Carving out seismic data from noise

Julián L. Gómez[12] and Danilo R. Velis[1], [1] Facultad de Ciencias Astronómicas y Geofísicas, Universidad Nacional de La Plata, and CONICET, Argentina, [2] YPF-Tecnología S.A.

This paper was prepared for presentation during the 18th International Congress of the Brazilian Geophysical Society held in Rio de Janeiro, Brazil, 16-19 October 2023.

Abstract

Machine learning offers a vast and growing landscape for digital signal processing, including seismic data. Very recently, probabilistic diffusion methods were shown to lead to unprecedented results in image and data synthesis, surpassing state-of-the-art algorithms such as the well-known generative adversarial networks, while offering more robust training schemes. Denoising score-matching aims to approximate the score function of the unknown and complex probability density function of the data using a deep neural network. The score function is the gradient with respect to data of the logarithm of the data-generating distribution. Once the vector field of the score function is learned, a sampling process based on gradient ascent will draw fair samples belonging to the original data-generating distribution. In this work we make use of Denoising Score-Matching and Annealed Langenvin Dynamics to get samples from the approximated distribution. In this framework, the score learning and data sampling involve a diffusion process, where the original and unknown data distribution are contaminated by different levels of additive noise. This process maps the original data distribution to a simpler shape, like a canonical Gaussian distribution. From the simpler distribution, samples are easily obtained and then transformed to samples from the true distribution reversing the noising process. Denoising score matching offers the possibility of achieving true manifold interpolation, where the interpolated traces belong to the same approximated data-generating distribution from the training set. We recall that the manifold hypothesis suggests that most field data lives only on a subspace of the ambient space where it is registered. We investigate the impact of the sampling parameters in the final results, and the number of required seismic traces needed to learn a good representation of the underlying data distribution. With the goal of applying denoising score matching on seismic applications, we device a series of tests to validate the technique in order to apply it to field data denoising and interpolation. We check the consistency of the produced samples in the Fourier domain and investigate constraints to produce more realistic results with respect to the given training data. We consider a situation where computing time and hardware are limited to a small budget, striving to show results that can be generated with minimum hardware requirements. The results of our experiments of seismic data generation are promising.