Landmark Alternating Diffusion

Sing-Yuan Yeh[1], Hau-Tieng Wu[2], Ronen Talmon[3], Mao-Pei Tsui[1]
[1]*National Taiwan University,* [2]*New York University,* [3]*Technion - Israel Institute of Technology*
[1]d10948003@ntu.edu.tw

Introduction

Alternating Diffusion (AD) [2] is a commonly applied diffusion-based sensor fusion algorithm. We propose Landmark AD (LAD), a variation of Alternating Diffusion, inspired by ROSELAND [1], to enhance computational efficiency while maintaining the benefits of AD. Theoretical analyses of LAD are provided, and its effectiveness is demonstrated in automatic sleep stage annotation using two EEG channels.

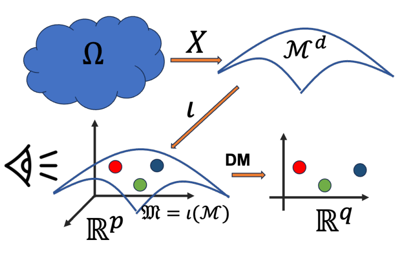
Manifold Setting

Assume the data is located on a manifold $\mathcal{M}$ and the sensor collects data by ι. Hence, the dataset $\{s_i\}_{i=1}^n \subset \mathbb{R}^p$.

(a) $\iota : \mathcal{M} \to \mathbb{R}^p$ a smooth d-dim manifold.

(b) $\mathcal{M}$-valued random variable $X : (\Omega, \mathcal{F}, \mathbb{P}) \to \mathcal{M}$ induced $\mu = X_*\mathbb{P}$.

(c) Assume $d\mu$ is absolutely continuous w.r.t. dV. Denote p.d.f. $p = \frac{d\mu}{dV}$.

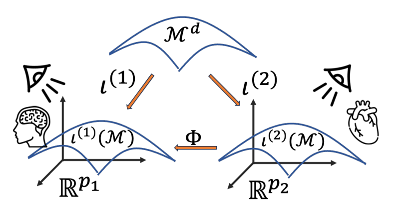


Preliminary: Alternating Diffusion (AD)

1. Definition of a common manifold model

(a) Two sensors collect data $\{(r_i, s_i)\}_{i=1}^n$ from common manifold $\mathcal{M}$ via $\iota^{(\ell)} : \mathcal{M} \to \mathbb{R}^{p_\ell}$ where $\ell = 1, 2$.

(b) Assume $d\mu$ is absolutely continuous w.r.t. $dV^{(\ell)}$. Denote p.d.f. $p^{(\ell)} = \frac{d\mu}{dV^{(\ell)}}$.



2. Algorithm of AD

1. Sample n pairs $\{(r_i, s_i)\}_{i=1}^n \subset \mathbb{R}^{p_1} \times \mathbb{R}^{p_2}$ by two seneors.
2. Construct affinity matrices $W_{ij}^{(\ell)}$ where $\ell = 1, 2$.
3. Let degree matrices $D_{ii}^{(\ell)} = \sum_{i=1}^n W_{ij}^{(\ell)}$.
4. Define alternative Markov matrix $M = D^{(1)^{-1}} W^{(1)} D^{(2)^{-1}} W^{(2)}$.

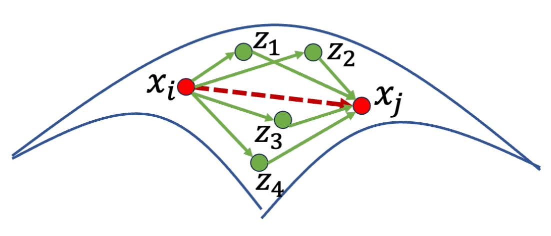
3. Computational complexity of AD

The computational complexity of AD is $\mathcal{O}(n^3)$ where n is the number of data.

Landmark Diffusion

Landmark diffusion [1] reduces computational demands by using a smaller landmark set $\mathcal{Z}$ for eigendecomposition instead of the entire dataset. Denote $p_{\mathcal{Z}}$ is p.d.f. of landmark set $\mathcal{Z}$ with $|\mathcal{Z}| = m \ll n$.

$$\frac{1}{m}\sum_{k=1}^{m} \epsilon^{-d/2} K_\epsilon(x_i, z_k) K_\epsilon(z_k, x_j) = p_{\mathcal{Z}}(x_i) K_{2\epsilon}(x_i, x_j) + \mathcal{O}(\epsilon^{1/2}) + \mathcal{O}\left(\frac{\sqrt{\log(n)}}{n^{\beta/2}\epsilon^{d/4}}\right)$$



Algorithm of α-LAD

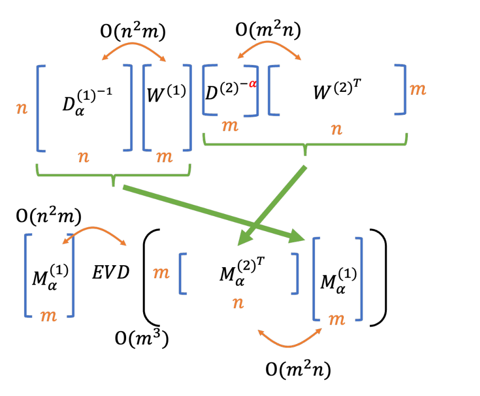
Input: A pair of datasets $\{(r_i, s_i)\}_{i=1}^n \subset \mathbb{R}^{p_1} \times \mathbb{R}^{p_2}$ are sampled simultaneously from two sensors and normalized constant $\alpha \in [0, 1]$.

1. Choose a landmark set $\mathcal{Z}$ with size m.
2. Build $n \times m$ affinity matrix $W_{ik}^{(\ell)}$.
3. $m \times m$ matrix $D_{ii}^{(2)} = \mathtt{diag}(W^{(2)} W^{(2)^\top} \mathbf{1}_m)$
4. $n \times m$ matrix $M_\alpha^{(2)} = W^{(2)} D^{(2)^{-\alpha}}$
5. $n \times n$ matrix $D_{\alpha;ii}^{(1)} = \mathtt{diag}(W^{(1)} M_\alpha^{(2)^\top} \mathbf{1}_n)$
6. $n \times m$ matrix $M_\alpha^{(1)} = D_\alpha^{(1)^{-1}} W^{(1)}$
7. EVD on $m \times m$ matrix $M_\alpha^{(2)^\top} M_\alpha^{(1)} = V \Lambda V^{-1}$
8. $U = M_\alpha^{(1)} V$ and choose top q vectors as U_q

Output: $e_i^\top \bar{\mathbf{U}}_q \bar{\mathbf{\Lambda}}_q^t$ as the embedded point of (s_i, r_i), $i = 1, \ldots, n$.

Illustration

Eigendecomposition of a smaller matrix has a time complexity of $O(m^3)$, while eigenvector "interpolation" recover it with a complexity of $O(n^2 m)$.



The overall computational complexity of α-LAD is $O(n^2 m)$ since we assume $m < n$.

Main Theorem

Based on the manifold setting of AD, suppose a function $f \in C^3(\mathcal{M})$ and $p \in C^2(\mathcal{M})$. Let $q_\alpha(x) := \frac{p_{\mathcal{Z}}^{(2)}(x)^{1-\alpha}}{p^{(2)}(x)^\alpha}$ where α is normalizer constant. Then, with probability $1 - \mathcal{O}(n^{-2})$, we have

$$\begin{aligned}
&\frac{1}{\epsilon}\left[\left(I_n - \left(D_\alpha^{(1)}\right)^{-1} W_\alpha^{(1)} M_\alpha^{(2)}\right)\boldsymbol{f}\right](i) \\
&= \frac{\mu_{2,0}^{(2)}}{2d}\Delta^{(2)} f(x_i) + \frac{\mu_{2,0}^{(1)}}{2d}\sum_{j=1}^{d}\lambda_j \nabla_{E_j E_j}^{(2)^2} f(x_i) + \frac{\mu_{2,0}^{(1)}}{d}\sum_{j=1}^{d}\lambda_j\left(\frac{\nabla_{E_j}^{(2)} p^{(2)}(x_i)}{p^{(2)}} + \frac{\nabla_{E_j}^{(2)} q_\alpha(x_i)}{q_\alpha(x_i)}\right)\nabla_{E_j}^{(2)} f(x_i) \\
&\quad + \frac{\mu_{2,0}^{(2)}}{d}\frac{\nabla^{(2)} p^{(2)}(x_i) \cdot \nabla^{(2)} f(x_i)}{p^{(2)}(x_i)} + \underbrace{\mathcal{O}(\epsilon^{1/2})}_{\text{Bias}} + \underbrace{\mathcal{O}\left(\frac{\sqrt{\log(n)}}{n^{1/2}\epsilon^{d/4+1}}\right)}_{\text{Variance}}.
\end{aligned}$$

where the rotation matrix is defined by

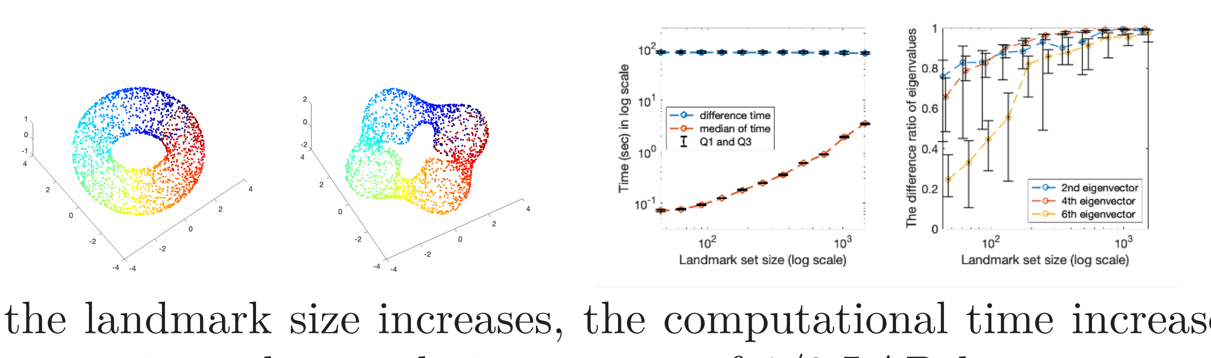
$$R_x = \left[d\exp_x^{(1)}\Big|_0\right]^{-1}\left[d\iota^{(1)}\right]^{-1}\nabla\Phi\left[d\iota^{(2)}\right]\left[d^{(2)}\exp_x^{(2)}\Big|_0\right].$$

Corollary

- Case 1 $\alpha = 0$: If $p^{(2)} = p_{\mathcal{Z}}^{(2)}$, $\iota^{(1)} = \iota^{(2)}$, then 0-LAD is Roseland.
- Case 2 $\alpha = 1/2$: If $p^{(2)} = p_{\mathcal{Z}}^{(2)}$, then $q_\alpha = 1$ and 1/2-LAD approach to AD.
- Case 3 $\alpha = 1$: 1-LAD is independent of $p_{\mathcal{Z}}$.

Speed up AD

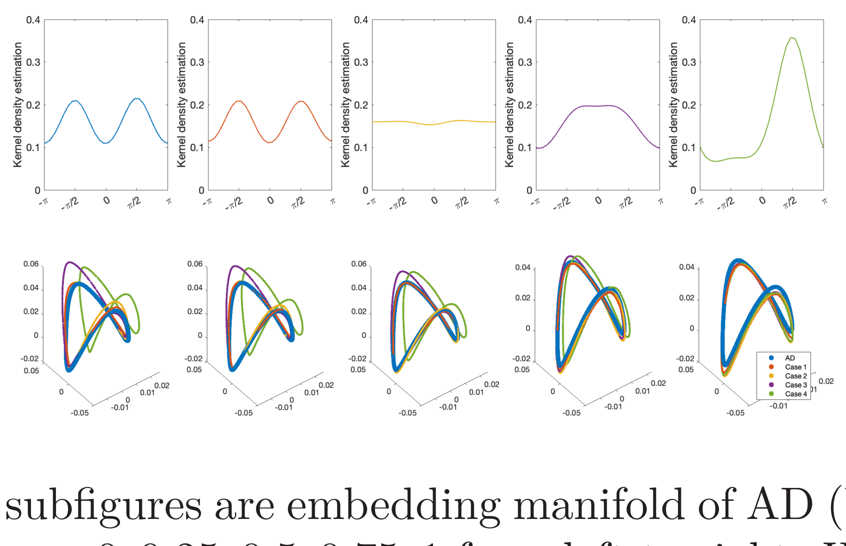
Sample 5000 pairs of data on the following two manifolds.



When the landmark size increases, the computational time increases, but the top eigenvalues and eigenvectors of 1/2-LAD better approximate those of AD.

Independent of the Landmark Distribution

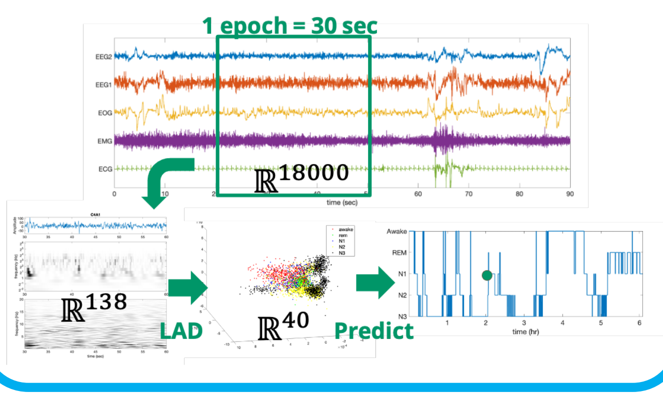
Sample 3000 pairs of points from the non-uniform PDF (blue on the left) and 1500 pairs of landmark points from the other distributions (red, yellow, purple, and green on the right).



The lower five subfigures are embedding manifold of AD (blue one) and α-LAD where $\alpha = 0, 0.25, 0.5, 0.75, 1$ from left to right. We see that as α approaches 1, the embeddings by LAD are more similar, but these embeddings are not necessarily similar to the embedding by AD.

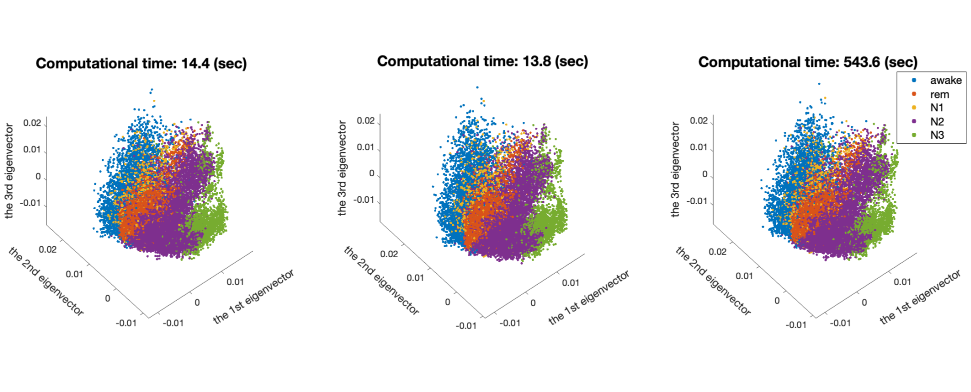
Sensor Fusion

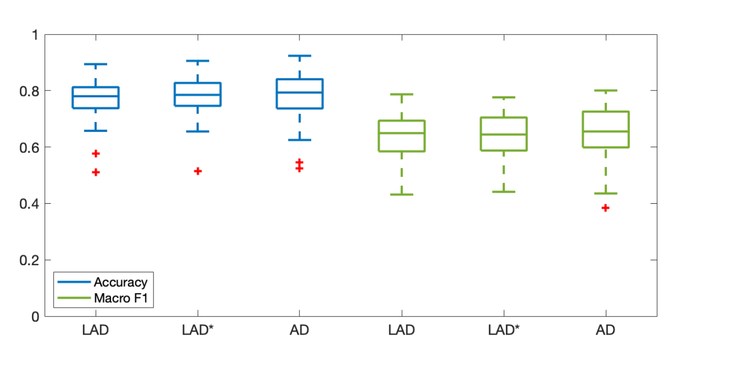
We demonstrate that 1/2-LAD serves as a computationally efficient alternative to AD for extracting common brain dynamics from EEG signals, enabling reliable automatic annotation of five sleep stages, including awake, REM, N1, N2, and N3.



Application to Sleep Stage Annotation

After applying 0.5-LAD, 0.5-LAD* (balanced landmark), and AD to 29,070 epochs, the top three eigenvectors from 0.5-LAD and 0.5-LAD* closely match those from AD (inner products are close to 1).





Boxplot results for leave-one-subject-out cross-validation of 40 patients. From left to right is LAD, LAD*, AD. With a p-value threshold of 0.05, both methods showed no significant difference in accuracy or macro F1 compared to AD, according to the Wilcoxon sign-rank test; however, computation time decreased from 9 minutes to 14 seconds.

References

[1] C. Shen and H.-T. Wu, *Scalability and robustness of spectral embedding: landmark diffusion is all you need*, IMAIAI, (2022).

[2] R. Talmon and H.-T. Wu, *Latent common manifold learning with alternating diffusion: Analysis and applications*, ACHA, (2019).

Acknowledgements

S.-Y. Yeh and M.-P. Tsui are supported in part by the National Science and Technology Council grants 112-2115-M-002-015.