

# **Mathematical Methods for Medical Image Computation**

**Allen Tannenbaum**

**Computer Science and Applied Mathematics/Statistics**

**Stony Brook University**

**[allen.tannenbaum@stonybrook.edu](mailto:allen.tannenbaum@stonybrook.edu)**

**<http://www.na-mic.org/Wiki/index.php/Algorithm:BU>**

# Medical Applications: IGT and IGS

- Mathematics to develop general-purpose algorithms and software that can be integrated into complete therapy/surgical delivery systems.
- Four main components of image-guided therapy (IGT): localization, targeting, monitoring and control.
- Develop robust algorithms for:
  - Segmentation - automated methods that create patient-specific models of relevant anatomy from multi-modal data.
  - Registration – automated methods that align multiple data sets with each other and with the patient.

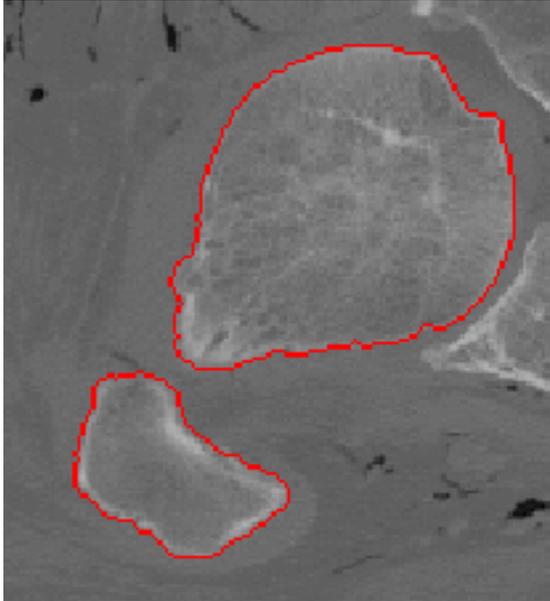
# Advanced Multimodality Image-Guided Operating (AMIGO) Suite



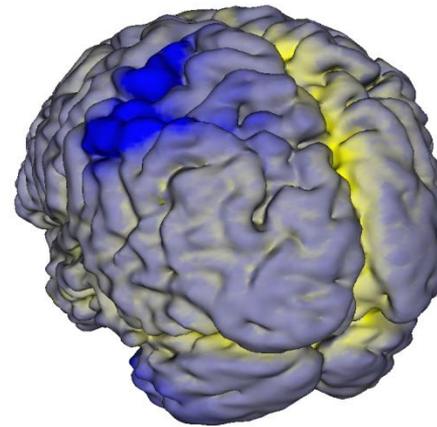
The AMIGO Suite is the nation's first integrated operating suite to offer immediate intra-procedural access to an extensive range of advanced imaging modalities. AMIGO's 5,700 square-foot space is divided into three interconnected procedure rooms housing real-time anatomic, functional, and molecular imaging, including 3T MRI, PET/CT, fluoroscopy, and ultrasound.

- Image Processing, Dynamics, and Control
- Evolving Shapes Statically and Dynamically
- Statistics, Shape, and Estimation
- Interactive Methods

# Shapes

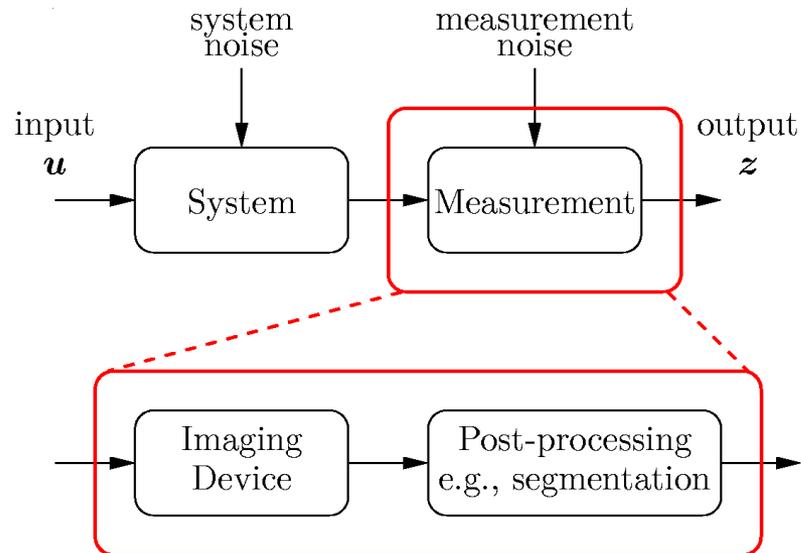


Closed curve



Closed surface

# Classical System Processing



## System:

plane, brain,  
heart, ...

## Measurement:

imaging + post-  
processing, camera,  
fMRI, MRI, ...

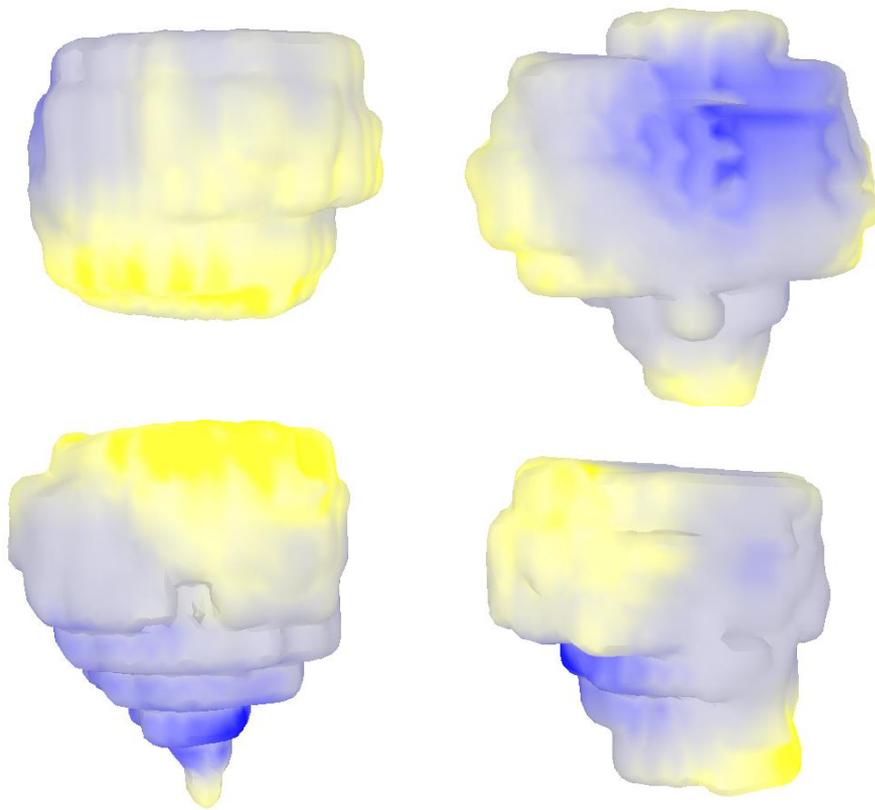
What happens if measurements change over time?

How to influence the system by measured output?

“How to combine image processing, control,  
and machine learning for medical image  
computation?”

# Examples of Shape Variation

Multiple patients

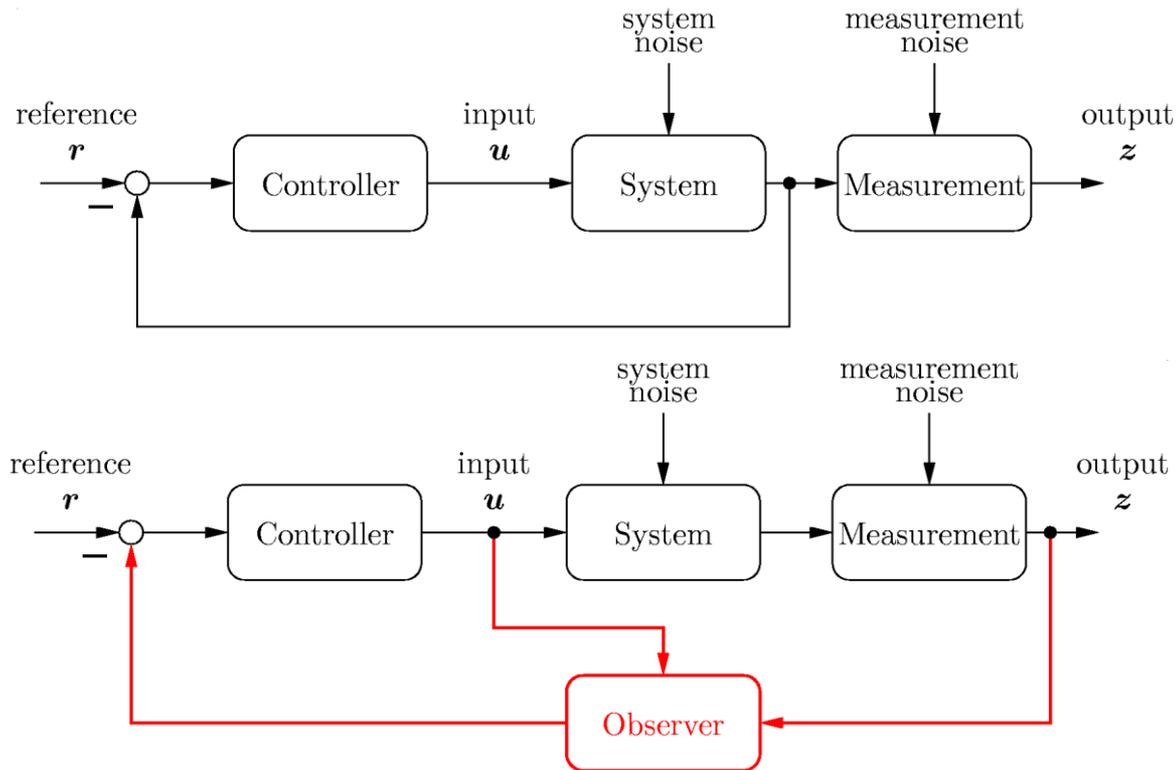


Temporal



[Dataset from C. Tempany MD, A. Szot MD, J. Zhang MD, S. Haker Ph.D.]

# Observer-based Feedback



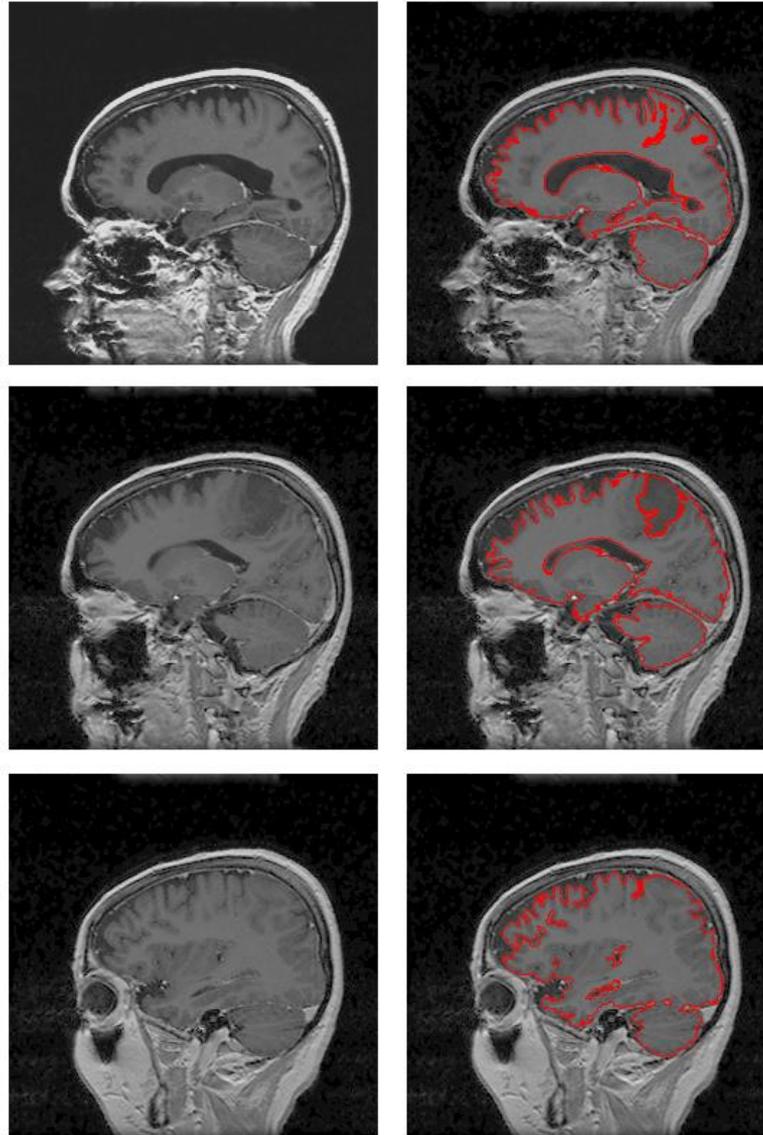
Observer = Filter + System knowledge

Main tools: Active Contours + Particle Filtering

# Geometric Active Contours

- Active Contour Method Using Geodesics, Minimal Surfaces, and Statistics
- Automatic Merging and Breaking (Topological Changes)
- Works for 2D, 3D, 4D
- Used to Segment Various Features: Texture, Intensity, Color, Shape

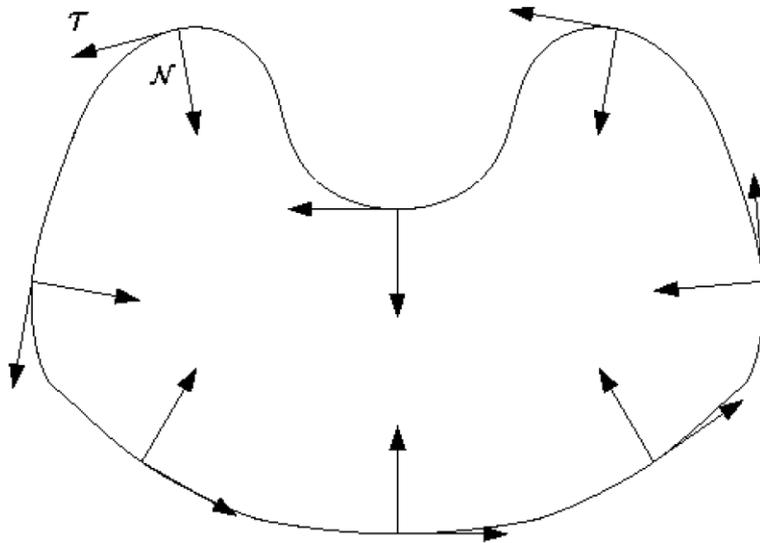
# Active Contours Find Cortical Surface



# Geometric Curve Evolution

The closed curve  $\mathcal{C}$  evolves according to

$$\mathcal{C}_t = (\mathbf{x}_t \cdot \mathcal{N}) \mathcal{N} + (\mathbf{x}_t \cdot \mathcal{T}) \mathcal{T}$$



$\mathcal{N}$  influences the curve's shape

$\mathcal{T}$  moves "particles" along the curve

How is the speed  $\mathbf{x}_t$  determined?

# Curve Evolution through Energy Minimization

Find curve that minimizes a given energy

$$C^* = \operatorname{argmin}_{C \in \mathcal{C}} E(C, I)$$

Static curve evolution

$$E(C, I) = \int_0^1 L(C, C_p, C_{pp}, \dots, I) dp$$

Dynamic curve evolution

$$E(C, I) = \int_t \int_0^1 L(C, C_p, C_{pp}, \dots, C_t, I) dp dt$$

# Geometric Dynamic Approach

Minimizing

$$E = \int_{t_0}^{t_1} \int_0^1 \left( \frac{1}{2} \mu \|C_t\|^2 - g(C) \right) \|C_p\| dp dt$$

by arclength  $ds = \|C_p\| dp$

results in the Euler-Lagrange equation

$$\mu C_{tt} = \underbrace{-\mu(T \cdot C_{ts})C_t - \mu(C_t \cdot C_{ts})T - \frac{1}{2}\|C_t\|^2 \mu \kappa \mathcal{N}}_{\text{dynamic part}} \underbrace{- (\mathcal{N} \cdot \nabla g)\mathcal{N} - g\kappa \mathcal{N}}_{\text{static part}}$$

image independent

image  
dependent

Next Step: Add Statistics, Shape and Estimation

# Bhattacharyya Distance: Statistics

$P_{\text{in}}(z)$  : normalized “interior” density

$P_{\text{out}}(z)$  : normalized “exterior” density

$$B = \int_{\mathcal{Z}} \sqrt{P_{\text{in}}(z)P_{\text{out}}(z)} dz$$

$z$  photometric variable (intensity, color vector, texture vector)

$0 \leq B \leq 1$  represents level of matching

cosine of “angle” between distributions

---

$I(x) : \mathbb{R}^2 \rightarrow \mathcal{Z}$  : image plane to photometric variable

$$P_{\text{in}}(z) = \frac{\int_{\omega} K(z - I(x)) dx}{\int_{\omega} dx}$$

non-parametric kernel density estimate of pdf of  $z$  similarly for  $P_{\text{out}}(z)$

$K$  : kernel function (e.g., Dirac, Gaussian, etc.)

# Bhattacharyya Flow

For curve evolution:

$P_{\text{in}}(z)$  : normalized density **inside** curve  $C$

$$P_{\text{in}}(z) = \frac{\int_{\Omega} K(z - I(x))H(-\phi(x))dx}{\int_{\Omega} H(-\phi(x))dx}$$

$\omega$  : enclosed region

$\Omega$  : whole image domain

$\phi(x)$  : level set function

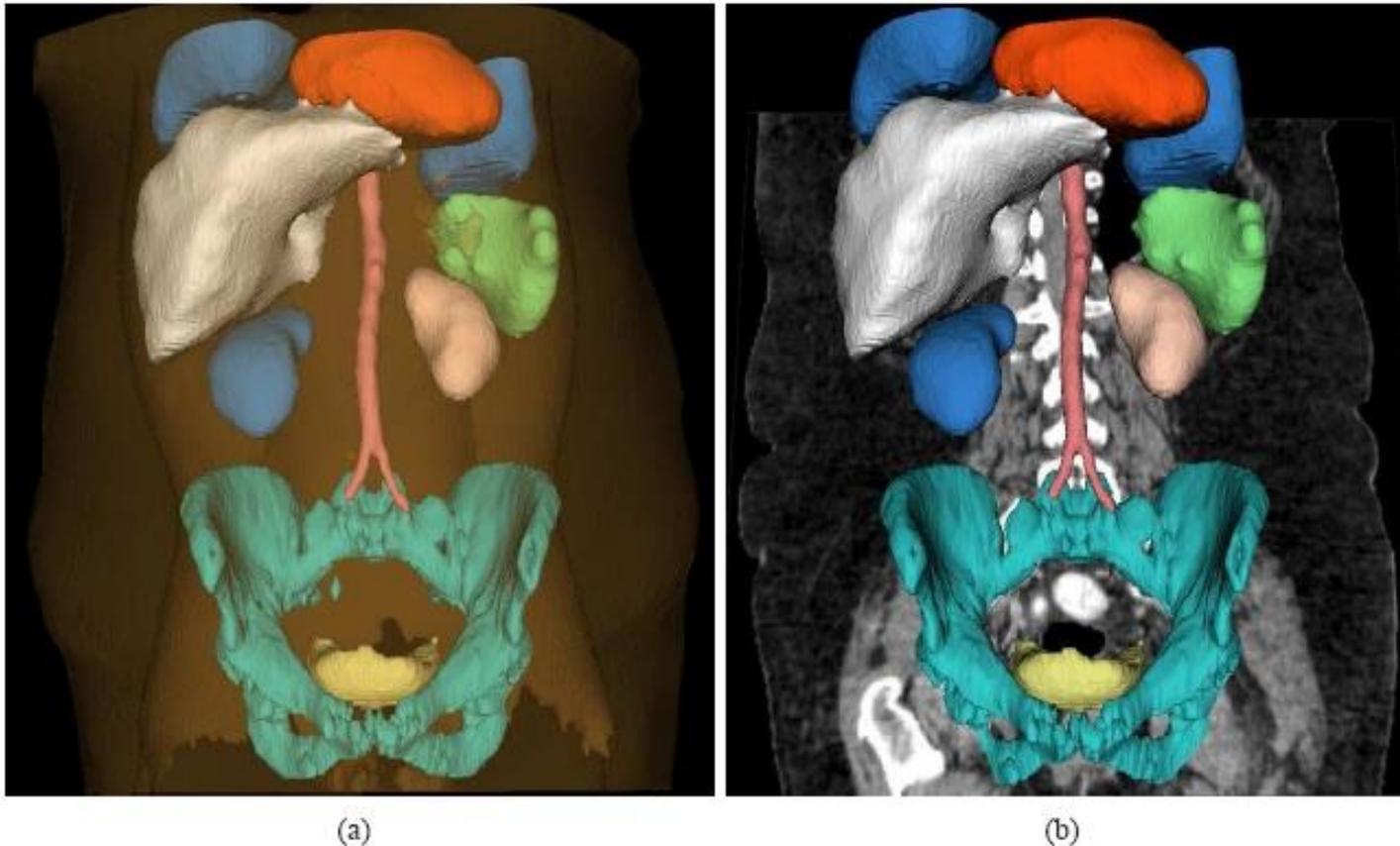
$H$  : Heaviside function

---

leads to the PDE:

$$\frac{\partial \phi(x, t)}{\partial t} = -\frac{B\delta_{\epsilon}(\phi)}{2} \left( \frac{1}{A_{\text{in}}} - \frac{1}{A_{\text{out}}} \right) - \frac{\delta_{\epsilon}(\phi)}{2} \times \int_{\mathcal{Z}} K(z - I(x)) \left( \frac{1}{A_{\text{in}}} \sqrt{\frac{P_{\text{in}}(z)}{P_{\text{out}}(z)}} - \frac{1}{A_{\text{out}}} \sqrt{\frac{P_{\text{out}}(z)}{P_{\text{in}}(z)}} \right)$$

# Multiple Structure Segmentation



Segmentation of heart, two lungs, liver, two kidneys, spleen, abdominal aorta, pelvis, bladder, skin/muscle/fat. The subplot (b) removes skin/muscle/fat but overlays the original image.

# Finsler Metrics: Shape

Given  $\psi : \mathbb{R}^n \times S^{n-1} \rightarrow \mathbb{R}_+$

$$\mathcal{C}(\Gamma) = \mathcal{L}(\Gamma) = \int_0^L \psi(\Gamma, \mathcal{T}) ds = \int_0^1 \psi\left(\Gamma, \frac{\Gamma_x}{|\Gamma_x|}\right) \cdot |\Gamma_x| dx$$

$L$  : length of  $\Gamma$

length function  $\psi$  defined on unit vectors

extended as positive homogeneous of degree 1 to all

$$F(p, v) = |v| \psi\left(x, \frac{v}{|v|}\right)$$

---

anisotropic length of  $\Gamma$ :

$$\mathcal{L}(\Gamma(\cdot, t)) = \int_0^1 F(\Gamma, \Gamma_x) dx$$

Homogeneity of  $F$ :

$$F(p, tv) = tF(p, v)$$

If  $F(p, v)^2$  convex, then  $F$  defines a *Finsler metric* on  $\mathbb{R}^n$

# Minimization: Gradient flow

- Computing the first variation of the functional  $C$ , the  $L_2$ -optimal  $C$ -minimizing deformation is:

$$\frac{\partial \Gamma}{\partial t} = -P_{\Gamma_s^\perp} \left( \nabla_p \Psi - \frac{\partial}{\partial s} \nabla_{\hat{d}} \Psi \right) + \Psi \Gamma_{ss}$$

projection (removes tangential component)

- The steady state  $\Gamma_\infty$  is *locally C-minimal*

$$C(\Gamma) = \int_{\Gamma} \Psi(\Gamma, \mathbf{T}) \, d\Gamma$$

$$\Psi : \mathbb{R}^n \times \mathbb{S}^{n-1} \rightarrow \mathbb{R}^+$$

$$(p, \hat{d}) \mapsto \Psi(p, \hat{d})$$

# Minimization: Dynamic programming

Consider a seed region  $S$  in  $R^n$ , define  
for all target points  $t$  in  $R^n$  the **value function**:

$$C_S^*(t) = \min_{\Gamma \in \mathcal{G}(S,t)} C(\Gamma) \quad \longrightarrow \text{curves between } S \text{ and } t$$

It satisfies the **Hamilton-Jacobi-Bellman equation**:

$$\begin{cases} C_S^* = 0 \text{ on } S \\ \max_{\hat{d} \in S^{n-1}} \{ \nabla C_S^*(t) \cdot \hat{d} - \psi(t, \hat{d}) \} = 0 \end{cases}$$

# Active Contours and Bayesian Statistics

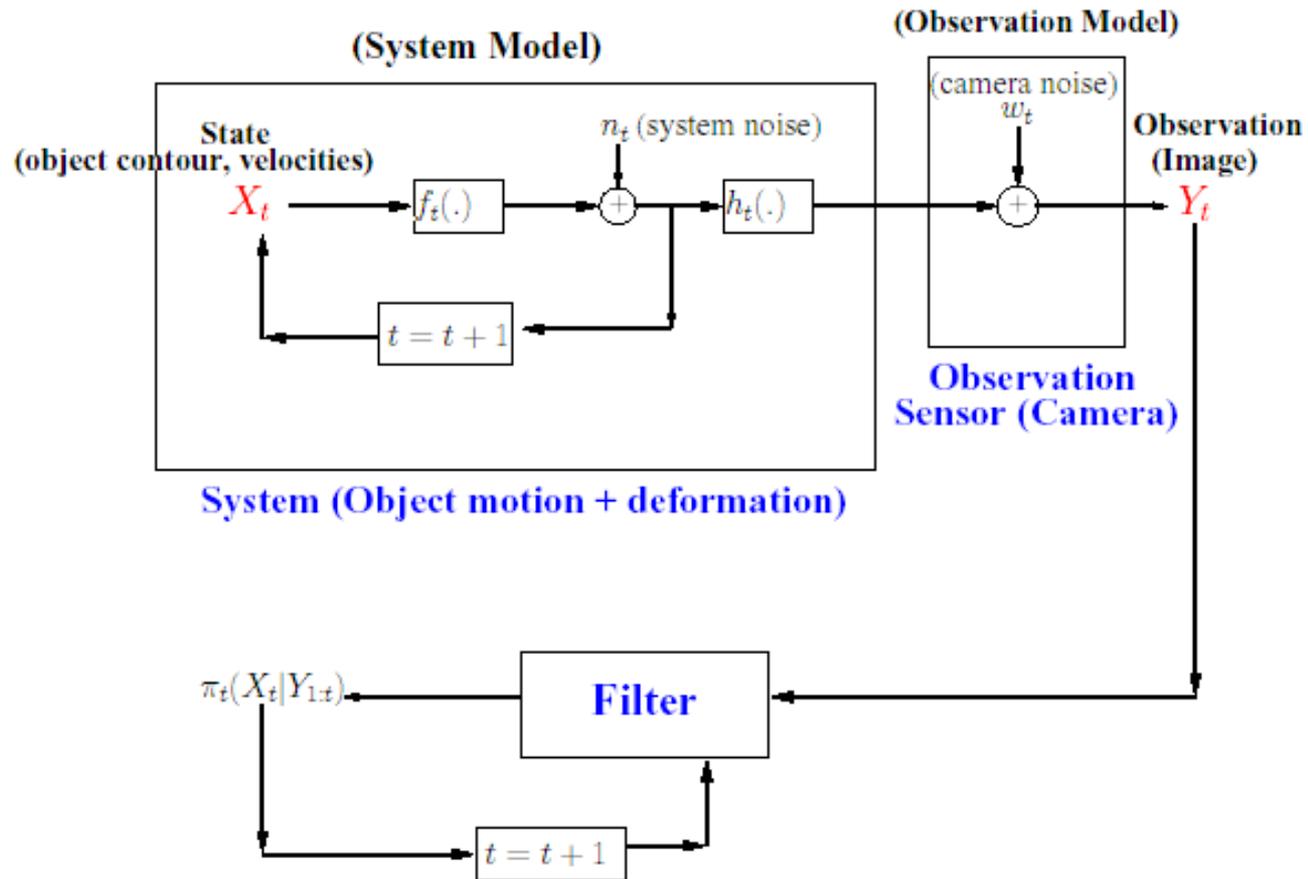
The active contour paradigm can be easily combined with Bayesian estimation. The underlying problem is to find an object  $x$  with prior  $p(x)$  using data  $z$ .

The posterior  $p(x|z)$  can be computed using Bayes' rule. The idea now is that the active contour serves as a prior model of the possible shapes and motions of the features of interest which we want to track. Filtering comes in by adding a dynamic system model to the prior and sensor models in the Bayesian approach. For example, in the linear case with Gaussian distributions and one uses the Kalman filter. For nonlinear models, one can use sigma-point or particle filtering.

Conformal factor is derived locally, based on edge computations.

A more flexible conformal metric is obtained when the metric is learned from the data and if the model incorporates non-local information. For this purpose, we have incorporated statistical methods into geodesic snakes.

# Particle Filters



## Filtering and Tracking

- **Filtering:** Estimating expected value of state  $X_t$  (and of any function of the state), given all observations until  $t$ ,  $Y_{1:t}$ .
- **Tracking:** Predicting the state at  $t$ , using observations until  $t - 1$
- **Complete Solution:** evaluate the tracking (prediction) and filtering (posterior) distributions at each  $t$  defined as:

$$\begin{aligned}\pi_{t|t-1}(dx) &= Pr(X_t \in dx | Y_{1:t-1}) : \text{Prediction} \\ \pi_t &\triangleq \pi_{t|t}(dx) = Pr(X_t \in dx | Y_{1:t}) : \text{Posterior}\end{aligned}$$

# Bayesian Tracking

Posterior over model parameters given an image sequence.

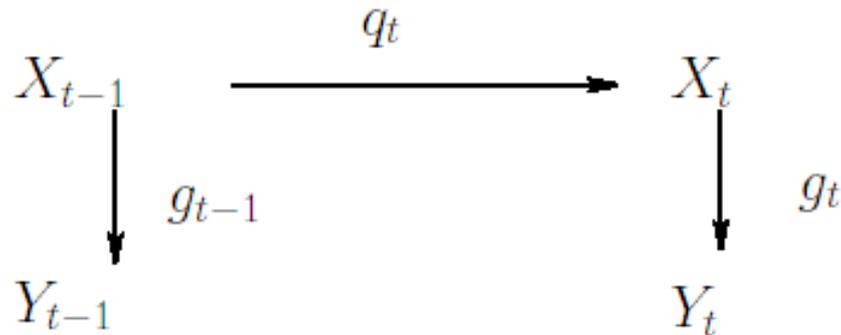
$$p(\mathbf{x}_t | \mathbf{Z}_t) = \kappa p(\mathbf{z}_t | \mathbf{x}_t) \int (p(\mathbf{x}_t | \mathbf{x}_{t-1}) p(\mathbf{x}_{t-1} | \mathbf{Z}_{t-1})) d\mathbf{x}_{t-1}$$

Temporal model (prior)

Posterior from previous time instant

Monte Carlo integration

## Notation



- **State transition model: ( $X_t$ : contour, velocities)**

$$X_t = f_t(X_{t-1}) + n_t, \quad q_t(X_t|X_{t-1}) = p_n(X_t - f_t(X_{t-1}))$$

- **Observation model: ( $Y_t$ : image)**

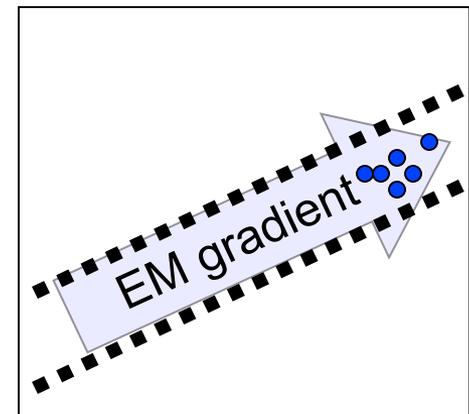
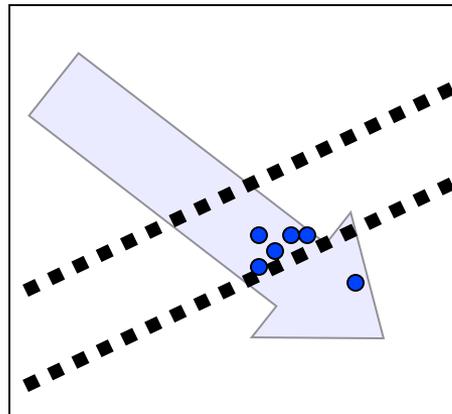
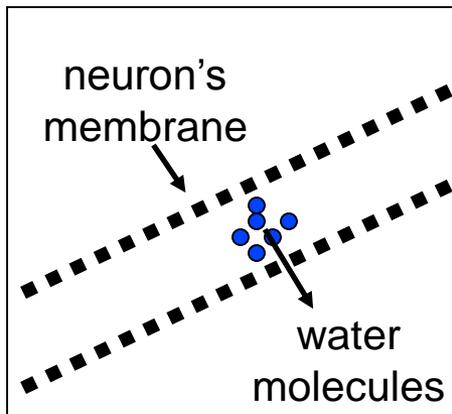
$$Y_t = h_t(X_t) + w_t, \quad g_t(Y_t|X_t) = p_w(Y_t - h_t(X_t))$$

- **Hidden Markov Model (HMM)**

# Application:

## Diffusion MRI tractography

- Diffusion MRI measures the diffusion of water molecules in the brain
- Neural fibers influence water diffusion
- **Tractography:** “recovering probable neural fibers from diffusion information”



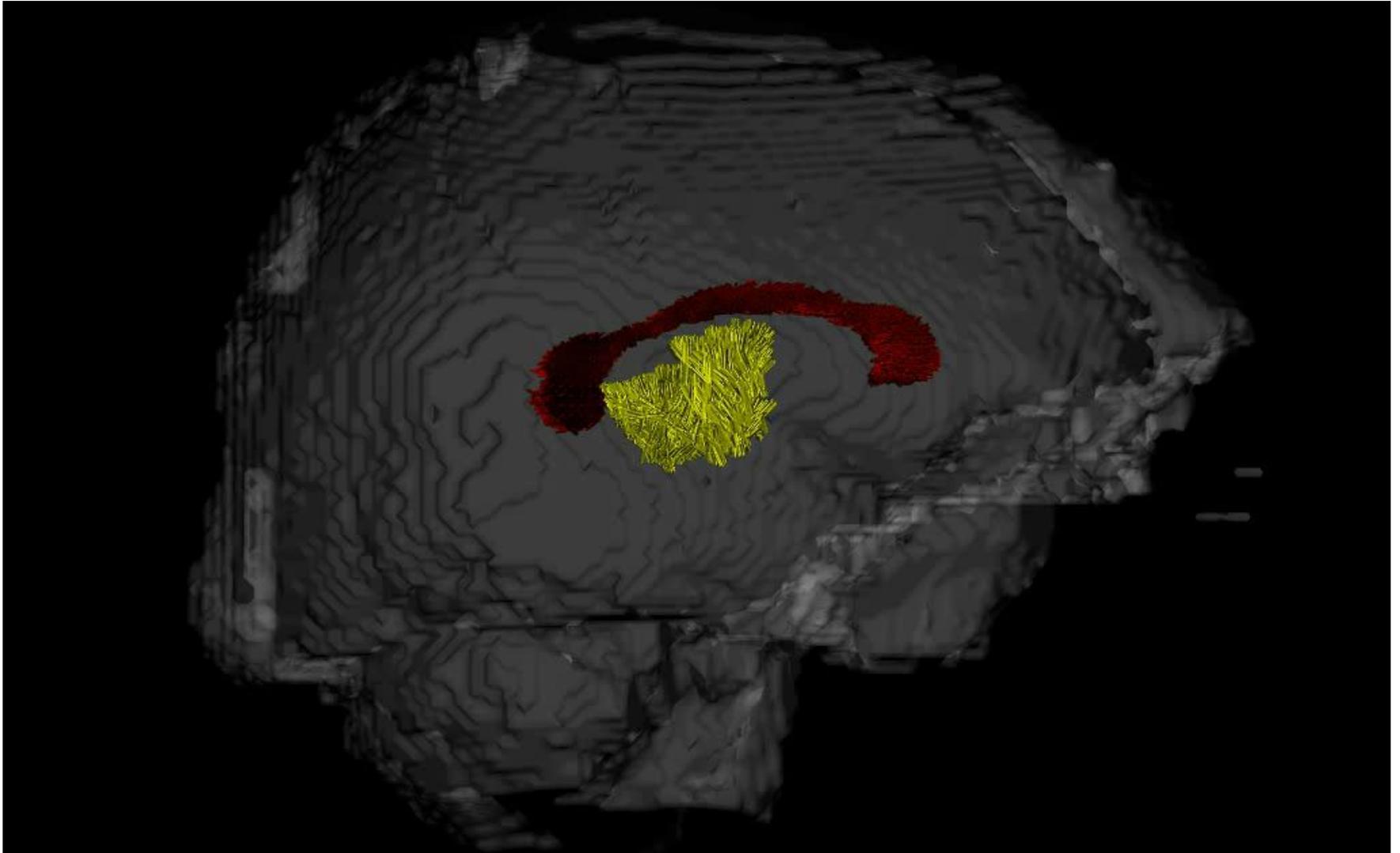
# Application: Diffusion MRI tractography (2)

- Diffusion MRI dataset:
  - Diffusion-free image:  $S(\cdot, \mathbf{0}) : \mathbb{R}^3 \rightarrow \mathbb{R}^+$
  - Gradient directions:  $\hat{\mathbf{k}}_i \in \mathbb{S}^2, i = 1 \dots N$
  - Diffusion-weighted images:  $S(\cdot, \hat{\mathbf{k}}_i) : \mathbb{R}^3 \rightarrow \mathbb{R}^+$
  - We choose:

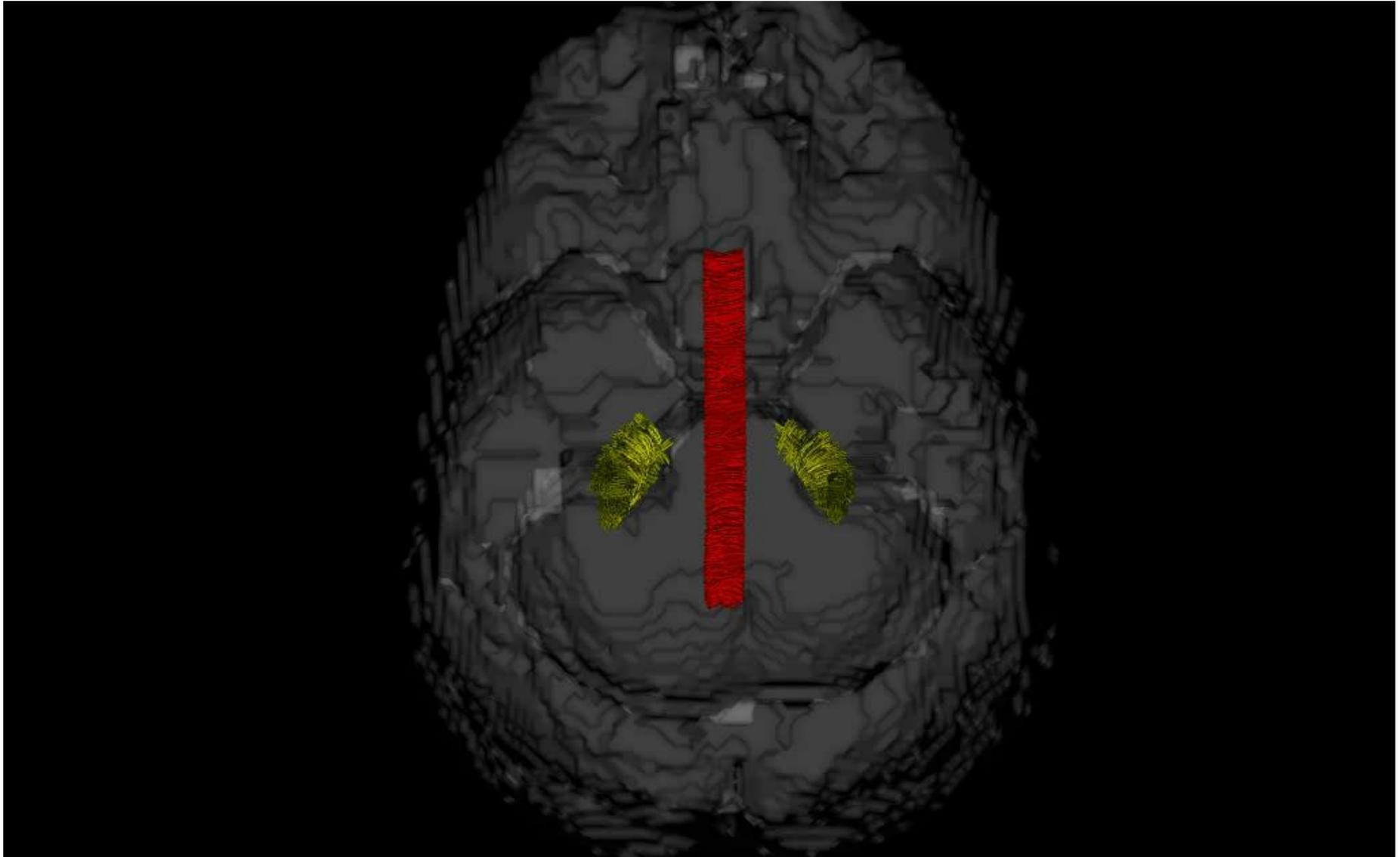
$$\psi(\mathbf{p}, \hat{\mathbf{d}}) = f\left(\frac{S(\mathbf{p}, \hat{\mathbf{d}})}{S(\mathbf{p}, \mathbf{0})}\right)$$

ratio = 1 if no diffusion  
< 1 otherwise

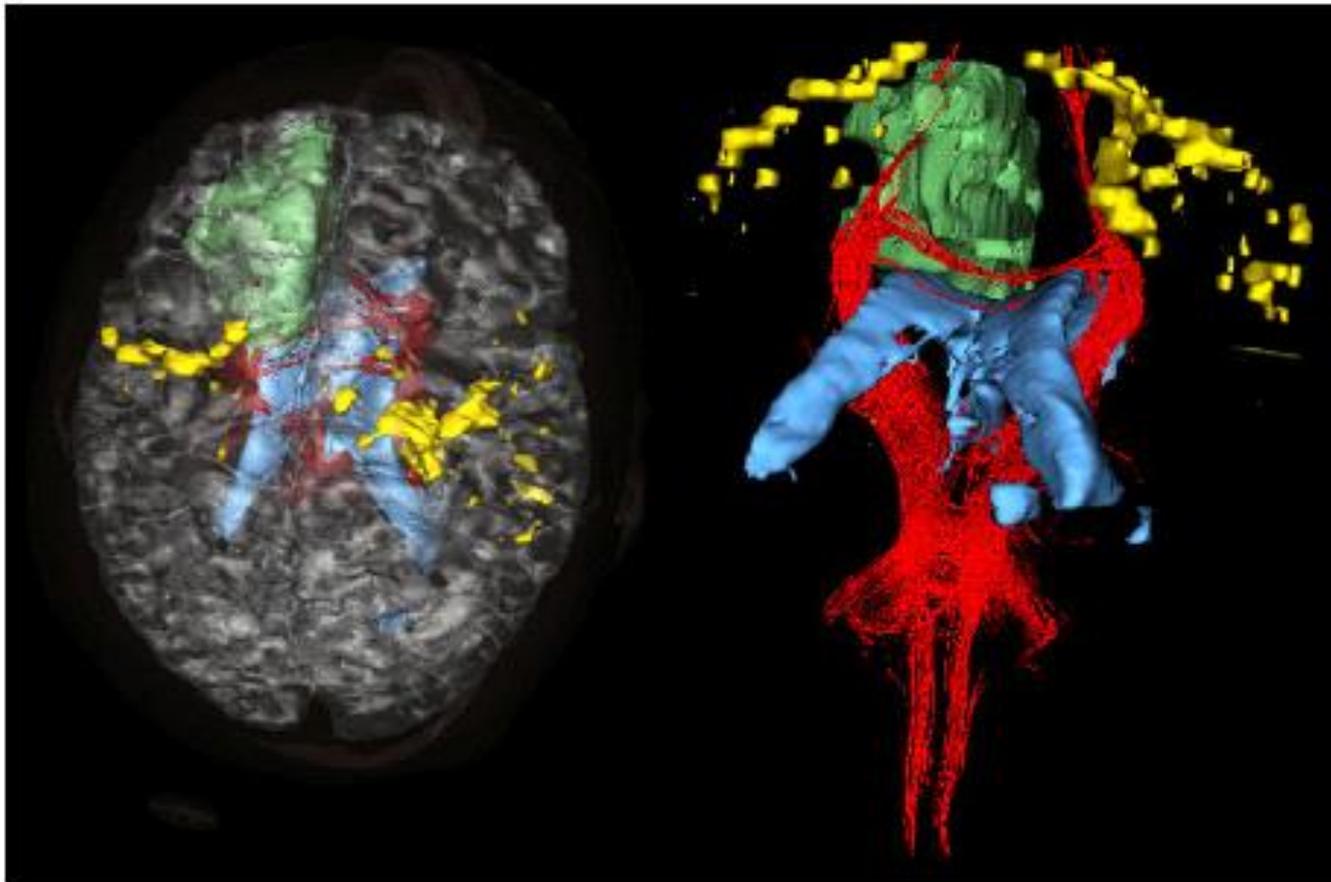
# Finsler Tract Growing: I



# Finsler Tract Growing: II

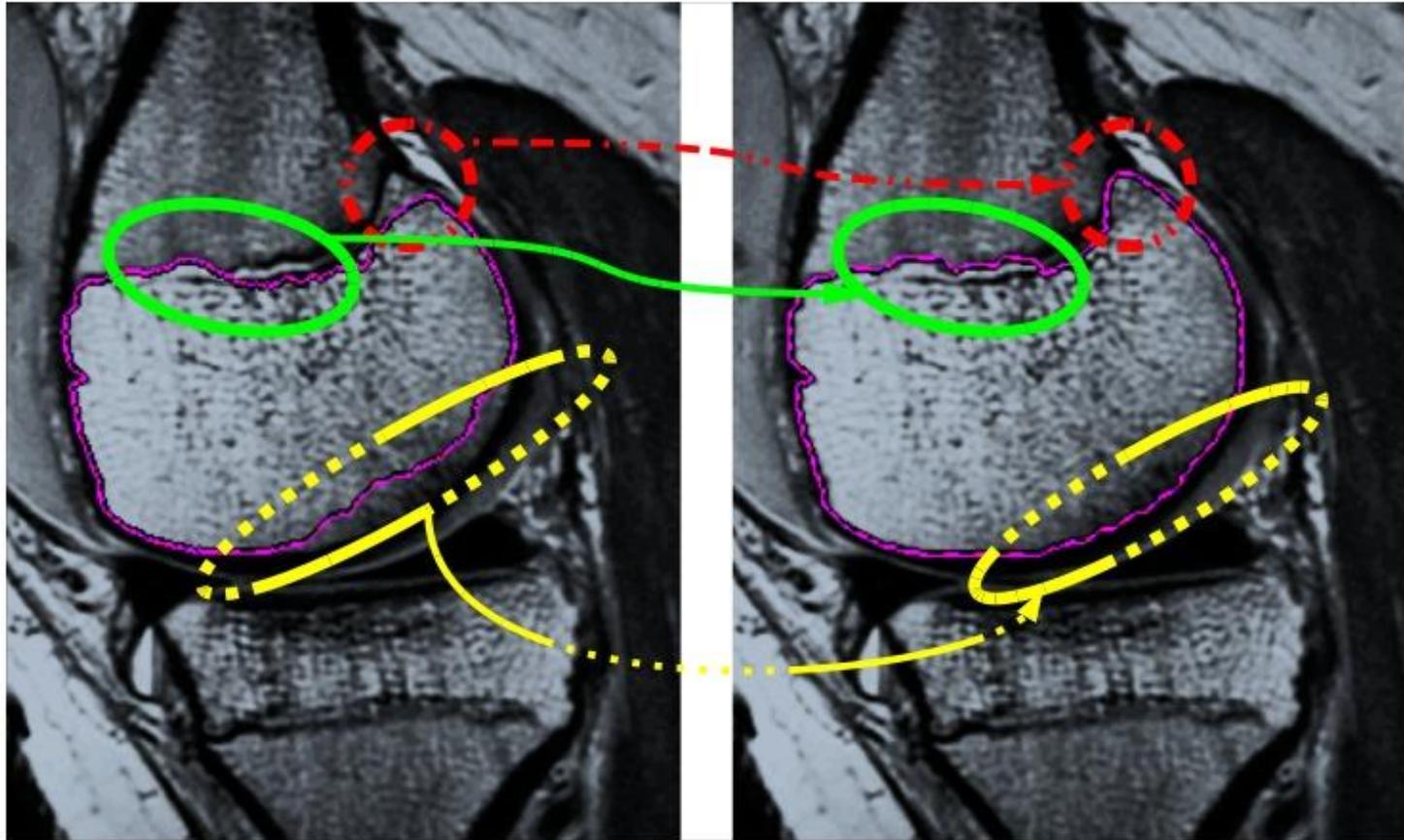


# fMRI and DTI for IGS



**Figure 8.4.6-1. Retrospective Example of fMRI for Neurosurgical Application**  
62-year-old female patient with left frontal hyperintense non-enhancing mass lesion  
Skin, Brain, Ventricles (blue) and Tumor (green) models from conventional MRI; fMRI  
activations (yellow) from pre-operative finger-taping experiment. Fiber tract indications  
(red) from Diffusion Tensor MRI.  
Imaging suggests that the tumor is in front of motor strip with involvement of  
supplementary motor area, with fibers from SMA piercing tumor in its posterior aspect.

# Expert Knowledge for Segmentation

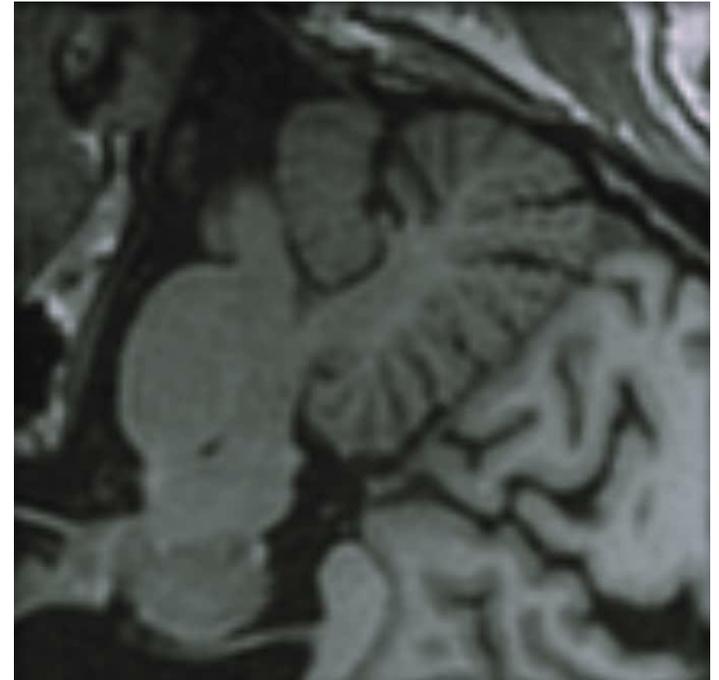
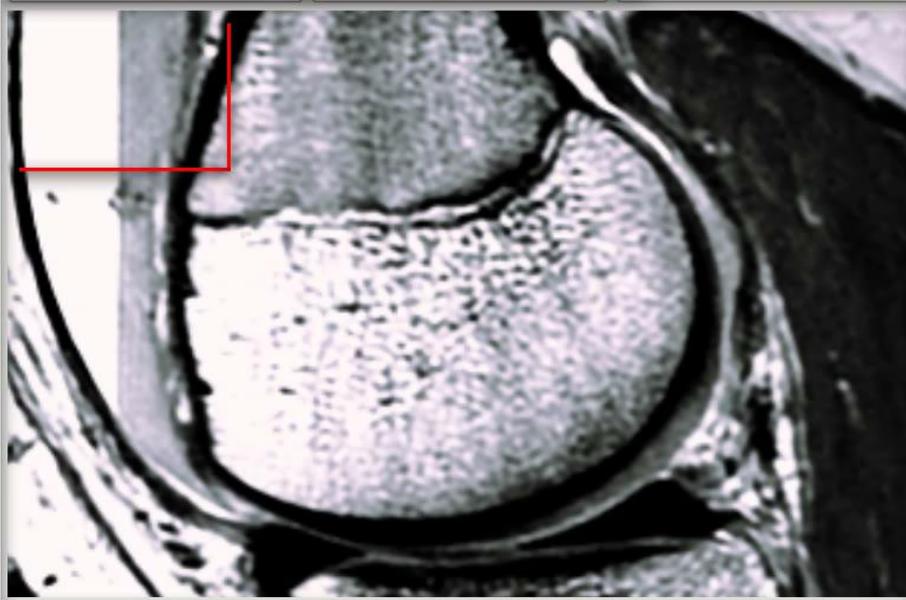


A motivating problem: measuring volume of *Epiphysis*, *Cartilage-Cap*, and *Physis* (growth-plate) during adolescence.

**Left:** automatic segmentation of Epiphysis.

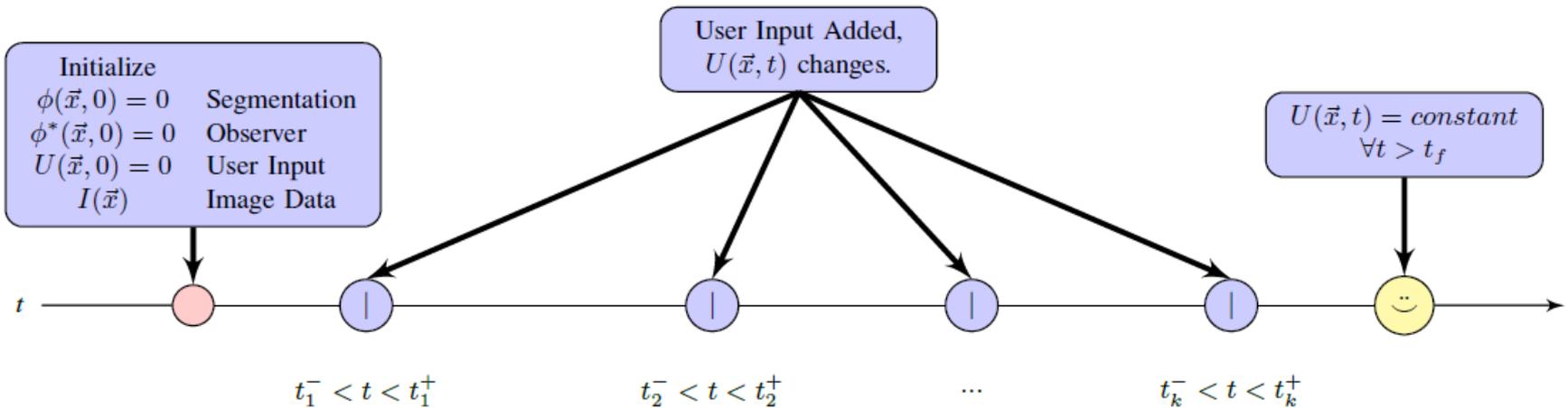
**Right:** augmentation by user-in-the-loop curve evolution.

# Video Demos



# Why Interactive Segmentation?

1. Ever more complex segmentation models will be slower and still not work in many cases.
2. Atlas-Based methods may not be applicable (trauma, unique growth stage, atlas is "To-Do")
3. Doctors & Med Students can use it easily, sole parameter is "editor brush size".



Above: Timeline of Interactive Segmentation System

# Formulation: Augmented Cost-Function (I)

$$\mathcal{E}(\phi) = \int_{\Omega} g(\phi, I) \|\nabla \phi\|_2 d\Omega$$

$$\frac{\partial U}{\partial t} = \begin{cases} \phi(\mathbf{x}, t_k^+) - \phi(\mathbf{x}, t_k^-) & \text{for } t_k^- < t < t_k^+ \\ 0 & \text{otherwise .} \end{cases}$$

$$\phi_t = \underbrace{G(\phi, I)}_{\text{nominal}} + \underbrace{H(\phi, \phi^*)}_{\text{control}}$$

$$e_U \doteq \phi^* - U$$

$$\hat{\phi} \doteq \phi^* - \phi$$

- Nominal cost function
  - minimizing this is "automatic algorithm"
- User-Input: changes to segmentation function at discrete intervals
- Augmented Curve Evolution
  - nominal plus user-driven term
    - Signal Definitions
      - User Input Error
      - Observer Error

# Formulation: Augmented Cost-Function (II)

$$\mathcal{F}[\phi^*] = \int_{\Omega} \frac{K_F}{2} \hat{\phi}^2 + \frac{K_U}{2} e_U^2 (\gamma U)^2 + \frac{\mu}{2} (\phi^{*2} + \|\nabla \phi^*\|_2^2) d\Omega .$$

$$\mathcal{H}[\phi] = \int_{\Omega} \frac{K_{\phi}}{2} (\hat{\phi}^2 + \|\nabla \hat{\phi}\|_2^2) + g(\phi, I) \|\nabla \phi\|_2 d\Omega .$$

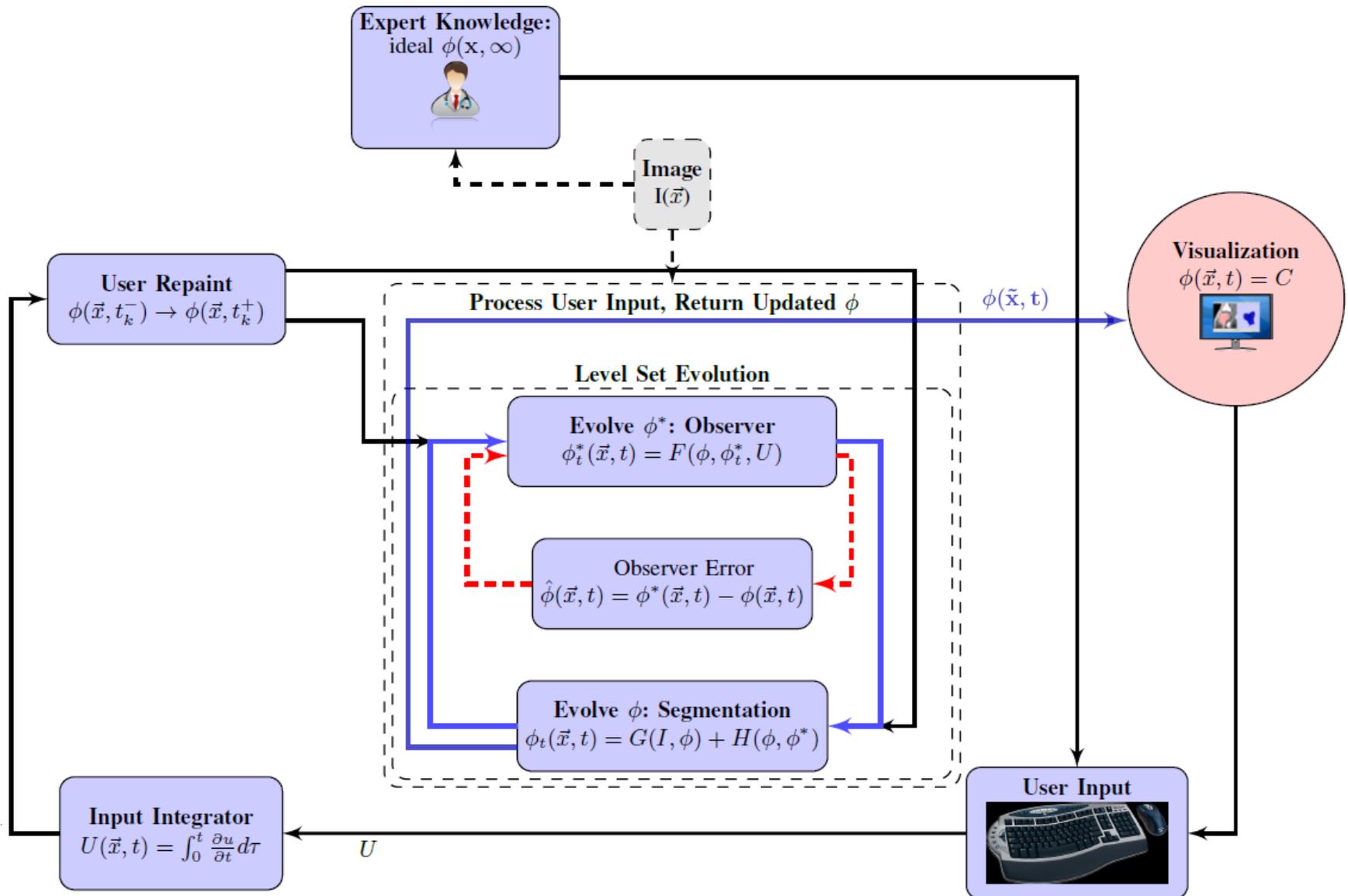
$$\phi_t^* = \left[ K_F \hat{\phi} + K_U e_U (\gamma U)^2 + \mu (\phi^* - \Delta \phi^*) \right]$$

$$\phi_t = G(\phi, I) + K_{\phi} (\hat{\phi} - \Delta \hat{\phi})$$

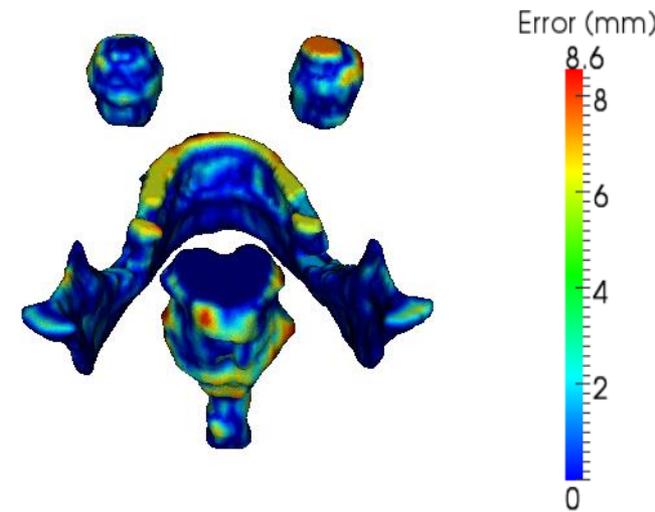
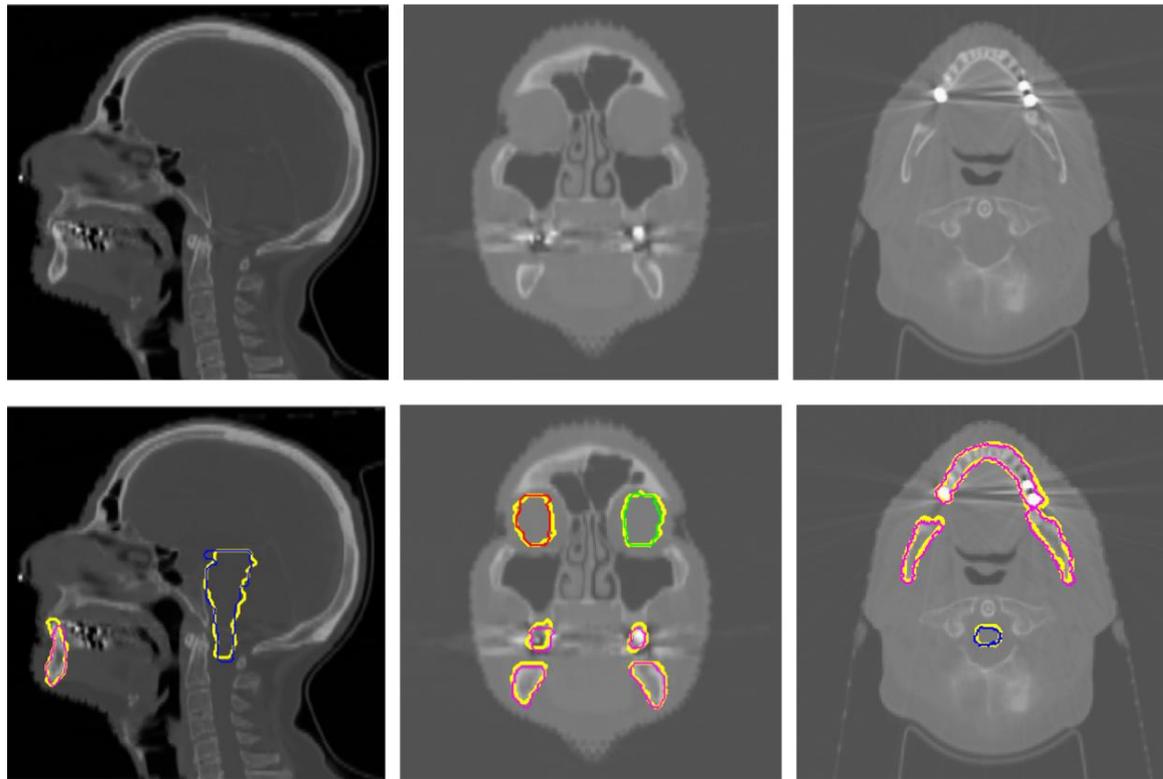
- Minimize F and H by gradient flow.
  - F filters the user input
    - enable sloppy clicking
    - accumulate U where user strongly disagrees with automatic algorithm
  - H balances automatic algorithm with observer error

- Observer Evolution
- Segmentation Evolution
  - real-time display update
  - enable human to generate U

# System Diagram with Feedback Loops

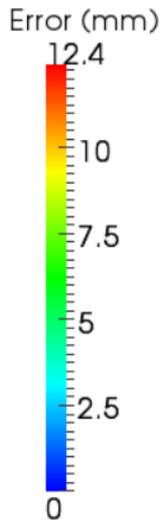
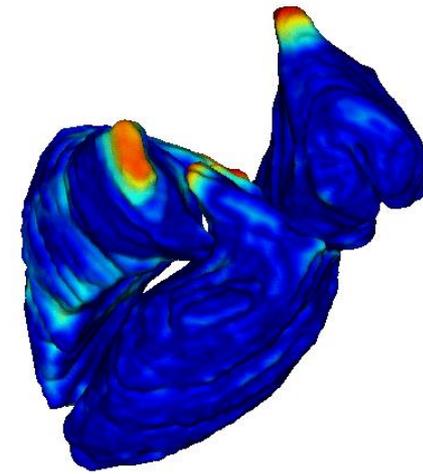
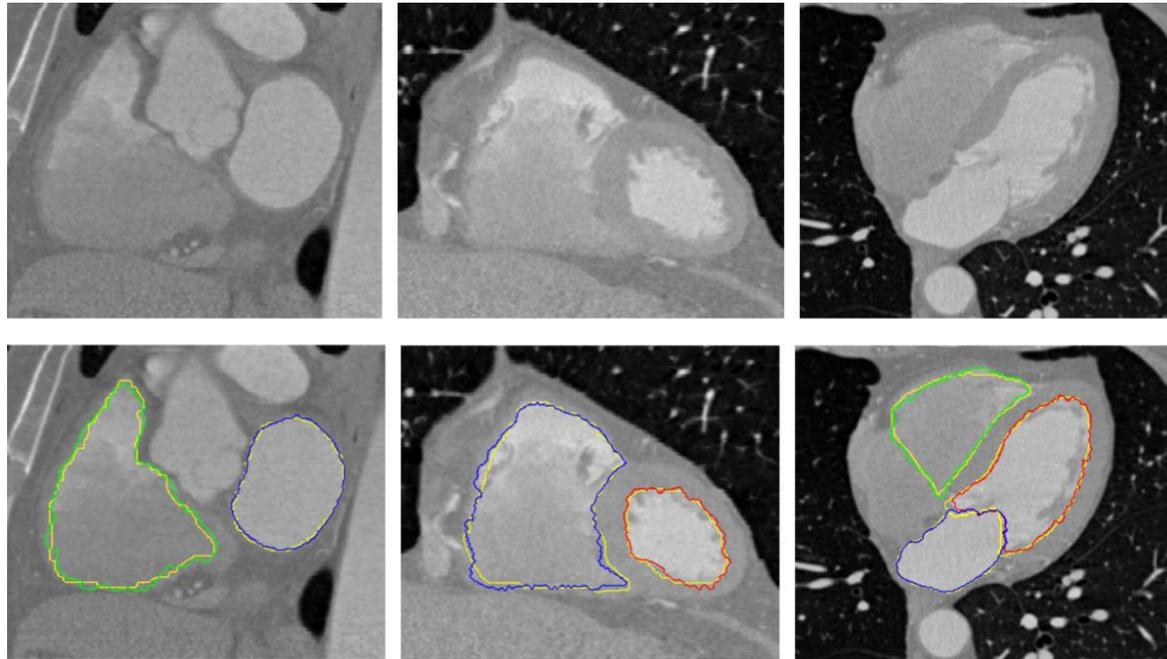


# 3D Volume: Head-Neck Image



Region-based active contour: Example of segmenting the left eye (red), right eye (green), brain stem (blue), and mandible (pink), superimposed over manual segmentations (yellow).

# 3D Volume: Cardiac Image

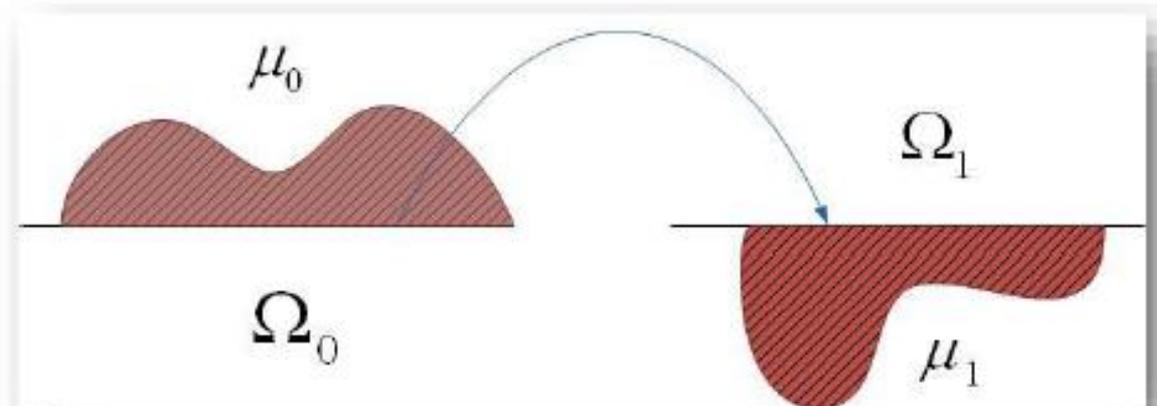


Distance-based clustering: Example of segmenting the left ventricle (red), right ventricle (green), and left atrium (blue), superimposed over manual segmentations (yellow).

# Optimal Mass Transport

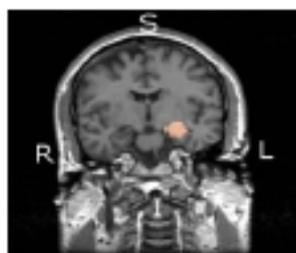
## Monge Transportation Cost (1781)

- ❑ Considers the engineer's problem of transporting a pile of soil or rubble to an excavation with the least amount of work.

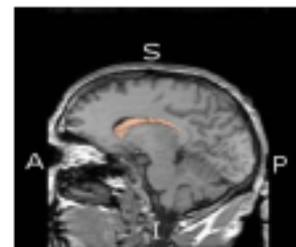
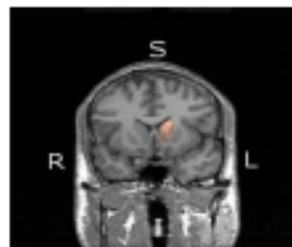


# Optimal Mass Transport Applications

- Econometrics, fluid dynamics, automatic control, statistical physics, shape optimization, expert systems, meteorology, spectral analysis, time-series analysis, and many more fields.
- Our interest here - Registration, shape analysis, visual tracking.



Left hippocampus data



Left caudate nucleus data



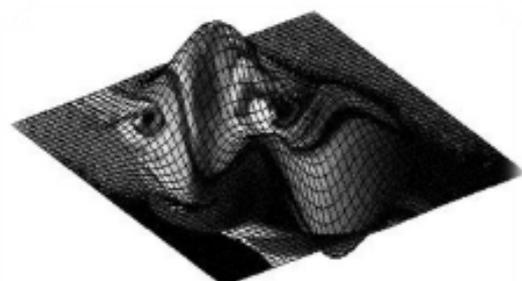
# Optimal Mass Transport (MKW)

Given two oriented Riemannian manifolds

$\Omega_0$  and  $\Omega_1$

with corresponding density functions

$\mu_0$  and  $\mu_1$



and the same  
amount of total mass:

$$\int_{\Omega_0} \mu_0(x) dx = \int_{\Omega_1} \mu_1(x) dx$$

## Transportation Cost Modern Formulation - Monge Kantorovich (MK)

Construct a smooth mapping:

$$u : (\Omega_0, \mu_0) \rightarrow (\Omega_1, \mu_1)$$

With *mass preserving (MP)* constraint:

$$\mu_0 = \det(\nabla u) \mu_1(u) \quad (\text{Jacobian equation})$$

so as to minimize the cost function:

$$M(u) = \int_{\Omega_0} \Phi(x, u(x)) \mu_0(x) dx$$

$\Phi(x, u(x))$  is a positive twice differentiable convex function.

# Kantorovich-Wasserstein Metric

Smooth mass preserving mapping:

$$u : (\Omega_0, \mu_0) \rightarrow (\Omega_1, \mu_1)$$

-- many solutions

Optimal (when it exists) provides a preferred geometry (like Riemann mapping on the plane).

Kantorovich-Wasserstein metric:

$$d_p(\mu_0, \mu_1)^p := \inf_u \int |u(x) - x|^p \mu_0(x) dx$$

# Algorithm for Optimal Transport-I

Subdomains with smooth boundaries and positive densities

$$\Omega_0, \Omega_1 \subset \mathbf{R}^d$$

$$\int_{\Omega_0} \mu_0 = \int_{\Omega_1} \mu_1$$

Consider diffeomorphisms mapping one density to another

$$\mu_0 = \det(\nabla \tilde{u}) \mu_1 \circ \tilde{u}$$

Satisfying the mass preservation property

-- we start from a particular such map  $u$

# Algorithm for Optimal Transport-II

Consider a smooth one parameter family of MP-maps:

$$\tilde{u} := u \circ s^{-1}, \quad s = s(\cdot, t), \quad \mu_0 = \det(\nabla s) \mu_0 \circ s$$

From the MP property and the construction of the path it follows

$$\tilde{u}_t = -\frac{1}{\mu_0} \nabla \tilde{u} \cdot \zeta, \quad \zeta = \mu_0 s_t \circ s^{-1}$$

$$\operatorname{div} \zeta = 0$$

# Algorithm for Optimal Transport-III

MK optimality requires that we minimize the functional

$$\begin{aligned} M(t) &= \int_{\Omega_0} \Phi(\tilde{u}(x, t) - x) \mu_0(x) \, dx \\ &= \int \Phi(u(y) - s(y, t)) \mu_0(y) \, dy, \quad x = s(y, t), \quad s^*(\mu_0(x) dx) = \mu_0(y) dy \end{aligned}$$

for which we take the first variation:

$$\begin{aligned} M'(t) &= - \int \langle \Phi'(u - s), \, s_t \rangle \mu_0 dy \\ &= - \int \langle \Phi'(\tilde{u}(x, t) - x), \, \mu_0 s_t \circ s^{-1} \rangle \, dx \\ &= - \int_{\Omega_0} \langle \Phi'(\tilde{u}(x, t) - x), \, \zeta \rangle \, dx \end{aligned}$$

# Algorithm for Optimal Transport-IV

**First choice:**

$$\zeta = \Phi'(\tilde{u} - x) + \nabla p$$

$$\operatorname{div} \zeta = 0$$

$$\zeta|_{\partial\Omega_0} \text{ tangential to } \partial\Omega_0$$

**This leads to the system of equations:**

$$\tilde{u}_t = -1/\mu_0 \nabla \tilde{u} \cdot (\Phi'(\tilde{u} - x) + \nabla p)$$

$$\Delta p + \operatorname{div} (\Phi'(\tilde{u} - x)) = 0, \quad \text{on } \Omega_0$$

$$\frac{\partial p}{\partial \vec{n}} + \vec{n} \cdot \Phi'(\tilde{u} - x) = 0, \quad \text{on } \partial\Omega_0$$

# Solution of L2 M-K and Polar Factorization

Specializing to quadratic cost:

$$\Phi(x) = \frac{|x|^2}{2}$$

leads to the following “non-local” gradient descent equation:

$$\tilde{u}_t = -1/\mu_0 \nabla \tilde{u} (\tilde{u} - \nabla \Delta^{-1} \operatorname{div}(\tilde{u}))$$

Motivation for the approach:

$$\tilde{u} = u \circ s^{-1} = \nabla w + \chi, \quad \operatorname{div}(\chi) = 0 \quad \text{Helmholtz decomp.}$$

The key idea is to push the fixed initial map  $u$  (thought of as a vector field) using the one-parameter family of MP maps in order to remove the divergence-free part!

$$u = \nabla w \circ s \quad \text{Polar factorization}$$

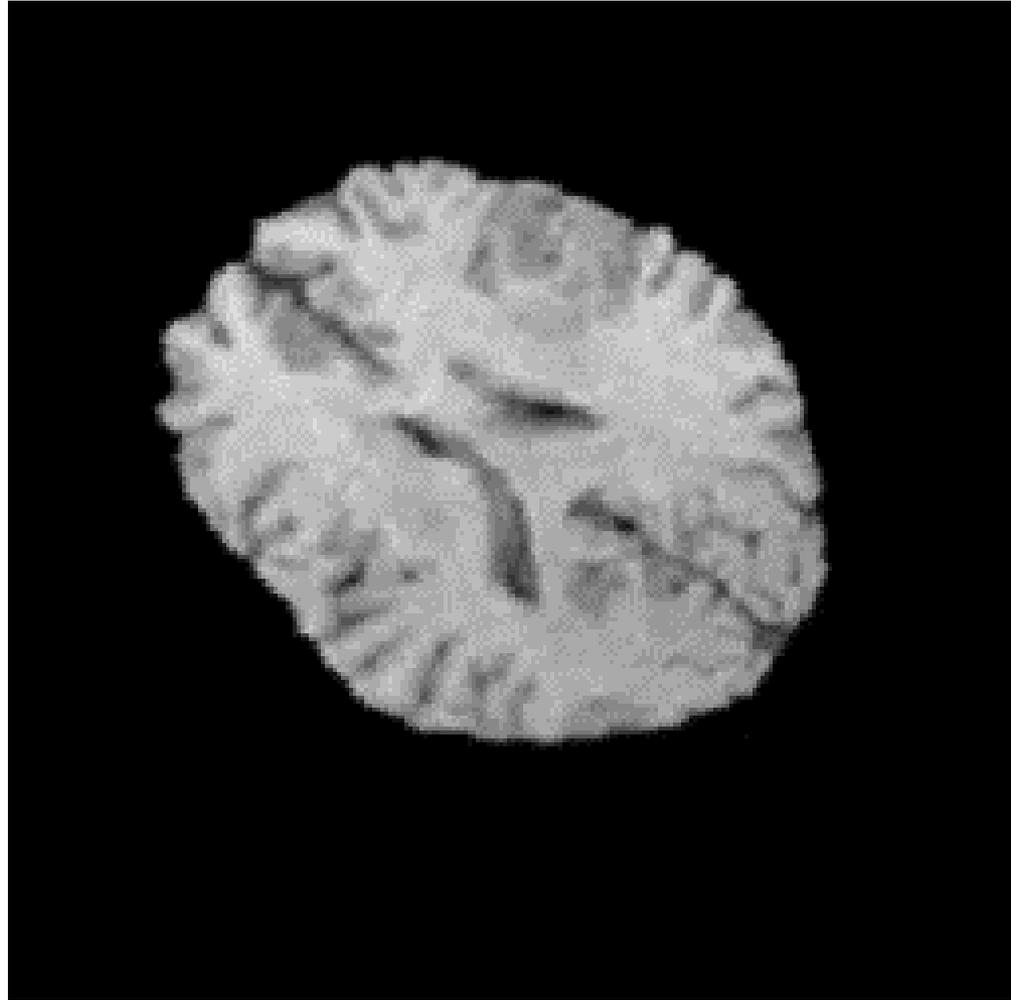
# Registration and Mass Transport

**Image registration is the process of establishing a common geometric frame of reference from two or more data sets from the same or different imaging modalities taken at different times.**

**Multimodal registration proceeds in several steps. First, each image or data set to be matched should be individually calibrated, corrected from imaging distortions, cleaned from noise and imaging artifacts. Next, a measure of dissimilarity between the data sets must be established, so we can quantify how close an image is from another after transformations are applied to them. Similarity measures include the proximity of redefined landmarks, the distance between contours, the difference between pixel intensity values. One can extract individual features that to be matched in each data set. Once features have been extracted from each image, they must be paired to each other. Then, a the similarity measure between the paired features is formulated can be formulated as an optimization problem.**

**We can use Monge-Kantorovich for the similarity measure in this procedure.**

# Brain Sag



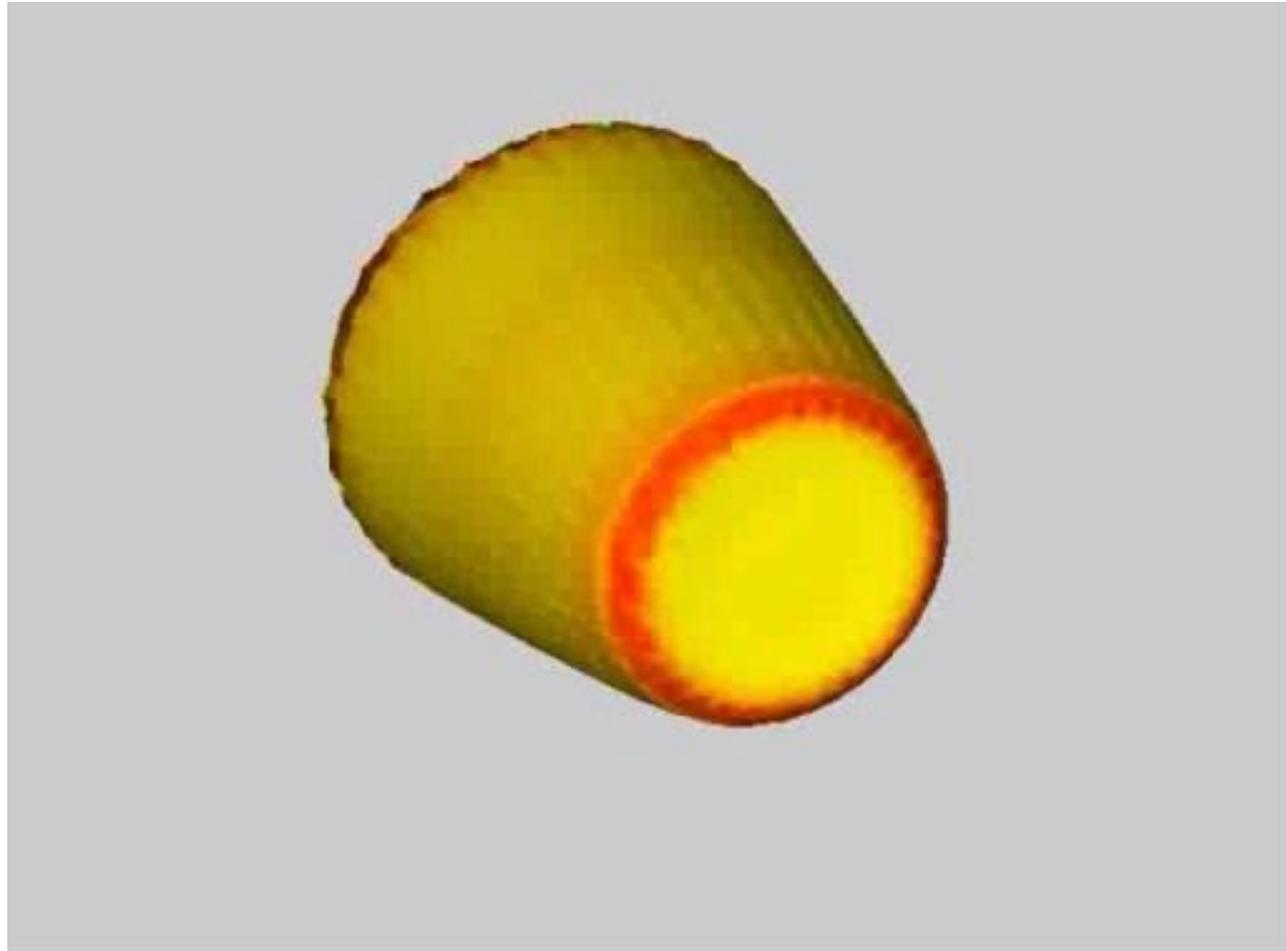
# Beating Heart



# Solar flare



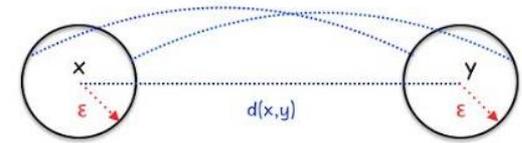
## Example of OMT Mapping on Spherical Shape



# Wasserstein Distance

## Motivation:

We begin by recording the basic definition of the  $L^p$ -Wasserstein distance from optimal transport theory that we will need below. Roughly speaking, on a metric measure space, one gets a natural distance on “small” balls around points or the “fuzzified” points.



Measuring Distance of “small balls”

## Defintion:

Let  $X$  be a metric measure space, equipped with distance  $d$ . Let  $m_i, i = 1, 2$  be two measures with the same total mass and finite  $p$ -th moment. A *coupling* between  $m_1$  and  $m_2$  is a measure  $m$  on  $X \times X$  such that

$$\int_y d m(x, y) = d m_1(x) \quad \int_x d m(x, y) = d m_2(x)$$

In other words, the marginals of  $m$  are  $m_1$  and  $m_2$ . Let  $\tilde{\mathcal{O}}(m_1, m_2)$  be the set of couplings between  $m_1$  and  $m_2$ . We then define the  $L^p$  Wasserstein distance as

$$W_p(m_1, m_2) := \inf_{m \in \tilde{\mathcal{O}}(m_1, m_2)} \left( \int d(x, y)^p d m(x, y) \right)^{1/p}$$

We are interested for cases of  $p = 1, 2$ . In particular, for the case of  $p=1$ , we can solve this distance very efficiently though a simple linear program.

# Wasserstein Distance

## Wasserstein 1-Metric:

Let  $\mu_1$  and  $\mu_2$  now be two discrete distributions with same total mass over  $n$  points, respectively, and let  $d(x,y)$  represent the distance between such samples (for the case of graphs, this is simply taken to be the hop distance). Then,  $W_1(\mu_1, \mu_2)$  may be described as follows:

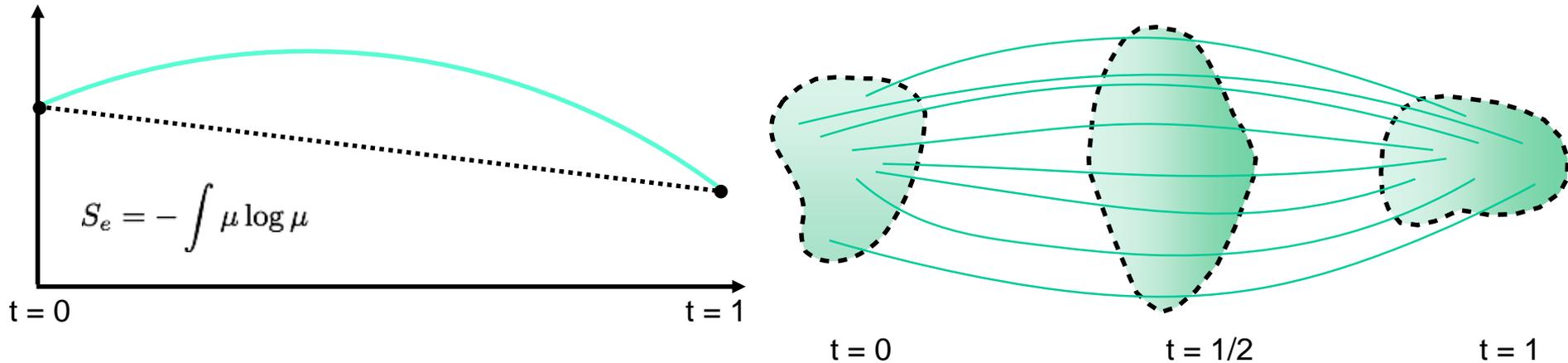
$$W_1(\mu_1, \mu_2) = \min \sum_{i,j=1}^n d(x_i, x_j) \mu(x_i, x_j)$$

where  $\mu(x, y)$  is a coupling (or flow) subject to the following constraints:

$$\begin{aligned} \mu(x, y) &\geq 0, \\ \sum_{j=1}^n \mu(x, y_j) &= \mu_1(x), \quad \forall x, \\ \sum_{i=1}^n \mu(x_i, y) &= \mu_2(y), \quad \forall y. \end{aligned}$$

The cost above finds the optimal coupling of moving a set of mass from distributions  $\mu_1$  to  $\mu_2$  with minimal “work” [4].

# Explaining Curvature to Boltzmann: Lazy Gas Experiment



If Ricci curvature is non-negative, then we have:

$$S_e(\mu_t) \geq (1 - t)S_e(\mu_0) + tS_e(\mu_1)$$

# Ricci Curvature and Entropy

Lott & Villani:

Let  $(X, d, m)$  denote a geodesic space, and set:

$$P(X, d, m) := \{\mu \geq 0 : \int_X \mu dm = 1\},$$

$$P^*(X, d, m) := \{\mu \in P(X, d, m) : \lim_{\varepsilon \searrow 0} \int_{\mu \geq \varepsilon} \mu \log \mu dm < \infty\}.$$

We define

$$H(\mu) := \lim_{\varepsilon \searrow 0} \int_{\mu \geq \varepsilon} \mu \log \mu dm, \text{ for } \mu \in P^*(X, d, m),$$

Which is the negative of the *Boltzmann entropy*  $S_e(\mu) := -H(\mu)$ ; note concavity of  $S_e$  is equivalent to the convexity of  $H$ . Then we say that  $X$  has *Ricci curvature bounded from below by  $k$*  if for every  $m_0, m_1 \in P(X)$  there exists a constant speed geodesic  $\mu_t$  with respect to the Wasserstein 2-metric connecting  $\mu_0$  and  $\mu_1$  such that

$$S_e(m_t) \geq tS_e(m_0) + (1-t)S_e(m_1) + \frac{kt(1-t)}{2} W(m_0, m_1)^2, \quad 0 \leq t \leq 1$$

This indicates the **positive correlation** of entropy and curvature that we will express as

$$DS_e \times DRic \geq 0$$

We now need to connect Ricci curvature and entropy to the notion of robustness (next slide) as well as define appropriate notions of curvature/entropy for discrete spaces (graphs).

# Curvature: Proxy for Robustness

## Recall Definition of Robustness:

- If we let  $p_e(t)$  denote the probability that the mean deviates by more than  $e$  at time  $t$  (with  $p_e(t) \rightarrow 0$  as  $t \rightarrow \infty$ ), then

$$R := \lim_{t \rightarrow \infty} \left( -\frac{1}{t} \log p_e(t) \right)$$

measures the decay rate.

## Fluctuation Theorem:

- In thermodynamics, it is well-known that entropy and rate functions from large deviations are closely related.

The Fluctuation Theorem is a realization of this fact for networks and can be expressed as:

$$DS_e \times DR \geq 0$$

This can now be further extended to be

$$DRic \times DR \geq 0.$$

- The Fluctuation Theorem has consequences for just about any type of network: **biological**, **communication**, **social**, or **neural**. In rough terms, it means that the ability of a network to maintain its functionality in the face of perturbations (internal or external), can be quantified by the correlation of activities of various elements that comprise the network.

## Network Entropy & Curvature:

- Given a Markov chain,  $\mu = (\mu_x)$ ,  $\sum_y \mu_x(y) = 1$ ,

**Network Entropy** can be defined as

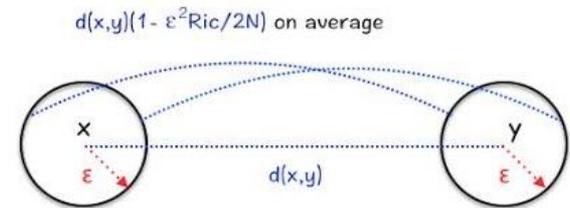
$$\bar{S}_e = \bar{a} \rho_x S_e(x) \quad S_e(x) = -\sum_y \mu_x(y) \log \mu_x(y)$$

- We now need an **appropriate definition of Ricci curvature** for a network.

# Ollivier-Ricci Curvature

## Motivation:

- We employ the notion of Ollivier-Ricci curvature motivated by adopting coarse geometric properties:
- Two very close points  $x$  and  $y$  with tangent vectors  $w$  and  $w'$ , in which  $w'$  is obtained by a parallel transport of  $w$ , the two geodesics will get closer if the curvature is positive.
- Distance between two small (geodesic balls) is less than the distance of their centers. Ricci curvature along direction  $x$ - $y$  reflects this, averaged on all directions  $w$  at  $x$ .



Pictorial Motivation for Ollivier Ricci Curvature

## Definition:

Formally, we define for  $(X,d)$  a metric space equipped with a family of probability measures  $\{\mu_x : x \in X\}$ , the *Ollivier-Ricci curvature*  $k(x,y)$  along the geodesic connecting  $x$  and  $y$  via

$$W_1(m_x, m_y) = (1 - k(x, y))d(x, y)$$

where  $W_1$  denotes the Wasserstein 1-metric defined previously and  $d(x,y)$  is the geodesic (hop) distance on a graph. For the case of weighted graphs, we set

$$d_x = \sum_y w_{xy}$$

$$m_x(y) := \frac{w_{xy}}{d_x}$$

and the sum is taken over all neighbors of  $x$  where  $w_{xy}$  denotes the weight of an edge connecting  $x$  and  $y$  (it is taken as zero if there is no connecting edge between  $x$  and  $y$ ). The measure  $\mu_x$  may be regarded as the distribution of a one-step random walk starting from  $x$ .

# Application Bottom Line (Precision Medicine)

*Targeted Therapies in the Fight Against Cancer*



Targeted Therapy: Drugs that block the growth and spread of cancer by interfering with specific molecules involved in the growth, progression, and spread of cancer

## Collaborator #1 (M.D. Anderson)

“Ewing Sarcoma needs 3-5 the targets for a particular patient to overcome drug resistance”

- Ludwig Group (Sarcoma Oncology)

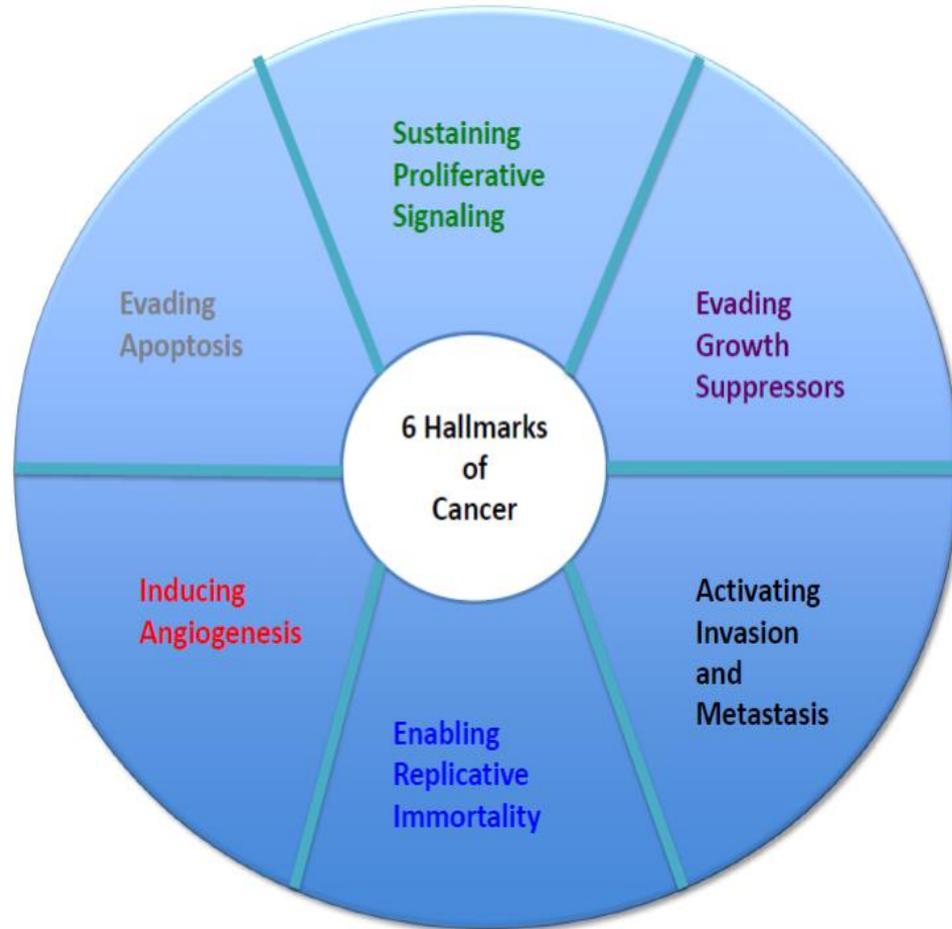
## Collaborator #2 (Sloan Kettering)

Yes, its safe to say to that drug selection is still very much a game of whack a mole”

- Baselga Group (Chief in Physician)

**Promising Yet Relatively Ineffective: Why, How To Improve?**

# Curvature: Cancer Hallmark?



Is Curvature a Cancer Hallmark?

# Drug Resistance

Drug-Sensitive



Parental



Drug-Resistant



$$R_{\text{DS}} < R_{\text{Untreated}} < R_{\text{DR}}$$

# Ewing Sarcoma: Test-Bed for Understanding Resistance

## Sarcomas have long been suspected to be connected to the immune system

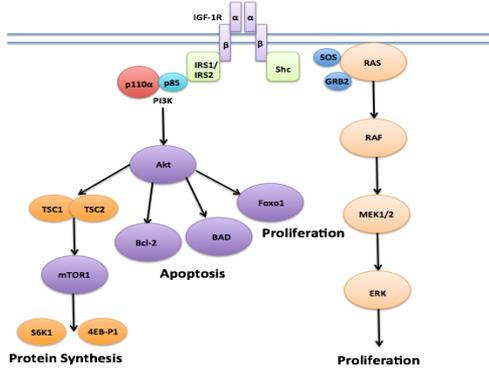


Fig 1: IGF-1R/mTOR Cascade Pathway

Motivation: Aggressive cancers offer testbed platform of understanding resistance in an adaptive setting as one can trace it back to a single aberrant molecular genetic aberration.

IGF-1R/mTOR: Recent discoveries shown that if one inhibits **both** IGF-1R/mTOR in Ewing Sarcoma (ES), the response/duration rate is **tripled** compared to single inhibition of IGF-1R.

Feedback & Robustness: On the one hand, increase in response rate is correlated to an increase in fragility while, on the other hand, the formation of IGF-1R/mTOR drug resistance ES samples points to increase in robustness (i.e., employing alternative feedback loops for continued metastasis)

**Validation:** *Given the complexity & cost of selecting drug/vaccine candidates, can we quantify (and therefore predict) pathway fragility in order to uncover a set of n-tuple targets that can disrupt modes of resistance?*

## Drug Resistant, Drug Sensitive, Untreated (Ewing Sarcoma) - w/ MD Anderson

- Obtained protein expression data from RPPA Panel for Ewing Sarcoma treated with IGF-1R Inhibitor (OSI-906/NVP-ADW 742)
- Increase in average curvature exhibited by resistance when compared to 72-hour and untreated samples

### **Initial treatment (72-Hours) gave positive response, then system adaptively built resistance to drug – How and why?**

	72-Hour	Untreated	Resistant
Average Curvature	0.0071	0.0329	0.0367
5 % Left Tail (Avg.)	-0.4619	-0.4428	-0.4123
1% Left Tail (Avg.)	-0.8535	-0.8246	-0.7822
Min Curvature	-1.9906	-1.7771	-1.4625

Table 1: Global Statistics w.r.t to Curvature/Robustness for Ewing Sarcoma



## Team, Collaborators, and Generalizable Results

- Partnered w/ MD Anderson & Memorial Sloan Kettering Cancer Center - **leaders in adaptive and immunotherapy**
- ES is our testbed for validation, but results will provide understanding key modes of resistance
- Advantages: access to valuable data, drug companies, and safety protocols as requested in RFI

# Initial Preliminary Results

Global Network Fragility via Ricci Curvature:

	72-Hour	Untreated	Resistant
Average Curvature <sup>(1)</sup>	0.0071	0.0329	0.0367
5 % Left Tail (Avg.)	-0.4619	-0.4428	-0.4123
1% Left Tail (Avg.)	-0.8535	-0.8246	-0.7822
Min Curvature	-1.9906	-1.7771	-1.4625

Notes:

- We quantify that resistant tumors are more robust 72-hour/untreated via curvature.
- The most fragile case is the 72-Hour
- This coincides with our initial hypothesis and with our previous cancer studies

Local Protein Interaction Fragility via Scalar Curvature:

	72-Hour	Untreated	Resistant
mTor	1.7404	3.5812	0.4081
MEK	1.2259	1.7511	1.8042

Notes:

- We noticed all (direct/indirect) pathways to mTor become "fragile" during resistant and 72-hour case
- MEK pathways becomes more robust in resistant case
- We caution these local results are too preliminary to draw convulsive evidence

# Concluding Remarks

- Interactive Control Methods
  - Dynamic Active Contours for Segmentation
  - Finsler Geometry
  - Bayesian Statistics (Particle Filtering)
  
- Optimal Mass Transport
  - Registration
  - Wasserstein Distance