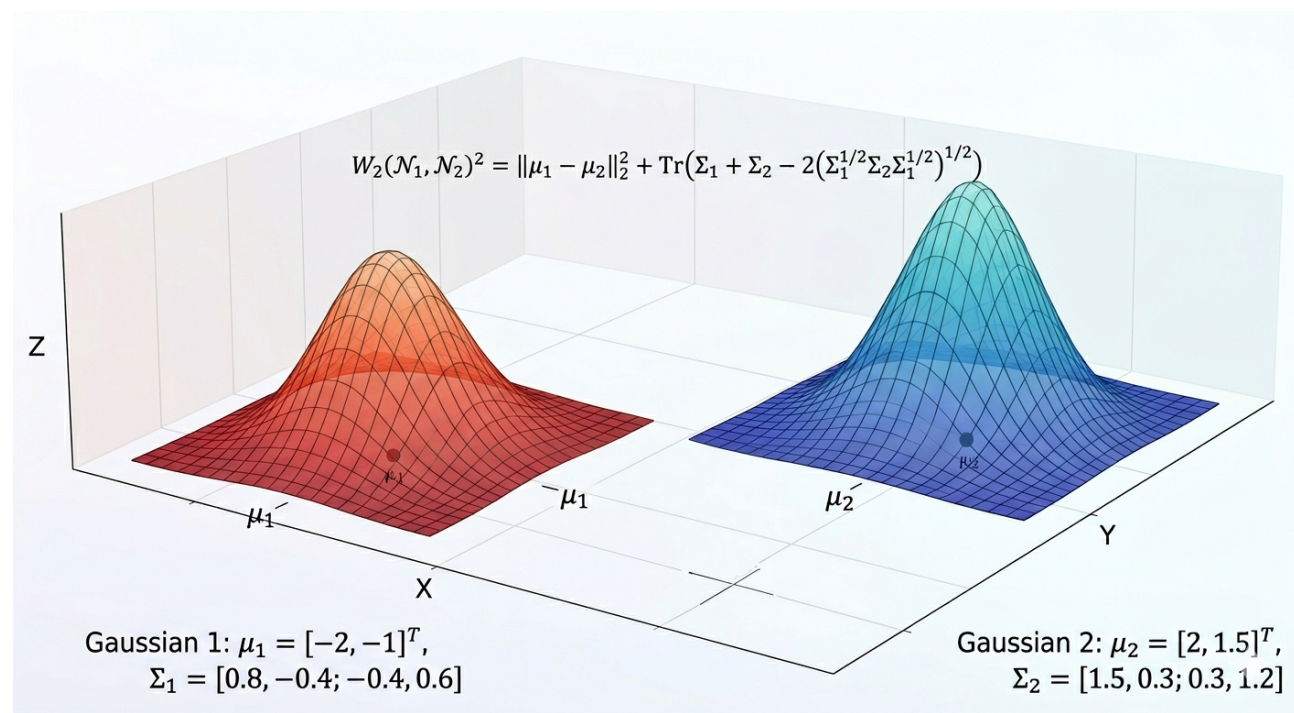


Bures-Wasserstein Manifold

The **Bures-Wasserstein manifold** \mathbb{S}_{++}^d is the space of symmetric positive definite matrices.



Why BW?

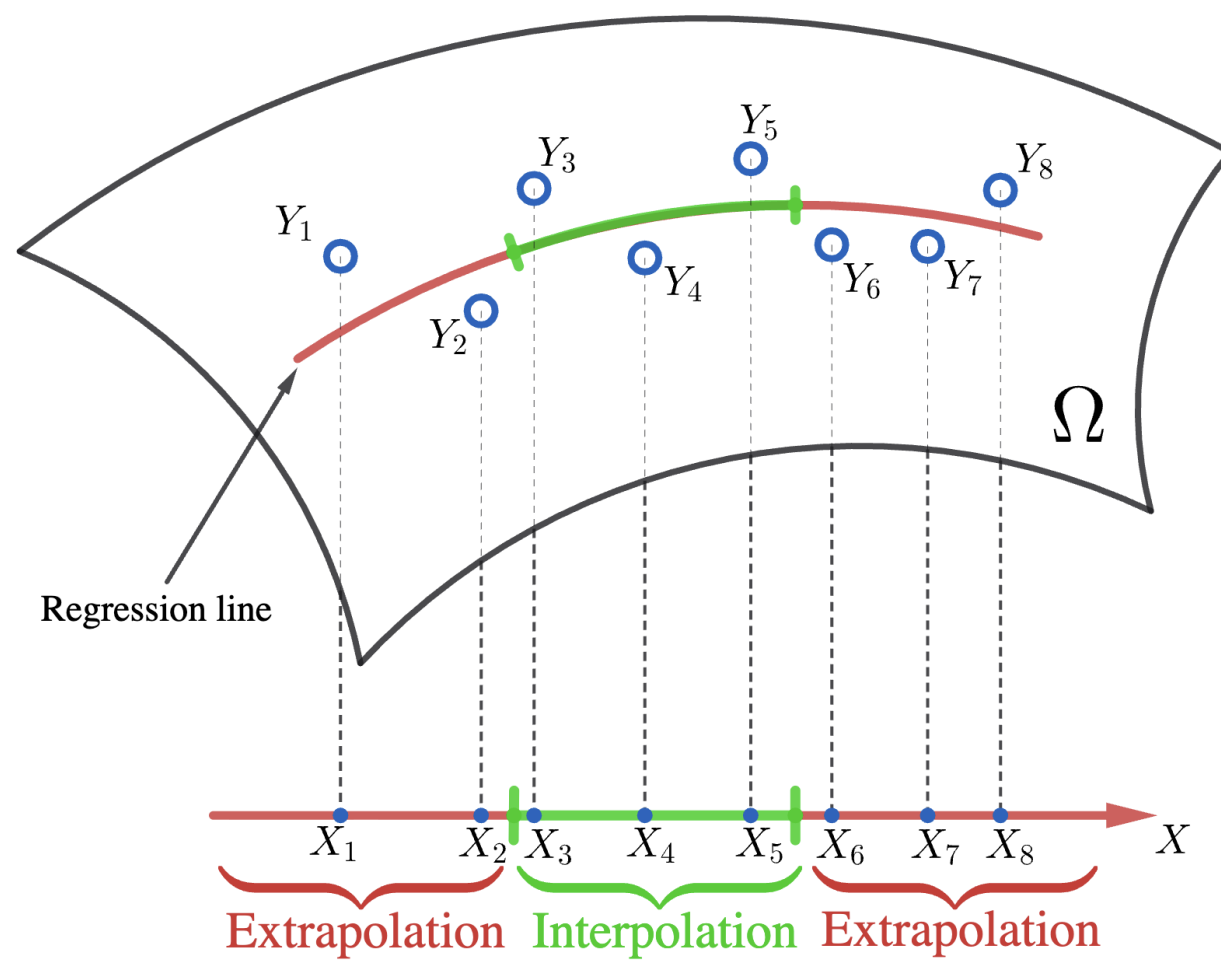
- BW avoids Frobenius “swelling effect”,
- BW does **NOT** require **matrix logarithms**, unlike affine-invariant (Fisher-Rao) and log-Euclidean metrics.

Fréchet Regression

Global Fréchet Regression: Given n independent samples $(X_k, Y_k) \sim \mathcal{F}$,

$$\hat{m}_G(x) = \arg \min_{\omega \in \Omega} \frac{1}{n} \sum_{k=1}^n s_{G,k}(x) d^2(Y_k, \omega), \quad (1)$$

where $s_{G,k}(x) = 1 + (X_k - \bar{X})^\top \hat{\Sigma}^{-1} (x - \bar{X})$.



Problem statement

Problem: Given n SPD matrices Σ_k and their weights $\lambda_k \in \mathbb{R}$, for $k \in \{1, \dots, n\}$, such that $\sum_{k=1}^n \lambda_k = 1$, we want to solve

$$\begin{aligned} \min_{S \in \mathbb{S}_{++}^d} F(S) &:= \sum_{k=1}^n \lambda_k W_2^2(S, \Sigma_k) & (2) \\ &= \sum_{i \in \mathcal{I}} \lambda_i^+ W_2^2(S, \Sigma_i) - \sum_{j \in \mathcal{J}} \lambda_j^- W_2^2(S, \Sigma_j), \end{aligned}$$

for $\lambda_i^+, \lambda_j^- > 0, \mathcal{I} = \{k : \lambda_k > 0\}, \mathcal{J} = \{k : \lambda_k < 0\}$.

Minimizer Existence

Example (Negative weights): Let $\lambda_1=2, \lambda_2=-1, \Sigma_1 = I, \Sigma_2 = 9I$, and $f(S) := \lambda_1 W_2^2(S, \Sigma_1) + \lambda_2 W_2^2(S, \Sigma_2)$. Then, the gradient $\nabla f(S) \succ 0$.

Theorem: Let $\Sigma_1, \dots, \Sigma_n \in \mathbb{S}_{++}^d$ and $\lambda_1, \dots, \lambda_n \in \mathbb{R}$ with $\sum_{k=1}^n \lambda_k = 1$. If the **Spectral Dominance of Positive Weights** condition holds, i.e.,

$$\sum_{i \in \mathcal{I}} \lambda_i^+ \sqrt{\lambda_{\min}(\Sigma_i)} > \sum_{j \in \mathcal{J}} \lambda_j^- \sqrt{\lambda_{\max}(\Sigma_j)}, \quad (3)$$

then Problem (2) admits a solution.

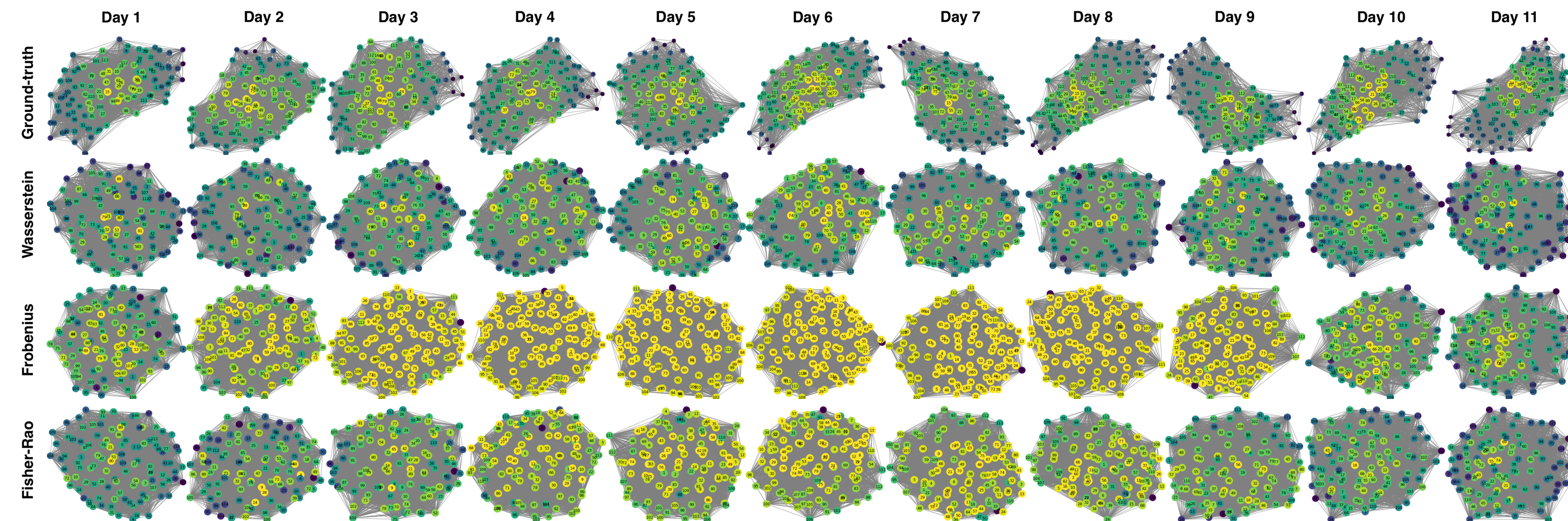
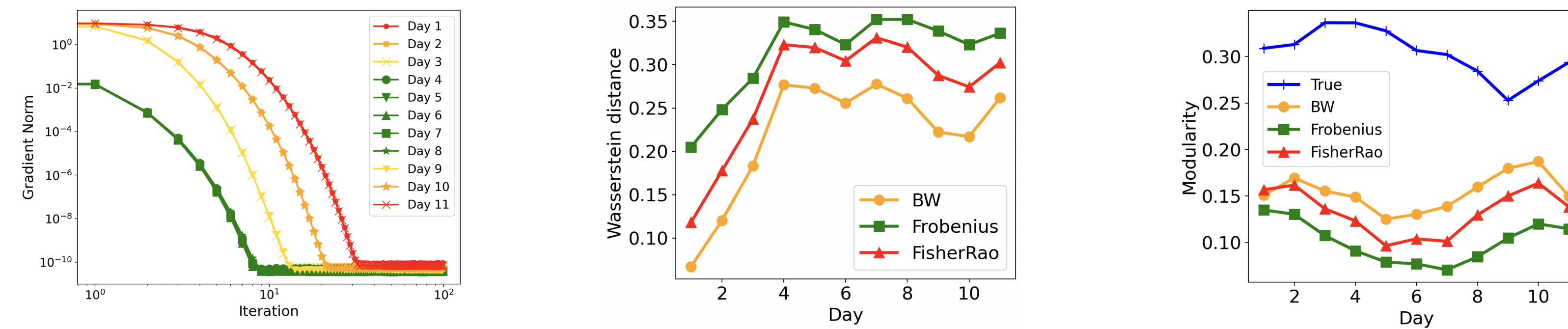
Minimizer Unique Existence

Theorem: Let $\Sigma_k \in \mathbb{S}_{++}^d$, for $k \in \{1, \dots, n\}$, $\lambda := \min \lambda_{\min}(\Sigma_k)$ and $\mathcal{S}_\lambda = \{\Sigma \in \mathbb{S}_{++}^d : \lambda_{\min}(\Sigma) \geq \lambda\}$. Denote $\mu_+ = \sum_i \lambda_i^+, \mu_- = \sum_j \lambda_j^-$. Assume that there exist $\rho > 0, r > 0$ and $\Sigma_0 \in \mathcal{S}_\lambda$ such that

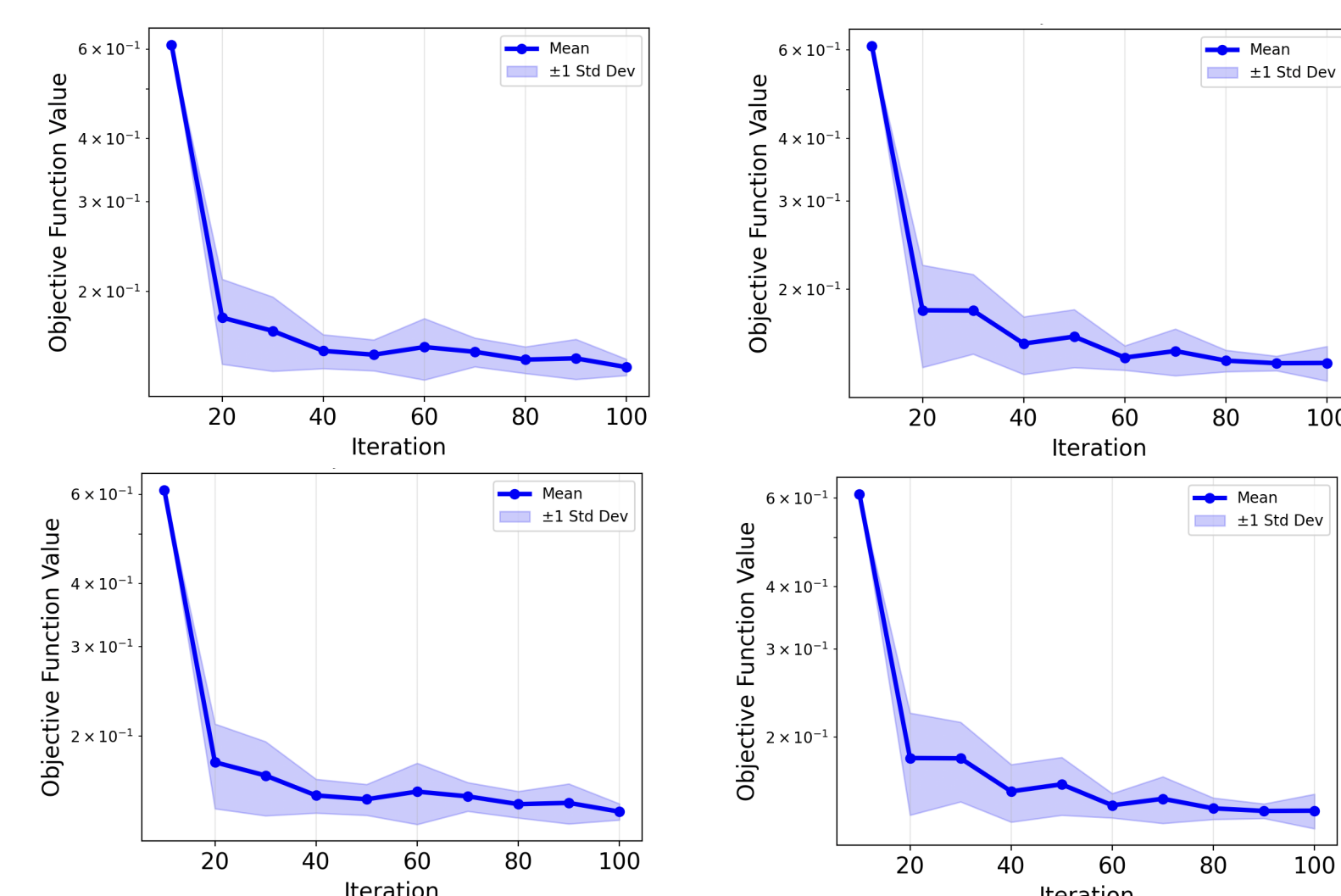
- $\Sigma_k \in B_r(\Sigma_0)$ for all $k \in \{1, \dots, n\}$, where $B_r(\Sigma_0)$ is the geodesic ball centered at Σ_0 with radius r ,
- $r < \rho / (\mu_+ + \mu_-)$,
- $\rho < \sqrt{\lambda} / 2$,
- $B_\rho(\Sigma_0) \subset \mathcal{S}_\lambda$, and
- $\mu_+ / \mu_- > (2\rho\sqrt{\Lambda^+}) / (\tanh(2\rho\sqrt{\Lambda^+}))$.

Then, there is a **unique minimizer** in $B_\rho(\Sigma_0)$.

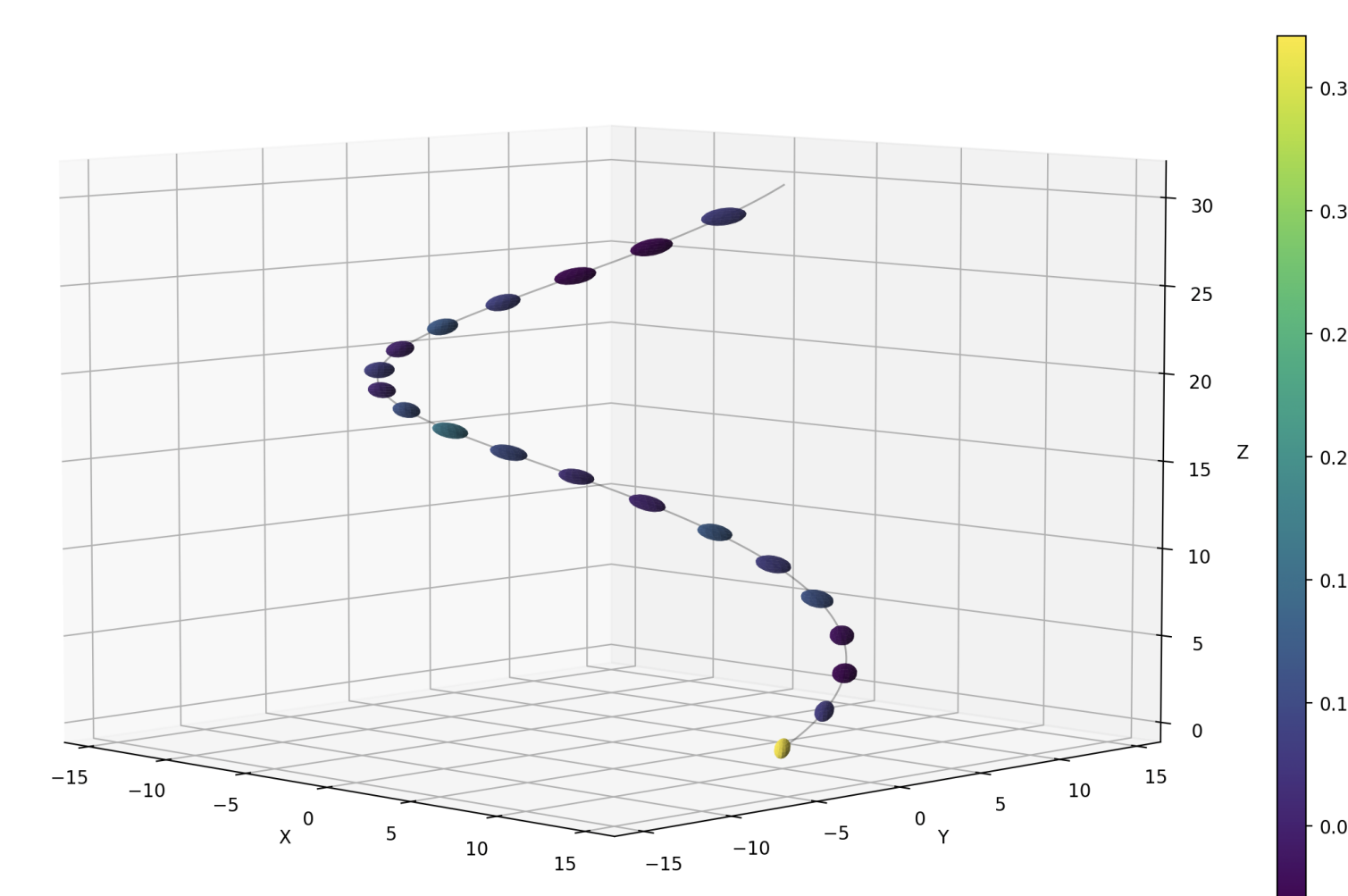
Network Regression with Ant Social Networks



Large-scale Diffusion Tensor Imaging (DTI)



Objective Values over 100 iterations of RSGD



Visualization of tensors

Additional Properties

Bounded Region: If the condition (3) holds, then:

$$\begin{aligned} \left(\sum_{i \in \mathcal{I}} \lambda_i^+ \sqrt{\lambda_{\min}(\Sigma_i)} - \sum_{j \in \mathcal{J}} \lambda_j^- \sqrt{\lambda_{\max}(\Sigma_j)} \right)^2 I &\prec S_*, \\ \left(\sum_{i \in \mathcal{I}} \lambda_i^+ \sqrt{\lambda_{\max}(\Sigma_i)} - \sum_{j \in \mathcal{J}} \lambda_j^- \sqrt{\lambda_{\min}(\Sigma_j)} \right)^2 I &\succ S_*. \end{aligned}$$

No local maximum: S_* is not a local maximum.

Algorithm

Algorithm 1 General BWGD

- Input:** SPD matrices Σ_k , weights λ_k , initial S_0 , step-size η , epochs T .
- for** $t = 1, 2, \dots, T$ **do**
- $\tilde{S}_t = (1 - \eta)I + \eta \sum_{k=1}^n \lambda_k \text{GM}(S_{t-1}^{-1}, \Sigma_k)$
- $\tilde{S}_t = \tilde{S}_t S_{t-1}^{-1} \tilde{S}_t$
- $S_t = \tilde{S}_t S_{t-1} \tilde{S}_t$
- $\tilde{S}_t = \exp_{S_{t-1}}(-\eta \nabla_{\text{bw}} F(S_{t-1}))$
- end for**
- Return** $\bar{S} = S_T$

NOTE: Convergence rate $\mathcal{O}(1/T)$ with $\eta \leq 1/L$. With condition (3), Algorithm 1 is **projection-free**

Large-scale Reformulation

$$\begin{aligned} F(S) &= \sum_{i \in \mathcal{I}} \lambda_i^+ W_2^2(S, \Sigma_i) - \sum_{j \in \mathcal{J}} \lambda_j^- W_2^2(S, \Sigma_j) \\ &= \sum_{i \in \mathcal{I}, j \in \mathcal{J}} \frac{\lambda_i^+ \lambda_j^-}{\mu_+ \mu_-} (\mu_+ W_2^2(S, \Sigma_i) - \mu_- W_2^2(S, \Sigma_j)) \\ &= \sum_{i \in \mathcal{I}, j \in \mathcal{J}} \frac{\lambda_i^+ \lambda_j^-}{\mu_+ \mu_-} f_{ij}(S), \end{aligned}$$

for $\lambda_i^+, \lambda_j^- > 0, \mu_+ := \sum_i \lambda_i^+, \mu_- := \sum_j \lambda_j^-, S \in \mathbb{S}_{++}^d$.

Stochastic Gradient:

$$\nabla f_{ij}(S) = I - (\mu_+ \text{GM}(S^{-1}, \Sigma_i) - \mu_- \text{GM}(S^{-1}, \Sigma_j)).$$

Conditions for Pairwise Formulation:

$$(\mu_+) \min_{i \in \mathcal{I}} \sqrt{\lambda_{\min}(\Sigma_i)} > (\mu_-) \max_{j \in \mathcal{J}} \sqrt{\lambda_{\max}(\Sigma_j)}$$