Interpolation on the manifold of K component GMMs

Hyunwoo J. Kim[1], Nagesh Adluru[1], Monami Banerjee[2], Baba C. Vemuri[2], Vikas Singh[1]

http://pages.cs.wisc.edu/~hwkim/projects/k-gmm

[1]University of Wisconsin-Madison [2]University of Florida

The motivating problem

- Interpolation (weighted mean) in the K-GMM space.
- Applications: diffusion weighted imaging, statistical machine learning.

What's new?

- **This paper : interpolation of k-GMMs in the K-GMM space w.r.t ℓ_2-distance and KL-divergence by numerically efficient schemes (GD, EM, and Modified-EM)**
- Connections of our EM algorithm with functional spaces of Gaussians and the mixture of Gaussian processes.

Where is this needed?

- Compressing a K-GMM to a K'-GMM ($K > K'$).
- Clustering K Gaussian distributions with K' Gaussian mean functions ($K > K'$).
- Identifying the shortest path in the K-GMM space between two K-GMMs.
- Inpainting, registration, and spatial transformation tasks which involve interpolation.
- For vector samples $X \in \mathbf{R}^d$, our approach may address (a) noisy observations with known Gaussian error, (b) ill-conditioned covariance matrices by a small number of high dimensional samples or too many clusters.

PDF interpolation with reparameterization and distance

The mean of probability densities $\{f_i\}_{i=1}^N$ is given by

$$\hat{f} = \arg\min_{f \in \text{PDF}} \sum_{i=1}^{N} w_i d(\Phi(f), \Phi(f_i))^2 \quad (1)$$

where $\Phi(\cdot)$ is a mapping function for parameterizing the given probability densities, $d(\cdot,\cdot)$ is a distance metric and w_i is a weight for f_i.

Distance metrics and dissimilarity measure

ℓ_2-distance

$$\|f_1 - f_2\|_2 = \left(\int_X |f_1(x) - f_2(x)|^2 d\mu(x)\right)^{1/2}. \quad (2)$$

where f_1 and f_2 are PDFs.

KL-divergence

$$D(f_1\|f_2) = \int f_1(x) \log\frac{f_1(x)}{f_2(x)} dx. \quad (3)$$

where f_1 and f_2 are PDFs.

ℓ_2-distance with the normalized PDFs

$$d_{n\text{-}\ell_2}(f_1, f_2)^2 = \int (f_1'(x) - f_2'(x))^2 dx = 2(1 - \int_{\mathcal{X}} f_1'(x) f_2'(x) dx). \quad (4)$$

where $f_i'(x) = f_i(x)/\|f_i(x)\|_2$.

Geodesic distance on the Hilbert sphere

$$d_{geo}(f_1, f_2) = \cos^{-1}\langle f_1', f_2'\rangle_2 = \cos^{-1}\left(\int_{\mathcal{X}} f_1'(x) f_2'(x) dx\right) \quad (5)$$

where f_1' and f_2' are normalized PDFs or square-root parameterizations of f_1 and f_2.

ℓ_2-mean of k-GMMs

The ℓ_2-mean (arithmetic mean) of $\{\mathcal{F}_n\}_{n=1}^N$ minimizes the sum of squared ℓ_2-distances to each $\mathcal{F}_i \in \mathbf{F}$,

$$\bar{\mathcal{F}} = \arg\min_{\mathcal{G}} \sum_{n=1}^{N} \|\mathcal{F}_n - \mathcal{G}\|_2^2 = \frac{\sum_{n=1}^N \mathcal{F}_n}{N} = \underbrace{\sum_{n=1}^{N}\sum_{j=1}^{K}}_{N\times K \text{ components}} \frac{\pi_n^j}{N} \mathcal{N}(\mu_n^j, \Sigma_n^j) \notin \mathbf{G}^{(K)} \quad (6)$$

where $\pi_n^j, \mu_n^j, \Sigma_n^j$ are the parameters for j_{th} component of the n_{th} GMM.

Interpolation on the manifold of K-GMMs

ℓ_2-distance

$$\mathcal{G}^* = \arg\min_{\mathcal{G} \in \mathbf{G}^{(K)}} \sum_{n}^{N} \|\mathcal{F}_n - \mathcal{G}\|_2^2 \quad (7)$$

KL-divergence

$$\mathcal{G}^* = \arg\min_{\mathcal{G} \in \mathbf{G}^{(K)}} \sum_{n=1}^{N} D(\mathcal{F}_n \| \mathcal{G}) \quad (8)$$

Lemma 1

The mean of functions $\{\mathcal{F}_n\}_{n=1}^N$ w.r.t. ℓ_2 metric is the closest $\mathcal{G}^*$ to the ℓ_2-mean $\bar{\mathcal{F}} = \sum_n^N \frac{\mathcal{F}_n}{N}$.

$$\mathcal{G}^* = \arg\min_{\mathcal{G} \in \mathbf{G}^{(K)}} \sum_{n}^{N} \|\mathcal{F}_n - \mathcal{G}\|_2^2 = \arg\min_{\mathcal{G} \in \mathbf{G}^{(K)}} \|\bar{\mathcal{F}} - \mathcal{G}\|_2^2 \quad (9)$$

KL-divergence and cross entropy minimization

$$\arg\min_{\mathcal{G} \in \mathbf{G}^{(K)}} D(\bar{\mathcal{F}}\|\mathcal{G}) = \arg\min_{\mathcal{G} \in \mathbf{G}^{(K)}} \int \bar{\mathcal{F}}(x) \log\frac{\bar{\mathcal{F}}(x)}{\mathcal{G}(x)} dx = \arg\min_{\mathcal{G} \in \mathbf{G}^{(K)}} -\int \bar{\mathcal{F}}(x) \log \mathcal{G}(x) dx. \quad (10)$$

Lemma 2

Given GMM $f(x) := \sum_i^L \pi_i f_i(x)$, where $f_i(x)$ is a Gaussian distribution, the minimum cross entropy / KL-divergence between $f(x)$ and an unknown single Gaussian $g := \mathcal{N}(x; \mu, \Sigma)$ is

$$(\mu^*, \Sigma^*) = \arg\min_{\mu, \Sigma} H(f(x), \mathcal{N}(x; \mu, \Sigma)), \quad (11)$$

where $H(\cdot,\cdot)$ is the cross entropy, $\mu^* = \mathbb{E}_{f(x)}[x]$ and $\Sigma^* = \mathbb{E}_{f(x)}[(x - \mu^*)(x - \mu^*)^T]$.

EM algorithm for minimizing cross entropy

E-step: Let $\Theta = \{w_j, \mu_j, \Sigma_j\}_{j=1}^K$, $\bar{\mathcal{F}}(x) = \sum_{i=1}^{NK} \pi_i f_i(x)$ and X_i be a set of points with density function $f_i(x)$.

$$\gamma_{ij} := p(z_i = j | X_i, \Theta) = \frac{w_j \exp\left[-H(f_i, g_j)\right]}{\sum_{j'}^K w_{j'} \exp[-H(f_i, g_{j'})]}$$

$H(f_i, g_j)$ is analytically obtained as,

$$\frac{1}{2}\{k \log 2\pi + \log|\Sigma_j| + \text{tr}[\Sigma_j^{-1}\Sigma_i] + (\mu_i - \mu_j)^T \Sigma_j^{-1}(\mu_i - \mu_j)\}$$

M-step:

$$w_j = \frac{\sum_{i=1}^{NK} \pi_i \gamma_{ij}}{\sum_{j'=1}^{K}\sum_{i'=1}^{NK} \pi_{i'}\gamma_{i'j'}}, \quad \mu_j = \mathbb{E}_{\bar{\mathcal{F}}'(x)}[x] = \sum_{i=1}^{NK} \pi_i' \mu_i$$

$$\Sigma_j = \mathbb{E}_{\bar{\mathcal{F}}'(x)}[(x - \mu_j)(x - \mu_j)^T] = \sum_{i=1}^{NK} \pi_i' \Sigma_i + \sum_{i=1}^{NK} \pi_i' (\mu_i - \mu_j)(\mu_i - \mu_j)^T \quad (12)$$

where $\bar{\mathcal{F}}' = \sum_{i=1}^{NK} \pi_i' f_i(x)$, and $\pi_i' = \frac{\pi_i \gamma_{ij}}{\sum_i \pi_i \gamma_{ij}}$, for fixed j.

Modified EM algorithm for limited GPMMs

E-step: Estimate the responsibilities of data PDFs $\{f_i\}$ to components of our model,

$$\gamma_{ij} = \frac{w_j C_j^{-1} \exp\left(-\frac{1}{2C_j^2}\|f_i - g_j\|_2^2\right)}{\sum_{k=1}^K w_k C_k^{-1} \exp\left(-\frac{1}{2C_k^2}\|f_i - g_k\|_2^2\right)} \quad (13)$$

M-step: Maximize cross entropy given assignments over model parameters (a weight w_j, mean function $\mathcal{N}(\mu_j, \Sigma_j)$ and a covariance function C_j).

$$C_j^2 = \sum_{i=1}^{NK} \gamma_{ij}\pi_i \|f_i - g_j\|_2^2 / \sum_{i'=1}^{NK} \gamma_{i'j}\pi_{i'} \quad (14)$$

w_j and μ_j, Σ_j are updated using (12).

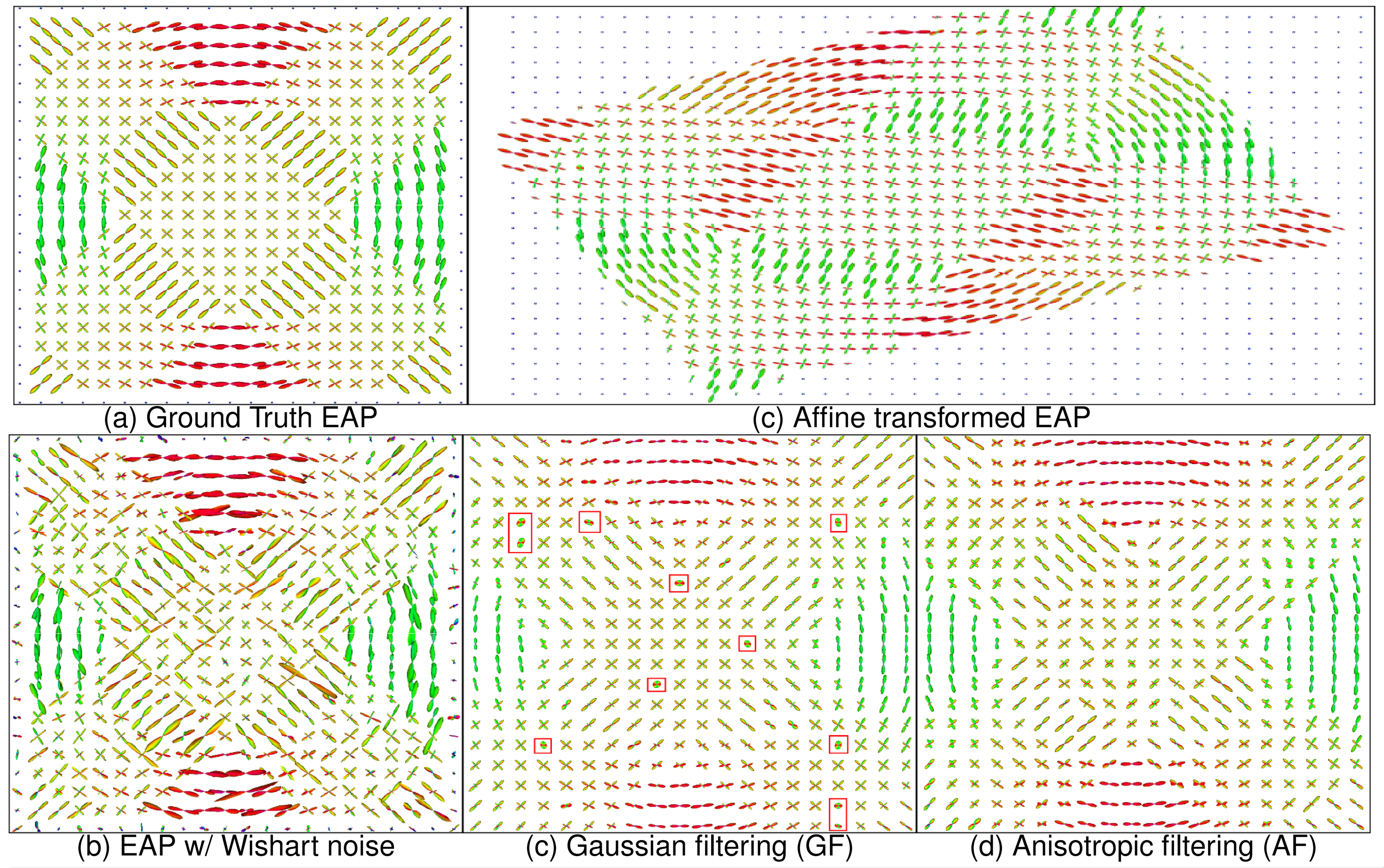
Ensemble average propagator (EAP)

Ensemble average propagator $P(R\mathbf{r})$ describes the probability of diffusion displacements of water molecules at radius R. Assuming antipodal (radial) symmetries (i.e., $E(q\mathbf{u}) = E(-q\mathbf{u})$ and $P(R\mathbf{r}) = P(-R\mathbf{r})$), the following relationship holds

$$P(R\mathbf{r}) = \int_{\mathbb{R}^3} E(q\mathbf{u}) \cos(2\pi q R \mathbf{u}^T \mathbf{r}) dq\mathbf{u}. \quad (15)$$

where $\mathbf{u}, \mathbf{r}$ are unit vectors in $\mathbb{R}^3$, q is proportional to the amplitude of the magnetic field gradient along $\mathbf{u}$. The EAP is a PDF whose domain is $\mathbb{R}^3$.

EAP spatial transformation and Denoising



(a) Ground Truth EAP (c) Affine transformed EAP

(b) EAP w/ Wishart noise (c) Gaussian filtering (GF) (d) Anisotropic filtering (AF)

Peak preserving complexity reduction

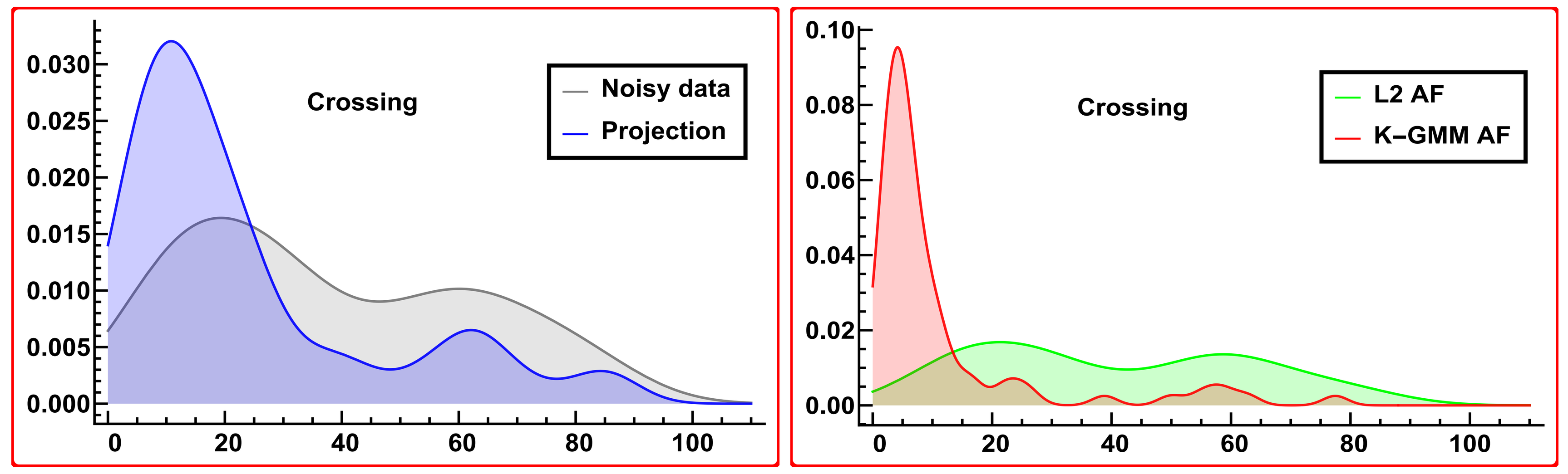


Figure: The distributions of angular deviations of the peaks.(Left) Comparison bet/ projected and noisy data. (Right) Comparison bet/ AF with K-GMM (ours) and ℓ_2 interpolation.

Interpolation path on K-GMM space w.r.t ℓ_2 distance

The objective function to identify the path is given by

$$\min_{\{\mathcal{G}_t \in \mathbf{G}^{(K)}\}_{t=1}^T} \sum_{t=0}^{T} \|\mathcal{G}_t - \mathcal{G}_{t+1}\|_2^2$$

where $\mathcal{G}_0$ ($\mathcal{G}_{T+1}$ resp.) is $\mathcal{G}_{start}$ ($\mathcal{G}_{end}$ resp.).

The approximate path length is given as

$$\lim_{T\to\infty} \sum_{t=0}^{T} \|\mathcal{G}_t - \mathcal{G}_{t+1}\|_2 = d$$

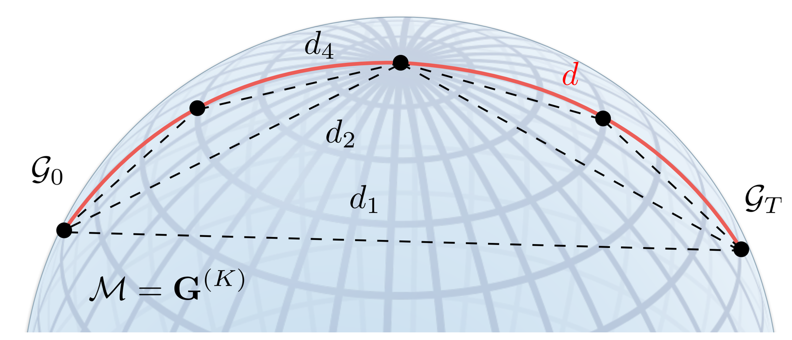


Figure: ℓ_2 distance between $\mathcal{G}_0$ and $\mathcal{G}_T$ is d_1. The approximate geodesic distance between two GMMs by the sum of ℓ_2 distances between many intermediate GMMs. $d_1 \le d_2 \le \ldots \le d$ and converge to the true distance d.

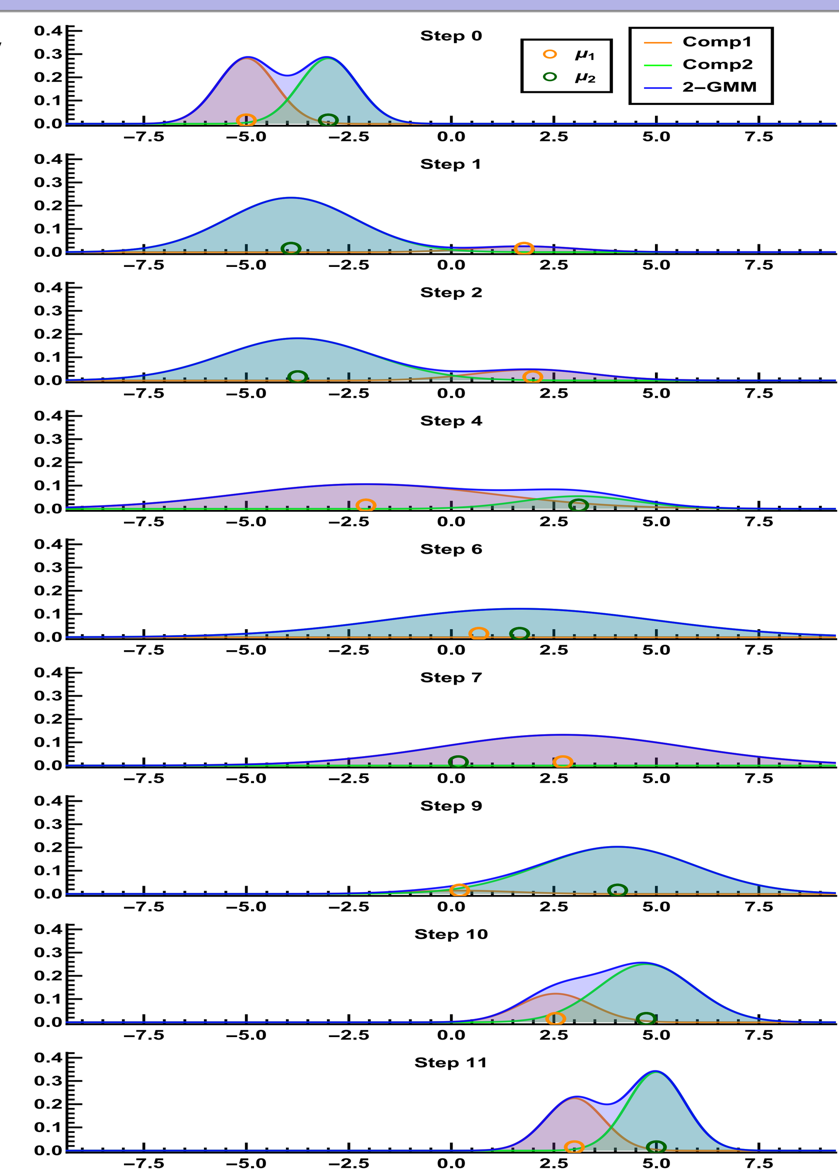


Figure: Interpolation path along 2-GMM manifold from top (GMM0) to bottom (GMM11).

International Conference on Computer Vision (ICCV) 2015

Research supported in part by NIH R01 AG040396, NS066340, NSF CAREER award 1252725,1UL1RR025011, P30 HD003352-45, UW CPCP NIH AI117924.