Gravity Inversion using Curve Sampling and Geometric Conditional Simulation

Ayres Fan
John W. Fisher III
Alan S. Willsky
Stochastic Systems Group
April 3, 2007
Outline

1. Overview
2. Curve evolution
3. Markov chain Monte Carlo
4. Curve sampling
5. Conditional simulation
6. 2.5D Segmentation
Overview

• Curve evolution attempts to find a curve $C$ (or curves $C_i$) that best segment an image (according to some model)

• Goal is to find a local minimum of energy functional $E(C)$ (view as a negative log likelihood) using gradient descent

• Instead of optimization, draw multiple samples from a probability distribution $p$ defined on the space of curves
  - Naturally handles multi-modal distributions
  - Avoid local minima
  - Higher-order statistics (e.g., variances)
  - Conditional simulation
Planar curves

• A curve is a function $\vec{C} : [0, 1] \rightarrow \mathbb{R}^2$

• We wish to minimize an energy functional with a data fidelity term and regularization term:
  \[ E(\vec{C}) = D(y|\vec{C}) + R(\vec{C}) \]

• This results in a gradient flow:
  \[ \frac{d\vec{C}}{dt}(p) = \vec{F}(p) \]

• We can write any flow in terms of the normal:
  \[ \frac{d\vec{C}}{dt}(p) = f(p)\vec{N}(p) \]
Euclidean curve shortening flow

• Let $E(\vec{C}) = \int_{\vec{C}} ds$

• This energy functional is smaller when $C$ is shorter so the gradient flow is the direction that minimizes the curve length the fastest

• Using Euler-Lagrange, we see

$$\frac{d\vec{C}}{dt}(p) = -\kappa(p)\vec{N}(p)$$

where $\kappa$ is curvature, $N$ is the outward normal
Level-Set Methods

- A curve is a function (infinite dimensional)

- A natural implementation approach is to use marker points on the boundary (snakes)
  - Reinitialization issues
  - Difficulty handling topological change

- Level set methods instead evolve a surface (one dimension higher than our curve) whose zeroth level set is the curve (Sethian and Osher)
Embedding the curve

- Force level set $\Psi$ to be zero on the curve
  \[
  \Psi(\vec{C}(p)) = 0 \\
  \forall p \in [0, 1]
  \]
- Chain rule gives us
  \[
  \frac{d\Psi}{dt} = -\frac{d\vec{C}}{dt} \cdot \nabla \Psi \\
  = -f||\nabla \Psi||
  \]
Popular energy functionals

• Geodesic active contours (Caselles et al.):

\[ E(\mathcal{C}) = \int_{\mathcal{C}} \frac{ds}{1 + |\nabla I|^2} \]

• Separating the means (Yezzi et al.):

\[ E(\mathcal{C}) = (\mu_{\text{in}} - \mu_{\text{out}})^2 \]

• Piecewise constant intensities (Chan and Vese):

\[ E(\mathcal{C}) = \int_{R_0} \int (y - \mu_0)^2 dx + \int_{R_1} \int (y - \mu_1)^2 dx + \alpha \int_{\mathcal{C}} ds \]
Examples
Gravity inversion

- Supplement to standard seismic data to segment bottom salt using an array of surface gravimeters (~$10^{-15}$ N accuracy)
- Subtract base effects (geoid, centrifugal force, etc.) to leave salt effects:

$$
\bar{g}(i; \vec{C}) = G \int \int_{\Omega} \frac{\rho(x; \vec{C}) \hat{r}_i(x)}{||\hat{r}_i(x)||^2} \, dx
$$

- Assume constant density inside and outside (using Heaviside function $H$):

$$
\rho(x; \vec{C}) = \rho_0 H(-\Psi(x))
$$

- Model energy as L2 estimation error (probability as Boltzmann distribution)

$$
E(\vec{C}) = \sum_{i=1}^{N_{array}} ||\bar{g}_{obs}(i) - \bar{g}(i; \vec{C})||^2 + \alpha \int_C ds
$$
Markov Chain Monte Carlo

\[ p(\vec{C}|y; S) \propto p(y|\vec{C}; S)p(\vec{C}; S) \]

- \( C \) is a curve, \( y \) is the observed image (can be vector), \( S \) is a shape model, data model usually iid given the curve

- We wish to sample from \( p(x|y; S) \), but cannot do so directly

- Instead, iteratively sample from a proposal distribution \( q \) and keep samples according to an acceptance rule \( a \). Goal is to form a Markov chain with stationary distribution \( p \)

- Examples include Gibbs sampling, Metropolis-Hastings
Metropolis-Hastings

- Metropolis-Hastings algorithm:
  - Start with $x^0$
  - At time $t$, generate candidate $\phi^t$ (from $q$ given $x^{t-1}$)
  - Calculate Hastings ratio:
    \[
    r^t = \frac{p(\phi^t)}{p(x^{t-1})} \cdot \frac{q(x^{t-1}|\phi^t)}{q(\phi^t|x^{t-1})}
    \]
  - Set $x^t = \phi^t$ with probability $\min(1, r^t)$, otherwise $x^t = x^{t-1}$
  - Go back to 2
Asymptotic Convergence

• We want to form a Markov chain such that its stationary distribution is \( p(x) \):

\[
p(x) = \int p(\phi) T(x|\phi) \, d\phi
\]

• For asymptotic convergence, sufficient conditions are:
  1) Ergodicity
  2) Detailed balance

\[
p(x_t^{-1}) q(\phi^t|x_t^{-1}) a(\phi^t|x_t^{-1}) = p(\phi^t) q(x_t^{-1}|\phi^t) a(x_t^{-1}|\phi^t)
\]
MCMC Curve Sampling

• Generate perturbation on the curve:
\[ \vec{C}'(s) = \vec{C}(s) + f(s)\vec{N}(s)\,dt \]

• Sample by adding smooth random fields:
\[ f \sim N(-\kappa + \gamma, \Sigma) \]

• \( \Sigma \) controls the degree of smoothness in field, \( \kappa \) term is a curve smoothing term, \( \gamma \) is an inflation term

• Mean term to move average behavior towards higher-probability areas of \( \mathbf{p} \)
Synthetic noisy image

- Piecewise-constant observation model:
  
  \[ y(x) = \mu(x) + n(x) \]

- Chan-Vese energy functional:
  
  \[ E(\vec{C}) = \iint_{R_0} (y - \mu_0)^2 \, dx + \iint_{R_1} (y - \mu_1)^2 \, dx + \alpha \oint_{\vec{C}} \, ds \]

- Probability distribution (T=2\sigma^2):
  
  \[ p(\vec{C}) = \frac{1}{Z} \exp\left(-\frac{E(\vec{C})}{T}\right) \]
Prostate in a Haystack
Most likely samples
Least likely samples
Confidence intervals
When “best” is not best

• In this example, the most likely samples under the model are not the most accurate according to the underlying truth

• 10%/90% “confidence” bands do a good job of enclosing the true answer

• Histogram image tells us more uncertainty in upper-right corner

• “Median” curve is quite close to the true curve

• Optimization would result in subpar results
Observed measurements

X- and z-components of gravity profile for synthetic salt body
Circle salt
Most likely samples
Least likely samples
Confidence intervals
Notable features

• Measurement points $<<$ image pixels, but we can do a reasonable job
• Much higher uncertainty at the bottom than the top (weaker measurements)
• Less uncertainty in middle than on sides
• Median of histogram not necessarily related to median of distribution
Conditional simulation

• In many problems, the model admits many reasonable solutions
• We can use user information to reduce the number of reasonable solutions
  ▪ Regions of inclusion or exclusion
  ▪ Partial segmentations
• Curve evolution methods largely limit user input to initialization
• With conditional simulation, we are given the values on a subset of the variables. We then wish to generate sample paths that fill in the remainder of the variables (e.g., simulating pinned Brownian motion)
• Can result in an interactive segmentation algorithm
Simulating curves

• Say we are given $C_s$, a subset of $C$ (with some uncertainty associated with it)

• We wish to sample the unknown part of the curve $C_u$

• One approach is to view as sampling from:

$$p(y|\bar{C})p(\bar{C}) = p(y|\bar{C})p(\bar{C}_u|\bar{C}_s)p(\bar{C}_s)$$

• Difficulty is being able to evaluate the middle term as theoretically need to integrate $p(C)$
Simplifying Cases

For special cases, evaluation of $p(\tilde{C}_u|\tilde{C}_s)$ is tractable:

- When $C$ is low-dimensional (can do analytical integration or Monte-Carlo integration)
- When $C_s$ is assumed to be exact
- When $p(C)$ has special form (e.g., independent, Markov)
- When willing to approximate

- When implementing conditional simulation, modify $q$ to be compatible with new conditional probability
More complex example

- Same x- and z-components of gravity
- Synthetic image with more complex geometry
Initialization
Most likely samples

Regular

Conditionally simulated
Marginal confidence bounds

Regular

Conditionally simulated
Optimization-based approach

- Gradient-flow of energy functional:
\[
\frac{dE}{d\vec{C}} = \sum_{i=1}^{N_{\text{array}}} -\frac{2G\rho_0 (\vec{g}_{obs} - \vec{g}(i, \vec{C})) \cdot \vec{r}_i(x)}{r_i^2(x)} - \alpha \kappa \vec{N}
\]
Aggregating samples
Adding additional constraints

Top salt constraint
With additional constraint
Marginal confidence bounds

- Top salt constraint
- With additional constraint
Principal Components Analysis

- Have vectors $x_1, x_2, \ldots, x_m$, linear combination forms a subspace $S$ in $\mathbb{R}^n$
- Find $k << m$ vectors that best explain the data
- One view: use singular value decomposition to find a basis for $S$ (with associated variance):
  \[
  X = U\Sigma V^T
  \]
  and keep $k$ largest eigenvectors in $U$, variance in $\Sigma$
- Vectors within sub-subspace are:
  \[
  x = \sum_{i=1}^{k} a_i \sigma_i u_i
  \]
Applying PCA to Shape Spaces

- Shape space depends on representation chosen
- Generally shapes do not form a linear space (manifold)
- If shapes are close to each other, manifold is fairly flat so PCA can find an appropriate local linear approximation
- One approach is to work in the space of level set functions with appropriate smoothness (e.g., C²)
- Many level sets map to one curve, so convert curves to level sets using signed distance function as canonical representation
First principal mode of variation

Top salt constraint

With additional constraint
Second principal mode of variation

Top salt constraint

With additional constraint
Future approaches

• More general density models (piecewise smooth)
• Full 3D volumes
• Additional features can be added to the base model:
  ▪ Uncertainty on the expert segmentations
  ▪ Shape models (semi-local)
  ▪ Exclusion/inclusion regions
  ▪ Topological change (through level sets)
  ▪ Better perturbations
Conclusion

• Computationally feasible algorithm to sample from space of curves
• Approximate detailed balance
• Demonstrated utility for robustness to noise, multimodal distributions, displaying uncertainty
• Can generate arbitrary shapes with relatively complex geometry
• Conditional simulation provides a natural framework to incorporate partial user segmentations on a slice-by-slice level
Piecewise smooth densities

- Previously we assumed C mapped to a piecewise constant $\rho$
- A more general model would be to allow $\rho$ to be smooth everywhere except on the curve (similar to Mumford-Shah):

$$E(\tilde{C}) = \sum_{i=1}^{N_{\text{array}}} \| \tilde{g}_{\text{obs}}(i) - \tilde{g}(i; \rho) \|^2 + \beta \int_{\Omega - \tilde{C}} \| \nabla \rho \|^2 \, dx + \alpha \int_\tilde{C} ds$$

- Usually solve using coordinate descent/EM approach
2.5D Approach

• In 3D world, natural (or built-in) partition of volumes into slices

• Assume Markov relationship among slices

• Then have local potentials (e.g., PCA) and edge potentials (coupling between slices)

• Naturally lends itself to local Metropolis-Hastings approach (iterating over the slices)
2.5D Model

- We can model this as a simple chain structure with pairwise interactions

- This admits the following factorization:

\[
p(Y|\vec{C}) = \prod_{i=1}^{n} p(y_i|\vec{c}_i)
\]

\[
p(\vec{C}) = \prod_{i=1}^{n} \psi_i(\vec{c}_i) \prod_{i=1}^{n-1} \psi_{i,i+1}(\vec{c}_i, \vec{c}_{i+1})
\]
Partial segmentations

- Assume that we are given segmentations of every other slice
- We now want to sample surfaces conditioned on the fact that certain slices are fixed
- Markovianity tells us that $c_2$ and $c_4$ are independent conditioned on $c_3$
Incorporating perpendicular slices

• $c_\perp$ is now coupled to all of the horizontal slices

• $c_\perp$ only gives information on a subset of each slice

• Specify edge potentials as, e.g.:

$$\psi_{i,\perp}(\vec{c}_i, \vec{c}_\perp) = \exp\left( - \int_0^{L_y} (\Gamma(x_0, y, i\triangle z) - \Gamma_\perp(x_0, y, i\triangle z))^2 dy \right)$$
Coverage/Detailed balance

• It is easy to show we can go from any curve $C_1$ to any other curve $C_2$ (shrink to a point)

• For detailed balance, we need to compute probability of generating $C'$ from $C$ (and vice versa)

\[
\vec{C}'(s) = \vec{C}(s) + f(s)\vec{N}(s)dt
\]

\[
\vec{C}(s) = \vec{C}'(s) + f'(s)\vec{N}'(s)dt
\]

• Probability of going from $C$ to $C'$ is the probability of generating $f$ (which is Gaussian) and the reverse is the probability of $f'$ (also Gaussian)
Approximations to $q$

- Relationship between $f$ and $f'$ complicated due to the fact that the normal function changes

- $f'$ does not always exist (given an $f$). Unknown what conditions on $f$ are necessary to guarantee existence.

- Various levels of exactness
  - Assume $\hat{N} = \hat{N}'$ (then $f' = -f$)
  - Locally-linear approximation
    $$f'(s) = -f(s) / \langle \hat{N}(s), \hat{N}'(s) \rangle$$
  - Trace along $\hat{N}$ (technical issues)

- Unknown how approximations affect convergence
Bias-corrected prostate

- “Expert” segmentation, add noise (simulate body coil image)
Learn probability densities

- Use histograms
- Learn pdf inside $p(y|1)$ and pdf outside $p(y|0)$ and assume iid given curve:

$$E(\hat{C}) = -\int_{\Omega} \log p(y(x)|\mathcal{H}(-\Psi(x))) \, dx + \alpha \oint_{\hat{C}} ds$$
Initialization
Results
Prostate-only cluster
Multimodality and convergence

- Natural multimodal distribution
- When starting near one mode, need a lot of time to traverse valley between modes
- Clustering helps with presenting results
- Interesting work to be done in learning dimensionality of manifold and local approximations