Thrust #3

Active Information Exploitation for Resource Management

Information-geometric Ideas in Sensor Management and Information Flow Quantification

Doug Cochran

Synopsis of Research Activities

- Information-geometric trajectory planning
  - Update on work from last year
  - Introduction to a related collaborative topic started this year
- Quantification of information flow
- Value of information sharing

Information-geometric Trajectory Planning*
Sensor Model

- In estimation of a parameter θ in a smooth parameter manifold M, the effect of a particular sensing action is to produce a log-likelihood $l(\theta)=\log p(x|\theta)$
- The Fisher information

$$F_\theta = \mathsf{E}_{p(\cdot|\theta)}\left[d_\theta l \otimes d_\theta l\right] = -\mathsf{E}_{p(\cdot|\theta)}\left[\nabla_\theta^2 l\right]$$

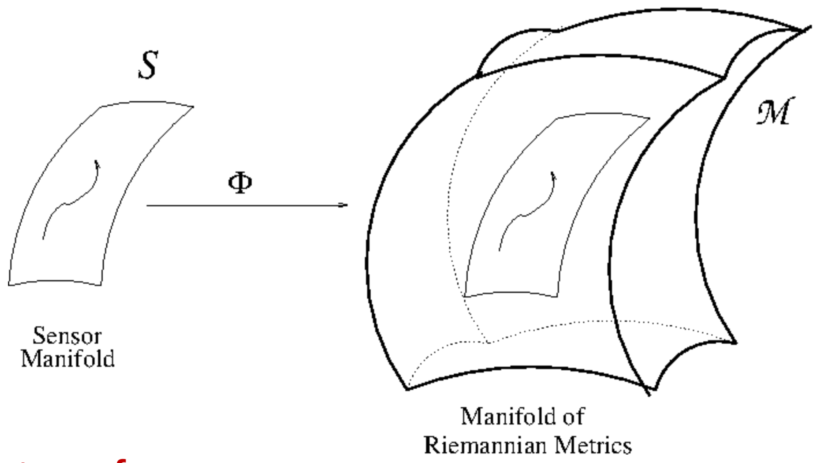
induces a Riemannian metric on M

> *The effect of choosing a sensing action / sensor configuration is to select a Riemannian metric for the parameter manifold* M

*Joint work with S. Howard and W. Moran

Information-geometric Trajectory Planning
Sensor Manifold

- The collection of all Riemannian metrics on a smooth manifold M is itself an infinite-dimensional (weak) Riemannian manifold M(M)
  - The geodesic structure of M has been studied outside the context of information geometry (Gil-Medrano & Michor, 1991)
- If the sensor configuration is parameterized by a smooth "sensor manifold" S, a particular configuration σ in S gives rise to a particular Riemannian metric g_σ on M
  - The mapping Φ taking σ to g_σ will be assumed to be smooth and one-to-one (immersion is enough)
  - Although **M** is infinite-dimensional, the trajectory planning takes place in a finite-dimensional sub-manifold, and…
  - This sub-manifold inherits a metric structure from M



Information-geometric Trajectory Planning
Geodesics

- Geodesics g in M minimize

$$E_g = \frac{1}{2}\int_0^1 \int_M \mathrm{Tr}(g^{-1}(\partial_t g) g^{-1}(\partial_t g))\, dF(\theta)\, dt$$

  - A variational approach was used by Gil-Medrano and Michor to obtain a differential equation for these curves
  - Classical differential geometric quantities (e.g., Christoffel symbols) appear in these formulae

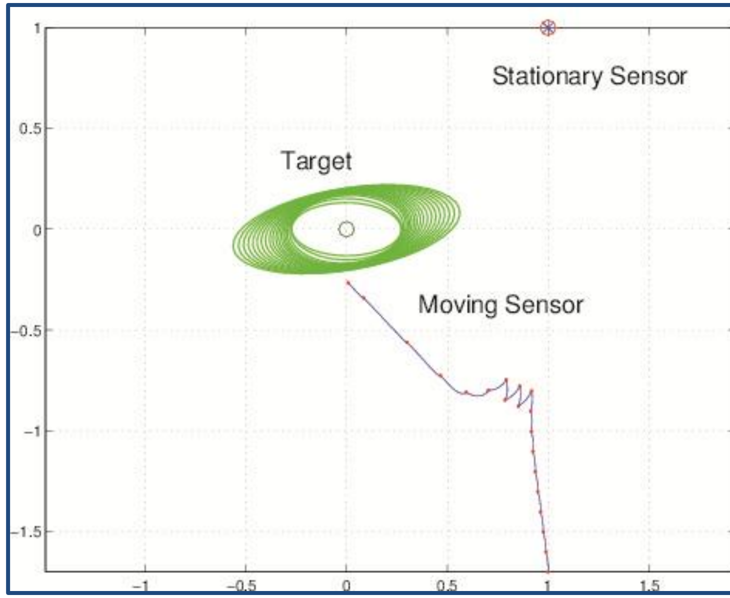
Last year: Showed divergences on M derived from KL (Fisher) and from symmetrized mutual information (Shannon) both give rise to the same Riemannian metric on M; i.e., the Gil Medrano-Michor metric

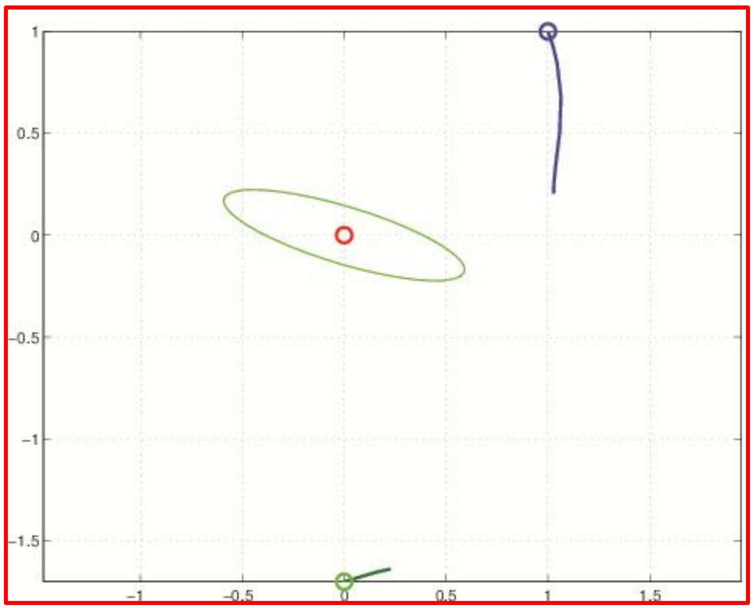
This year: Showed how to constrain the geodesic derivation to the sub-manifold $\Phi(S)$ to obtain geodesic sensor configuration trajectories in S

Information-geometric Trajectory Planning

Bearings-only emitter localization example

- One fixed and one mobile sensor platform (e.g., UAV) with bearings-only sensors seek to localize a stationary emitter
- Noise is independent von Mises on each sensor
- Target prior is Gaussian; covariance is updated with each measurement along the moving sensor trajectory





- Two mobile sensor platforms with different initial positions and constrained speeds
- Gaussian target prior
- Geometrically determined platform motion reduces error covariance determinant measurements are collected

VOI

Navigation in the Multinomial Family*

- Consider the family F of multinomial distributions on K outcomes and N=1 trial
  - $p \in$ F is characterized by a parameter vector $q=(q_1,\ldots, q_K)$ with each $q_k \geq 0$ and $\Sigma\, q_k = 1$; i.e., $q \in \Delta$, the standard simplicial face in $\mathbf{R}^N$
  - Equivalently, F is parameterized by

$$\tilde{q} = (\sqrt{q_1}, \ldots, \sqrt{q_K}) \in S$$

    where S is the part of the unit sphere in the non-negative orthant
  - These parameters are still suitable for N>1 trial, so information analysis does not change
- Consider a curve $q(t)$, $t \in [0,1]$ in Δ; the Fisher information for estimating $q(t)$ is

$$I(t) = \sum_{k=1}^{K} \left(\frac{1}{q_k(t)} \frac{d}{dt} q_k(t) \right)^2 q_k(t) = \sum_{k=1}^{K} \left(\frac{d}{dt} q_k(t) \right)^2 \left(\frac{1}{q_k(t)} \right)$$

*Joint work with A. Hero and B. Sadler

Navigation in the Multinomial Family

Equivalence of Hellinger and Information Distances

- On the other hand, the squared differential path length element in S is

$$\left|\frac{d}{dt}\tilde{q}(t)\right|^2 = \sum_{k=1}^{K}\left(\frac{d}{dt}\sqrt{q_k(t)}\right)^2 = \frac{1}{4}\sum_{k=1}^{K}\frac{1}{q_k(t)}\left(\left|\frac{d}{dt}q_k(t)\right|\right)^2 = \frac{1}{4}I(t)$$

- So the *information distance* between the density at q(0) and the one at q(t) is the shortest distance between points on the unit sphere; i.e., a great circle distance

$$d_I(p_{q(0)}, p_{q(t)}) = \arccos\left\langle \tilde{q}(0), \tilde{q}(t)\right\rangle$$

- The *Hellinger distance* (an f-divergence) is given by

$$d_H(p_{q(0)}, p_{q(t)}) = 2 - 2\left\langle \tilde{q}(0), \tilde{q}(t)\right\rangle$$

In this family, Hellinger distance (global) is monotonically related to information distance, which is differentially navigable by Fisher information

Quantification of Information Flow*
De Bruijn's Identity on **R**^{n}

- Given a probability density p on **R**^{n}, its *intrinsic Fisher information is*

$$F = -\int_{R^n} (\nabla^2 \log p) p \, dx = \int_{R^n} (d \log p \otimes d \log p) p \, dx$$

and its entropy is

$$H = \int_{R^n} p \log p \, dx$$

- If p evolves according to

$$\frac{\partial p}{\partial t} - \frac{1}{2} \nabla^2 p = 0$$

then

$$\frac{\partial H}{\partial t} = \frac{1}{2} \mathrm{Tr}(F)$$

*Joint work with S. Howard and W. Moran

Quantification of Information Flow
Intrinsic Fisher Information on $\mathbf{R}^n$

- Suppose we wish to estimate the location parameter $\theta \epsilon \mathbf{R}^n$ from a datum having pdf $p(x\text{-}\theta)$

- The Fisher information at $\theta = 0$ is

$$F = -\int_{R^n} \nabla_\theta^2 \log p(x-\theta)\, p(x-\theta)\, dx = -\int_{R^n} \nabla_\theta^2 \log p(x)\, p(x)\, dx$$

i.e., the intrinsic Fisher information of p describes how well p can be distinguished from a small translation of itself

I.V. Arnold (1966) pointed out the Euler equations for an inviscid incompressible fluid on a Riemannian manifold M arise from considering geodesic flows on $\mathrm{SDiff}(M)$, the Lie group of volume preserving diffeomorphisms on M. Does this view carry over to information dynamics?

Quantification of Information Flow
K-L Divergence on a Riemannian Manifold

- *Data space:* Orientable Riemannian n-manifold X with metric g and invariant volume form μ
- *Density:* $p:X \rightarrow \mathbf{R}^+$ with $\int p\mu = 1$
- Consider the parametric family of densities

$$\{p(\cdot\,|\,\phi) = p \circ \phi \,|\, \phi \in \mathrm{SDiff}\,(X)\}$$

- How well $p\circ\phi_1$ can be distinguished from $p\circ\phi_2$ given data $x \in X$ is described by the K-L divergence

$$D : \mathrm{Sdiff}(X) \times \mathrm{Sdiff}(X) \rightarrow \mathbf{R}$$

with

$$D(p\circ\phi_1 \| p\circ\phi_2) = \int_X \log\left(\frac{p\circ\phi_1}{p\circ\phi_2}\right) p\circ\phi_1\,\mu$$

Quantification of Information Flow

Intrinsic Fisher Information on a Riemannian Manifold

- Differentiating D, one obtains an expression for Fisher information at $\phi \in \mathrm{Sdiff}(X)$ and correspondingly an intrinsic Fisher information for p

$$F_{\mathrm{id}} = \int_X \nabla^2 \log p \; p\mu$$

and a scalar intrinsic Fisher information

$$f = \int_X \mathrm{Tr}(\nabla^2 \log p) \; p\mu = \int_X d \log p \wedge * \, dp$$

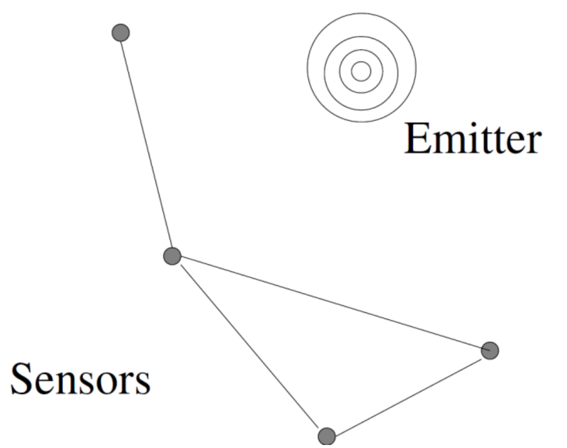
- Suppose $p(t,x)$ satisfies $\partial_t p - \Delta p = 0$ and define the entropy density by $h_t = p_t \log p_t \, \mu$. Then there is a de Bruijn identity on X for entropy H_t

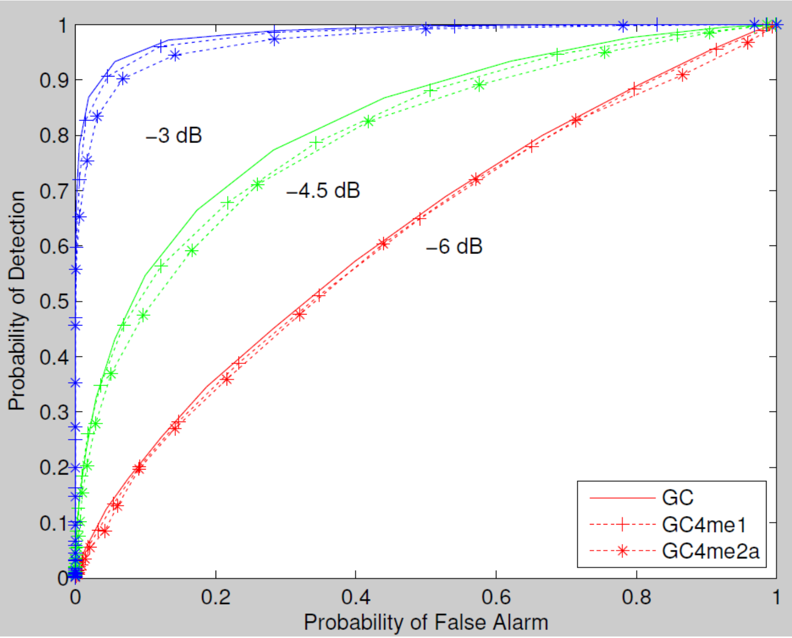
$$\partial_t H_t = \int_X \partial_t h_t = f_t$$

Value of Information Sharing in Networks

Synopsis / Update

- Last year, results were reported quantifying the value of individual edges in a network graph for achieving global registration in the presence of noise
  - The measures take the form of a “local Fisher information”
- This year, the value of information sharing was investigated in the context of detecting an uncharacterized emitter using a network of receivers





Research Directions

- Develop further information-geometric characterizations of VoI and apply them in sensor management and adaptive information collection

- Consider how understanding of intrinsic Fisher information and information diffusion can be used in modeling the evolution of uncertainty in unobserved systems

- Establish a theoretical framework that supports experimental results that indicate the value of network connectivity (e.g., actual data sharing vs. maximum entropy proxies for shared data)