La topologie algébrique à la rescousse de l'analyse de données

Jean-Philippe Chassé

Département de mathématiques et statistique
Université de Montréal

7 mars 2018

Plan

1. Topologie et géométrie
   - Définitions et différences
   - Branches
2. Un aperçu d'application
   - Complexe de Vietoris-Rips
   - Nombres de Betti
   - Persistance homologique

Plan

1. Topologie et géométrie
   - Définitions et différences
   - Branches
2. Un aperçu d'application
   - Complexe de Vietoris-Rips
   - Nombres de Betti
   - Persistance homologique

Que sont-ce que la topologie et la géométrie ?

Il n'y a pas de définition universellement acceptée de ce qu'est la topologie, et encore moins de ce qu'est la géométrie. Cependant, on pourrait généralement dire que

- La **topologie** est l'étude des propriétés globales d'espaces abstraits préservées sous les transformations continues.
- La **géométrie** est l'étude des propriétés locales d'espaces abstraits préservées sous les transformations respectant des conditions plus strictes que la simple continuité.

Invariants propres à la topologie et à la géométrie

Une bonne façon de différentier des structures mathématiques est de définir des invariants préservés sous les transformations d'intérêt ; alors si deux structures ont des invariants différents, ils seront différents.

De ce fait, une bonne façon de comprendre ce que la topologie et la géométrie étudient est de comprendre les invariants pertinents à ces sujets :

- **Invariants topologiques :** dimension d'une variété, genre d'une surface, compacité, connexité, etc.
- **Invariants géométriques :** distance entre deux points, courbure, angle à l'intersection de courbes, etc.

Une transformation continue ne préservant pas la courbure

Quelques branches de topologie et géométrie

- Topologie différentielle
- Topologie algébrique
- Topologie symplectique

- Géométrie riemannienne/pseudo-riemannienne
- Géométrie spectrale

Plan

1. Topologie et géométrie
   - Définitions et différences
   - Branches
2. Un aperçu d'application
   - Complexe de Vietoris-Rips
   - Nombres de Betti
   - Persistance homologique

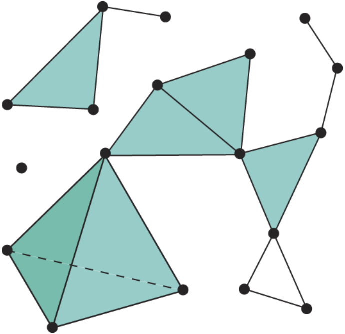
Simplexes et complexes simpliciaux

Un **p-simplexe** est une généralisation en p dimension d'un triangle. Par exemple,

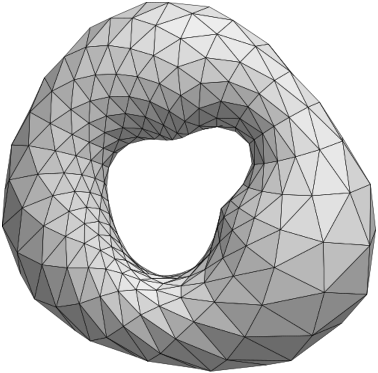
- un 0-simplexe est un point ;
- un 1-simplexe est un segment de droite ;
- un 2-simplexe est un triangle ;
- un 3-simplexe est un tétraèdre, etc.

Alors, un **complexe simplicial** K n'est qu'un ensemble de simplexes collés ensemble de sorte à ce que deux simplexes ne se rencontrent au plus qu'en une seule « face ». On appelle alors la **dimension** de K le plus grand n tel que K possède un n-simplexe.

Visualisation



(a) Un complexe simplicial



(b) Une triangulation d'un tore déformé

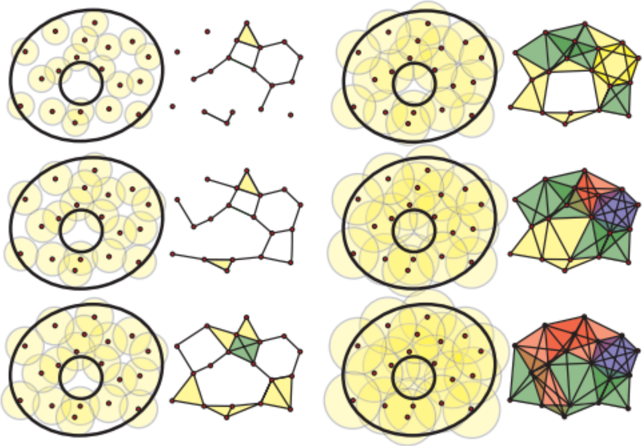
Complexe de Vietoris-Rips

Les complexes simpliciaux sont très utiles en topologie algébrique (surtout computationnelle) puisqu'ils permettent de faire des calculs explicites pour obtenir les invariants algébriques que l'on désire. Ainsi, on aimerait bien en associer un à nos données :

Le **complexe de Vietoris-Rips** K_ε associé à un ensemble de points $E = \{x_1, \dots x_k\}$ dans $\mathbb{R}^N$ est le complexe simplicial abstrait dont les sommets des p-simplexes sont les ensembles de $p + 1$ points $y_0, \dots, y_p$ dans E tels que

$$\mathrm{dist}(y_i, y_j) < \varepsilon \qquad \forall i, j \in \{0, \dots, p\} .$$

Exemple



Mais quel ε capture le mieux la topologie de nos données ?

La matrice bord

Pour trouver le bon ε, il nous serait utile de mesurer la différence entre deux K_ε. Pour ce faire, il nous faudra définir certains invariants algébriques pouvant mesurer cette différence.

Tout d'abord, on introduit une matrice permettant de mesurer comment les p-simplexes sont attachés aux $(p-1)$-simplexes pour un complexe simplicial donné K. Pour ce faire, on numérote les p-simplexes de 1 à n_p, et ce pour tout p. Alors, la **matrice bord** D_p est définie par

$$(D_p)_{ij} := \begin{cases} 1 & \text{si } i\text{-ième } (p-1)\text{-simplexe est une face du } j\text{-ième } p\text{-simplexe} \\ 0 & \text{sinon} \end{cases}$$

Nombres de Betti

Ceci nous permet de définir les **nombres de Betti** de K :

$$\beta_p(K) := c_p - \mathrm{rang}\ D_p - \mathrm{rang}\ D_{p+1}\ .$$

qui ne dépend clairement pas de l'ordre dans lequel on avait numéroté les simplexes. En fait, on peut montrer que c'est un invariant de l'espace topologique sous-jacent.

Pour ceux et celles se demandant où est l'algèbre :
On peut voir D_p comme une transformation linéaire $\mathbb{Z}_2^{c_p} \to \mathbb{Z}_2^{c_{p-1}}$ et alors,

$$\beta_p(K) = \dim \frac{\ker D_p}{\mathrm{im}\ D_{p+1}} = \dim H_p(K; \mathbb{Z}_2)\ ,$$

où $H_p(K; \mathbb{Z}_2)$ est un invariant algébrique de la plus haute importance : le p-ième groupe d'homologie de K avec coefficients dans $\mathbb{Z}_2$.

Interprétation des nombres de Betti

Les nombres de Betti en basse dimension s'interpréte relativement bien :

- $\beta_0(K)$ calcule le nombre de composante connexe de K ;
- $\beta_1(K)$ calcule le nombre de lacets qui ne sont pas continuement déformables un dans l'autre de K.

Par exemple,

$$\beta_0(S^2) = 1 \quad \text{et} \quad \beta_1(S^2) = 0$$

et

$$\beta_0(T^2) = 1 \quad \text{et} \quad \beta_1(T^2) = 2 \ .$$

Évolution des nombres de Betti selon ε

On peut finalement obtenir de l'information topologique pertinente sur nos données de façon algorithmique :

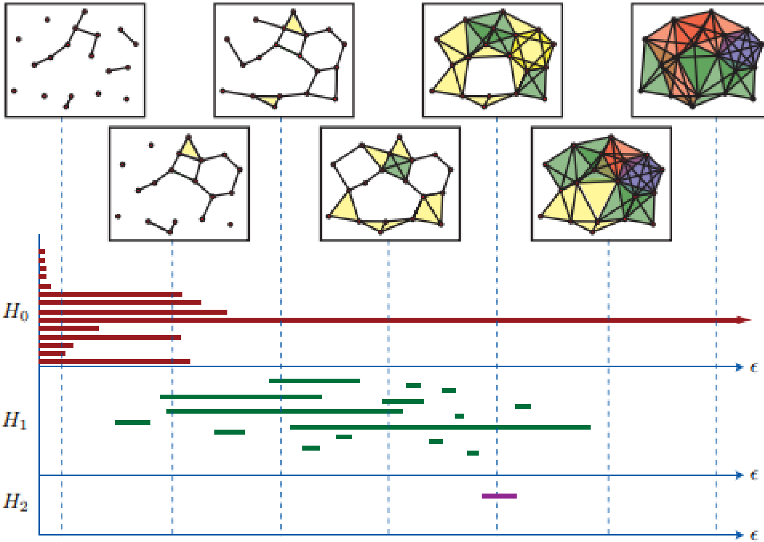
(I) Choisir un ensemble $\{\varepsilon_1, \ldots, \varepsilon_N\}$ représentant bien le spectre des distances entre les points.

(II) Associer à ces distances les complexes de Vietoris-Rips $K_{\varepsilon_1}, \ldots, K_{\varepsilon_N}$.

(III) Calculer les nombres de Betti de chacun de ces complexes.

(IV) Comparer $\beta_p(K_{\varepsilon_i})$ pour tous les i et déterminer lesquels sont persistants.

Retour à notre exemple



Ce qu'on peut faire de plus

Dans le contexte des complexes de Vietoris-Rips :

- Traquer chaque classe d'homologie (mène à l'homologie de persistence et aux code-barres).
- Considérer l'orientation des simplexes (mène aux coefficients dans $\mathbb{Z}$).

Dans le contexte plus théorique :

- Étendre la notion pour différentes variantes de l'homologie.
- Relier la « distance de goulot » des code-barres à la distance C^0 de deux fonctions de Morse ou la distance de Hofer de difféomorphismes hamiltoniens.

References

[1] Ghrist, R. Barcodes : The Persistent Topology of Data. *Bulletin of the American Mathematical Society*, 45(1), (2008), 61-75.

[2] Edelsbrunner, H. et Harer, J. Persistent Homology — a Survey. *Contemporary Mathematics*, 453, (2008), 257-282.

[3] Hatcher, A. *Algebraic Topology.* Cornell Math, 2002.