Visualizing Loss Functions as Topological Landscape Profiles

Caleb Geniesse* CGENIESSE@LBL.GOV
Lawrence Berkeley National Laboratory

Jiaqing Chen* JCHEN501@ASU.EDU
Arizona State University

Tiankai Xie* TXIE21@ASU.EDU
Arizona State University

Ge Shi GESHI@UCDAVIS.EDU
University of California, Davis

Yaoqing Yang YAOQING.YANG@DARTMOUTH.EDU
Dartmouth College

Dmitriy Morozov DMOROZOV@LBL.GOV
Lawrence Berkeley National Laboratory

Talita Perciano TPERCIANO@LBL.GOV
Lawrence Berkeley National Laboratory

Michael W. Mahoney MMAHONEY@STAT.BERKELEY.EDU
ICSI, LBNL, and UC Berkeley

Ross Maciejewski RMACIEJE@ASU.EDU
Arizona State University

Gunther H. Weber GHWEBER@LBL.GOV
Lawrence Berkeley National Laboratory

Editors: List of editors' names

Abstract

In machine learning, a loss function measures the difference between model predictions and ground-truth (or target) values. For neural network models, visualizing how this loss changes as model parameters are varied can provide insights into the local structure of the so-called loss landscape (e.g., smoothness) as well as global properties of the underlying model (e.g., generalization performance). While various methods for visualizing the loss landscape have been proposed, many approaches limit sampling to just one or two directions, ignoring potentially relevant information in this extremely high-dimensional space. This paper introduces a new representation based on topological data analysis that enables the visualization of higher-dimensional loss landscapes. After describing this new topological landscape profile representation, we show how the shape of loss landscapes can reveal new details about model performance and learning dynamics, highlighting several use cases, including image segmentation (e.g., UNet) and scientific machine learning (e.g., physics-informed neural networks). Through these examples, we provide new insights into how loss landscapes vary across distinct hyperparameter spaces: we find that the topology of the loss landscape is simpler for better-performing models; and we observe greater variation in the shape of loss landscapes near transitions from low to high model performance.

Keywords: Topological data analysis, loss landscapes, model diagnosis

* Equal contribution.

1. Introduction

A central aim of machine learning (Simonyan and Zisserman, 2014; He et al., 2016; Krizhevsky et al., 2017; Vaswani et al., 2017; Devlin et al., 2018; Liu et al., 2019) is to learn the underlying structure of data. This learning process is governed by a *loss function*, denoted as $\mathcal{L}(\theta)$, where θ is the set of parameters (or weights) defining, e.g., a neural network. The loss function measures the difference between the outputs of a neural network and ground-truth values. In this way, the loss reflects how good (or bad) the current weights are at making correct predictions and how to adjust these weights during training. Given the important role that the loss function plays during learning, examining it with respect to a neural network's weights—by visualizing the so-called *loss landscape*—can provide valuable insights into both network architecture and learning dynamics (Goodfellow et al., 2014; Im et al., 2016; Li et al., 2018; Yao et al., 2020; Martin and Mahoney, 2021; Martin et al., 2021; Yang et al., 2022b, 2021; Zhou et al., 2023; Sakarvadia et al., 2024; Khan et al., 2024). Indeed, the loss landscape has been essential for understanding other aspects of deep learning, including generalizability (Cha et al., 2021; Yang et al., 2021) and robustness (Kurakin et al., 2016; Djolonga et al., 2021; Yang et al., 2022a). In addition, the loss landscape has been characterized in the context of scientific machine learning, e.g., to understand why different physics-informed architectures and loss functions are often brittle, exhibiting failure modes, and are hard to optimize (Krishnapriyan et al., 2021; Rathore et al., 2024; Xie et al., 2024).

Despite its promise and appeal, *loss landscape visualization* is a complex and often bespoke process. Indeed, exploring and extracting insights from a loss landscape—which is inherently high-dimensional, with as many dimensions as the number of parameters in the model—is challenging to do, especially when trying to visualize directly on a two-dimensional screen. Most efforts to date have focused on projecting the loss function down to one or two dimensions. Goodfellow et al. (2014) proposed a random-direction-based approach, where model parameters are interpolated along a one-dimensional path to see how the loss changes. Im et al. (2016) later introduced an extension of this method which involves projecting the loss landscape onto a two-dimensional space using barycentric interpolation between triplets of points and bilinear interpolation between quartets of points. Li et al. (2018) continued improving the resolution of loss landscapes by introducing filter-wise normalization to remove the scaling effects incurred by previous approaches. A more sophisticated approach to visualizing the loss landscape leverages the Hessian to define more relevant directions along which the model can be interpolated. More recently, Yao et al. (2020) used the top two Hessian eigenvectors as directions, thereby capturing more important changes in the underlying loss landscape. While various methods have been proposed, most applications have limited sampling to just one or two directions. Importantly, by restricting the sampling of loss landscapes to two dimensions, whether it be using random or Hessian-based directions, we ignore potentially informative information captured by additional dimensions (e.g., the eigenvectors associated with the dominant eigenvalues of the Hessian matrix).

Towards characterizing higher-dimensional loss landscapes, here we take inspiration from topological data analysis (TDA). Specifically, we use a merge tree to encode the critical points of an n-dimensional neural network loss landscape, and we represent the merge tree as a topological landscape profile. The merge tree allows us to capture important features in an arbitrary-dimensional loss landscape; and by using the topological landscape profile, we are able to re-represent this information in two dimensions. Note, we first explored applications of TDA in our previous work (Xie et al., 2024). There, we focused on quantifying loss landscapes and developing new topology-based metrics, but the sampling was limited to two dimensions. Here, we focus on developing a new visual representation of the loss landscape which allows us to visualize higher-dimensional loss landscapes. We demonstrate the utility of our new topological landscape profile representation by exploring higher-dimensional loss landscapes, i.e., sampling along more directions and representing these higher-dimensional subspaces as topological landscape profiles. This approach allows us to extract more information from the additional dimensions we consider. While our approach technically can work with arbitrary dimensional loss landscapes, in practice we are limited by sampling. As such, here we limit ourselves to three and four-dimensional loss landscapes.

We demonstrate the versatility of our new topological profile representations of loss landscapes and our complementary visualization tool through several use case scenarios. Through these examples, we show the many different ways our tool can be used to extract insights about neural network models based on our topological landscape profiles and by comparing loss landscapes across different hyperparameters. In doing so, we also provide new insights into how loss landscapes vary across distinct hyperparameter spaces, finding that (1) the topology of loss landscapes is simpler for better-performing models, and (2) this topology often exhibits greater variability near transitions from low to high model performance. For example, for the scientific machine learning models we study here, we observe a sharp transition from low to high error as one of the physical parameters is increased. We find that models with lower error have smoother and more funnel-like loss landscapes, whereas models with higher error have flatter (but rougher) and more bowl-like loss landscapes. Along this transition from low to high error, models have more variably shaped loss landscapes (i.e., different shapes are observed across different random seeds).

2. Background

2.1. Topological Data Analysis

Topological data analysis (TDA) aims to reveal the global underlying structure of data. TDA is particularly useful for studying high-dimensional data or functions, where direct visualization (in two or three dimensions) is inherently not possible. We leverage ideas and algorithms from TDA to study the global structure of the loss function—that is, the shape of the so-called loss landscape. Much of TDA is based on the more general idea of "connectedness." In the context of a loss function, we are interested in the number of minima (i.e., unique sets of parameters for which the loss is locally minimized) and how "prominent" they are (i.e., measuring how many other sets of neighboring parameters have a higher loss than the parameter set that minimizes the loss function). Such information can be obtained from a persistence diagram (i.e., captured by the zero-dimensional persistent homology) and the so-called merge tree.

A *merge tree* (Carr et al., 2003; Heine et al., 2016) tracks connected components of sub-level sets $L^{-}(v) = \{x \in D; x \leq v\}$ as a threshold, v, is increased. The merge tree encodes changes in the loss landscape as nodes in a tree-like structure. The local minima