Beyond Mode Hunting

Joint work with SAMSI Nonlinear WG members:
Bobrowski, Kim, Marron, Noh, and Sommerfeld

Giseon Heo

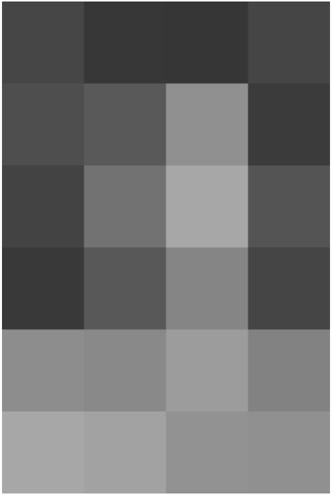
CANSSI-SAMSI Workshop-Fields Institute

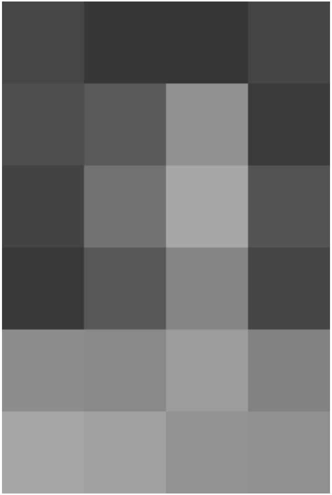
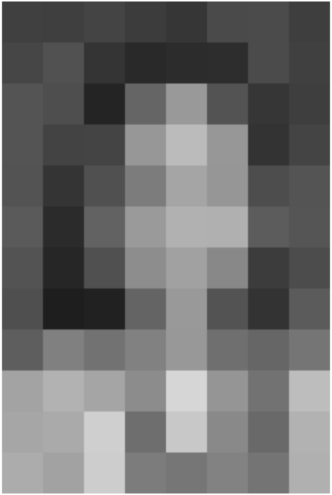
May 23, 2014

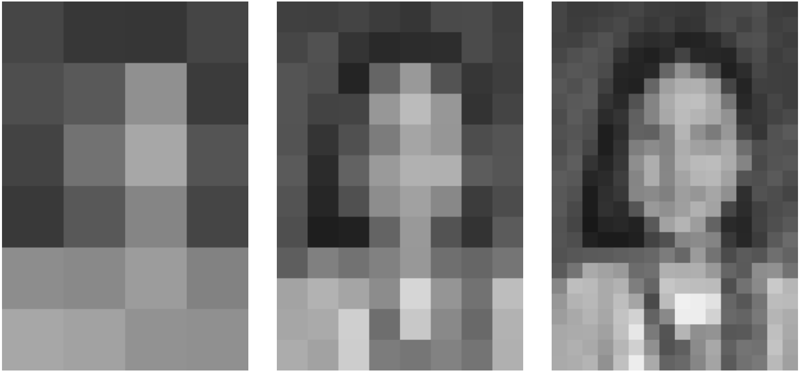
Outline

1. Brief literature review
   - Scale-space theory in computer vision
   - SiZer in statistics
   - Persistent homology in computational topology
2. From 1D to 2D and higher
3. Application to real data: persistence landscape and hypothesis test

Picture at different scale of resolution





Scale-space for signals [Witkin (1983), Koenderink (1984), Lindeberg (1994)]

Given a signal $f : \mathbb{R}^d \to \mathbb{R}$, the scale-space representation $u : \mathbb{R}^d \times \mathbb{R}^+ \to \mathbb{R}$ defined such that the representation at zero scale is equal to the original signal

$$u(x, 0) = f(x),$$

and the representation at coarser scales are given by convolution of the signal with Gaussian kernels of increasing bandwidth

$$u(x, t) = g(x, t) * f(x) = \int_{\mathbb{R}^d} f(y) \frac{e^{-||y-x||^2/2t}}{|2\pi\, t|^{d/2}}\, dy,$$

Gaussian mean $x \in \mathbb{R}^d$, variance matrix $t\mathbf{I}_{d\times d}$.

Connection with heat equation

The scale-space representation can equivalently be defined as the solution to heat equation with initial condition $u(x, 0) = f(x)$.

$$\frac{\partial}{\partial t} u(x, t) = \frac{1}{2} \Delta u(x, t) := \frac{1}{2} \sum_{i=1}^{d} \frac{\partial^2}{\partial x_i^2} u(x, t)$$

Two properties

- Non-enhancement of local extrema

  At a certain scale $t_0 \in \mathbb{R}^+$, a point $x_0 \in \mathbb{R}^\mathrm{d}$ is a local maximum for the mapping $x \mapsto u(x, t_0)$, then $\Delta u(x_0, t_0) < 0$, which means $\frac{\partial}{\partial t} u(x_0, t_0) < 0$.

  A hot spot will not become warmer and a cold spot will not become cooler (true for all dimensions).

- Non-creation of new features (Causality)

  Fine-scale features disappear monotonically with increasing scale. (true for only $d = 1$)

Nonparametric curve estimation in statistics

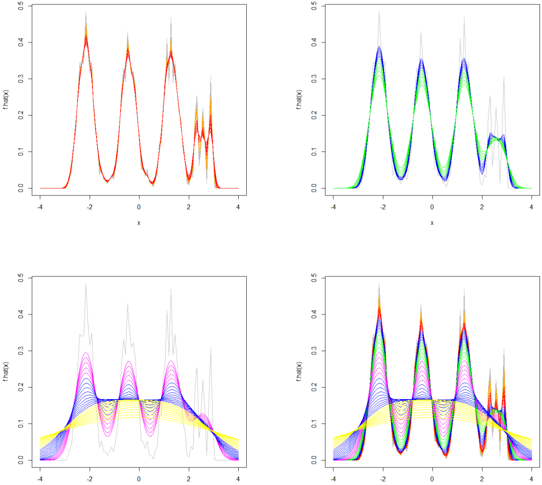
The kernel density estimator based on data $x_1, \ldots, x_n$ is

$$\hat{f}(x,h) = \frac{1}{nh}\sum_{i=1}^{n} K(\frac{x - x_i}{h}),\ x \in \mathbb{R}, h \in \mathbb{R}^+$$

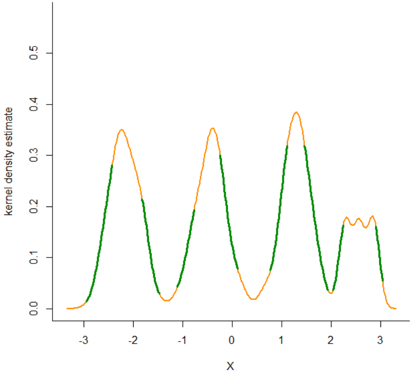
Note: With Gaussian kernel, $K(x) = (1/\sqrt{2\pi})\exp(-x^2/2)$, **?** proved causality.

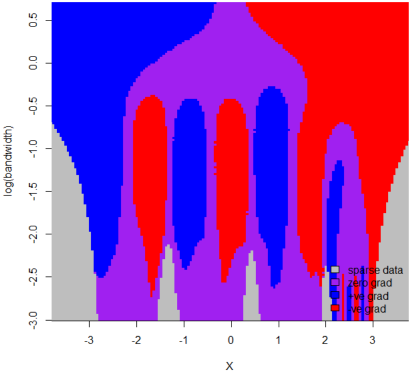
Scale space surface [Chaudhuri and Marron (2000)]

$$\{\hat{f}(x, t) : x \in \mathbb{R},\ t \in \mathbb{R}^+\}$$



SiZer map (SIgnificant ZERo crossings of derivatives) [Chaudhuri and Marron (1999)]





Persistent homology

Sublevel sets $\mathbb{R}_t = \{x \in \mathbb{R} \mid f(x) \leq t\}$, for each $t \in \mathbb{R}$

- As we increase t, the **connectivity** of $\mathbb{R}_t$ remains the same, except when we pass a critical value.
- At a local minimum the sublevel set adds a new component.
- At a local maximum two components merge into one.

Persistence diagram of a Morse function

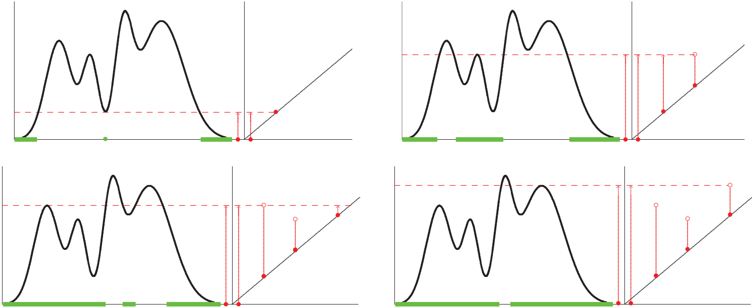
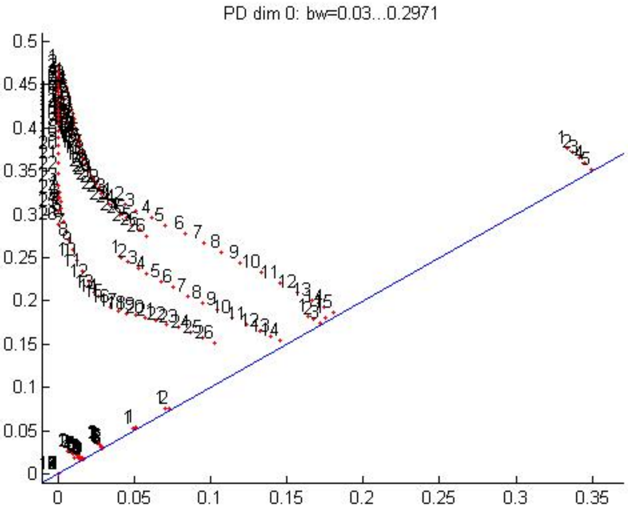


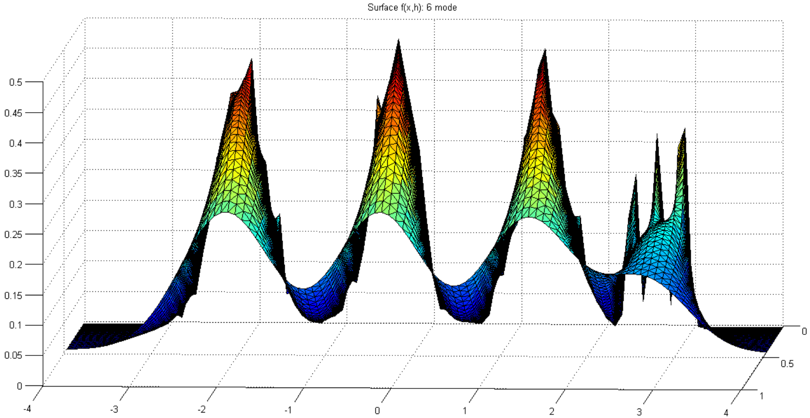
Figure : (Top left) Two components are born at boundary points and a new component is born at a local minimum. (Top right) The components which appeared at local minima have merged at a local maximum. (Bottom) While there are three short-lived components, two components persist ($\beta_0 = 2$). It is therefore likely that the data set is sampled from bimodal distribution.

Persistence diagram of 6 Gaussian mixture as bandwidth increases

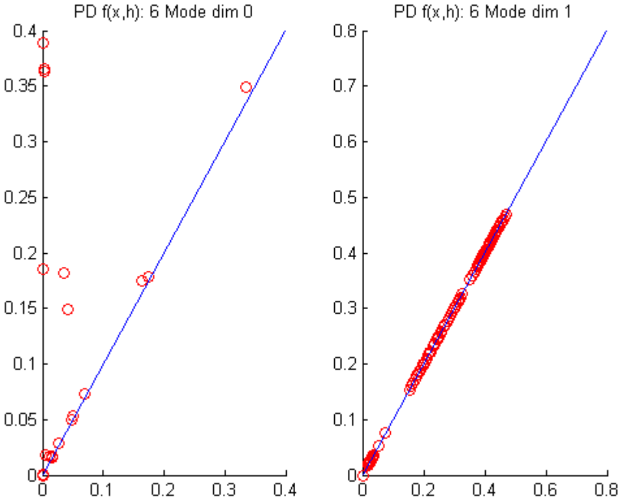


Movie by Max Sommerfeld

Density surface $\hat{f}(x, h)$



Persistence diagram for $\hat{f}(x, h)$



Confidence band

Theorem

Let $X_1, \ldots, X_n$ be i.i.d random variables. The kernel density estimator based on data $x_1, \ldots, x_n$, is $\hat{f}(x,h) = \frac{1}{nh}\sum_{i=1}^{n} K(\frac{x-x_i}{h})$, where $x \in \mathbb{R}$ and $h \in H = (0, \infty)$. Set $f(x,h) = E(\hat{f}(x,h))$

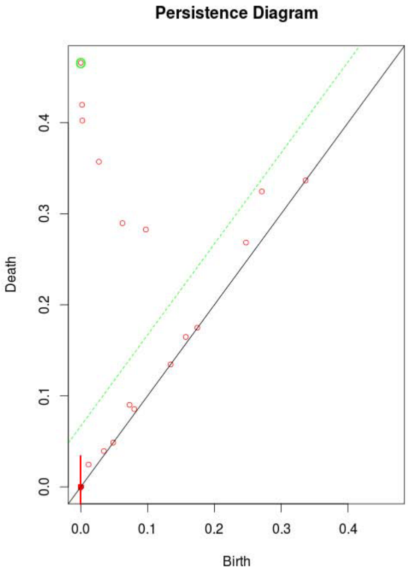
- *(Stability theorem–Cohen-Steiner et al. (2007)) Let $\hat{Dgm}$ and Dgm be corresponding persistence diagram of $\hat{f}$ and f.*

$$d_B(\hat{Dgm}, Dgm) \leq ||\hat{f} - f||_\infty := \operatorname*{ess\,sup}_{\omega} \sup_{x,h} ||\hat{f}(x,h,\omega) - f(x,h)||,$$

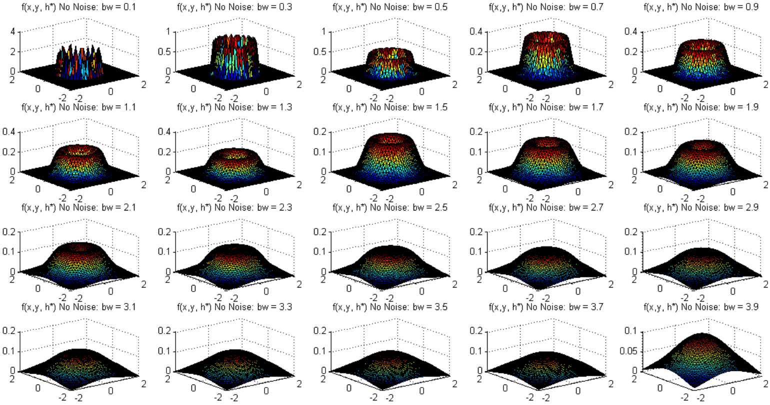
  where d_B is a bottleneck distance between persistence diagrams.

- *(Fasy et al. (2013))*
  $\mathbb{P}(d_B(\hat{Dgm}, Dgm) > c_n) \leq \mathbb{P}(||\hat{f} - f||_\infty > c_n) = \alpha$

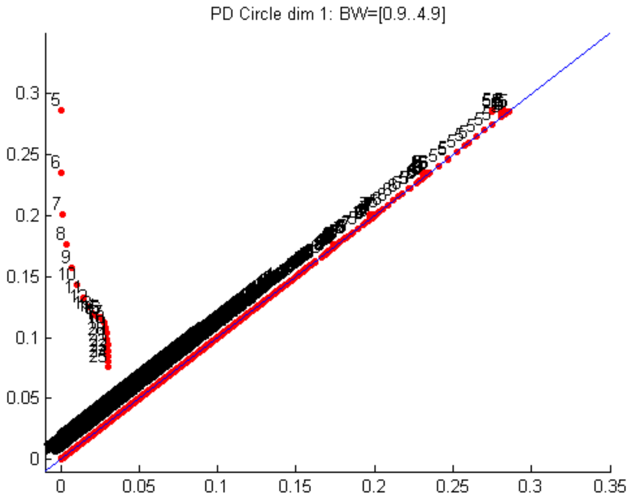
Significance $\hat{f}(x, h)$ by Sommerfeld



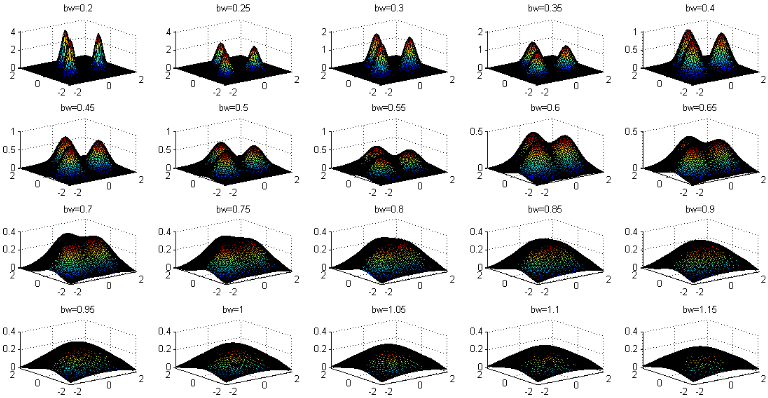
Gaussian kernel density on a circle



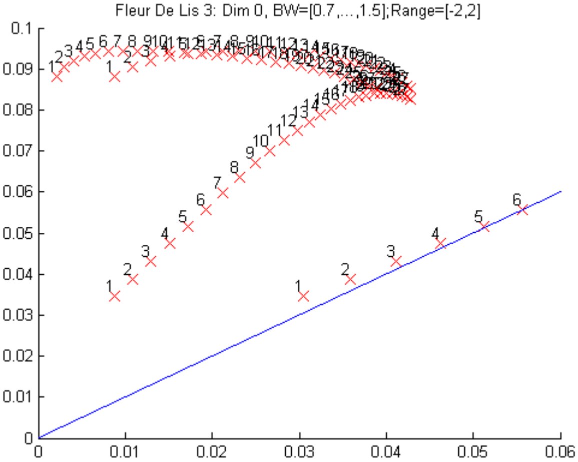
Persistence diagram



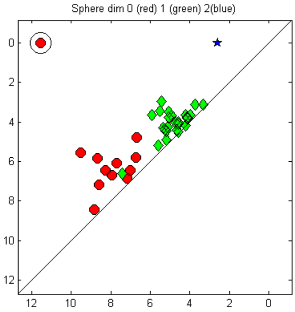
Density surface of Fleur De Lis

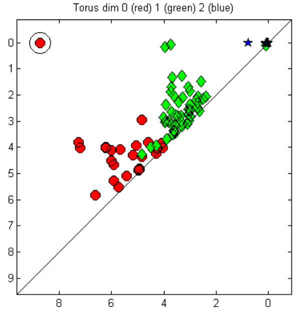


Persistence Diagram of Fleur De Lis



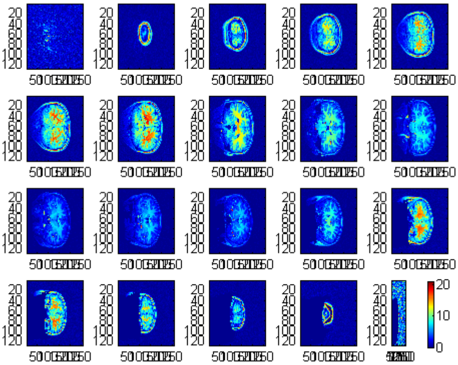
Persistence diagrams of sphere and torus: dim 0, 1 and 2 (Bobrowski)



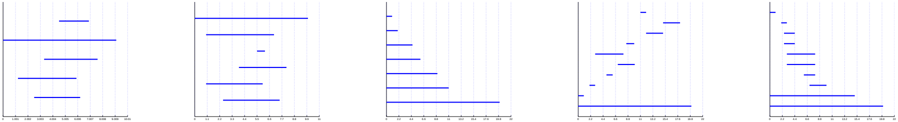


Statistical inference motivated by Brain data: "Normal" versus ADHD

- Intensities represent different tissue material, say grey matter, white matter and CSF.
- "Expect" some intensity different for some ROIs between two groups.



Recent development in persistent homology: Statistics with descriptors



- How do we calculate the mean and variance?
- Can we apply it to hypothesis testing?

New descriptor [Bubenik (2012)] Statistical topology using persistence landscapes

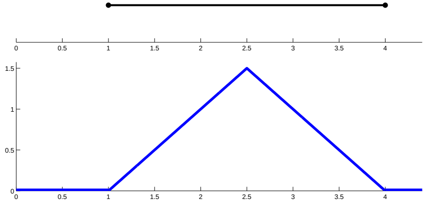


Figure : For (a, b), define $f_{(a,b)} : \mathbb{R} \to \mathbb{R}$ by $f_{(a,b)}(t) = \min(t - a, b - t)_+$

Figure : For (a, b), define $f_{(a,b)} : \mathbb{R} \to \mathbb{R}$ by $f_{(a,b)}(t) = \min(t - a, b - t)_+$

Figures prepared by Violeta Kovacev-Nikolic

Persistence Landscape

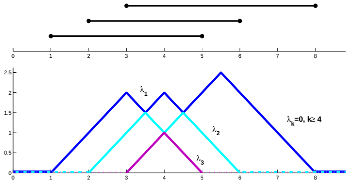


Figure : For $\{(a_i, b_i)\}_{i=1}^m$, $\left\{\lambda(k,t) = k^{th} \textit{ largest value of } \{f_{(a_i,b_i)}(t)\}_{i=1}^m\right\}$

Figure : For $\{(a_i, b_i)\}_{i=1}^m$, $\left\{\lambda(k,t) = k^{th} \textit{ largest value of } \{f_{(a_i,b_i)}(t)\}_{i=1}^m\right\}$

- The persistence landscapes are functions from $\mathbb{N} \times \mathbb{R} \to \mathbb{R}$, and are bounded and nonzero on a bounded domain.
- Hence, persistence landscapes belong to $\mathrm{L}^{\mathrm{p}}(\mathbb{N} \times \mathbb{R})$ with the metric induced by p-integrable functions, which is a separable Banach space.
- In separable Banach space, for any continuous linear function f, the random variable $f(\lambda(k, t))$ satisfies SLLN and CLT.

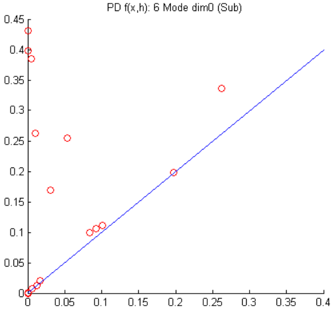
Tests

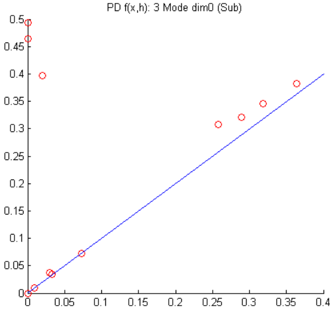
- t-test: $\sum_{k=1}^{K} \left[\int |\lambda_k^A(t) - \lambda_k^B(t))|^p \, dt \right]^{1/p}$
- Multivariate test (Hotellings T^2 test):

  Consider a vector, $(\int |\lambda_1^A - \lambda_1^B|, \int |\lambda_2^A - \lambda_2^B|, \ldots, \int |\lambda_k^A - \lambda_k^B|)$,

  where k is chosen so that, $k << n_1 + n_2 - 2$.

6 vs. 3 Modes Gaussian mixture: p-value= 0.0015.





References

Bubenik, P. (2012), "Statistical topology using persistence landscapes. arXiv:1207.6437," .

Chaudhuri, P. and Marron, J. (1999), "SiZer for exploration of structures in curves," *Journal of the American Statistical Association*, 94, 807–823.

— (2000), "Scale space view of curve estimation," *Annals of Statistics*, 28, 408–428.

Cohen-Steiner, D., Edelsbrunner, H., and Harer, J. (2007), "Stability of persistence diagrams," *Discrete and Computational Geometry*, 37, 103–120.

Fasy, B., Lecci, F., Rinaldo, A., Wasserman, L., Balakrishnan, S., and Singh, A. (2013), "Statistical inference for persistent homology: confidence sets for persistence diagrams. arXiv:1303.7117v2," .

Koenderink, J. J. (1984), "The structure of images," *Biological Cybernetics*, 50, 363–370.

Lindeberg, T. (1994), *Scale-Space Theory in Computer Vision*, The Netherlands: Kluwer Academic Publisher.

Witkin, A. P. (1983), "Scale-space filtering," in *Proceedings of the 8th International Joint Conference on Artificial Intelligence*, pp. 1019–1022.

Thanks to

- SAMSI LDHD orginizers: **Peter Kim** and **Ezra Miller**
- SAMSI TDA wg member: **Paul Bendich**
- SAMSI Graphical Models wg member: **Helene Massam**
- SAMSI Hilbert Manifolds wg member: **Victor Patrangenaru**
- **NSERC** and **McIntyre Memorial Funding** in the School of Dentistry

UNIVERSITY OF ALBERTA