Manifold-valued Dirichlet Processes

Hyunwoo J. Kim[1], Jia Xu[1], Baba C. Vemuri[2], Vikas Singh[1]

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

http://pages.cs.wisc.edu/~hwkim/projects/dp-mglm

The motivating problem

- **Non-parametric Bayesian** construction relating multivariate covariates to "manifold-valued" responses.
- Extension of locally-defined models on Riemannian manifolds to "global" models.

What's new?

- DPMM for manifold-valued data e.g., sphere [Straub et al., AISTAT 2015], SPD [Cherian et al. CVPR 2011]
- Inspired by DP-GLM in Euclidean space [Hannah et al., JMLR 2011] and Multivariate general linear model (MGLM) on manifolds [Kim et al. CVPR 2014]
- **This paper : DP-MGLM for Riemannian manifolds (SPD(n)) w.r.t. the GL-invariant metric**
- **Main technical highlight:** New Hamiltonian Monte Carlo (HMC) algorithm for parameters on manifolds and tangent spaces $(B, V) \in \mathcal{M} \times T_B\mathcal{M}^d$.
- New distribution proposed for $(B, V) \in \mathcal{M} \times T_B\mathcal{M}^d$.
- A direct by-product is "clustering based on nonlinear correlation" between covariates and response.

Euclidean and Riemannian MGLMs: Basic Operation

Operation	Subtraction	Addition	Distance	Mean	Covariance
Euclidean	$\overrightarrow{x_ix_j} = x_j - x_i$	$x_i + \overrightarrow{x_jx_k}$	$\lVert\overrightarrow{x_ix_j}\rVert$	$\sum_{i=1}^n \overrightarrow{\bar{x}x_i} = 0$	$\mathbb{E}\left[(x_i-\bar{x})(x_i-\bar{x})^T\right]$
Riemannian	$\overrightarrow{x_ix_j} = \mathrm{Log}(x_i, x_j)$	$\mathrm{Exp}(x_i, \overrightarrow{x_jx_k})$	$\lVert\mathrm{Log}(x_i, x_j)\rVert_{x_i}$	$\sum_{i=1}^n \mathrm{Log}(\bar{x}, x_i) = 0$	$\mathbb{E}\left[\mathrm{Log}(\bar{x}, x_i)\mathrm{Log}(\bar{x}, x_i)^T\right]$

Euclidean GLM

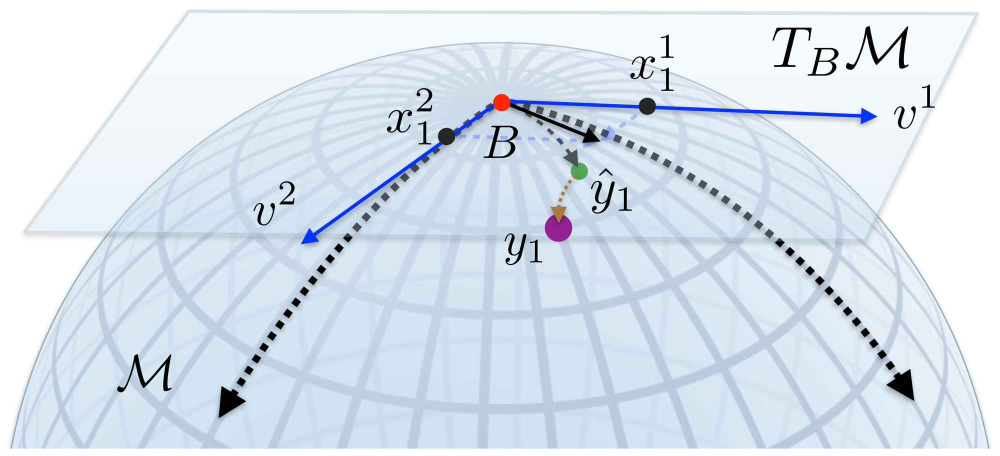
$$y = \alpha + \beta^1 x^1 + \beta^2 x^2 + \ldots + \beta^d x^d + \epsilon,$$

where $x \in \mathbf{R}^d$, $y \in \mathbf{R}^{d'}$ and $\beta \in \mathbf{R}^{d'}$.

Riemannian MGLM

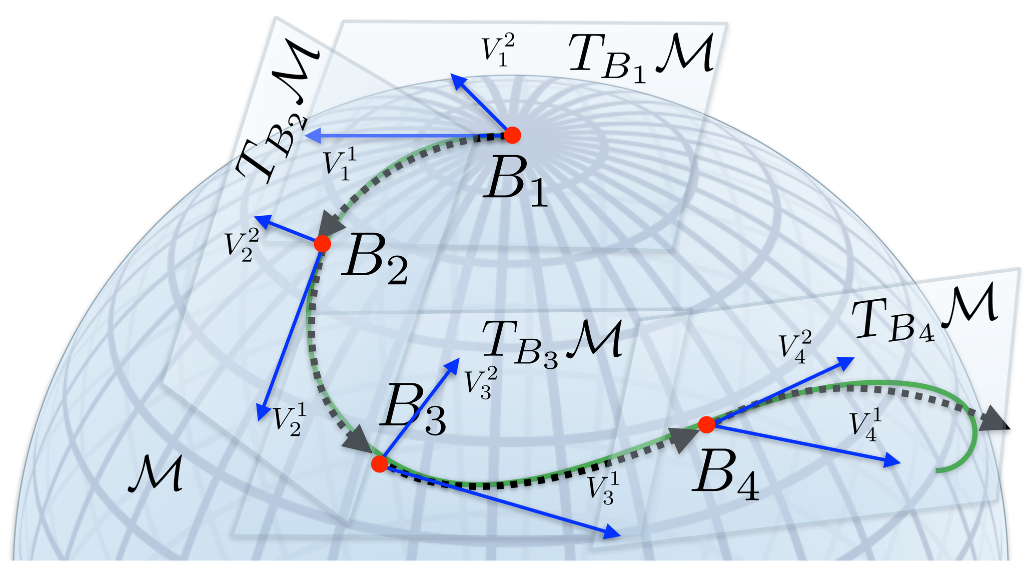
$$y = \mathrm{Exp}\left(\mathrm{Exp}\left(B, V^1x^1 + V^2x^2 + \ldots + V^dx^d\right), \epsilon\right),$$

where $x \in \mathbf{R}^d$, $y \in \mathcal{M}$, $B \in \mathcal{M}$, and $V^j \in T_B\mathcal{M}$.



DP-MGLM on manifolds

$$y = \mathrm{Exp}\left(\mathrm{Exp}\left(B_i, V_i^1x_i^1 + V_i^2x_i^2 + \ldots + V_i^dx_i^d\right), \epsilon\right) \quad (1)$$



DP-MGLM on manifolds

$$G \sim DP(G_0, \nu)$$
$$\theta_i = (\theta_{x_i}, \theta_{y_i}) | G \sim G$$
$$x_i | \theta_{x_i} \sim f_x(\theta_{x_i})$$
$$y_i | x_i, \theta_{y_i} \sim MGLM(x_i, \theta_{y_i})$$

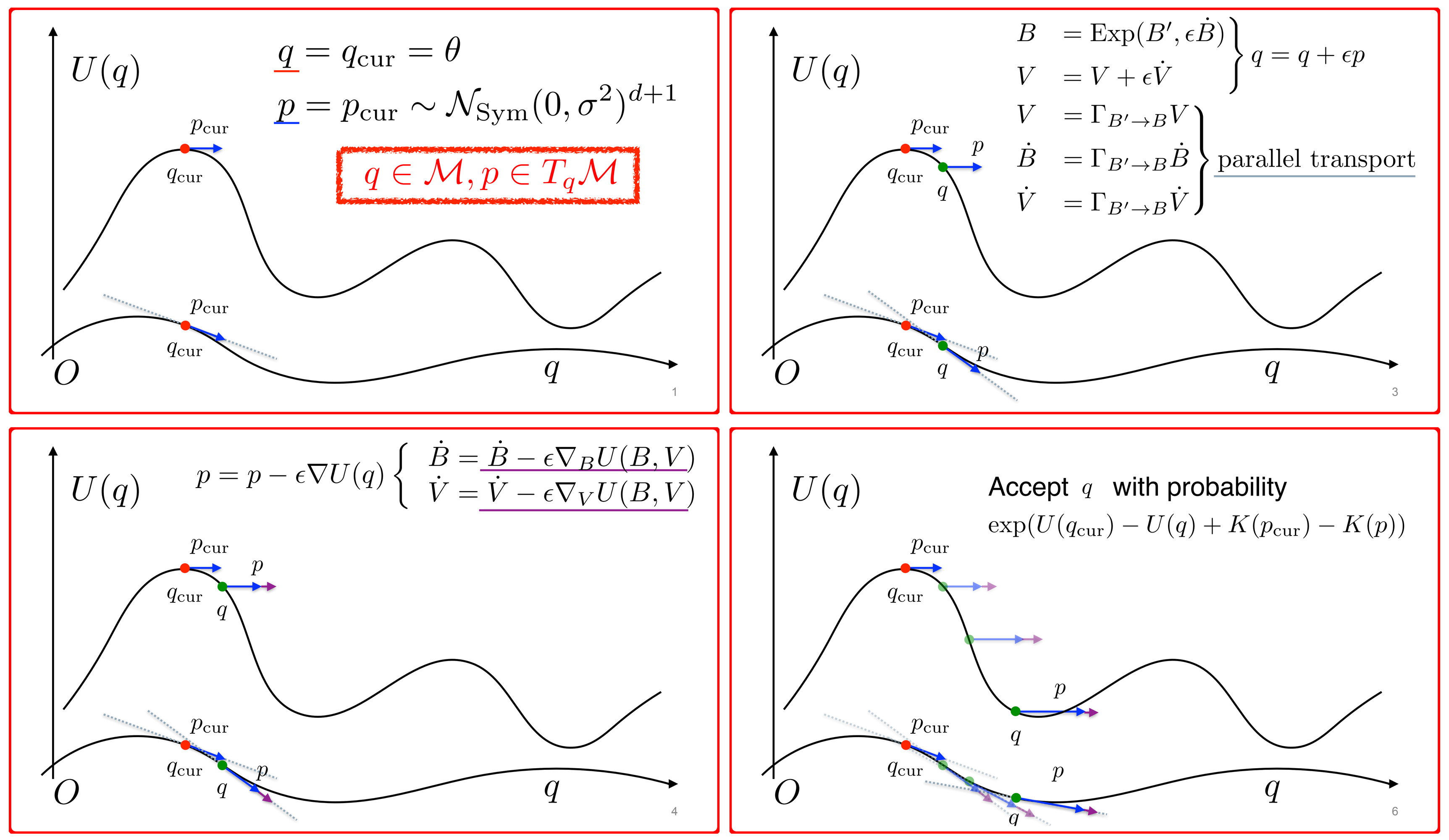
$$G \sim DP(G_0, \nu)$$
$$\theta_i := (\underbrace{\mu_x, \sigma_x^2}_{\theta_{x_i}}, \underbrace{B, \mathbf{V}}_{\theta_{y_i}}) | G \sim G,$$
$$x_i | \theta_{x_i} \sim \mathcal{N}(\theta_{x_i}, \sigma_{x_i}^2),$$
$$y_i | x_i, \theta_{y_i}, \sigma_{y_i}^2 \sim \mathcal{N}_{SPD}(\hat{y}_i, \sigma_{y_i}^2),$$
$$\text{where } \hat{y}_i = \mathrm{Exp}(B_i, \mathbf{V}_i x),$$

Hamiltonian Monte Carlo Sampling (HMC)

$$H(q,p) = U(q) + K(p)$$
$$U(q) := -\log f(q) \text{ and } K(p) := \frac{1}{2}p^T M^{-1} p$$

where q is parameters to sample, p is momentum vector, $M \succ 0, M^T = M$ is "mass matrix" [Duane et al., "Hybrid monte carlo.", Physics letters B, 1987].

HMC algorithm for DP-MGLM on Riemannian manifolds



Input: $(B_{cur}, \mathbf{V}_{cur}) \in \mathcal{M} \times T_B\mathcal{M}^n$, Leapfrog parameters $\epsilon \in \mathbf{R}_{++}, L \in \mathbf{Z}_{++}$
Output: $(B_{next}, \mathbf{V}_{next}) \in \mathcal{M} \times T_B\mathcal{M}^n$
Sample $(\dot{B}_{cur}, \dot{\mathbf{V}}_{cur}) \in T_B\mathcal{M} \times T_B\mathcal{M}^n$ from independent normal distribution w.r.t. Riemannian metric.
Initialize $(B, \mathbf{V}, \dot{B}, \dot{\mathbf{V}}) \leftarrow (B_{cur}, \mathbf{V}_{cur}, \dot{B}_{cur}, \dot{\mathbf{V}}_{cur})$
$\dot{B} \leftarrow \dot{B} - \frac{\epsilon}{2}\nabla_B U(B, \mathbf{V})$ and $\dot{\mathbf{V}} \leftarrow \dot{\mathbf{V}} - \frac{\epsilon}{2}\nabla_{\mathbf{V}} U(B, \mathbf{V})$
for $i \in \{1, \cdots, L\}$ **do**
  $B' \leftarrow B$, $B \leftarrow \mathrm{Exp}(B, \epsilon\dot{B})$, $V \leftarrow V + \epsilon\dot{V}$
  $(\mathbf{V}, \dot{B}, \dot{\mathbf{V}}) \leftarrow (\Gamma_{B' \to B}\mathbf{V}, \Gamma_{B' \to B}\dot{B}, \Gamma_{B' \to B}\dot{\mathbf{V}})$
/* Parallel transport */
  if i is not L **then**
    $\dot{B} \leftarrow \dot{B} - \epsilon\nabla_B U(B, \mathbf{V})$ and $\dot{\mathbf{V}} \leftarrow \dot{\mathbf{V}} - \epsilon\nabla_{\mathbf{V}} U(B, \mathbf{V})$
  end if
end for
$\dot{B} \leftarrow \dot{B} - \frac{\epsilon}{2}\nabla_B U(B, \mathbf{V})$ and $\dot{\mathbf{V}} \leftarrow \dot{\mathbf{V}} - \frac{\epsilon}{2}\nabla_{\mathbf{V}} U(B, \mathbf{V})$
Accept $(B, \mathbf{V})$ with probability
$\min[1, \exp(H(\dot{B}_{cur}, \dot{\mathbf{V}}_{cur}, B_{cur}, \mathbf{V}_{cur}) - H(\dot{B}, \dot{\mathbf{V}}, B, \mathbf{V}))]$

Lemma

Let $(B, V) \in SPD(n) \times Sym(n)$ be a sample drawn using (2) , then V is Normally distributed w.r.t. a GL-invariant metric at the tangent space $T_B\mathcal{M}$ at B. For each B, the probability density function of V is proportional to $\exp(-\frac{1}{2}\lVert V \rVert_B^2)$ at $T_B\mathcal{M}$, when $\mu_V = 0$.

$$B | \mu_B, \sigma_B^2 \sim \mathcal{N}_{SPD}(B | \mu_B, \sigma_B^2), V | \mu_V, B \sim \mathcal{N}_{Sym}(V | \mu_V, B) \quad (2)$$

Synthetic Experiments: Riemannian DP-MGLM vs. MGLM

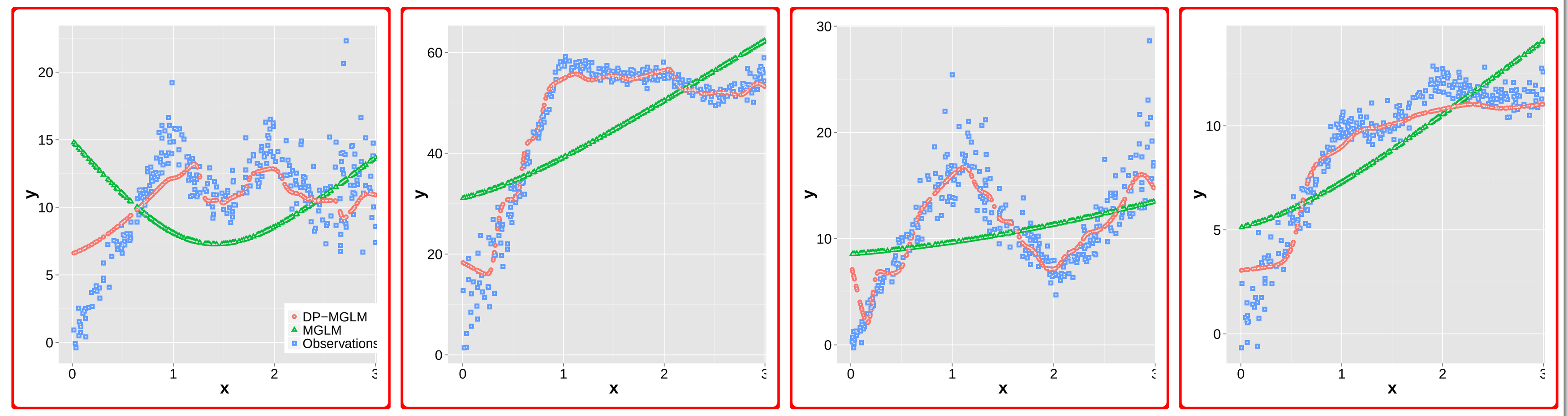


Figure : The figure shows the models fitted in response $y \in \mathcal{M}$ versus covariate $x \in \mathbf{R}$. The prediction of DP-MGLM is shown using a single sample from the posterior, $\theta^{(i)}$. To visualize the response variable $Y \in$ SPD(3), we project the variables onto the axis obtained by PGA (y-axis). The x-axis is the covariate $\boldsymbol{x} \in \mathbf{R}$. Red and blue correspond to our predictions and the measurements respectively.

Synthetic Experiments: Clustering by nonlinear correlation

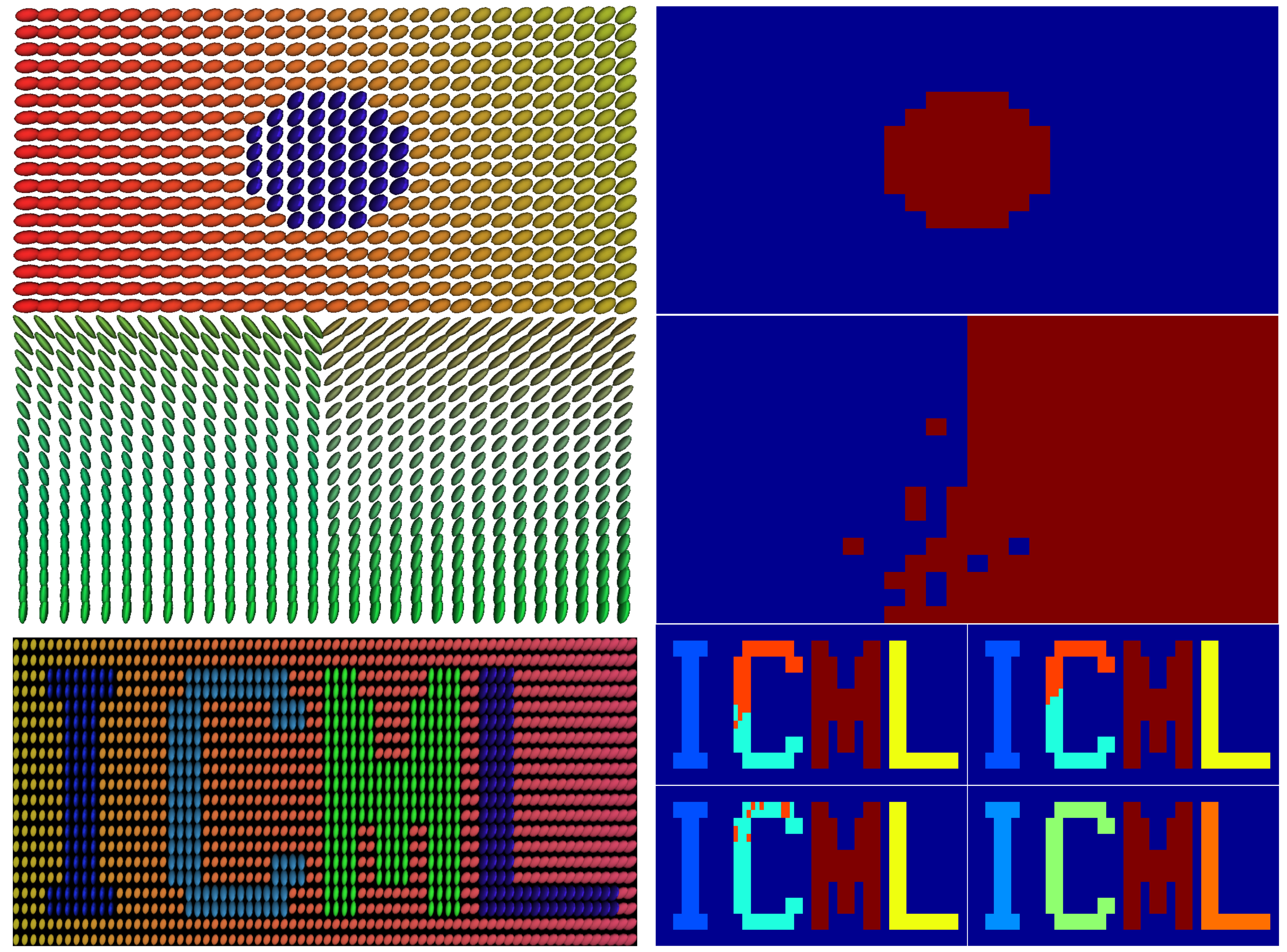


Figure : (Column 1) shows SPD(3) response variables in each voxel; the corresponding covariates are the grid positions (horizontal, vertical coordinates). (Column 2) shows a clustering result. A clustering is shown based on the posterior samples $\theta^{(i)}$.

Synthetic Experiments: Clustering effect

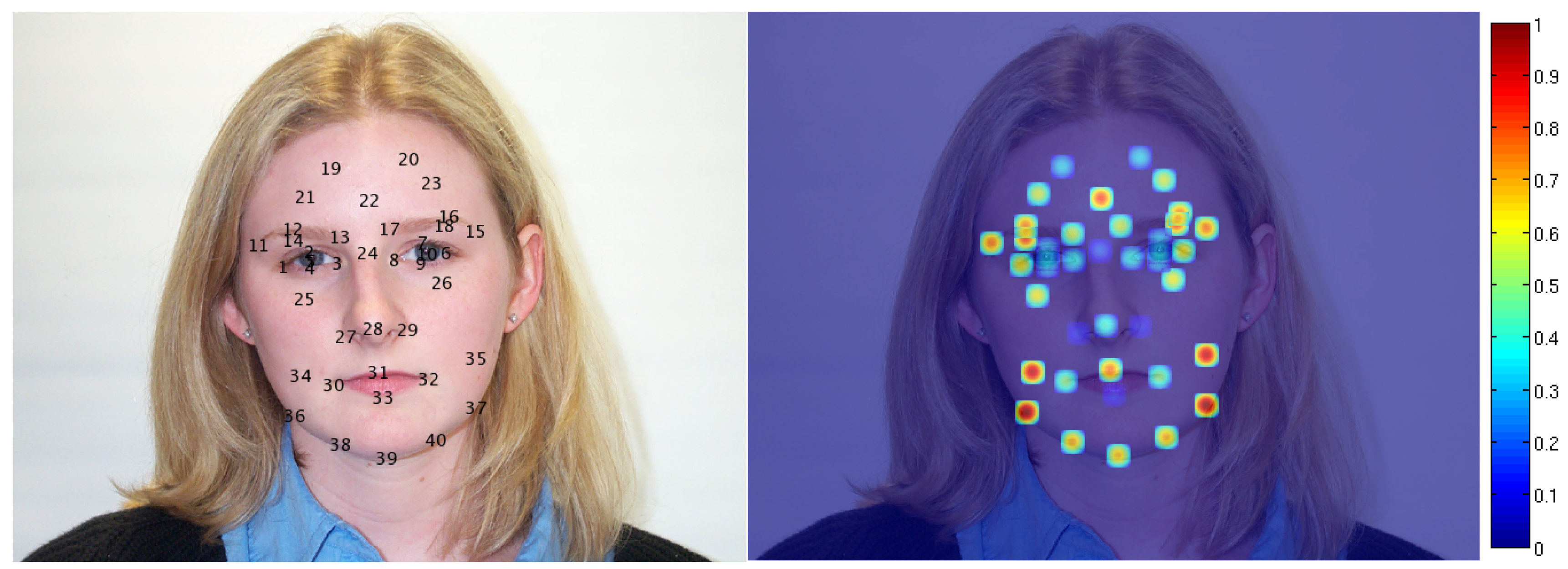


Figure : Left image shows 40 landmarks (indexed by numbers) on an example image . Right image of the bottom row shows correlation magnitude of the landmark's variation with age as a heat map. (Best viewed in color.)

International Conference on Machine Learning (ICML) 2015

Research supported in part by NIH R01 AG040396, NS066340, NSF CAREER award 1252725, UW CPCP NIH AI117924.