Topological Aspects of Optimal Fusion of Multimodal Sensor Data

John Baillieul
Intelligent Mechatronics Lab
Boston University
johnb@bu.edu

ACC Workshop on Tony Bloch
Motivation

Completely Integrable Hamiltonian Systems and Total Least Squares Estimation©
by Anthony Michael Bloch


The thesis applied symplectic geometry and especially Morse-Bott theory to problems in statistical estimation theory.
Climbing Information Gradients — how to use multimodal sensors to best advantage

Talk Outline

1. How information theory and differential topology might be connected
2. A brief introduction to topological persistence
3. Some aspects of robotic information gathering
4. How to understand the tradeoff between speed and accuracy in the acquisition of knowledge
Talk Outline:

• Introduction - why did we begin to think about topology and information?
• The information theory of functions
• Motion primitives for robotic reconnaissance
• Reconnaissance as information acquisition
• The topology of unknown fields
• Data induced partitions and topology induced partitions
• Topology guided information acquisition
Talk Themes:

- Information
- Topology
- Control


What is information?
What is information?

The number of “yes-no” questions that must be answered to answer a given question.
What is information?

Typical (simple) question decomposition:

Where is Shenyang?
What is information?

Where is Shenyang?
What is information?

Where is Shenyang?

Is he here?
What is information?

Where is Shenyang?

Is he here?
What is information?

Where is Shenyang?
What is information? Shannon’s answer (three properties)

Suppose an experiment can have any of $n$ outcomes, and the probabilities of the outcomes are $p_1, p_2, \ldots, p_n$. If there is a measure of the amount of information contained in observing an outcome, say $H(p_1, p_2, \ldots, p_n)$, it is reasonable to assume that $H$ satisfies the following.

1. $H$ should be continuous in the $p_i$.
2. If all the $p_i$ are equal, $p_i = \frac{1}{n}$, then $H$ should be a monotonic increasing function of $n$. With equally likely events there is more choice, or uncertainty, when there are more possible events.
3. If a choice be broken down into two successive choices, the original $H$ should be the weighted sum of the individual values of $H$.

\[
H(p_1, p_2, \ldots, p_n) = - \sum_{i=1}^{n} p_i \log_2 p_i
\]
What is information?

Shannon’s third property:

\[ H(\frac{1}{3}, \frac{2}{3}) + \frac{2}{3} H(\frac{1}{2}, \frac{1}{2}) = \log_2 3 \]
Motivation for connecting information theory and topology

- The desire to understand decision making in tasks involved in search, surveillance, reconnaissance.
- The desire to understand the cognitive psychology of spatial knowledge.
- The desire to understand what humans and animals perceive to be of interest in visual images and other spatially varying data fields.
- The desire to understand compression and sensor fusion for continuous data.
Animals are attracted to feature-rich backgrounds - the case of bats
Animals are attracted to feature-rich backgrounds - the case of Manduca sexta

Tom Daniels Lab, U. Washington
Functions and mappings as information channels

Suppose $X$ is a random variable taking on values $X_1, \ldots, X_N$ with probabilities $p_1, \ldots, p_N$.

The *entropy*

$$H(X) = - \sum_{k=1}^{N} p_k \log_2 p_k$$

measures the amount of information needed to for knowledge of $X$. 
Functions and mappings as information channels

A function \( f : X \rightarrow Y \) is a communication channel that provides information in the range \( Y \) about the structure of the domain \( X \).

The function \( f \) is informative about \( X \) if the mutual information

\[
I(X ; f(X)) = H(f(X))
\]

is large.

Theorem. \( H(f(X)) \leq H(X) \).
Functions and mappings as information channels

**Theorem.** The function that maximizes information preservation, $H(f(X))$, minimizes the conditional entropy $H(X | f(X))$.

**Proof.**

\[
I(X, f(X)) = H(X) - H(X | f(X)) = H(f(X)).
\]
Functions and mappings as information channels - diversity and noise

Among all functions \( f : X \rightarrow \{0,1\} \)

\[
H(p) = -p \log_2 p - (1 - p) \log_2 (1 - p)
\]

maximizes \( H(f(X)) \).

\( \ p_1 < p_2 < p_3 \)
Mathematical Quantification of “Being Interesting”

Things are interesting if they are not predictable.

A function $f : X \rightarrow Y$ maximally preserves information if its image reflects the diversity of $X$.

$$C(f, \mathcal{V}) = - \sum_{j=1}^{m} \frac{\mu(f^{-1}(V_j))}{\mu(X)} \log_2 \frac{\mu(f^{-1}(V_j))}{\mu(X)}$$
The information contained in a random field

We are interested in more than merely whether the field is spatially varying.
A critical point of a scalar field $f : X \to \mathbb{R}$ is a point $x$ such that $d_x f = 0$.

If $x$ is a critical point, $f(x)$ is the corresponding critical value.

For each $t \in \mathbb{R}$, the corresponding sublevel set is

$$\mathbb{R}_t = f^{-1}(-\infty, t].$$
The topology-induced partition
Let $X \subseteq \mathbb{R}^m$ be a compact, connected, simply connected domain, and let $[a, b]$ be the image of $X$ under $f : \mathbb{R}^m \to \mathbb{R}$. Consider the partition

$$a = x_0 < x_1 < \cdots < x_m = b.$$ 

The corresponding partition $\mathcal{V}_n = \bigcup_{j=1}^m \{ \text{cc} (f^{-1} ([x_{j-1}, x_j])) \}$ of $X$ is called the data-induced partition.

The data-induced partition will be deemed to be interesting or not according to the metric

$$H_k = - \sum_{j=0}^{n} A(V_j^k) \log_2 A(V_j^k).$$
Level crossings, excursion sets, and the height map of random fields
Critical Level Sets Are Essential Objects in Potential Field Reconnaissance

- Index zero critical point
- Index one critical point
- Index two critical point
The topology-induced partition

**Definition:** A *critical level set* of $f$ on a compact domain is a connected component of values

$$
\xi(c^*) = \text{cc}(\{ r \in X | f(r) = c^* \}),
$$

$f(r^*) = c^*$, with $r^*$ being a critical point.

**Notation:** The set of all *critical level sets* is denoted $\text{Cr}(f,X)$

**Definition:** The *topology induced partition* is the domain partition

$$
\mathcal{M}(f,X) = \text{cc}(X \setminus \text{Cr}(f,X))
$$
Partition Entropy

To each partition, we associate a measure

$$H(\alpha) = - \sum_{A_i \in \alpha} \mu(A_i) \log_2 \mu(A_i)$$

called the partition entropy.

The conditional entropy of $\alpha$ conditioned on $\beta$ is

$$H(\alpha|\beta) = \sum_{B_j \in \beta} \mu(B_j) H(\alpha|B_j)$$

$$= - \sum_{B_j \in \beta} \mu(B_j) \sum_{A_i \in \alpha} \frac{\mu(A_i \cap B_j)}{\mu(B_j)} \log_2 \frac{\mu(A_i \cap B_j)}{\mu(B_j)}$$

$$= - \sum_{B_j \in \beta} \sum_{A_i \in \alpha} \mu(A_i \cap B_j) \log_2 \frac{\mu(A_i \cap B_j)}{\mu(B_j)}.$$
Let $\alpha, \beta, \gamma$ be partitions of a domain $X$.

(i) $0 \leq H(\alpha|\beta) \leq H(\alpha)$ with $H(\alpha|\beta) = 0 \iff \beta$ is a refinement of $\alpha$.

(ii) $\beta$ a refinement $\alpha \Rightarrow H(\alpha|\gamma) \leq H(\beta|\gamma)$ "<"
if $\beta$ a proper refinement provided $\gamma$ not a refinement $\beta$.

(iii) $\gamma$ a refinement of $\beta \Rightarrow H(\alpha|\beta) \geq H(\alpha|\gamma)$.

(iv) $\beta \equiv X \Rightarrow H(\alpha|\beta) = H(\alpha)$. 
How is the data-induced partition related to the topology-induced partition?

How is the information content of A different from that of B?
Random fields as information channels - diversity and noise

Some peaks are more significant than others.

How can we quantify feature *significance*?

How do we respect diversity and reject noise?

1. Height.
2. *Topological persistence*.
3. *Information utility*. 
Random fields as information channels - diversity and noise

As the dimension of $X$ increases, the topology of the sublevel sets

$$R_t = f^{-1}(-\infty, t]$$
ecomes more complex.
Random fields as information channels - diversity and noise

How can we quantify feature significance?

1. Height of critical values.
2. Topological persistence.
3. Information utility.
Random fields as information channels - diversity and noise

How can we quantify feature significance?

Topological persistence is defined in terms of the topology of sublevel sets \( R_t = f^{-1}(-\infty, t] \).
Let regular values $x_j$ bracket the $m$ critical values of $f: X \to \mathbb{R}$.

Case dim $X=1$, persistence tracks

$$\beta_0(\mathbb{R}_{x_j}) = \text{rank } \mathcal{H}_0.$$  

Case dim $X>1$, persistence tracks

$$\beta_p(\mathbb{R}_{x_j}) = \text{rank } \mathcal{H}_p.$$
Morse-Smale fields as information channels - diversity and noise

Topological persistence looks at the topology of sublevel sets.
A minimum is paired with and destroyed by a “negative saddle”:
A minimum can also be paired with a “mixed saddle”:

\[ \beta_0 \quad - \quad - \quad \beta_1 \quad + \quad + \]
A minimum can also be paired with a “mixed saddle”:

\[ \beta_0 \quad \cdot \cdot \cdot \]

\[ \beta_1 \quad +\quad + \]
When $t^*$ is a “positive saddle”, $f^{-1}[-\infty, t]$ topology changes: $\beta_1 \leftrightarrow +$
A positive saddle is paired with and eventually destroyed by a local max: $\beta_1$ --
Let $X \subset \mathbb{R}^m$ be a compact, connected, simply connected domain, and let $[a, b]$ be the image of $X$ under $f : \mathbb{R}^m \rightarrow \mathbb{R}$. Consider the partition

$$a = x_0 < x_1 < \cdots < x_m = b.$$ 

The corresponding partition $\mathcal{V}_n = \bigcup_{j=1}^{m} \{ \text{cc} \left( f^{-1} ([x_{j-1}, x_j]) \right) \}$ of $X$ is called the \textit{data-induced partition}.

The \textit{data-induced partition} will be deemed to be interesting or not according to the metric

$$H_k = - \sum_{j=0}^{n} A(V_j^k) \log_2 A(V_j^k).$$
Information Acquisition as a Search Metric

\[ H_k = - \sum_{j=0}^{n} A(V_j^k) \log_2 A(V_j^k) \]

If \( H_{k+1} - H_k > \eta \), continue exploiting.
Motion Primitives for Reconnaissance of Random Scalar Fields

\[ b^{iso}(r_0) = \text{level set contour passing through } r_0 \]

\[ b^{grad}(r_0) = \text{gradient contour passing through } r_0 \]

\[ B_k = \{b_1, \ldots, b_k\} = \text{motion program sequence} \]

sequence where \( b_i \in \{b^{iso}, b^{grad}\} \)

\[ S(B_k) = \{\xi_1, \ldots, \xi_k\} = \text{contours corresponding to } B_k \]
The data-induced partition of random scalar fields

\[ B_k = \{ b_1, \ldots, b_k \} = \text{motion program sequence} \]

sequence where \( b_i \in \{ b^{iso}, b^{grad} \} \)

\[ S(B_k) = \{ \xi_1, \ldots, \xi_k \} = \text{contours corresponding to } B_k \]

The data-induced partition is a proxy for the topology-induced partition:

\[ \mathcal{V}_n := \mathcal{V}(S(B_n)) = cc(X \setminus S(B_n)) \]
Reconnaissance of Potential Fields Defined on 2-dimensional Domains

• Map level sets

• Map steepest ascent/descent curves
Robotic Search of an Unknown Magnetic Field
The Conditional Entropy of the TIP Given the Data Induced Partition

\[ H(\mathcal{M}|\mathcal{V}_n) := - \sum_{M_i \in \mathcal{M}} \sum_{V_n^j \in \mathcal{V}_n} \frac{\mu(M_i \cap V_n^j)}{\mu(X)} \log_2 \frac{\mu(M_i \cap V_n^j)}{\mu(V_n^j)} \]

**Theorem:** Given the topology induced partition \( \mathcal{M} \) and the data induced partition \( \mathcal{V}_n \),

\[ 0 \leq H(\mathcal{M}|\mathcal{V}_n) \leq H(\mathcal{M}). \]

\[ H(\mathcal{M}|\mathcal{V}_n) = 0 \text{ if and only if } \mathcal{V}_n \text{ is a refinement of } \mathcal{M}. \]
How the Topology Induced Partition is Related to Topological Entropy

**Theorem:** (Baronov, 2010) Let $\mathcal{V}_N$ be the domain partition of $f : X \to [0, 1]$ corresponding a uniform partition of the range into subintervals of length $1/N$. Let $\mathcal{M}$ be the topology induced partition of the same function. Define:

$$
\delta_j = \frac{\sup_{R \in M_j} f(R) - \inf_{R \in M_j} f(R)}{\sup_{R \in X} f(R) - \inf_{R \in X} f(R)}.
$$

Then

$$
\lim_{N \to \infty} (H(f, \mathcal{V}_N) - \log_2 N) \leq H(\mathcal{M}) + \sum_{i=1}^{n} \frac{\mu(M_i)}{\mu(X)} \log_2 \delta_i.
$$
Sensing depth with event cameras
Sensing depth
Sensing depth
Sensing depth
Sensing depth
Sensing depth
Sensing depth
Sensing depth
Sensing depth through motion

\[ \tau = \frac{r(t)}{\dot{r}(t)} \text{ is a proxy for distance} \]
Sensing depth—with event camera
Information and the topology of unknown fields

• One interpretation of Baronov’s theorem is that the critical sets of an unknown field encode the essential information that can be obtained through exploration and mapping.

• This raises the question, will humans engaged in reconnaissance focus on discovering these topological characteristics?

• Another question is whether it is possible to design reconnaissance strategies aimed at discovering the topological characteristic of an unknown field.
Related Work — Information Based Image Segmentation

1. Reconnaissance of time-varying fields
2. Concepts of sensor fusion

Given multiple sensor fields $f_1, \ldots, f_N$ on a common domain $X$, partition $X$ into subdomains $Y_1, \ldots, Y_N$ such that on each subdomain $Y_j$ the $j$-th sensor field $f_j$ is maximally informative.

Sensor-fusion: Form composite

$$f(x) = f_j(x) \quad \text{if } x \in Y_j$$
Enhanced perception from sensor fusion

Visual spectrum

Infrared spectrum
Enhanced perception from sensor fusion
Conclusion

- Topological methods married to information-theoretic concepts are useful in designing control algorithms for rapid robotic reconnaissance.
- The same methods appear to be useful in data compression of multiband images.
- These methods also provide a baseline for studying human performance in directing reconnaissance and in studying visual cognitive styles.
- Current research is aimed at understanding how to extend this circle of ideas to time-varying fields.
Thanks Tony

For making our lives richer over the past 60 years.
Thanks Tony

For making our lives richer over the past 60 years—and thanks in advance for the next 60.