Automatic fault extraction based on Topological Data Analysis
Master Internship + PhD proposal

Julien Tierny[1], Mélanie Plainchault[2]
[1] *CNRS, Sorbonne-Université, Paris, France*
[2] *Total SA, Pau, France*
julien.tierny@sorbonne-universite.fr, melanie.plainchault@total.com

Keywords: Data Analysis, Deep Learning (post-processing), Fault Extraction, Geosciences, Morse-Smale Complex.

1 CONTEXT

Total SA is a French Oil & Gas company whose activity is shared between upstream (exploration and hydrocarbon extraction) and downstream aspects (production and sell of gasoline, lubricants etc).

For hydrocarbon extraction, a whole field study methodology is used in order to maximize the quantity of recovered oil and gas. This methodology requires the design of a structural model of the underground, see for instance Figure 1. It is composed of horizons, corresponding to iso-time sediment deposition, and faults representing breaking events in the strati-graphic layers.

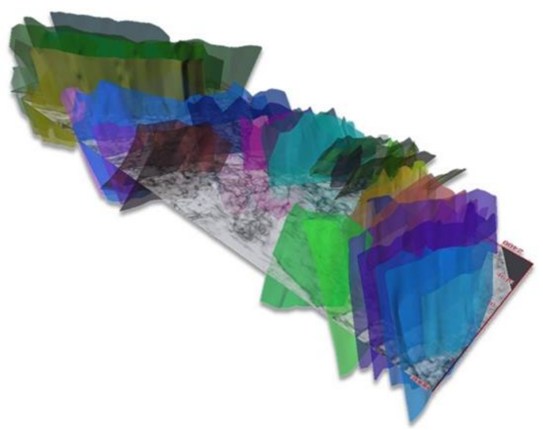

Figure 1: Fault model on a time-slice of the coherency seismic attribute, each fault surface is represented by a colored trimesh. *Courtesy of Sismage-CIG Team*

These horizons and faults are extracted by geophysicists manually picking through seismic images presented on Figure 2.

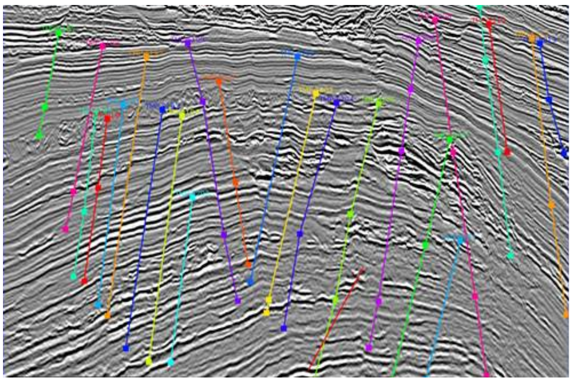

Figure 2: Fault sticks which have been picked by hand on a seismic image. Sediment horizons correspond to the black and white lines in the back ground, while faults appear as discontinuities in these horizons. Each stick corresponds to the presence of a fault on this seismic image. *Courtesy of Sismage-CIG Team*

This work is particularly tedious, and any automatic extraction method can be highly valuable to ease and accelerate this phase. For instance, David Hale in [1] proposes an automatic fault extraction 3D method from a fault likelihood metric which is based on crease surfaces [2]. Some other extraction method focus on using one or several seismic attributes in order to precisely define this fault position in the data set [3].

2 RESEARCH PROBLEM

For this research work, we will focus on a fault probability presence volume data set which is represented on Figure 3 where high probabilities are yellow spots. Such probability volumes are produced by deep-learning algorithms in the context of a collaborative project involving Total and Google. While these algorithms manage to produce highly relevant fault presence probability estimates, their interpretation for geometrical analysis remains challenging. In particular, geophysicists would like to extract an explicit representation of these faults (in the form of a triangular surface, to perform various measurements on them: size, curvature, etc.) as well as a higher level understanding of their global structure (how faults intersect and connect together). However, no off-the-shelf algorithm exists for such a post-processing of these deep-learning results.

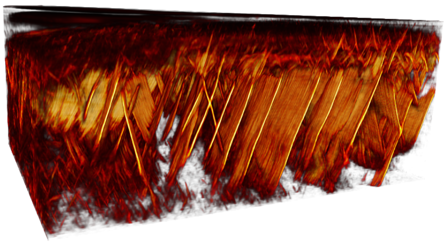

Figure 3: Volumic representation of fault probability (yellow stops represent the highest probability locations)

In this work, we want to explore how Topological Data Analysis [4] can be used for the exploitation of feature presence probability fields generated by deep-learning algorithms, in the context of fault extraction for geosciences. In particular, we would like to focus on the Morse-Smale complex [5], which is a topological object that is, in principle, well suited for the extraction of surfaces locally maximizing a scalar function [6] (here the presence probability field). While our preliminary experiments confirm the relevance of this research directions, many research questions remain open.

In this research work, we will focus on:

1. How to exploit the Morse-Smale complex to extract the network of fault surfaces:
   - Designing algorithms to coherently gather surfaces patches from the Morse-Smale complex in order to recover the fault surfaces;
   - Designing algorithms to extract the global structure of the set of fault surfaces;
   - Developing, in collaboration with the geoscientists, new representations of networks of fault surfaces (possibly including physical properties), as well as new methods for their interpretation.
2. How to make this approach scale for real-life datasets used at Total:
   - Designing algorithms for the simplification of the Morse-Smale complex, to account for the presence of noise in the probability fields;
   - Designing algorithms capable of handling large-scale probability fields (several dozens gigabytes in size), possibly out-of-core or in a distributed manner;
   - Designing time-efficient algorithms (possibly shared-memory parallel);
   - Applying all the designed algorithms on real-life use cases.

3 ORGANIZATION

Ideally, we would like to organize this project into a master-2 level internship followed, in case of success, by a Ph.D. thesis.

Note that the master-2 level internship is not a requirement and that candidates already holding a master-2 degree are invited to apply directly to the Ph.D. thesis.

3.1 Master Internship

The master-2 internship would last from 16 to 24 weeks, depending on the availability of the student (standard French academic allowance: 500 euros per month). The internship would mostly focus on the development of a prototype for fault extraction based on the Morse-Smale complex (item 1 of the bullet list, previous section) on synthetic data sets resembling fault probability fields. The internship could be organized as follows:

1. Review of the literature in Topological Data Analysis [4], especially regarding Persistent Homology and Morse-Smale complexes.
2. Design of an algorithm for the extraction of locally maximizing surfaces in a probability field (based on the Morse-Smale complex);
3. Design of an algorithm for the extraction of the global structure of the set of maximizing surfaces;
4. Experiments on manufactured synthetic examples created from a ground truth.

3.2 Ph.D. Thesis

In case of success of the internship, we would like to continue this research into a Ph.D. thesis (3 years exactly in France). The thesis could be organized as follows:

1. In-depth review of the literature in Topological Data Analysis [4] (with a focus on Persistent Homology and Morse-Smale complexes);
2. Preliminary use-case study on a selected real-life example of small size;
3. Exploration of the following research questions:
   - How to make this approach more time efficient?
   - How to make this approach scale to real-life data sets of large size?
4. Full size case study on real-life data sets in collaboration with geophysicists;
5. Exploration of the following perspective questions:
   - How to help pre-process the training data for the deep learning approach generating the probability field?
   - How to generalize this approach to other problems involving surface presence probability estimates (for example: density estimations of LIDAR point clouds)?

4 ENVIRONMENT

This PhD will be co-supervised by Julien Tierny [7] and Mélanie Plainchault, who already co-supervised a Ph.D. thesis previously, on topological methods for material sciences [8, 9, 10, 11]. It will be a CIFRE doctoral program promoting research collaboration between universities and companies, see [12] for more information.

Research time will be shared between the computer science department (LIP6) of Sorbonne University (downtown Paris – Jussieu subway station – France) and Total (Pau, France) in order to take benefit from both the academic environment and the feedback from end users, i.e. geoscientists. The balance between academic and company time is adjustable and will be decided in collaboration with the student. The internship will take place at Sorbonne University, with possible visits to Total. Travels to visit the California-based Total team producing the probability fields at Google may need to be considered. This work will lead to publications and participations to international conferences (such as IEEE VIS [13]). Most of the developed code will be released open-source in TTK library [14].

5 APPLICATION

We are looking for a highly motivated student, with strong C++ programming skills, a clear interest for Topological Data Analysis and its applications, as well as a good English (spoken/written) level. Some background in geosciences would be a plus.

To apply, candidates are invited to send us their CV and a short cover letter by email to julien.tierny@sorbonne-universite.fr and melanie.plainchault@total.com.

Note that the master-2 level internship is not a requirement and that candidates already holding a master-2 degree are invited to apply directly to the Ph.D. thesis.

REFERENCES

[1] D. Hale, "Methods to compute fault images, extract fault surfaces, and estimate fault throws from 3d seismic images," *Geophysics*, 2013.

[2] T. Schultz, H. Theisel, and H. P. Seidel, "Crease surfaces: From theory to extraction and application to diffusion tensor MRI," *IEEE Transactions on Visualization and Computer Graphics*, 2009.

[3] M. Bahorich and S. Farmer, "3-d seismic discontinuity for faults and stratigraphic features: The coherence cube," *The Leading Edge*, 1995.

[4] H. Edelsbrunner and J. Harer, *Computational Topology: an Introduction.*

[5] A. Gyulassy, P. T. Bremer, B. Hamann, and V. Pascucci, "A practical approach to Morse-Smale complex computation: Scalability and Generality," *IEEE Transactions on Visualization and Computer Graphics (Proc. of IEEE VIS)*, 2008.

[6] A. Gyulassy, N. Kotava, M. Kim, C. D. Hansen, H. Hagen, and V. Pascucci, "Direct Feature Visualization Using Morse-Smale Complexes," *IEEE Transactions on Visualization and Computer Graphics*, 2012.

[7] J. Tierny, "Homepage," https://julien-tierny.github.io/.

[8] M. Soler, "Large Data Reduction and Structure Comparison with Topological Data Analysis," Ph.D. dissertation, Sorbonne University, 2019, https://hal.archives-ouvertes.fr/tel-02171190.

[9] M. Soler, M. Plainchault, B. Conche, and J. Tierny, "Topologically controlled lossy compression," in *Proc. of IEEE PacificVis*, 2018,

https://julien-tierny.github.io/stuff/papers/soler_pv18.pdf.

[10] ——, "Lifted Wasserstein Matcher for Fast and Robust Topology Tracking," in *IEEE Symposium on Large Data Analysis and Visualization*, 2018, Best Paper Honorable Mention Award, https://arxiv.org/pdf/1808.05870.

[11] M. Soler, M. Petitfrere, G. Darche, M. Plainchault, B. Conche, and J. Tierny, "Ranking Viscous Finger Simulations to an Acquired Ground Truth with Topology-Aware Matchings," in *IEEE Symposium on Large Data Analysis and Visualization*, 2019, Best Paper Award, https://julien-tierny.github.io/stuff/papers/soler_ldav19.pdf.

[12] ANRT, "CIFRE Program," http://www.anrt.asso.fr/fr/cifre-7843.

[13] IEEE, "VIS conference," http://ieeevis.org/.

[14] J. Tierny, G. Favelier, J. A. Levine, C. Gueunet, and M. Michaux, "The Topology ToolKit," *IEEE Transactions on Visualization and Computer Graphics (Proc. of IEEE VIS)*, 2017, Best Paper Honorable Mention Award, https://topology-tool-kit.github.io/.