Advanced Optimization in Deep Learning

2.1 The Problem of Poor Conditioning

August 4, 2025

Beyond Vanilla Gradient Descent

- Vanilla gradient descent can be slow and unstable, especially in deep networks.
- Advanced optimizers are designed to address these shortcomings.
- Today, we'll focus on a key challenge: **poor conditioning**.

Navigating the Error Surface

The error surface of a neural network is often shaped like a long, narrow valley. This is what we call an **ill-conditioned** or **poorly conditioned** problem.

Figure: An example of an elongated error surface, as seen in Optimization I.pdf, p. 18.

Why a Single Learning Rate Fails

- In a poorly conditioned error surface, the curvature is very different along different directions.
- This means that the optimal learning rate also varies depending on the direction.
- A single learning rate will cause:
  - **Slow progress** along the shallow directions.
  - **Oscillations** along the steep directions.
- This makes the optimization process inefficient and unstable.

Conclusion

- Poor conditioning is a major challenge in optimizing deep neural networks.
- It arises from the complex, elongated shape of the error surface.
- A single learning rate is not sufficient to navigate this landscape effectively.
- This motivates the need for more advanced optimization techniques, which we will explore next.