Optimization by Gaussian Smoothing
with Application to Geometric Alignment

PhD Final Exam
Hossein Mobahi
Advised by Prof. Yi Ma
Department of Computer Science
University of Illinois at Urbana-Champaign

Hossein Mobahi (UIUC)
Outline

- Contributions
- Optimization & Homotopy
- Smoothing
- Asymptotic Analysis
- Transformation Kernels
- Application to 2D Alignment
- Application to 3D Alignment
- Conclusion
A Formal Framework for Asymptotic Analysis
Easy to Check Condition for Asymptotic Convexity
Simple Form for Asymptotic Minimizer
Closed Form Kernels for Efficiently Smoothing Alignment
Objective
Formulation & Evaluation of 2D Alignment by Smoothing Method
Formulation & Evaluation of 3D Point Cloud Alignment by Smoothing Method
Nonconvex optimization difficult in general
Pressure to approximate by convex models
Real world problems have regularity... may lead to tractable solutions
Exploiting such structures: SOS, DC, smoothness, etc.
Smooth deformation of an easy problem into the actual problem, while tracing the solution

- Consider solving \( f : \mathcal{X} \to \mathcal{Y} \)
- Embed \( f(x) \) into a parameterized family \( g(x, t) \), where \( g : \mathcal{X} \times \mathcal{T} \to \mathcal{Y} \) and \( \mathcal{T} = \{ t \mid 0 \leq t \leq 1 \} \)
  - \( g(x, 0) \) should be easy to solve
  - \( g(x, 1) = f(x) \)
Homotopy Continuation

\[ f(x) \triangleq \begin{cases} \end{cases} \]

\[ k_\sigma(x) \triangleq \frac{1}{(\sqrt{2\pi\sigma})^{\text{dim}(x)}} e^{-\frac{||x||^2}{2\sigma^2}} \]

\[ g(x, t) \triangleq tf(x) + (1 - t)(x^2 + y^2) \]

\[ g(x, t) \triangleq [f \ast k_{\frac{1}{t-1}}](x) \]

Hossein Mobahi (UIUC)  Optimization by Smoothing for Alignment  November 5th, 2012  6 / 45
Where Does Smoothing Come From?

Describing Smoothing by PDEs

- Evolution of a function in a region $\mathcal{X}$ over time.
- Initial condition $g(x, \sigma \to 0) = f(x)$.
- Boundary condition on $\mathcal{X}$ if any

Heat Equation

$$\frac{\partial}{\partial \sigma} g = \sigma \Delta g$$

Schrodinger’s Equation

$$\frac{\partial}{\partial \sigma} g = i\sigma \Delta g$$
Where Does Smoothing Come From?

Why Heat Kernel for Optimization?

If $\mathcal{X} = \mathbb{R}^n$ and $k(x; \sigma^2)$ is isotropic Gaussian, solution of heat equation is:

$$g(x; \sigma) = [f \ast k(\cdot; \sigma^2)](x).$$

Kills high frequencies, hence suppresses brittle local minima.
Consider $f(x) = e^{-\frac{t^2}{2\epsilon^2}} - e^{-\frac{t^2\epsilon^2}{2}}$ for $\epsilon > 0$

This function resembles the $\ell_0$ norm much better than $\ell_1$. Except at its tip, it is concave everywhere. However, it is asymptotically convex!
Known As

- Graduated Optimization (Computer Vision) [Blake & Zisserman 87]
- Optimization by Diffusion Equation (Chemistry) [Piela 89]
- Optimization by Homotopy Continuation (Numerical Computing) [Watson 88]
- Deterministic Annealing (Machine Learning) [Rose 98]

Despite its long age and mathematical roots, there is little understanding about its fundamental aspects.
A real-valued continuous function \( f(x) \) is called asymptotically convex if the following statement holds:

\[
\forall M > 0, \ \exists \sigma^*(M), \ \forall x_1 \in B(0, M), \ x_2 \in B(0, M), \ a \in [0, 1] \\
\sigma \geq \sigma^*(M) \Rightarrow g(ax_1 + (1-a)x_2; \sigma) \leq ag(x_1; \sigma) + (1-a)g(x_2; \sigma)
\]
Asymptotic Convexity Example

For $f(x) =$ we have $M = 1, \sigma^* \approx 0.9$
Theoretical Contributions

Asymptotic Convexity
Under mild conditions, any function \( f(x) \) satisfying
\[-\infty < \int_{\mathbb{R}^n} f(x) \, dx < 0 \]
is asymptotically convex.

Asymptotic Minimizer
Under mild conditions, any function \( f(x) \) that is asymptotically convex with
\( \int_{\mathbb{R}^n} f(x) \neq 0 \) has the asymptotic minimizer at the center of mass:
\[ x^* = \frac{\int_{\mathbb{R}^n} x f(x) \, dx}{\int_{\mathbb{R}^n} f(x) \, dx} \]

Nice Property
Both conditions are “derivative-free”. 
Show that \( f(x) = e^{-\frac{(x-1)^2}{0.1}} - e^{-x^2} \) is asymptotically convex and find its asymptotic minimizer.

\[
\int_{\mathbb{R}} f(x) \, dx \approx -1.21195, \text{ thus } f(x) \text{ is asymptotically convex. Also } \\
x^* = \frac{\int_{\mathbb{R}} xf(x) \, dx}{\int_{\mathbb{R}} f(x) \, dx} \approx -0.46247.
\]
Taxonomy for Functions $\{f : \mathbb{R}^n \rightarrow \mathbb{R}\}$
Applications to Image Alignment

Alignment as optimization

$$\theta^* = \arg \min_\theta \int_X \left( f_1(\tau(x; \theta)) - f_2(x) \right)^2 dx$$

This is non-convex in variable $\theta$. 
Linearization

Linearize $f_1(x + d)$ around $d = 0$ to get a convex quadratic.

$$\hat{d} = \arg \min_d \int_X \left( f_1(x) + d^T \nabla f_1(x) - f_2(x) \right)^2 dx$$

$$|\hat{f}_1(x) - f_1(x)| \leq \frac{\Lambda}{2} \|d\|^2$$
Problem of Lucas-Kanade for Other Models

Consider scaling transformation $\boldsymbol{x} \rightarrow s\boldsymbol{x}$. Linearization of $f_1(s\boldsymbol{x})$ around $s = 1$:

\[
\hat{f}_1(s\boldsymbol{x}) = f_1(\boldsymbol{x}) + (s - 1)(\boldsymbol{x}^T \nabla f_1(\boldsymbol{x}))
\]

\[
\hat{s} = \arg \min_s \int_\mathcal{X} \left( f_1(\boldsymbol{x}) + (s - 1)(\boldsymbol{x}^T \nabla f_1(\boldsymbol{x})) - f_2(\boldsymbol{x}) \right)^2 d\boldsymbol{x}
\]

\[
| \hat{f}_1(s\boldsymbol{x}) - f_1(s\boldsymbol{x}) | \leq \frac{\Lambda(s - 1)^2}{2} \| \boldsymbol{x} \|^2
\]

Error grows in $\| \boldsymbol{x} \|^2$ as well!
Smoothing the objective automatically produces spatially varying blurs without any hack.

People in computer vision have realized the advantage of spatially varying kernels for matching on heuristic basis, e.g. (Berg & Malik 2001)
Smoothing the Objective for Alignment

Toy Example: 1-d Scale Alignment

- **Actual Task** \( \int_{\mathbb{R}} (f_1(ax) - f_2(x))^2 \mathbb{I}_{\|x\| \leq 1} \, dx \)

- **Signal Smoothing** \( \int_{\mathcal{X}} \left( \mathbb{I}_{\|x\| \leq 1} \left( [f_1 \ast k_{\sigma}](ax) - [f_2 \ast k_{\sigma}](x) \right) \right)^2 \, dx \)

- **Objective Smoothing** \( k_{\sigma} \overset{\Theta}{\ast} \int_{\mathcal{X}} \left( \mathbb{I}_{\|x\| \leq 1} \left( f_1(\cdot \times x) - f_2(x) \right) \right)^2 \, dx \)(a)
Efficient Computation of Smoothed Objective

Question

- Assume $\mathcal{X} = \mathbb{R}^n$ and $\Theta = \mathbb{R}^m$
- Given a domain transformation $\tau: \mathcal{X} \times \Theta \rightarrow \mathcal{X}$.

Is there any $u_{\tau,\sigma}: \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$ satisfying the following integral equation?

$$\forall f : [f(\tau(x, \cdot)) \ast k(\cdot; \sigma^2)](\theta) = \int_{\mathcal{X}} f(y) u_{\tau,\sigma}(\theta, x, y) dy$$

Applied Contributions

We do find such $u$ and call it a transformation kernel.
Kernels for Common Transformations

<table>
<thead>
<tr>
<th>Name</th>
<th>$\tau(x, \theta)$</th>
<th>$u_{\tau,\sigma}(\theta, x, y)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Translation</td>
<td>$x + d$</td>
<td>$k(\tau(x, \theta) - y; \sigma^2)$</td>
</tr>
<tr>
<td>Translation+Scale</td>
<td>$a^T x + d$</td>
<td>$K(\tau(x, \theta) - y; \sigma^2 \text{diag}([1 + x_i^2]))$</td>
</tr>
<tr>
<td>Affine</td>
<td>$Ax + b$</td>
<td>$k(\tau(x, \theta) - y; \sigma^2 (1 + |x|^2))$</td>
</tr>
<tr>
<td>Homography</td>
<td>$\frac{1}{1+c^T x}(Ax + b)$</td>
<td>$n = 2 : p(\theta, x, y, \sigma) e(\theta, x, y, \sigma)$</td>
</tr>
</tbody>
</table>
Correctness

<table>
<thead>
<tr>
<th>Name</th>
<th>( \tau(x, \theta) )</th>
<th>( u_{\tau, \sigma}(\theta, x, y) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Translation</td>
<td>( x + d )</td>
<td>( k(\tau(x, \theta) - y; \sigma^2) )</td>
</tr>
<tr>
<td>Translation+Scale</td>
<td>( a^T x + d )</td>
<td>( K(\tau(x, \theta) - y; \sigma^2 \text{ diag}([1 + x_i^2])) )</td>
</tr>
<tr>
<td>Affine</td>
<td>( Ax + b )</td>
<td>( k(\tau(x, \theta) - y; \sigma^2 (1 + |x|^2)) )</td>
</tr>
<tr>
<td>Homography</td>
<td>( \frac{1}{1 + c^T x}(Ax + b) )</td>
<td>( n = 2 : p(\theta, x, y, \sigma) e(\theta, x, y, \sigma) )</td>
</tr>
</tbody>
</table>

Heat Equation

\[
\sigma \Delta_\theta [f(\tau(x, \cdot)) \ast k(\cdot; \sigma)](\theta) = (\partial / \partial \sigma) [f(\tau(x, \cdot)) \ast k(\cdot; \sigma)](\theta)
\]

Initial Condition

\[
\lim_{\sigma \to 0^+} \int_{\mathcal{X}} f(y) u_{\tau, \sigma}(\theta, x, y) \, dy = f(\tau(x, \theta))
\]
### Comparison against Geometric Blur

<table>
<thead>
<tr>
<th>Name</th>
<th>$\tau(x, \theta)$</th>
<th>$u_{\tau,\sigma}(\theta, x, y)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Translation</td>
<td>$x + d$</td>
<td>$k(\tau(x, \theta) - y; \sigma^2)$</td>
</tr>
<tr>
<td>Translation+Scale</td>
<td>$a^T x + d$</td>
<td>$K(\tau(x, \theta) - y; \sigma^2 \text{diag}(1 + x^2_i))$</td>
</tr>
<tr>
<td>Affine</td>
<td>$Ax + b$</td>
<td>$k(\tau(x, \theta) - y; \sigma^2 (1 + |x|^2))$</td>
</tr>
<tr>
<td>Homography</td>
<td>$\frac{1}{1+c^T x} (Ax + b)$</td>
<td>$n = 2 : p(\theta, x, y, \sigma) e(\theta, x, y, \sigma)$</td>
</tr>
</tbody>
</table>

### Geometric Blur Kernel (Berg & Malik 2001)

$$u_\sigma(x, y) = k(y - x; \sigma^2 \|x\|^2)$$

- Heuristic
- Identity Transformation
- Singular at Origin
Alignment By Smoothing Algorithm

Original Objective

\[ h(\theta) \triangleq \int_X f_1(\tau(x, \theta)) f_2(x) \, dx \]

Smoothed Objective

\[ z(\theta, \sigma) = \int_X \left( f_2(x) \left[ f_1(\tau(x, \cdot)) \ast k(\cdot, \sigma^2) \right](\theta) \right) \, dx \]

\[ = \int_X \left( f_2(x) \left( \int_X f_1(y) u_{\tau, \sigma}(\theta, x, y) \, dy \right) \right) \, dx \]

Blurred \( f_1 \)

Inner Product
Computation with Homography Kernel

Approximation

- Model image as piecewise constant so that,

$$\int_{\mathcal{X}} f_1(y) u_{\tau,\sigma}(\theta, x, y) \, dy = \sum_{i=1}^{W} \sum_{j=1}^{H} F_1(i, j) \int_{y_i}^{y_i+y_j} \int_{y_j}^{y_j+y_i} u_{\tau,\sigma}(\theta, x, y) \, dy.$$  

- Now computing the integral transform amounts to computing

$$\int_{\mathcal{X}_{ij}} u_{\tau,\sigma}(\theta, x, y) \, dy.$$  

- We use Laplace approximation to compute integral above.
Qualitative 2D Alignment Results

A 3D Reconstruction Scenario

- In reality, a scene barely consists of a **single planar** surface
- A real example: 3D reconstruction of an octagonal building
  - Given eight uncalibrated and widely separated images
  - Each image covers a pair of adjacent facades
  - Segment an image into piecewise planar regions
  - Rectify a segment by a single homography
  - **Dense pixel-wise match** across pairs of facades
  - Bundle adjustment
Qualitative 2D Alignment Results

A 3D Reconstruction Scenario

- Each segmented region may not share the same location and scale across images
- Refine their location and scale by alignment in scale+displacement space
- Use smoothed objective of normalized cross correlation (NCC)
Qualitative 2D Alignment Results

A 3D Reconstruction Scenario
\[
(\theta^*, c^*) = \text{arg min}_{\theta, c} \sum_{i=1}^{m} \sum_{j=1}^{n} c_{i,j} \| \tau(p_i, \theta) - q_j \|_2^2
\]

s.t. \quad \forall j \in \{1, \ldots, n\} \quad \sum_{i=1}^{m} c_{i,j} = 1

\forall i \in \{1, \ldots, m\} \quad \forall j \in \{1, \ldots, n\} \quad c_{i,j} \in \{0, 1\}
(\theta^*, c^*) = \arg \min_{\theta, c} \sum_{i=1}^{m} \sum_{j=1}^{n} c_{i,j} \| \tau(p_i, \theta) - q_j \|^2 \\
\text{s.t.} \quad \forall j \in \{1, \ldots, n\} \quad \sum_{i=1}^{m} c_{i,j} = 1 \\
\forall i \in \{1, \ldots, m\} \forall j \in \{1, \ldots, n\} \quad c_{i,j}(1 - c_{i,j}) = 0
Approximate by Quadratic Penalty

\[
(\hat{\theta}, \hat{c}) = \arg \min_{\theta, c} h(\theta, \{c_{i,j}\})
\]

\[
h(\theta, c) \triangleq \epsilon \left( \sum_{i=1}^{m} \sum_{j=1}^{n} c_{i,j} \| \tau(p_i, \theta) - q_j \|^2 \right)
\]

\[
= \sum_{j=1}^{n} \left( 1 - \sum_{i=1}^{m} c_{i,j} \right)^2 + \sum_{i=1}^{m} \sum_{j=1}^{n} c_{i,j}^2 (1 - c_{i,j})^2
\]

\(\epsilon\) is a small positive number.
3D Point Cloud Alignment

Smoothed Objective

\[ z(\theta, c; \sigma) \triangleq \left[ \left( \left[ h(\cdot, \cdot) \ast k(\cdot; \sigma^2) \right](c) \right) \ast k(\cdot; \sigma^2) \right](\theta) \]

\[ = \epsilon \sum_{i=1}^{m} \sum_{j=1}^{n} c_{i,j} \left( \| \tau(p_i, \theta) - q_j \|^2 + 3\sigma^2 (1 + \| p_i \|^2) \right) \]

\[ + \sum_{j=1}^{n} \left( 1 - \sum_{i=1}^{m} c_{i,j} \right)^2 \]

\[ + \sum_{i=1}^{m} \sum_{j=1}^{n} (c_{i,j} - 1)^2 c_{i,j}^2 + 6\sigma^2 (c_{i,j} - \frac{1}{2})^2 \]

We assume \( \tau \) is an affine transformation, i.e. \( \tau(p; (A, b)) \triangleq Ap + b \)
The objective is asymptotically convex.

Its asymptotic minimizer is the simple form $A^* = O$, $b^* = 0$, and $c^* = \frac{1}{2}$. 
Set $d = 1$, $\epsilon = 0.01$, $P = \{p_1, p_2\}$, and $Q = \{q_1, q_2\}$, where $p_1 = q_1 = -1$ and $p_2 = q_2 = 1$.

The points are already aligned, thus fix $A$ and $b$ to their optimal values $A = I$ and $b = 0$.

Fix $c_{1,2}$ and $c_{1,2}$ to their optimal value $c^*_{1,2} = 1 - c_{1,1}$ and $c^*_{2,2} = 1 - c_{2,1}$.

This leaves us with only two variables $c_{1,1}$ and $c_{2,1}$, whose optimal solution must be $c^*_{1,1} = 1$ and $c^*_{2,1} = 0$. 
At the non-smoothed function ($\sigma = 0$), global minimum is near $(c_{1,1}, c_{2,1}) = (1, 0)$ and three local minima near $(0, 1), (0, 0)$ and $(1, 1)$.

For large enough $\sigma$, the landscape becomes convex, with minimizer around $(c_{1,1}, c_{2,1}) = (\frac{1}{2}, \frac{1}{2})$, as anticipated by the asymptotic minimizer result.

The path of minimizer form large to small $\sigma$ converges to the global minimum.
3D Point Cloud Alignment

Qualitative Results

- Using point cloud data of some objects in Stanford 3D.
- Comparison against Iterative Closest Point (ICP) algorithm; that is alternation between:
  - Given transformation, establish correspondence between pair of points (of the two clouds)
  - Given correspondence, optimization the alignment transformation between the point clouds.
Qualitative Results

- Top: Input $P$, which is a rotated version of $Q$.
- Middle: Transformed $P$ to match $Q$ using ICP.
- Bottom Row: Transformed $P$ to match $Q$ using proposed method.
3D Point Cloud Alignment

Qualitative Results
- Top: Input $P$, which is a rotated version of $Q$.
- Middle: Transformed $P$ to match $Q$ using ICP.
- Bottom Row: Transformed $P$ to match $Q$ using proposed method.
3D Point Cloud Alignment

Qualitative Results

- Top: Input $\mathcal{P}$, which is a rotated version of $\mathcal{Q}$.
- Middle: Transformed $\mathcal{P}$ to match $\mathcal{Q}$ using ICP.
- Bottom Row: Transformed $\mathcal{P}$ to match $\mathcal{Q}$ using proposed method.
Conclusion & Future Directions

Optimization by Smoothing

Contributions
- Rigorous definitions for asymptotic convexity.
- Derivative free test for asymptotic convexity.
- Derivative free form for asymptotic minimizer.

Future Directions
- Asymptotic analysis of non-Gaussian smoothing.
- Conditions that guarantee a traceable path.
- Conditions that guarantee reaching the global minimum.
**Conclusion & Future Directions**

**Alignment**

- **Contributions**
  - Showing that traditional Gaussian image blurring (e.g. Lucas-Kanade) is not suitable for non-displacement motions.
  - Derivation of spatially varying kernels required for objective smoothing.
  - Formulation of 2D and 3D alignment via objective smoothing.

- **Future Directions**
  - Exploring potential connections between our kernels and blur kernels for deblurring or motion from blur.
  - Heuristic spatially-varying kernels showed success in object detection & recognition. Our kernels may provide a principled framework for developing such kernels.
  - Exploiting smoothness and localized form of our kernels to compute integral transforms faster.
Thank you!