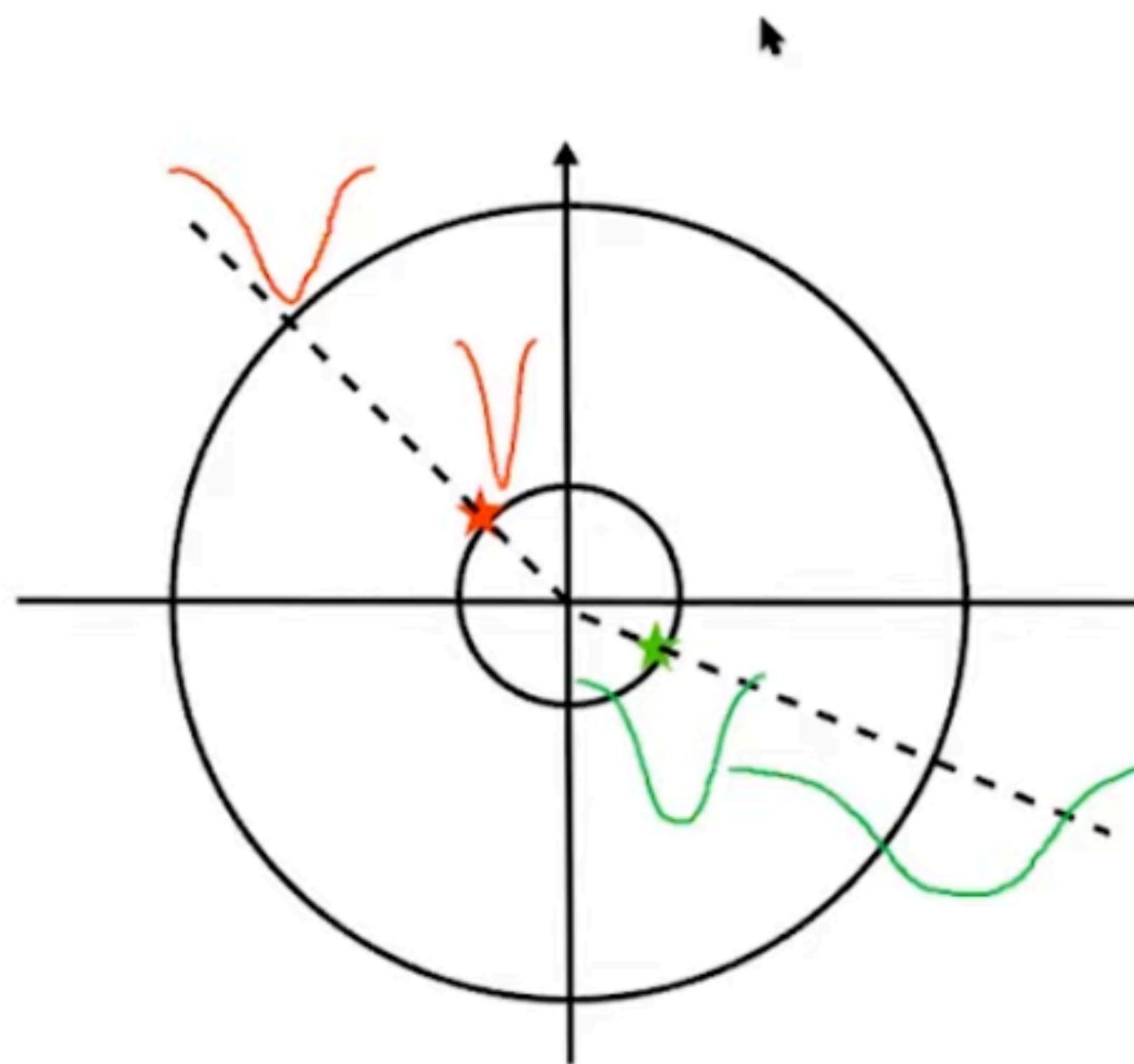


Оптимизация с помощью SGD

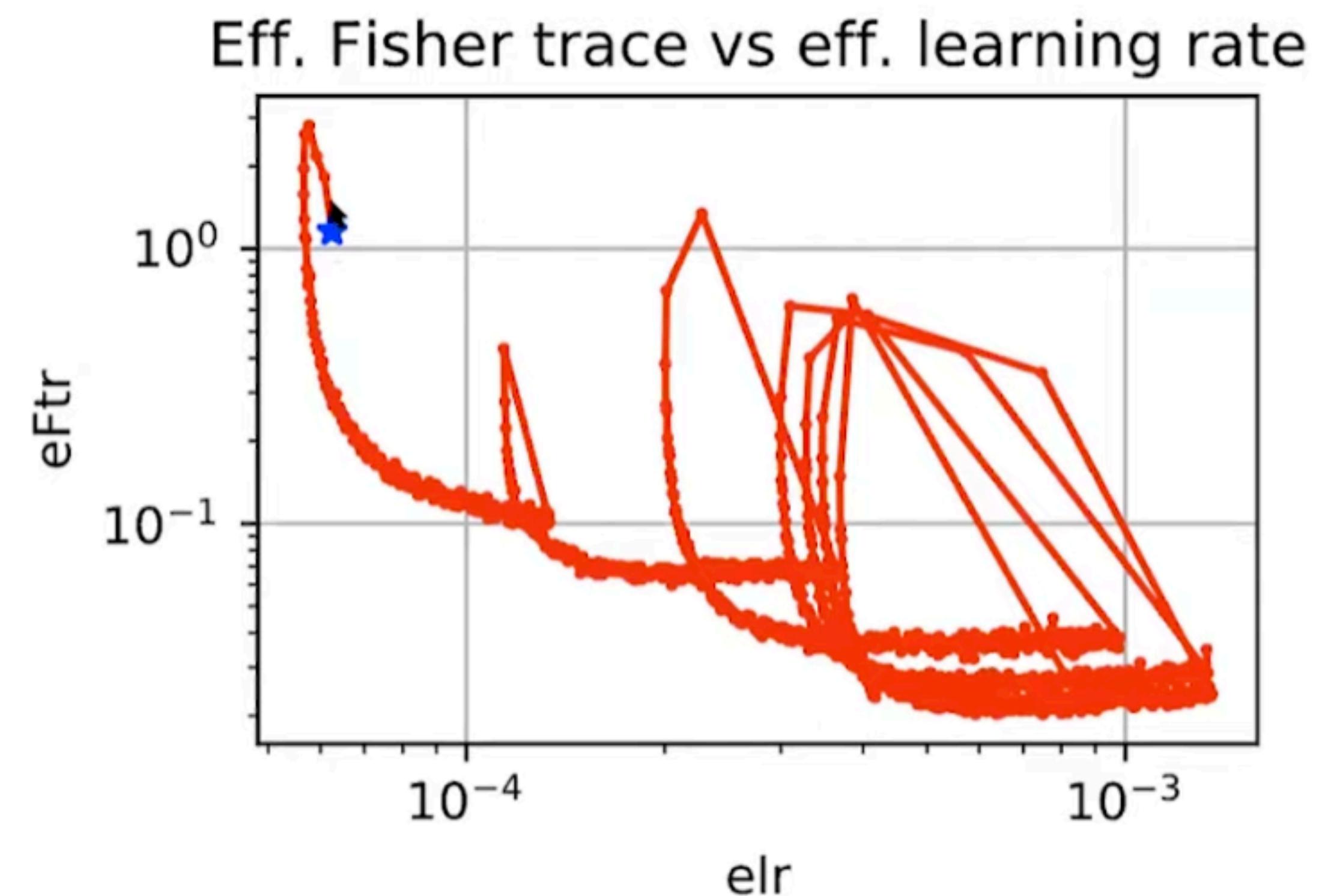
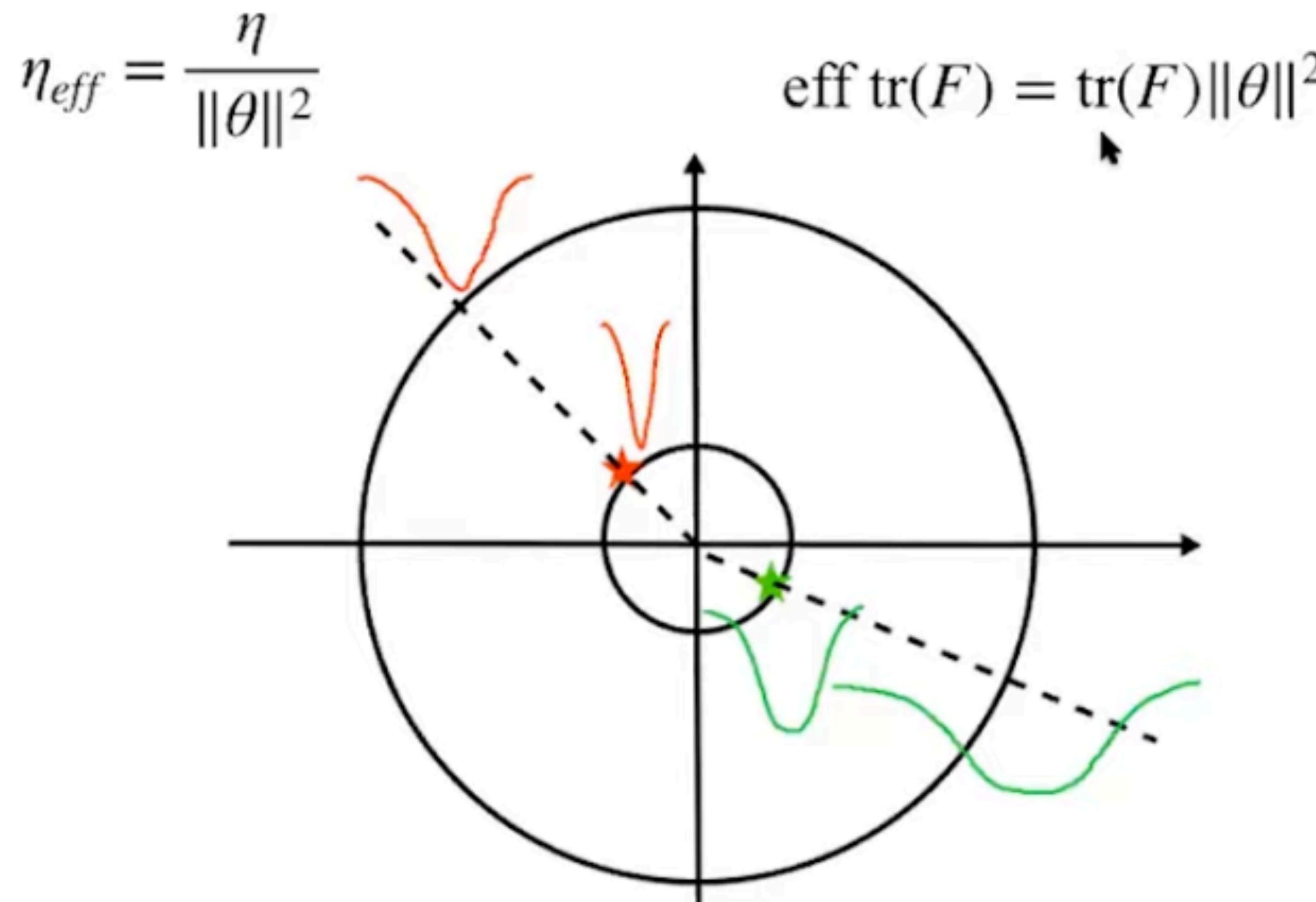


Масштабная инвариантность

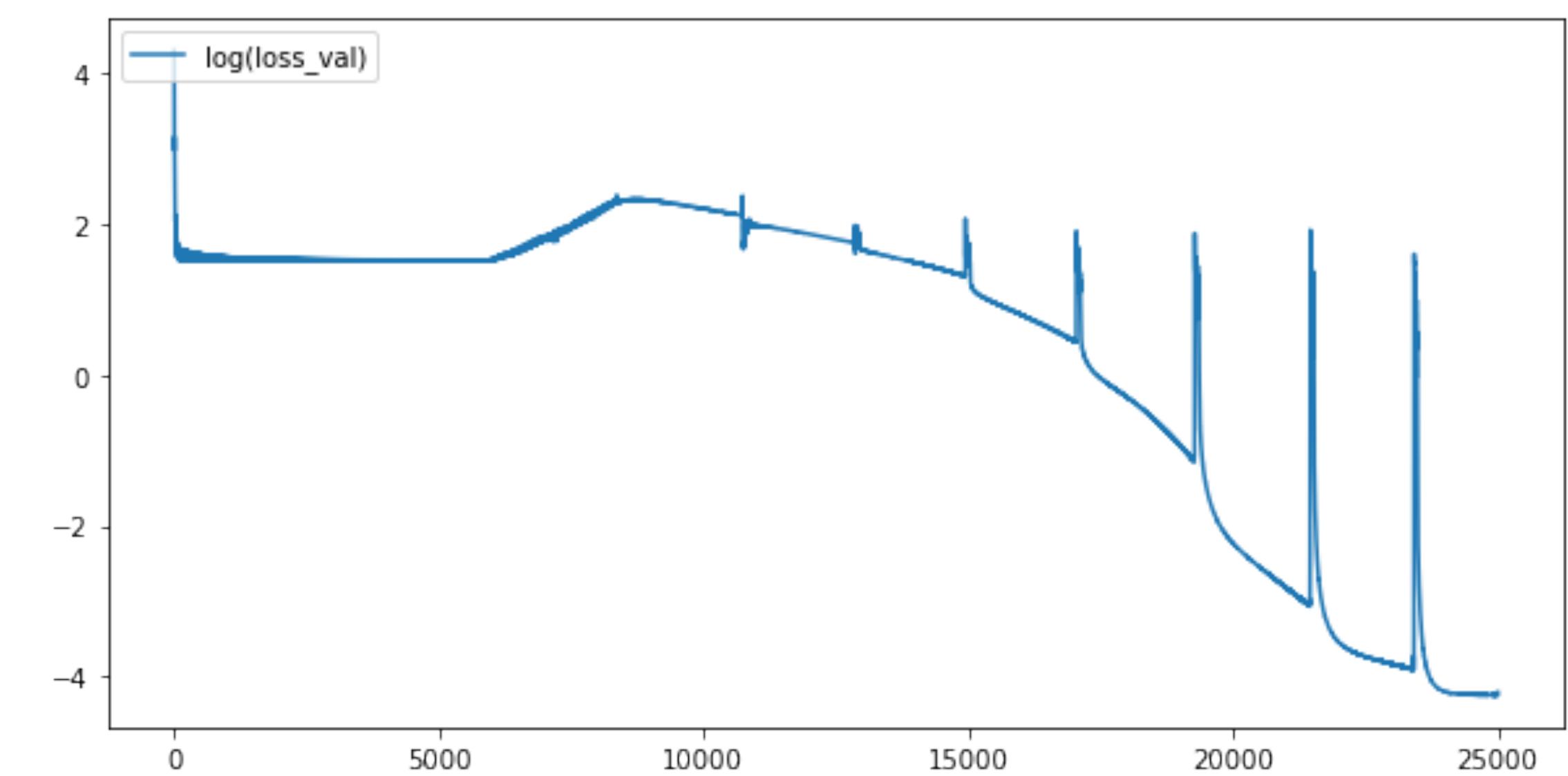
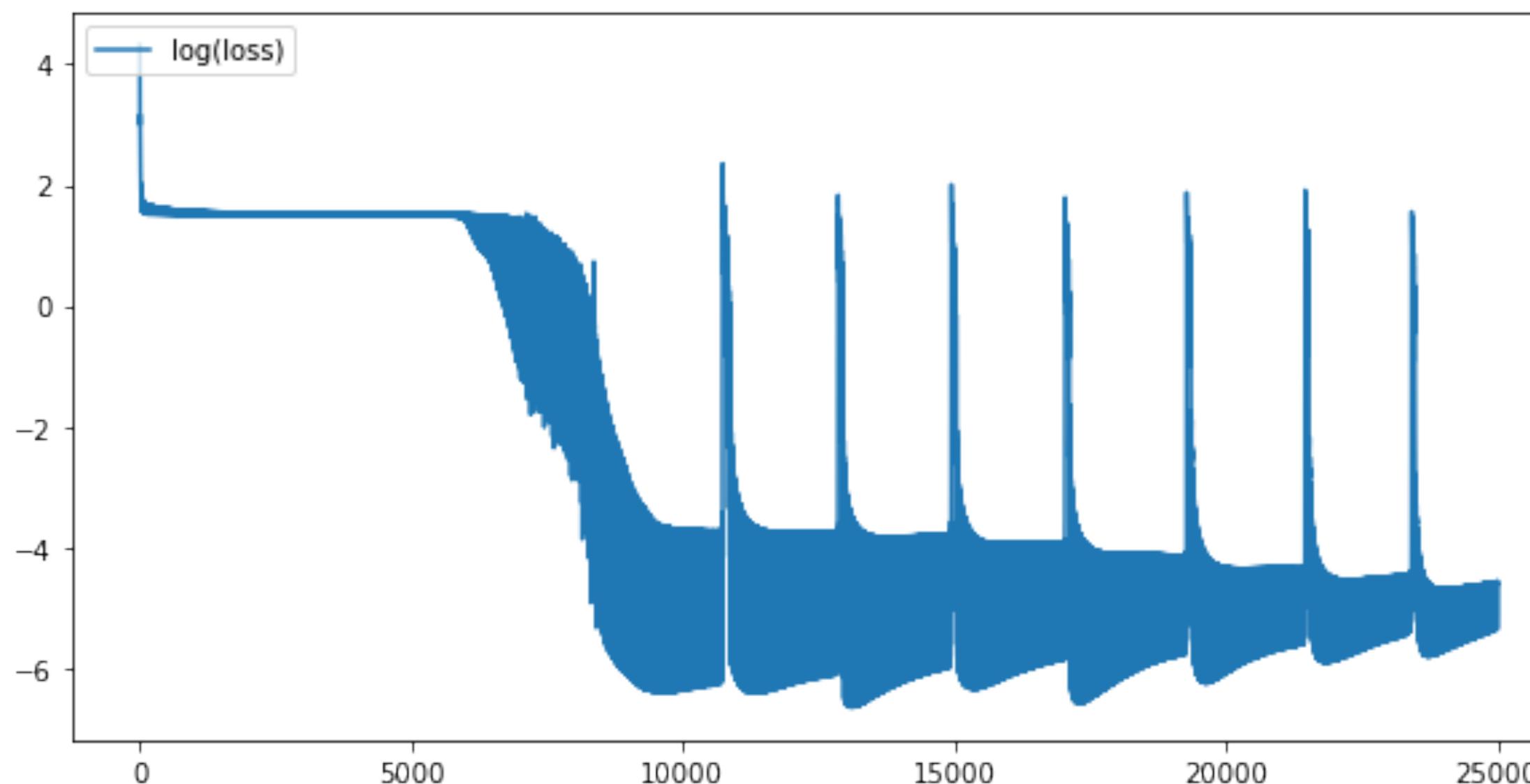
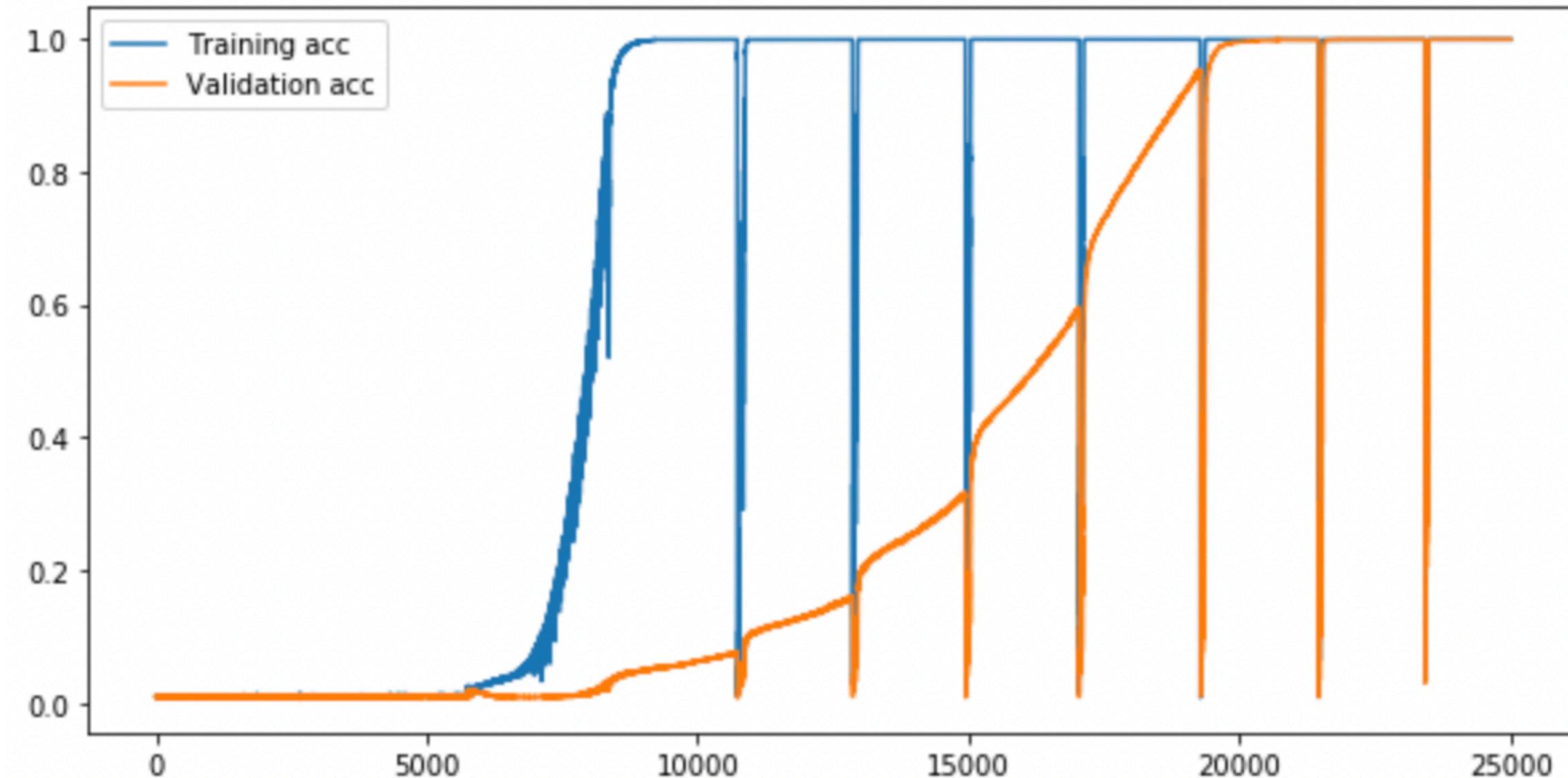
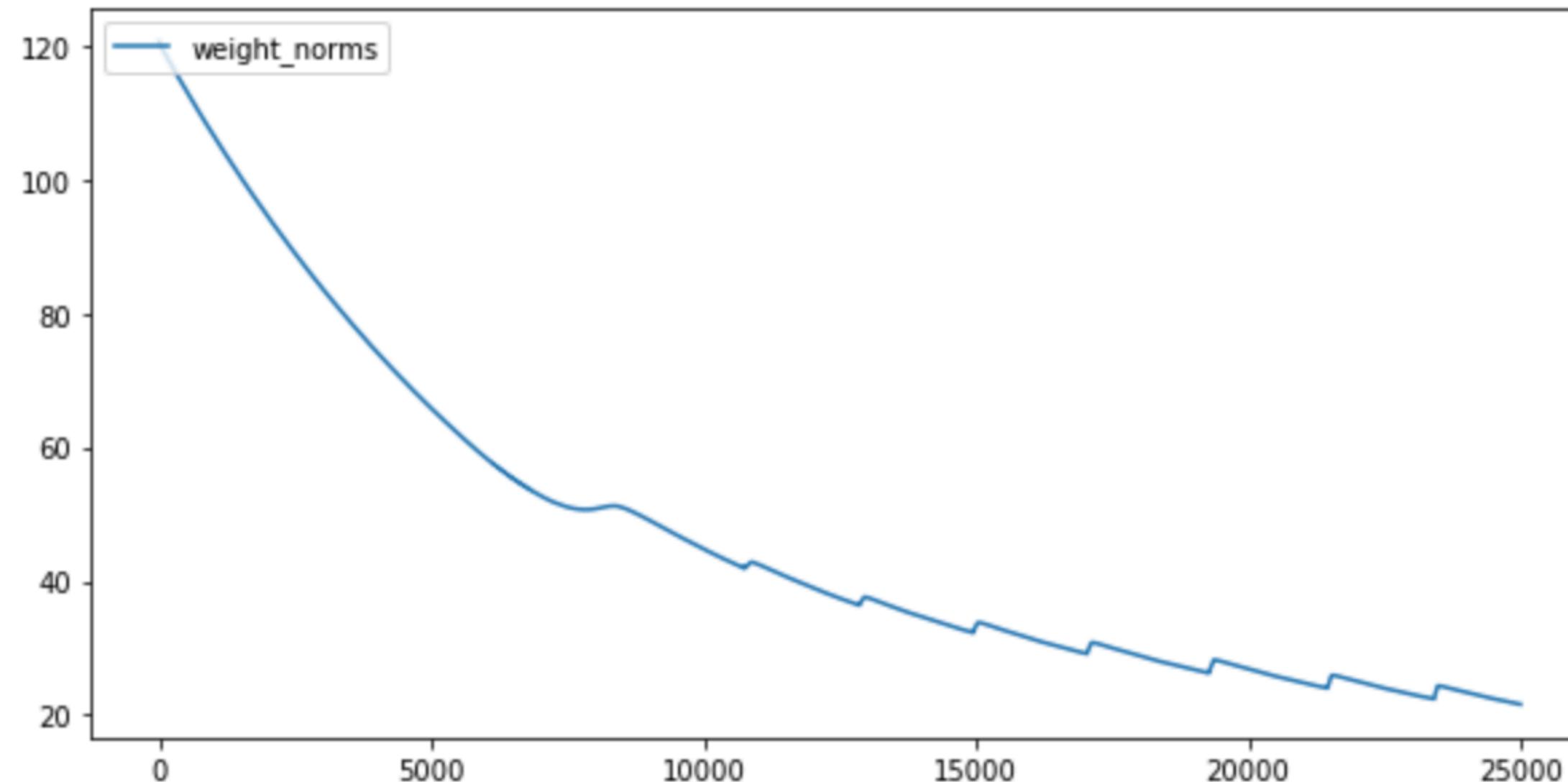
$$\log p(y|x, \theta) = \log p(y|x, C\theta), \quad \forall C > 0$$



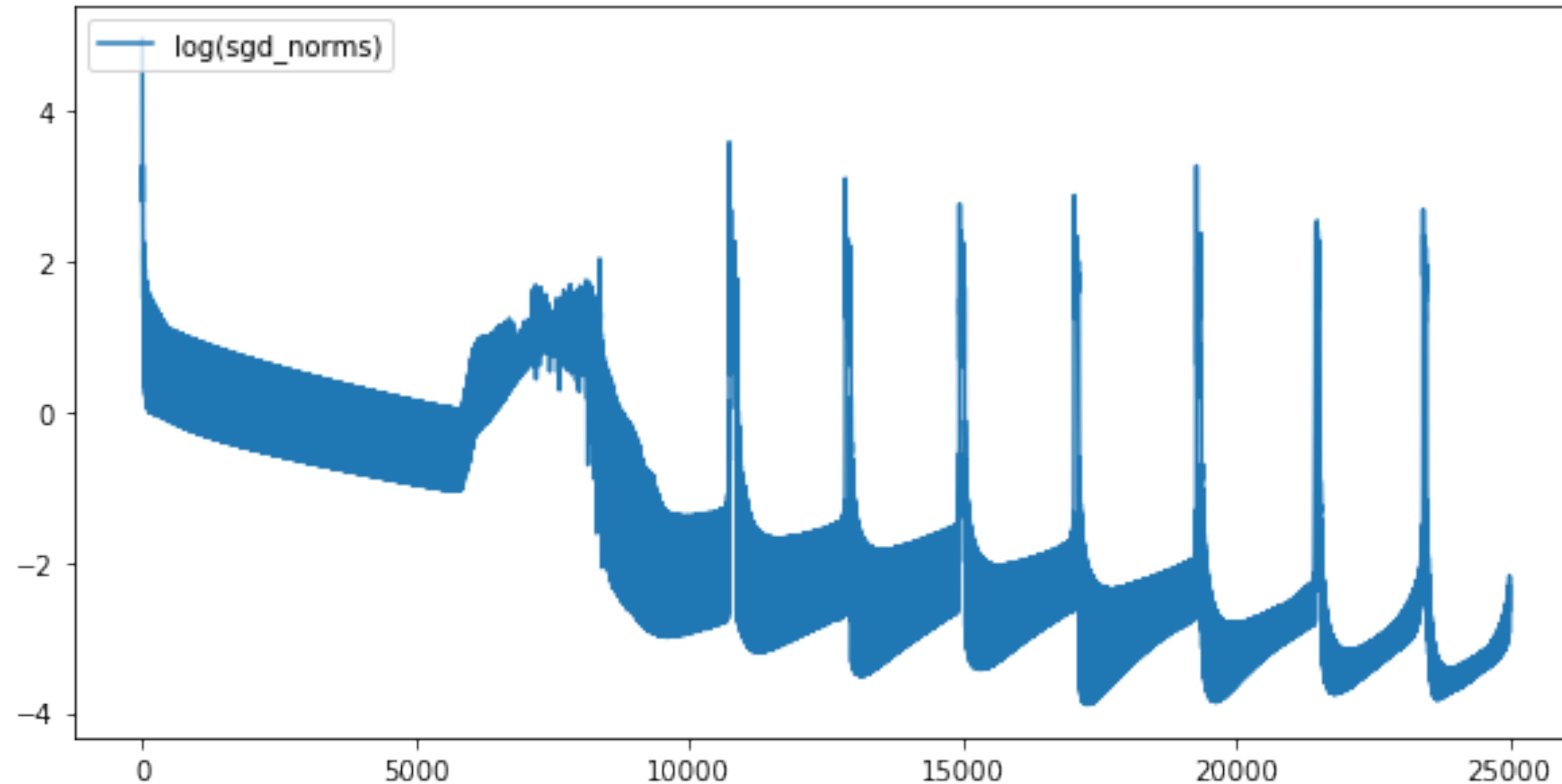
Масштабная инвариантность



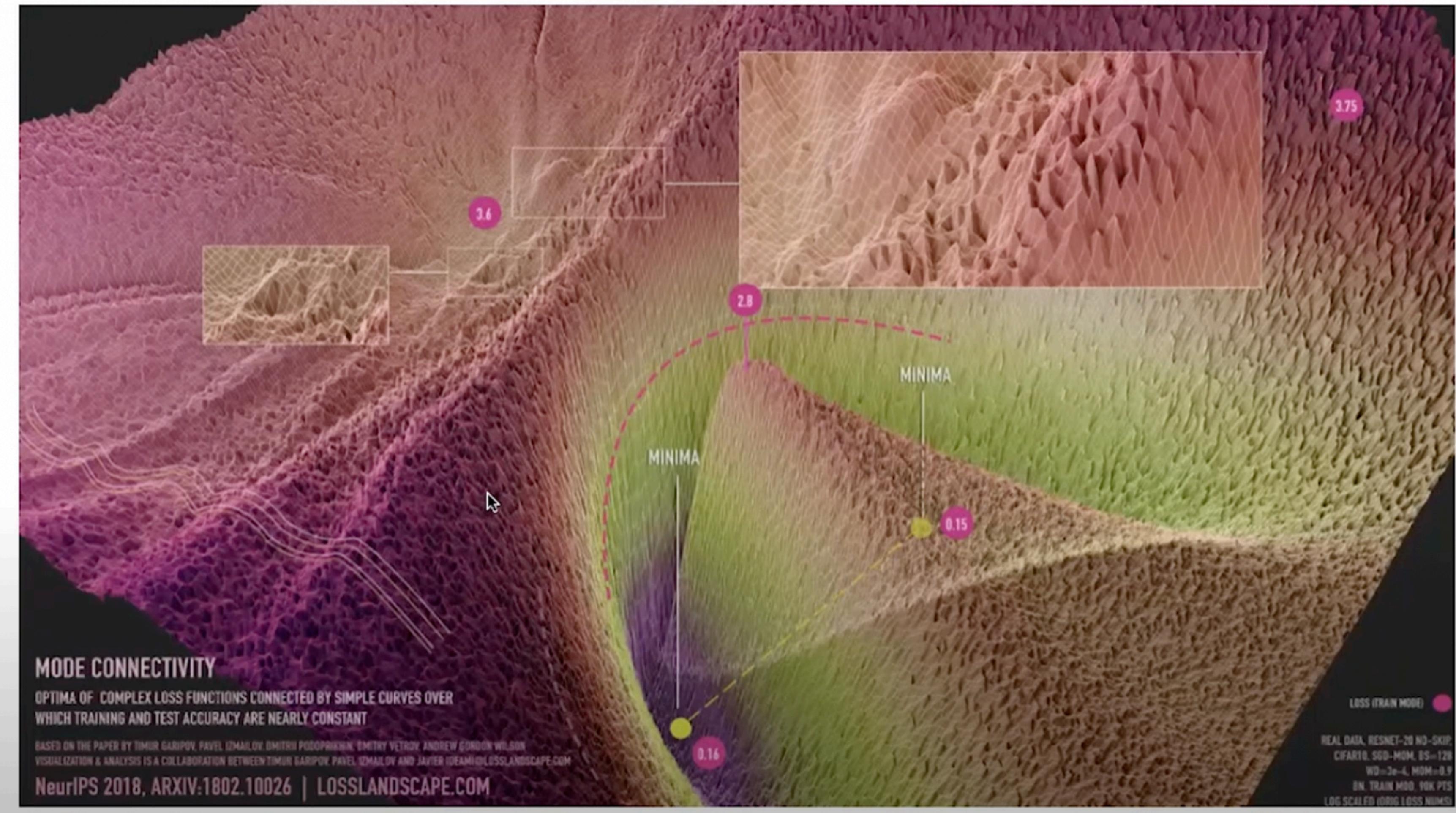
Оптимизация с помощью SGD



Оптимизация с помощью SGD

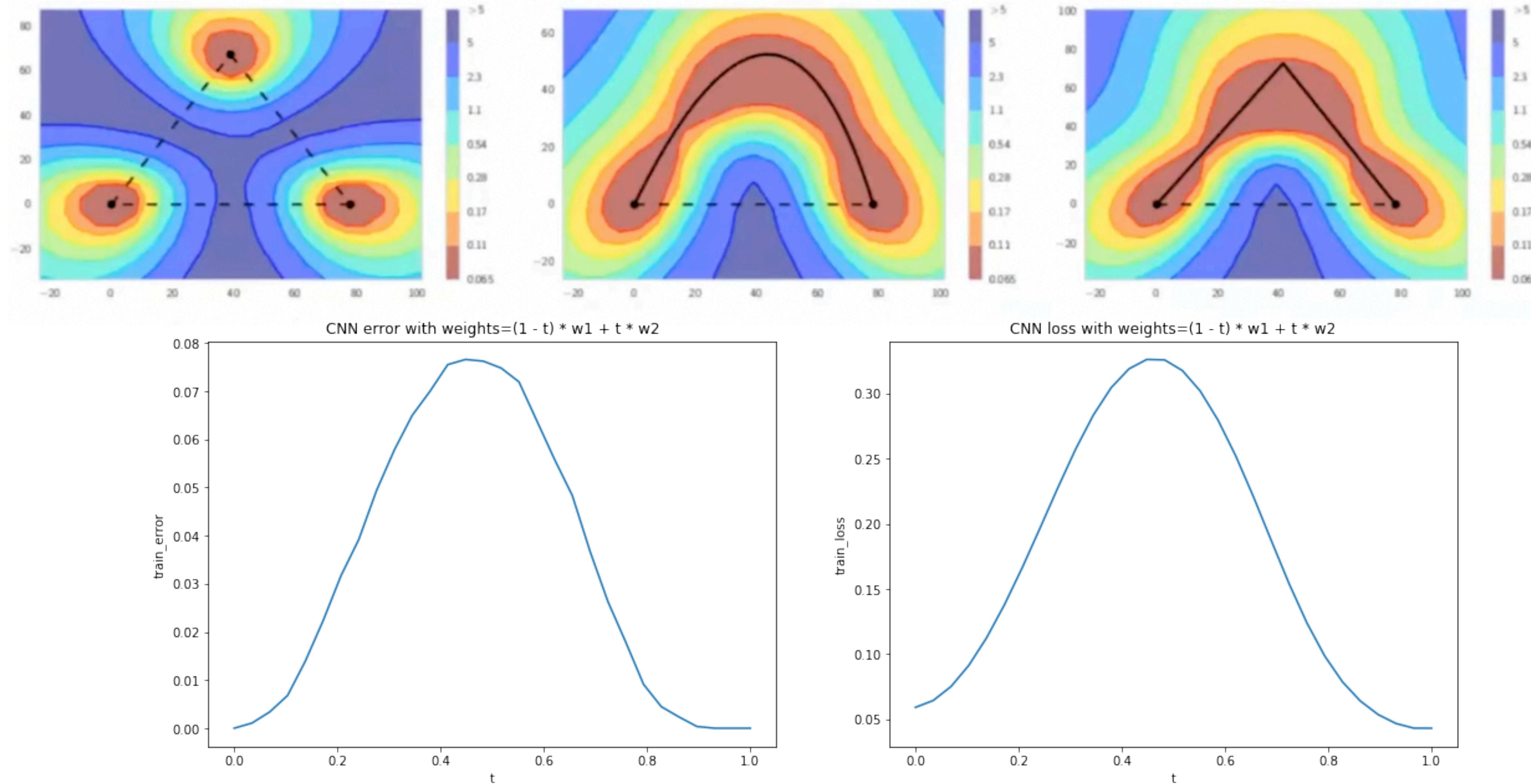


Связанность мод

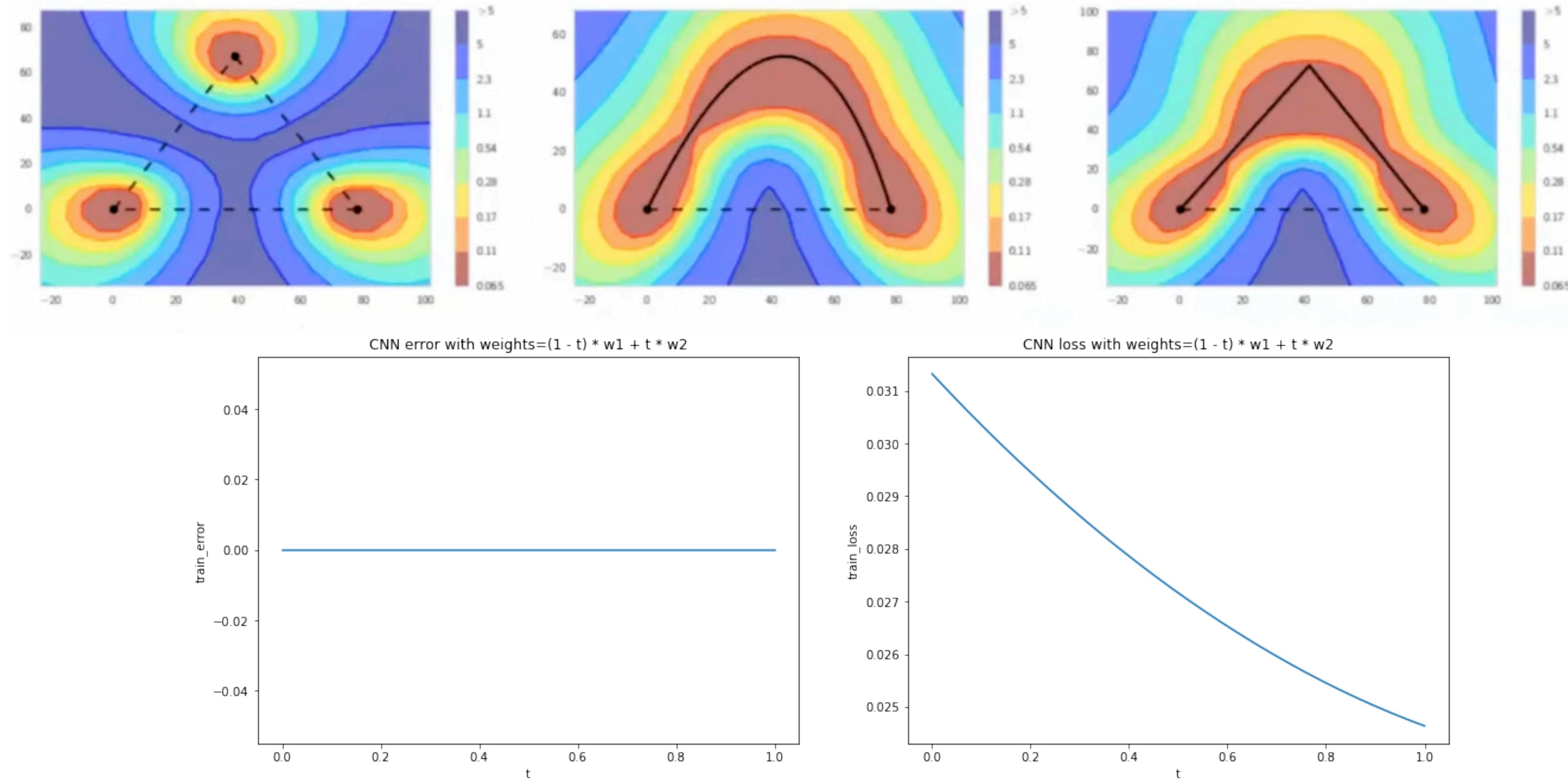


<https://arxiv.org/abs/1802.10026>

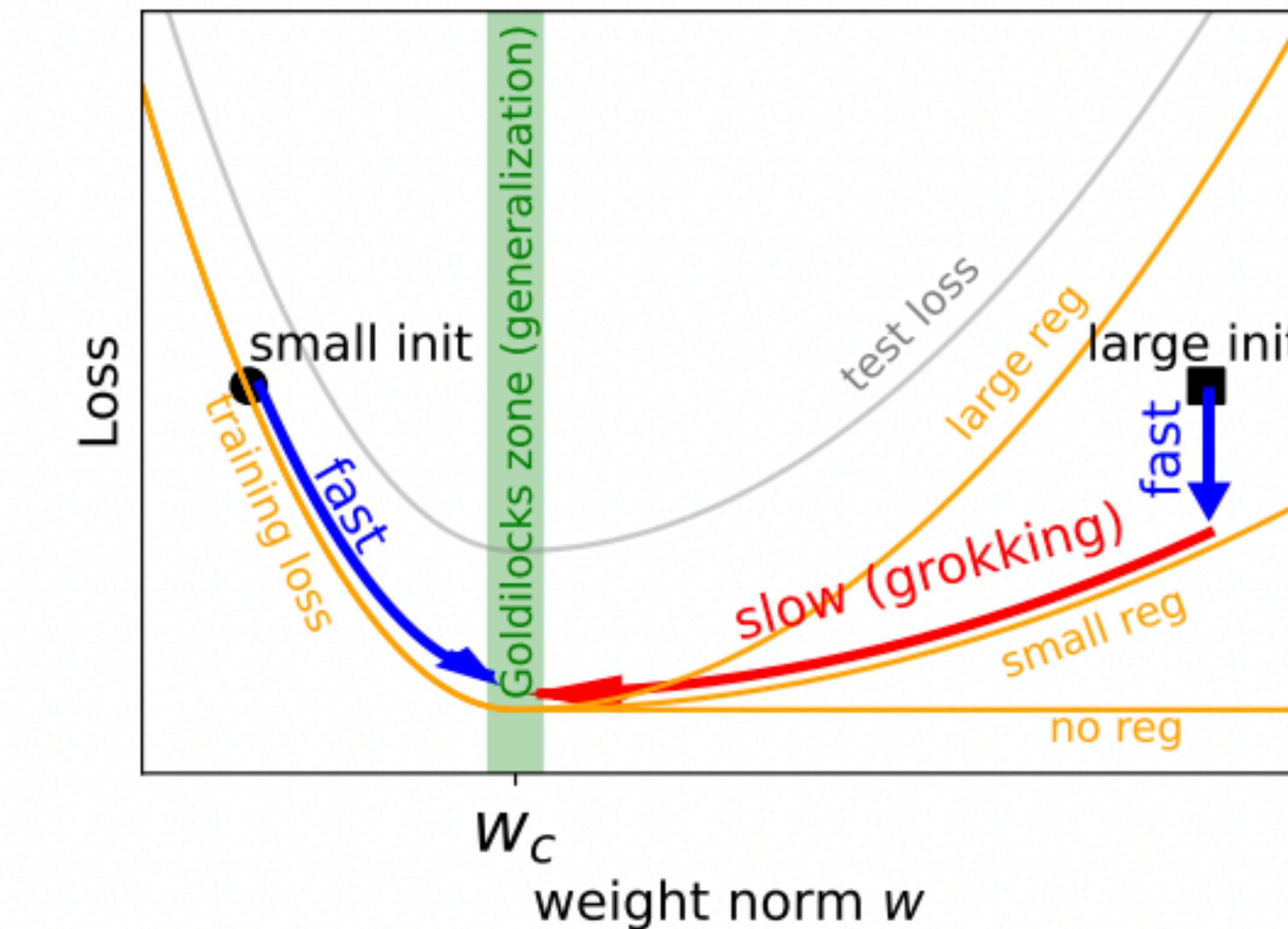
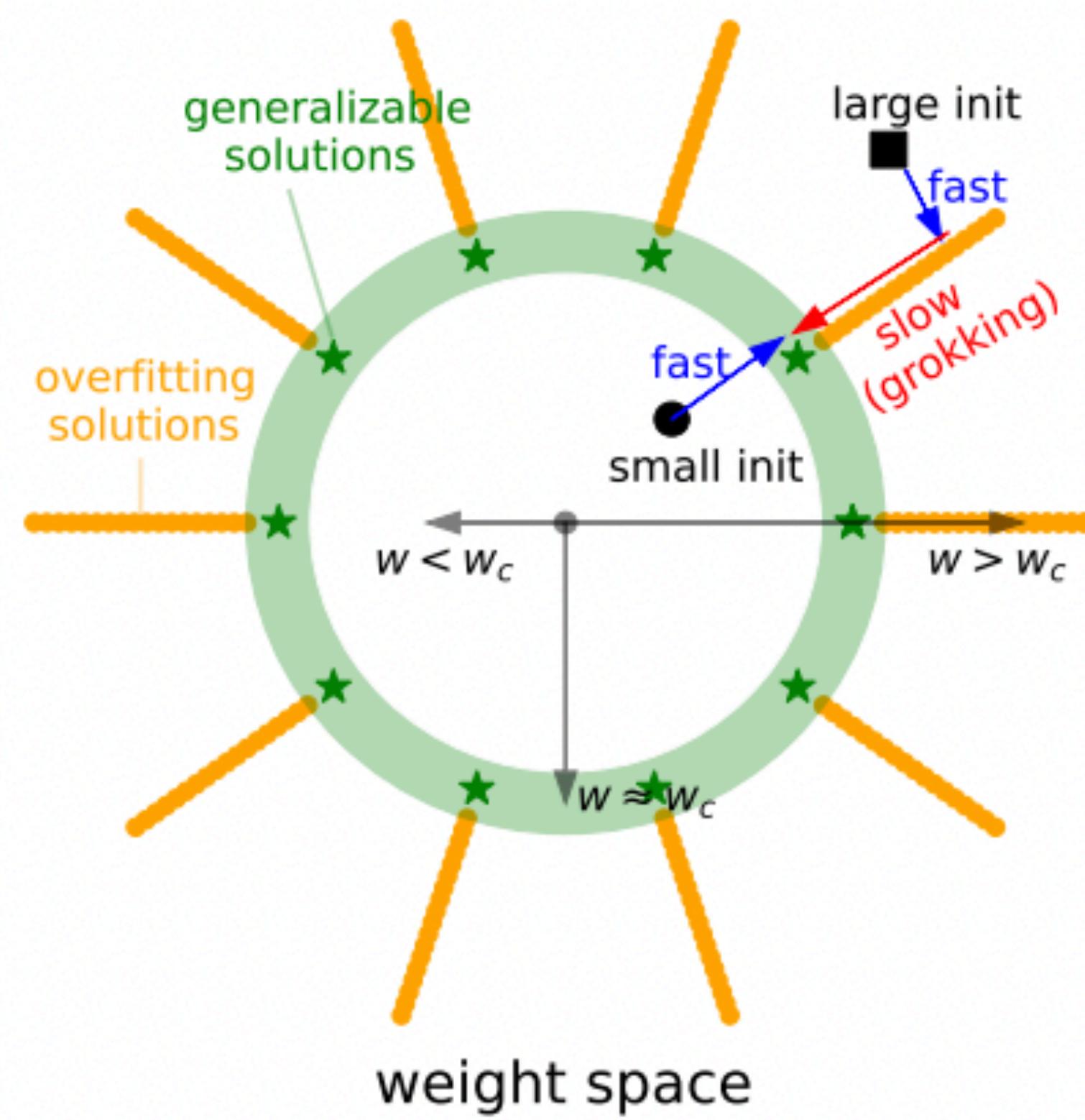
Связанность мод



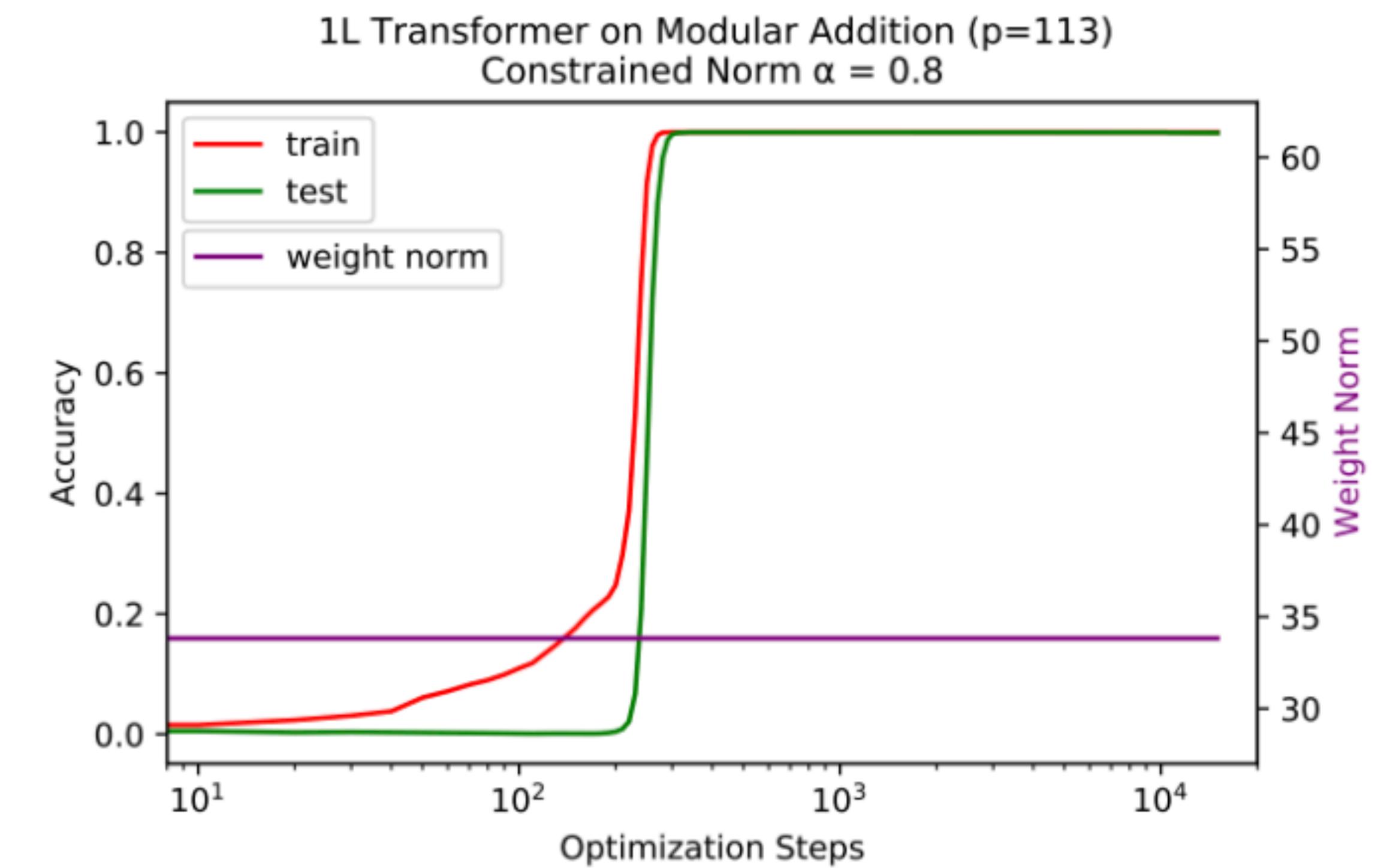
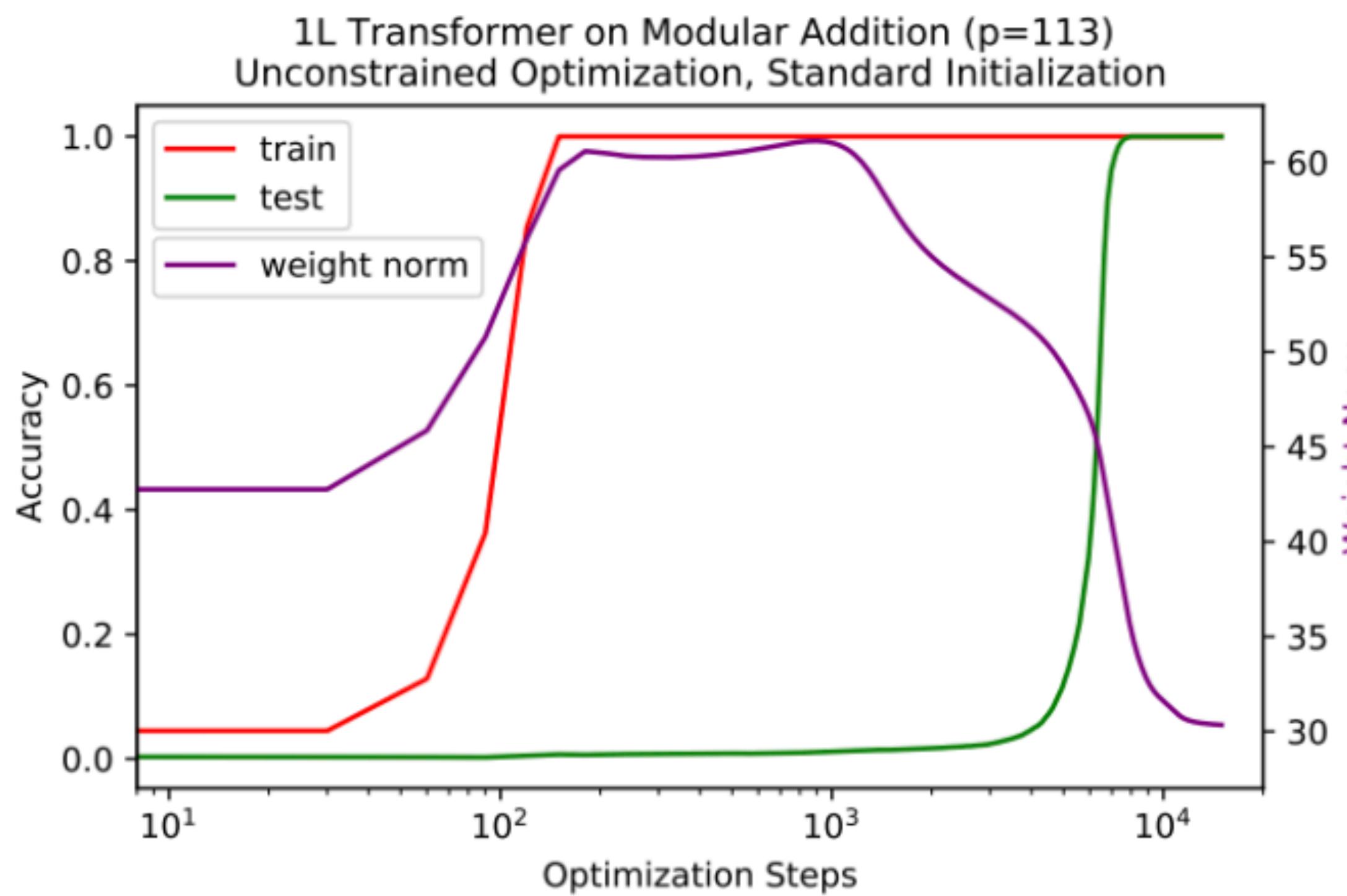
Связанность мод



OMNIGROK



OMNIGROK



<https://openreview.net/pdf?id=zDiHoIWa0q1>

END