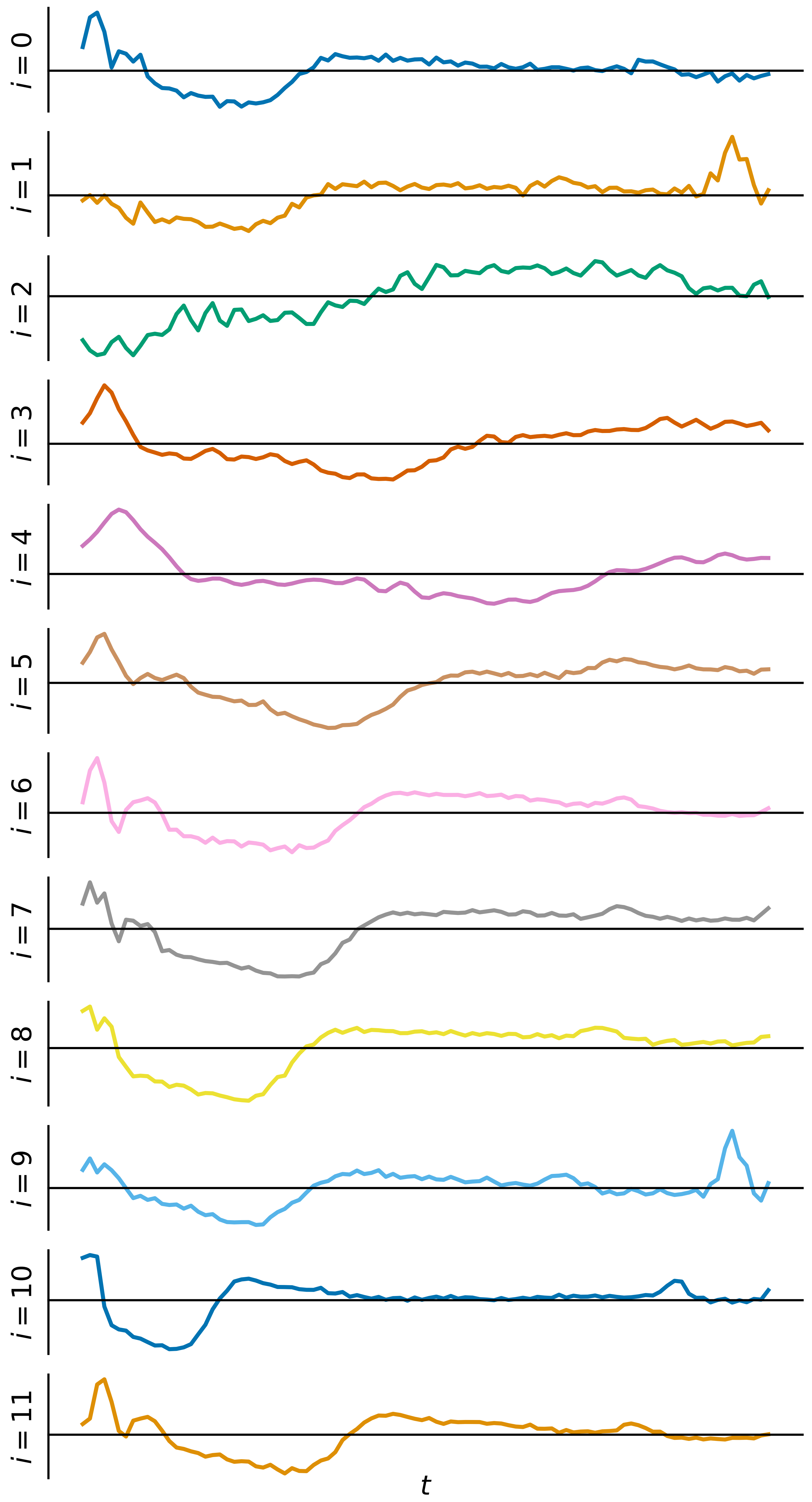
Nonlinear Sequence Data Embedding by Monotone Variational Inequality

Jonathan Y. Zhou, Yao Xie
School of Industrial and Systems Engineering
Georgia Institute to Technology

ICLR 2025 | Apr. 24-28 | Singapore EXPO

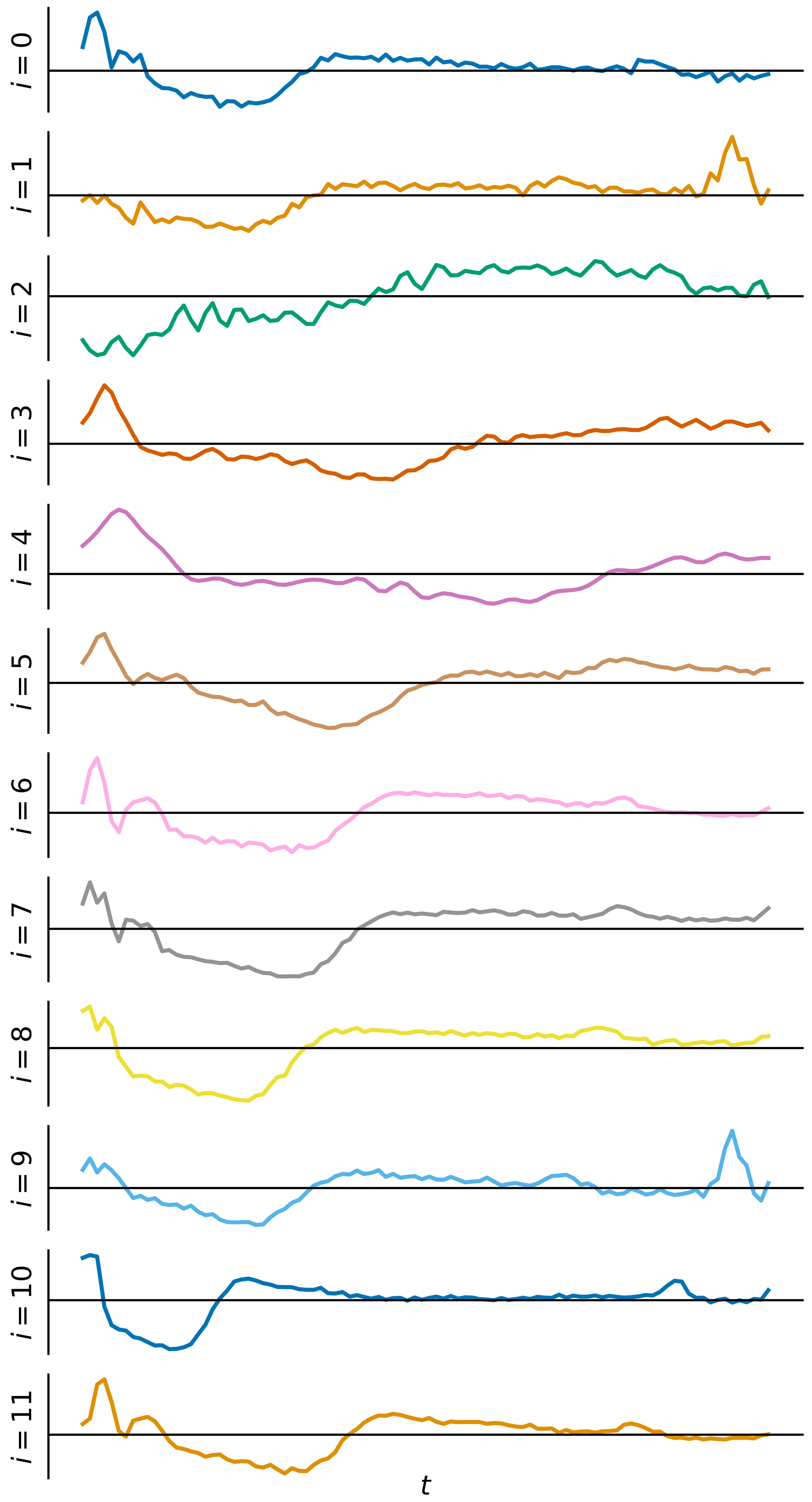
Georgia Tech College of Engineering
H. Milton Stewart School of
Industrial and Systems Engineering

Problem Setup



Problem Setup

Natural, societal, and engineering worlds contain **collections of sequence data**.

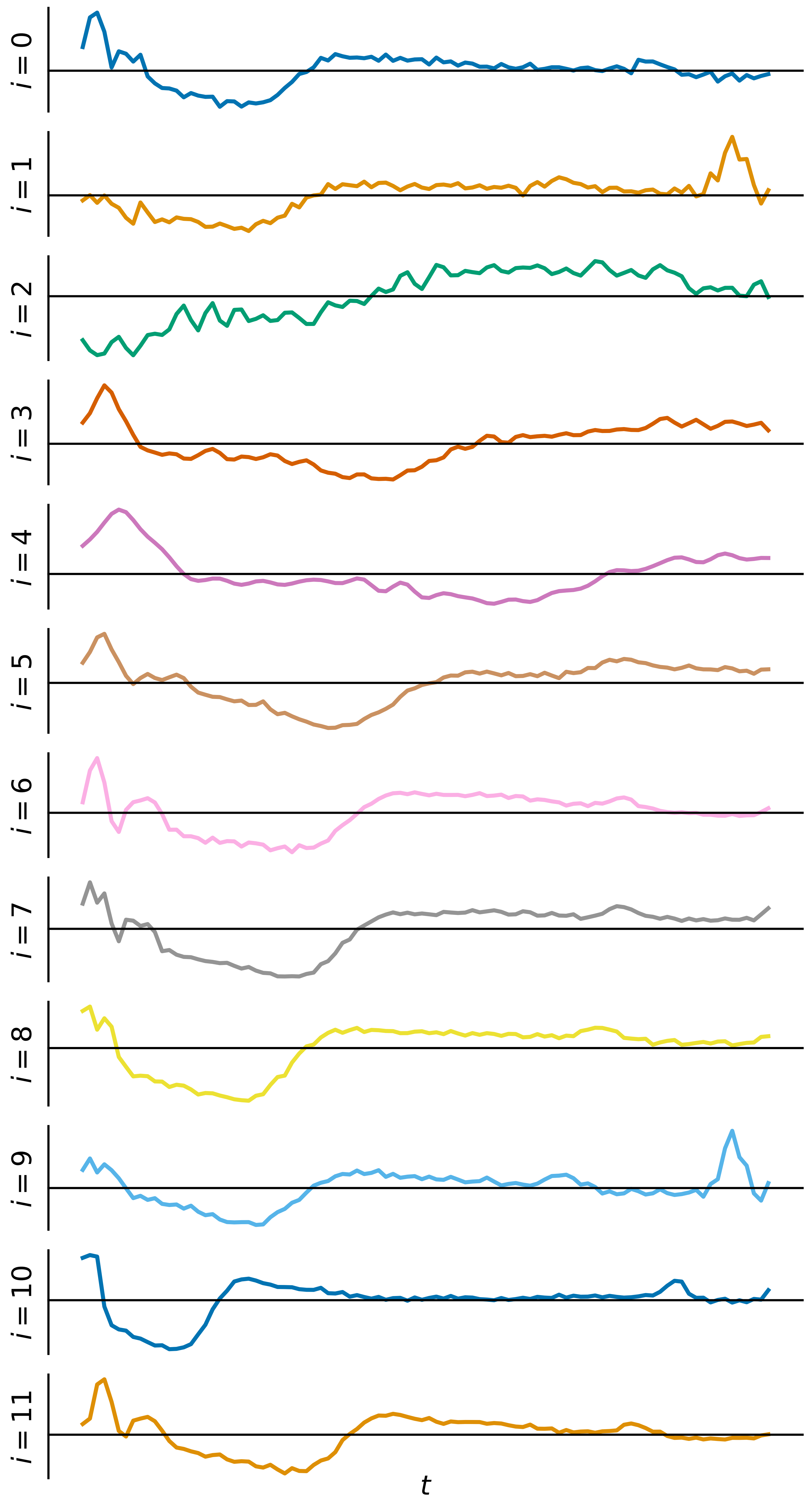


Problem Setup

Natural, societal, and engineering worlds contain **collections of sequence data.**

Data from **generative process**

$$\forall i, \boldsymbol{\Theta}_i \sim \mathscr{D}; \quad \forall (i,t), \mathbf{x}_{i,t} \sim F_{\boldsymbol{\Theta}_i}[\mathscr{H}_{i,t}]\,.$$



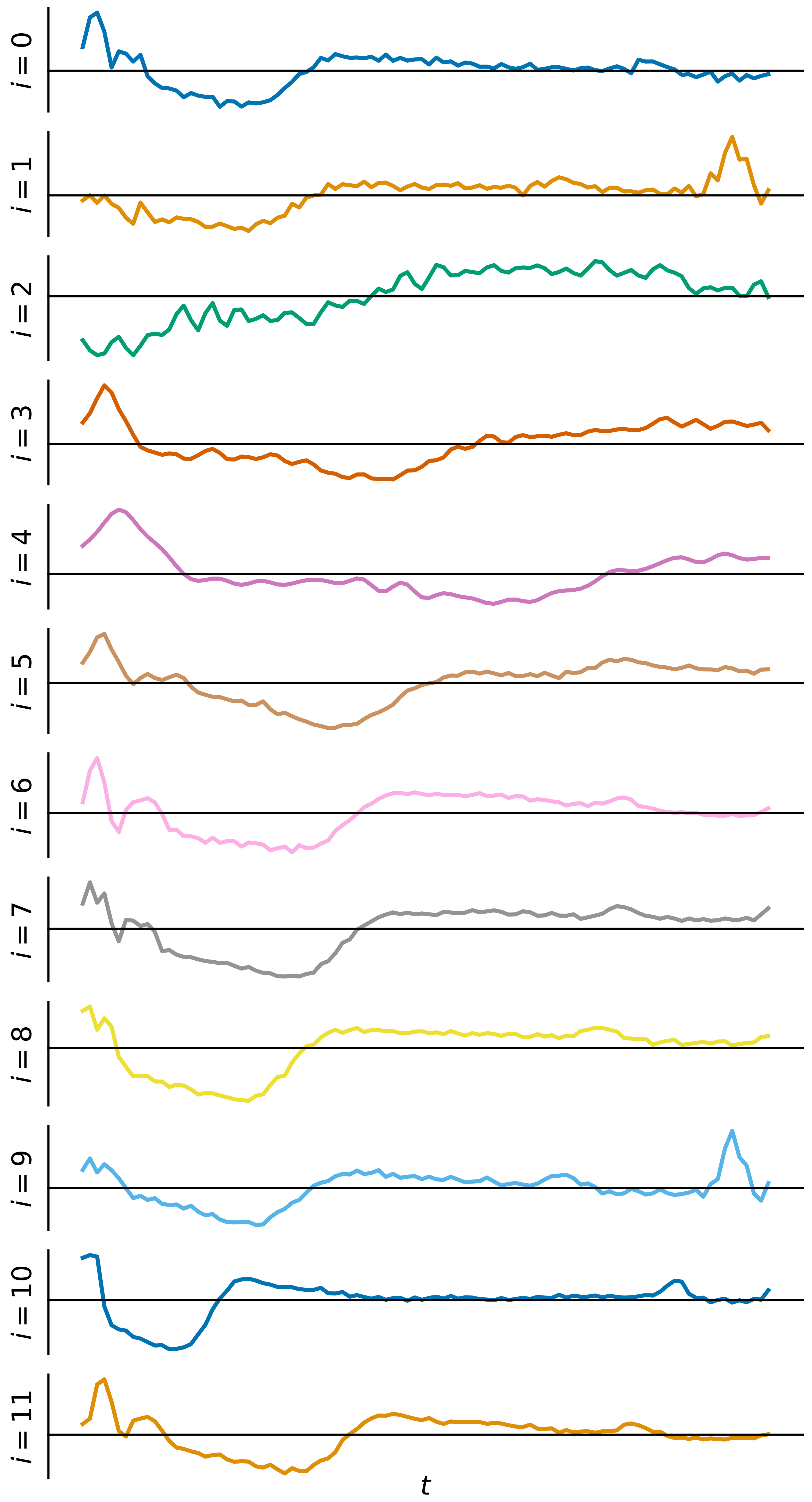
Problem Setup

Natural, societal, and engineering worlds contain **collections of sequence data.**

Data from **generative process**

$$\forall i, \boldsymbol{\Theta}_i \sim \mathscr{D}; \quad \forall (i,t), \mathbf{x}_{i,t} \sim F_{\boldsymbol{\Theta}_i}[\mathscr{H}_{i,t}] \,.$$

Individual sequence dynamics ($\boldsymbol{\Theta}_i$) drawn from **common domain.**



Problem Setup

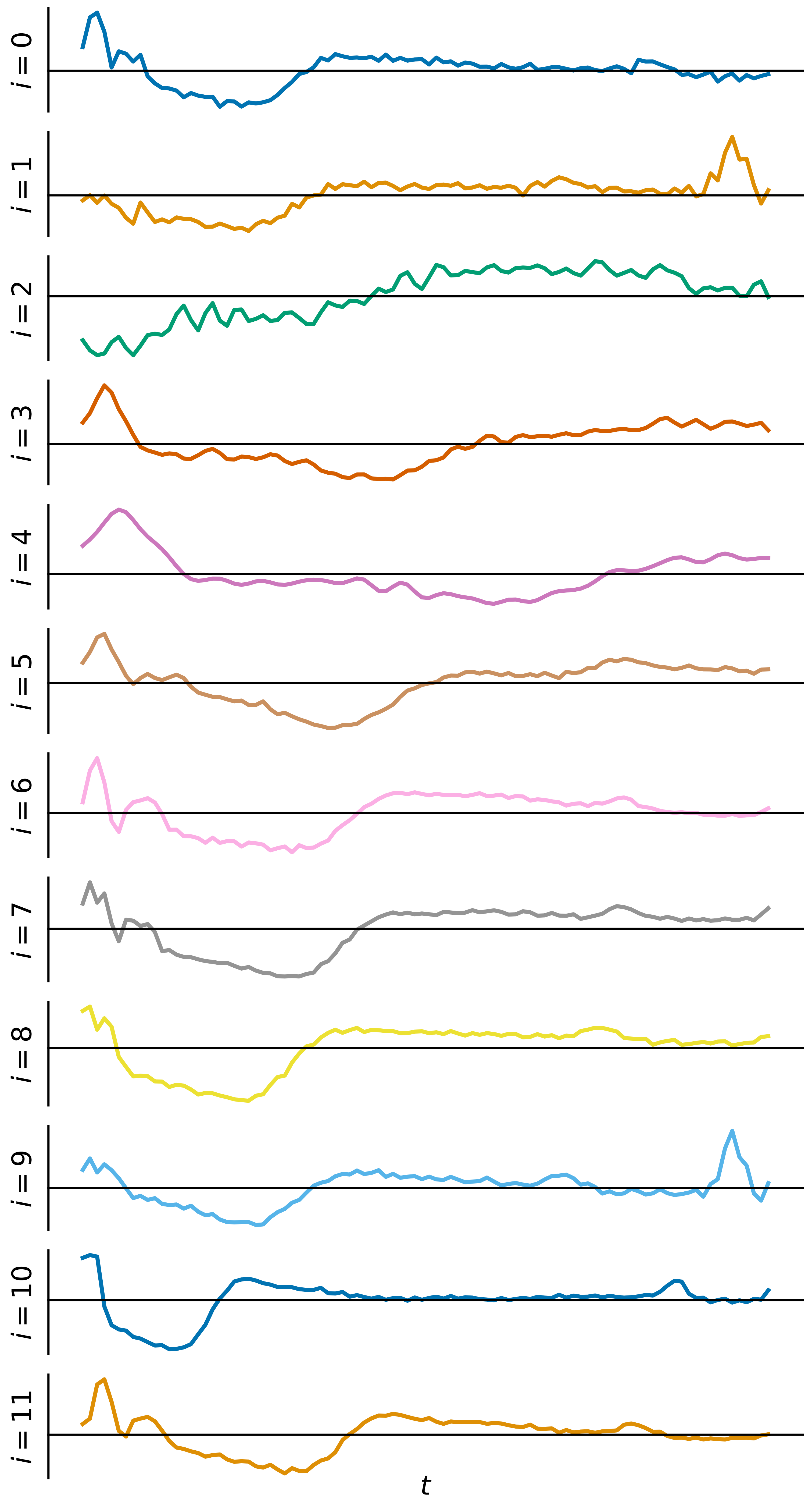
Natural, societal, and engineering worlds contain **collections of sequence data.**

Data from **generative process**

$$\forall i, \boldsymbol{\Theta}_i \sim \mathscr{D}; \quad \forall (i,t), \mathbf{x}_{i,t} \sim F_{\boldsymbol{\Theta}_i}[\mathscr{H}_{i,t}] \,.$$

Individual sequence dynamics ($\boldsymbol{\Theta}_i$) drawn from **common domain.**

State variable $\mathbf{x}_{i,t} \in \mathbb{R}^C$ depends on:



Problem Setup

Natural, societal, and engineering worlds contain **collections of sequence data.**

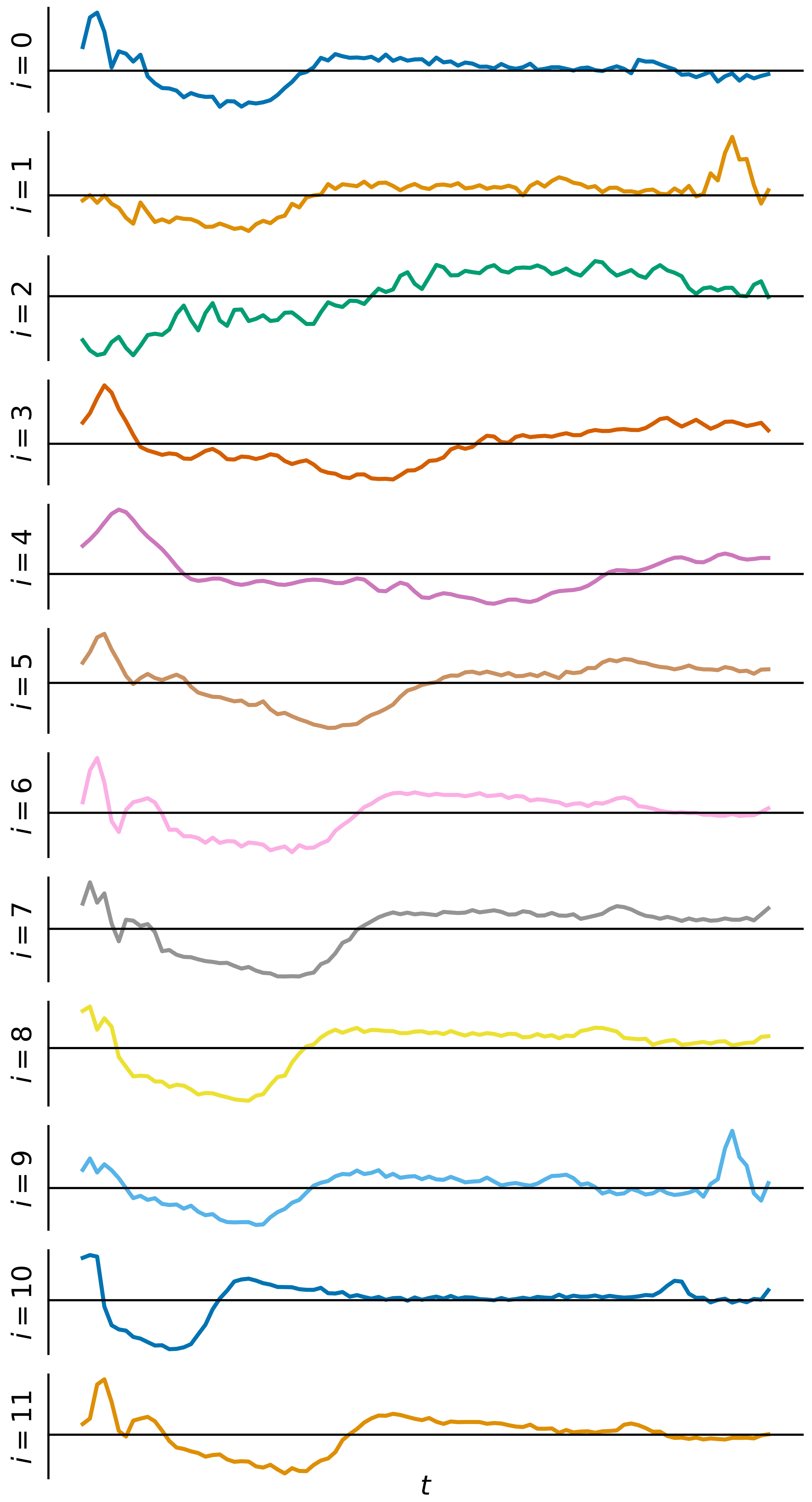
Data from **generative process**

$$\forall i, \boldsymbol{\Theta}_i \sim \mathscr{D}; \quad \forall (i,t), \mathbf{x}_{i,t} \sim F_{\boldsymbol{\Theta}_i}[\mathscr{H}_{i,t}]\,.$$

Individual sequence dynamics ($\boldsymbol{\Theta}_i$) drawn from **common domain.**

State variable $\mathbf{x}_{i,t} \in \mathbb{R}^C$ depends on:

1. Past event history:
   $\mathscr{H}_{i,t} := \{\mathbf{x}_{i,s} \mid s < t\}$



Problem Setup

Natural, societal, and engineering worlds contain **collections of sequence data.**

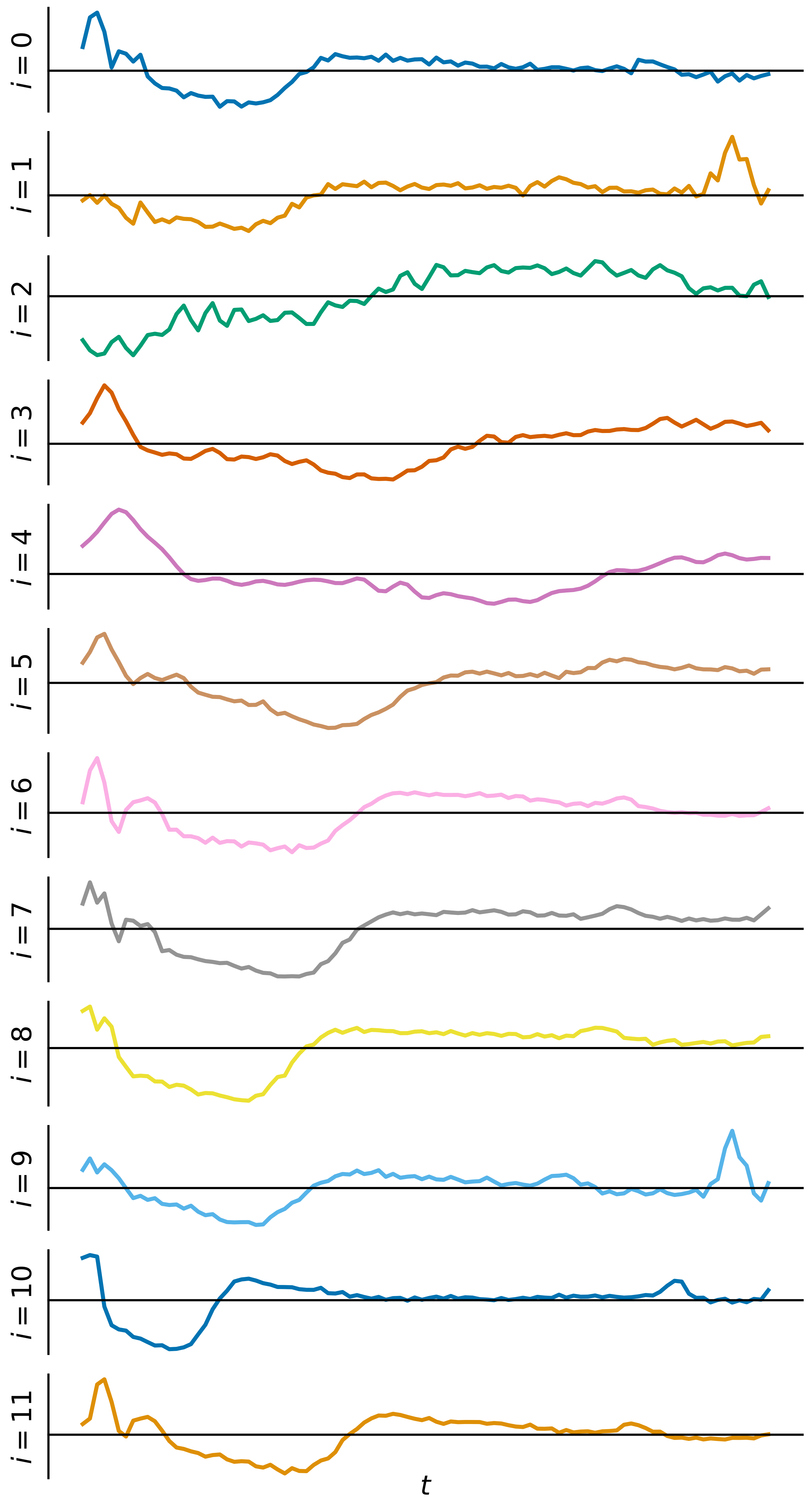
Data from **generative process**

$$\forall i, \boldsymbol{\Theta}_i \sim \mathscr{D}; \quad \forall (i,t), \mathbf{x}_{i,t} \sim F_{\boldsymbol{\Theta}_i}[\mathscr{H}_{i,t}] \,.$$

Individual sequence dynamics ($\boldsymbol{\Theta}_i$) drawn from **common domain.**

State variable $\mathbf{x}_{i,t} \in \mathbb{R}^C$ depends on:

1. Past event history: $\mathscr{H}_{i,t} := \{\mathbf{x}_{i,s} \mid s < t\}$
2. Distribution $F_{\boldsymbol{\Theta}_i}$ dynamic unique to each sequence.



Observation Model

Observation Model

Model sequence dynamics via **autoregression**

$$\mathbb{P}[\mathbf{x}_{i,t} \mid \mathscr{H}_{i,t}] = F_{\mathbf{\Theta}_i}[\mathscr{H}_{i,t}] \implies \mathbb{E}[\mathbf{x}_{i,t} \mid \mathscr{H}_{i,t}] = f_{\mathbf{\Theta}_i}(\mathscr{H}_t)\,.$$

Observation Model

Model sequence dynamics via **autoregression**

$$\mathbb{P}[\mathbf{x}_{i,t} \mid \mathscr{H}_{i,t}] = F_{\mathbf{\Theta}_i}[\mathscr{H}_{i,t}] \implies \mathbb{E}[\mathbf{x}_{i,t} \mid \mathscr{H}_{i,t}] = f_{\mathbf{\Theta}_i}(\mathscr{H}_t)\,.$$

Goal: Learn Θ_i (parameters for a model) for all $i \in [N]$.

Observation Model

Model sequence dynamics via **autoregression**

$$\mathbb{P}[\mathbf{x}_{i,t} \mid \mathscr{H}_{i,t}] = F_{\mathbf{\Theta}_i}[\mathscr{H}_{i,t}] \implies \mathbb{E}[\mathbf{x}_{i,t} \mid \mathscr{H}_{i,t}] = f_{\mathbf{\Theta}_i}(\mathscr{H}_t)\,.$$

Goal: Learn Θ_i (parameters for a model) for all $i \in [N]$.

Typical assumption in sequence modeling: Observations $\mathbf{x}_{i,t}$ *come from repeated realizations of the same random source* ($F_{\mathbf{\Theta}_i}$ consistent across all i)

Observation Model

Model sequence dynamics via **autoregression**

$$\mathbb{P}[\mathbf{x}_{i,t} \mid \mathscr{H}_{i,t}] = F_{\mathbf{\Theta}_i}[\mathscr{H}_{i,t}] \implies \mathbb{E}[\mathbf{x}_{i,t} \mid \mathscr{H}_{i,t}] = f_{\mathbf{\Theta}_i}(\mathscr{H}_t)\,.$$

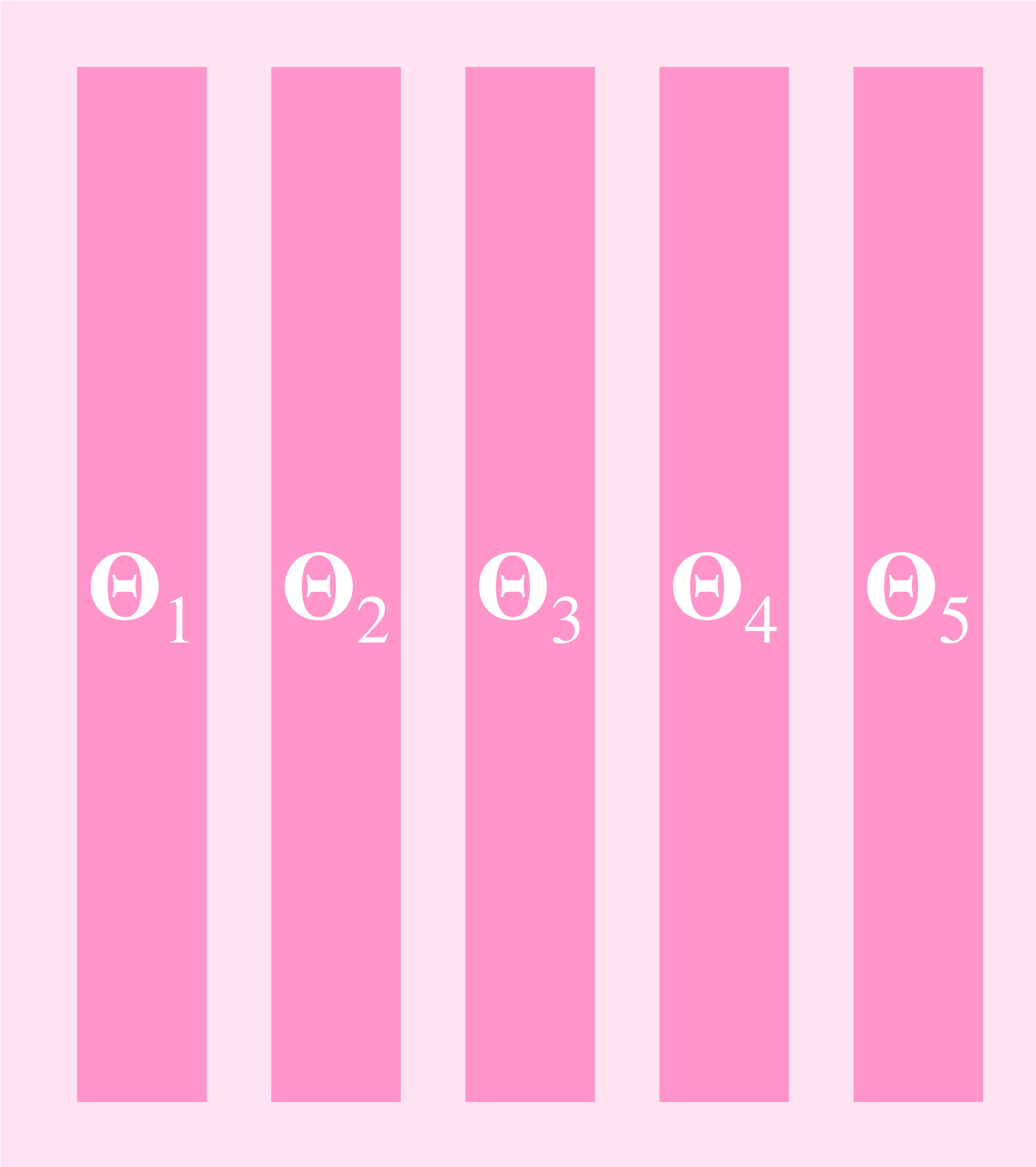
Goal: Learn Θ_i (parameters for a model) for all $i \in [N]$.

Typical assumption in sequence modeling: Observations $\mathbf{x}_{i,t}$ *come from repeated realizations of the same random source* ($F_{\mathbf{\Theta}_i}$ consistent across all i)

Reality: Each *sequence's dynamics may be highly distinct* (e.g., sick vs healthy)

Low Rank Sequence Learning

Low Rank Sequence Learning

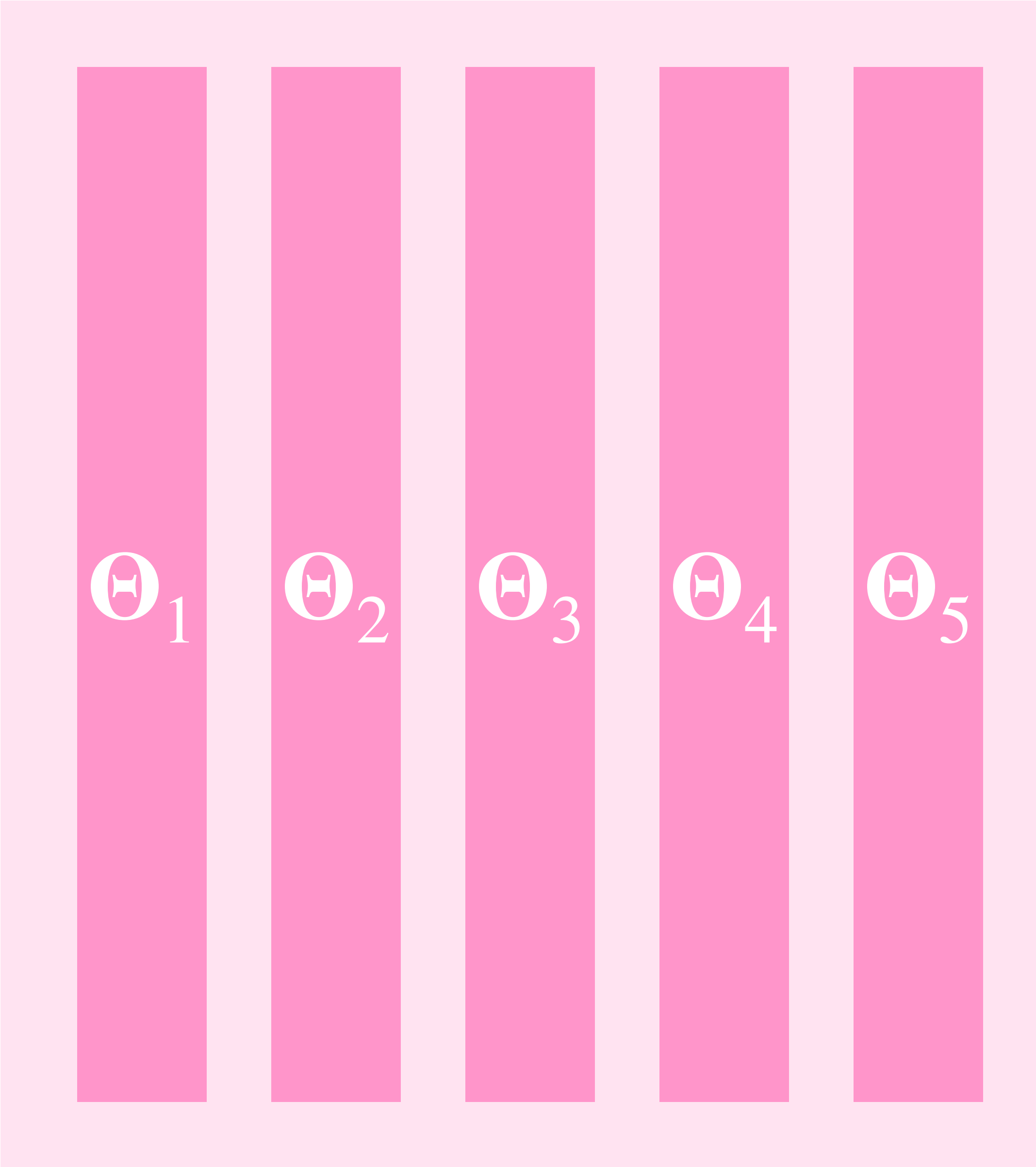


Θ_i: individual sequence parameter

Low Rank Sequence Learning

Constrain $\mathbf{B}$ to be of low rank:

$$\|\mathbf{B}\|_* = \sum_i \sigma_i(\mathbf{B}) \leq \lambda.$$

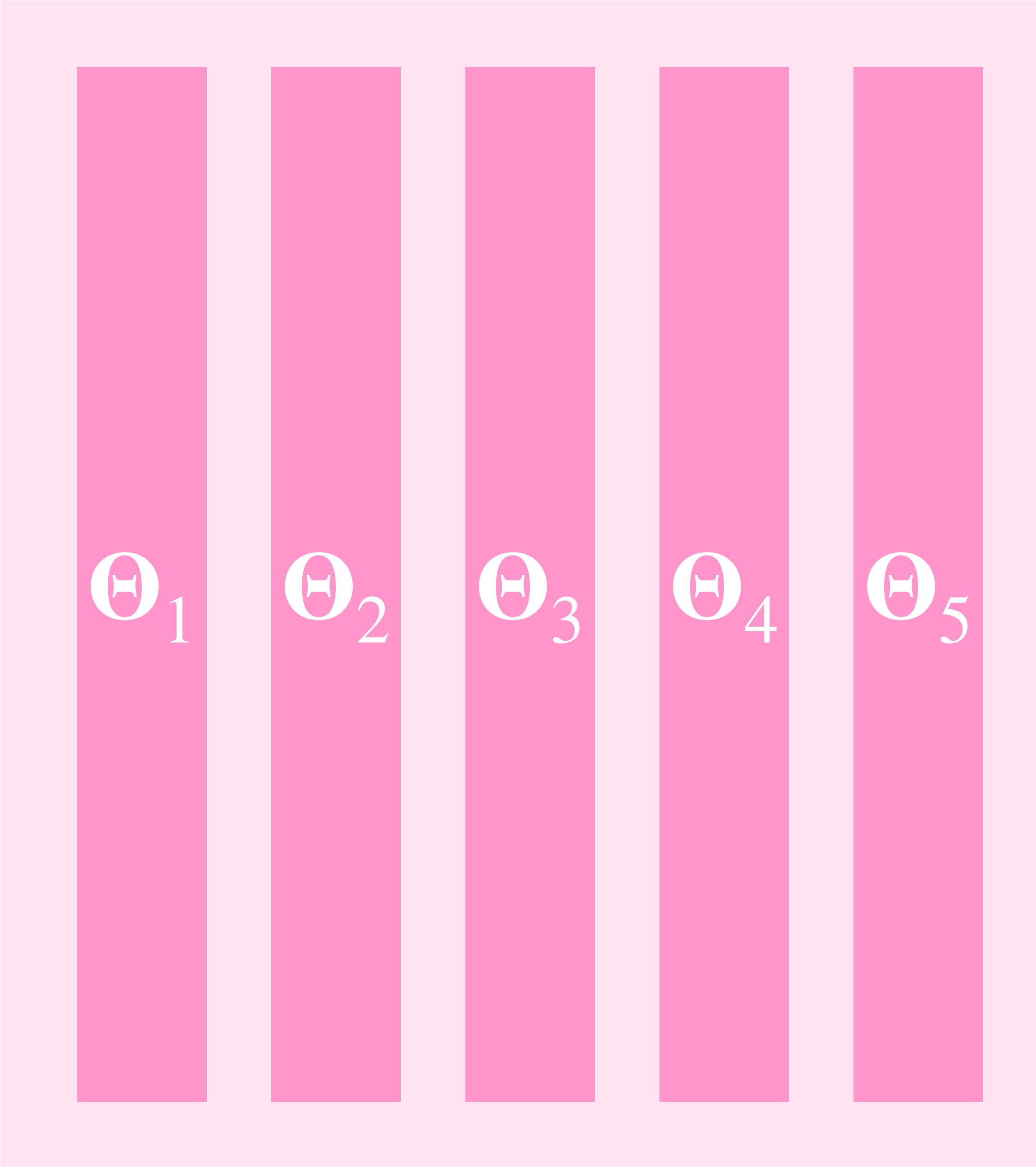


$\boldsymbol{\Theta}_i$: individual sequence parameter

Low Rank Sequence Learning

Constrain $\mathbf{B}$ to be of low rank:

$$\|\mathbf{B}\|_* = \sum_i \sigma_i(\mathbf{B}) \leq \lambda.$$



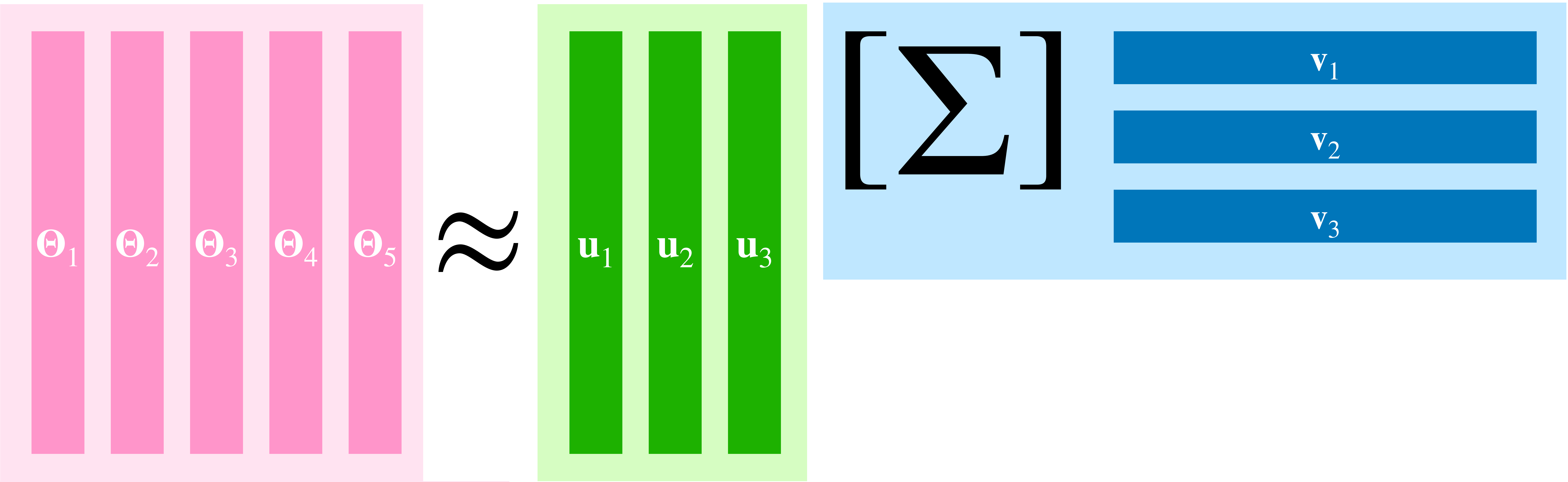
$\approx$

$\mathbf{\Theta}_i$: individual sequence parameter

Low Rank Sequence Learning

Constrain $\mathbf{B}$ to be of low rank:

$$\|\mathbf{B}\|_* = \sum_i \sigma_i(\mathbf{B}) \leq \lambda.$$

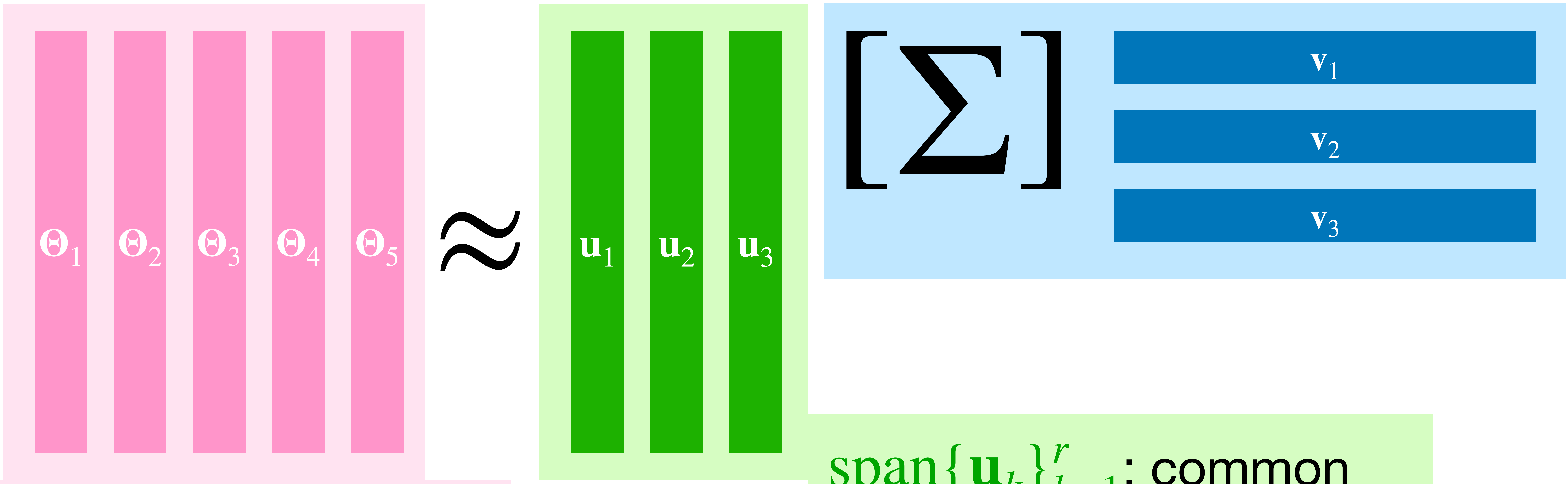


$\mathbf{\Theta}_i$: individual sequence parameter

Low Rank Sequence Learning

Constrain $\mathbf{B}$ to be of low rank:

$$\|\mathbf{B}\|_* = \sum_i \sigma_i(\mathbf{B}) \leq \lambda.$$



$\mathbf{\Theta}_i$: individual sequence parameter

$\mathrm{span}\{\mathbf{u}_k\}_{k=1}^r$: common domain of the sequences

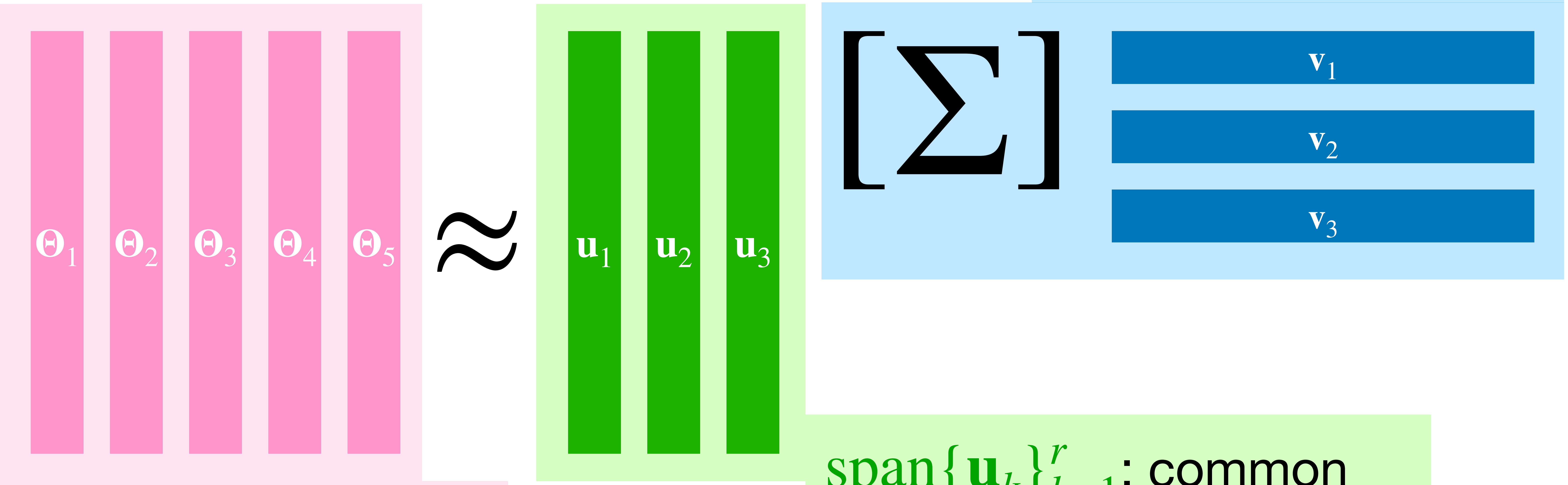
Low Rank Sequence Learning

Constrain $\mathbf{B}$ to be of low rank:

$$\|\mathbf{B}\|_* = \sum_i \sigma_i(\mathbf{B}) \leq \lambda.$$

$\mathbf{C} := \mathbf{\Sigma V}^* = [\mathbf{c}_i]_{i=1}^N \in \mathbb{R}^{r \times N}$

r-dimensional representation each sequence



$\mathrm{span}\{\mathbf{u}_k\}_{k=1}^r$: common domain of the sequences

$\mathbf{\Theta}_i$: individual sequence parameter

Observation Model

Observation Model

Parametric form of $F_{\boldsymbol{\Theta}_i}$ and $f_{\boldsymbol{\Theta}_i}$

$$\mathbb{E}[\mathbf{x}_{i,t} \mid \mathscr{H}_{i,t}] = f_i(\mathscr{H}_{i,t}) = \phi(\boldsymbol{\Theta}_i \xi_{i,t})$$

$$\xi_{i,t} = \mathrm{vec}(1, \{\mathbf{x}_{i,t-s}\}_{s=1}^{d}) \in \mathbb{R}^{Cd+1} \text{ (Past } d \text{ events)}$$

Observation Model

Parametric form of $F_{\mathbf{\Theta}_i}$ and $f_{\mathbf{\Theta}_i}$

$$\mathbb{E}[\mathbf{x}_{i,t} \mid \mathscr{H}_{i,t}] = f_i(\mathscr{H}_{i,t}) = \phi(\mathbf{\Theta}_i \xi_{i,t})$$

$$\xi_{i,t} = \text{vec}(1, \{\mathbf{x}_{i,t-s}\}_{s=1}^{d}) \in \mathbb{R}^{Cd+1} \text{ (Past } d \text{ events)}$$

ϕ monotone **link function**

- e.g., $\sigma(\cdot)$ for symbolic sequences, $\exp(\cdot)$ count process

Loosest choice to maintain problem convexity

Least Squares and Matrix Sensing

Least Squares and Matrix Sensing

Goal: Recover **true parameters** $\mathbf{B} = [\mathrm{vec}(\boldsymbol{\Theta}_i)]_{i=1}^{N}$ via observations $[\mathbf{x}_{i,t}]_{i\in[n],t\in[T]}$ obeying $\mathbb{E}[\mathbf{x}_{i,t} \,|\, \mathscr{H}_{i,t}] = \eta(\boldsymbol{\Theta}_i \xi_{i,t})$

Least Squares and Matrix Sensing

Goal: Recover **true parameters** $\mathbf{B} = [\text{vec}(\boldsymbol{\Theta}_i)]_{i=1}^{N}$ via observations $[\mathbf{x}_{i,t}]_{i\in[n],t\in[T]}$ obeying $\mathbb{E}[\mathbf{x}_{i,t} \mid \mathscr{H}_{i,t}] = \eta(\boldsymbol{\Theta}_i \xi_{i,t})$

$$\min_{\mathbf{B}\in\mathbb{R}^{d\times N}} \sum_{i,t} \|\mathbf{x}_{i,t} - \phi\left(\boldsymbol{\Theta}_i \xi_{i,t}\right)\|_2^2 \quad \text{s.t.} \quad \|\mathbf{B}\|_* \leq \lambda \qquad \textbf{(Least Squares)}$$

Least Squares and Matrix Sensing

Goal: Recover **true parameters** $\mathbf{B} = [\text{vec}(\mathbf{\Theta}_i)]_{i=1}^{N}$ via observations $[\mathbf{x}_{i,t}]_{i\in[n],t\in[T]}$ obeying $\mathbb{E}[\mathbf{x}_{i,t} \,|\, \mathscr{H}_{i,t}] = \eta(\mathbf{\Theta}_i \xi_{i,t})$

$$\min_{\mathbf{B}\in\mathbb{R}^{d\times N}} \sum_{i,t} \|\mathbf{x}_{i,t} - \phi\left(\mathbf{\Theta}_i \xi_{i,t}\right)\|_2^2 \quad \text{s.t.} \quad \|\mathbf{B}\|_* \leq \lambda \qquad \textbf{(Least Squares)}$$

Low-rank matrix recovery:

Indirect samples to $\mathbf{B}$ through linear measurement operators $\mathscr{A}_t$

$$\mathscr{A}_t(\mathbf{B}) := \text{vec}([\mathbf{R}_i \xi_{i,t}]_{i=1}^{N}) : \mathbb{R}^{m\times N} \to \mathbb{R}^{CN}$$

$$\mathbf{y}_t := \text{vec}([\mathbf{x}_{i,t}]_{i=1}^{N}) \in \mathbb{R}^{CN}$$

Least Squares and Matrix Sensing

Goal: Recover **true parameters** $\mathbf{B} = [\text{vec}(\mathbf{\Theta}_i)]_{i=1}^N$ via observations $[\mathbf{x}_{i,t}]_{i\in[n],t\in[T]}$ obeying $\mathbb{E}[\mathbf{x}_{i,t} \,|\, \mathscr{H}_{i,t}] = \eta(\mathbf{\Theta}_i \xi_{i,t})$

$$\min_{\mathbf{B}\in\mathbb{R}^{d\times N}} \sum_{i,t} \|\mathbf{x}_{i,t} - \phi\left(\mathbf{\Theta}_i \xi_{i,t}\right)\|_2^2 \quad \text{s.t.} \quad \|\mathbf{B}\|_* \leq \lambda \qquad \textbf{(Least Squares)}$$

Low-rank matrix recovery:

Indirect samples to $\mathbf{B}$ through linear measurement operators $\mathscr{A}_t$

$$\mathscr{A}_t(\mathbf{B}) := \text{vec}([\mathbf{R}_i \xi_{i,t}]_{i=1}^N) : \mathbb{R}^{m\times N} \to \mathbb{R}^{CN}$$

$$\mathbf{y}_t := \text{vec}([\mathbf{x}_{i,t}]_{i=1}^N) \in \mathbb{R}^{CN}$$

Observation: When $\phi = \text{Id}$, least squares same as matrix recovery

$$\min_{\mathbf{B}} \ell(\mathbf{B}) := \sum_{t=d+1}^{T} \|\mathscr{A}_t(\mathbf{B}) - \mathbf{y}_t\|_2^2 \quad \text{s.t.} \quad \|\mathbf{B}\|_* \leq \lambda$$

Least Squares and Matrix Sensing

Goal: Recover **true parameters** $\mathbf{B} = [\mathrm{vec}(\mathbf{\Theta}_i)]_{i=1}^N$ via observations $[\mathbf{x}_{i,t}]_{i\in[n],t\in[T]}$ obeying $\mathbb{E}[\mathbf{x}_{i,t} \mid \mathscr{H}_{i,t}] = \eta(\mathbf{\Theta}_i \xi_{i,t})$

$$\min_{\mathbf{B}\in\mathbb{R}^{d\times N}} \sum_{i,t} \|\mathbf{x}_{i,t} - \phi\left(\mathbf{\Theta}_i \xi_{i,t}\right)\|_2^2 \quad \text{s.t.} \quad \|\mathbf{B}\|_* \leq \lambda \qquad \textbf{(Least Squares)}$$

Low-rank matrix recovery:

Indirect samples to $\mathbf{B}$ through linear measurement operators $\mathscr{A}_t$

$$\mathscr{A}_t(\mathbf{B}) := \mathrm{vec}([\mathbf{R}_i \xi_{i,t}]_{i=1}^N) : \mathbb{R}^{m\times N} \to \mathbb{R}^{CN}$$

$$\mathbf{y}_t := \mathrm{vec}([\mathbf{x}_{i,t}]_{i=1}^N) \in \mathbb{R}^{CN}$$

Problem: With nonlinear ϕ, least squares *loses convexity* and parameter recovery guarantees

Observation: When $\phi = \mathrm{Id}$, least squares same as matrix recovery

$$\min_{\mathbf{B}} \ell(\mathbf{B}) := \sum_{t=d+1}^{T} \|\mathscr{A}_t(\mathbf{B}) - \mathbf{y}_t\|_2^2 \quad \text{s.t.} \quad \|\mathbf{B}\|_* \leq \lambda$$

Least Squares and Matrix Sensing

Goal: Recover **true parameters** $\mathbf{B} = [\mathrm{vec}(\mathbf{\Theta}_i)]_{i=1}^N$ via observations $[\mathbf{x}_{i,t}]_{i\in[n],t\in[T]}$ obeying $\mathbb{E}[\mathbf{x}_{i,t} \,|\, \mathscr{H}_{i,t}] = \eta(\mathbf{\Theta}_i \xi_{i,t})$

$$\min_{\mathbf{B}\in\mathbb{R}^{d\times N}} \sum_{i,t} \|\mathbf{x}_{i,t} - \phi\left(\mathbf{\Theta}_i \xi_{i,t}\right)\|_2^2 \quad \text{s.t.} \quad \|\mathbf{B}\|_* \le \lambda \qquad \textbf{(Least Squares)}$$

Low-rank matrix recovery:

Indirect samples to $\mathbf{B}$ through linear measurement operators $\mathscr{A}_t$

$$\mathscr{A}_t(\mathbf{B}) := \mathrm{vec}([\mathbf{R}_i \xi_{i,t}]_{i=1}^N) : \mathbb{R}^{m\times N} \to \mathbb{R}^{CN}$$

$$\mathbf{y}_t := \mathrm{vec}([\mathbf{x}_{i,t}]_{i=1}^N) \in \mathbb{R}^{CN}$$

Problem: With nonlinear ϕ, least squares *loses convexity* and parameter recovery guarantees

$\Longrightarrow$ **monotone VI formulation**

Observation: When $\phi = \mathrm{Id}$, least squares same as matrix recovery

$$\min_{\mathbf{B}} \ell(\mathbf{B}) := \sum_{t=d+1}^{T} \|\mathscr{A}_t(\mathbf{B}) - \mathbf{y}_t\|_2^2 \quad \text{s.t.} \quad \|\mathbf{B}\|_* \le \lambda$$

Monotone VI

A **monotone vector field** on $\mathbb{R}^m$ with modulus of convexity β is a vector field $G : \mathbb{R}^m \to \mathbb{R}^m$ such that

$$\langle G(\mathbf{x}) - G(\mathbf{x}'), \mathbf{x} - \mathbf{x}' \rangle \geq \beta \|\mathbf{x} - \mathbf{x}'\| \qquad \forall\ \mathbf{x}, \mathbf{x}' \in \mathscr{X}$$

$\mathscr{X} \subseteq \mathbb{R}^m$ compact, convex set (e.g. nuclear ball)

Tuple $(G, \mathscr{X})$: **Monotone Variational Inequality Problem**

$$\forall \mathbf{x} \in \mathscr{X}, \langle G(\mathbf{x}), \mathbf{x} - \mathbf{x}^* \rangle \geq 0 \implies \mathbf{x}^* \text{ is a weak solution to } (G, \mathscr{X})$$

Monotone VI for Parameter Recovery

Monotone VI for Parameter Recovery

Goal: Construct a **Monotone VI** which has solution true parameters $\mathbf{B} = [\mathrm{vec}(\boldsymbol{\Theta}_i)]_{i=1}^N$

Monotone VI for Parameter Recovery

Goal: Construct a **Monotone VI** which has solution true parameters $\mathbf{B} = [\mathrm{vec}(\boldsymbol{\Theta}_i)]_{i=1}^N$

$$\mathscr{A}_t(\mathbf{B}) := \mathrm{vec}([\mathbf{R}_i \xi_{i,t}]_{i=1}^N) : \mathbb{R}^{m \times N} \to \mathbb{R}^{CN}$$

$$\mathscr{A}_t^*(\mathbf{y}) := [\mathrm{vec}([x_{i,c} \boldsymbol{\Theta}_i^T \mathbf{e}_c])_{c=1}^C)]_{i=1}^N : \mathbb{R}^{CN} \to \mathbb{R}^{(C^2 d + C) \times N}$$

$$\mathbf{y}_t := \mathrm{vec}([\mathbf{x}_{i,t}]_{i=1}^N) \in \mathbb{R}^{CN}$$

Monotone VI for Parameter Recovery

Goal: Construct a **Monotone VI** which has solution true parameters $\mathbf{B} = [\mathrm{vec}(\boldsymbol{\Theta}_i)]_{i=1}^N$

$$\mathscr{A}_t(\mathbf{B}) := \mathrm{vec}([\mathbf{R}_i \xi_{i,t}]_{i=1}^N) : \mathbb{R}^{m \times N} \to \mathbb{R}^{CN}$$

$$\mathscr{A}_t^*(\mathbf{y}) := [\mathrm{vec}([x_{i,c} \boldsymbol{\Theta}_i^T \mathbf{e}_c])_{c=1}^C)]_{i=1}^N : \mathbb{R}^{CN} \to \mathbb{R}^{(C^2 d + C) \times N}$$

$$\mathbf{y}_t := \mathrm{vec}([\mathbf{x}_{i,t}]_{i=1}^N) \in \mathbb{R}^{CN}$$

Consider now matrix-space vector field $\Psi : \mathbb{R}^{(C^2 d + C) \times N} \to \mathbb{R}^{(C^2 d + C) \times N}$

$$\boxed{\Psi(\mathbf{B}) = \mathbb{E}_t[\mathscr{A}_t^*(\phi(\mathscr{A}_t(\mathbf{B})) - \mathbf{y}_t)]}$$

Monotone VI for Parameter Recovery

Goal: Construct a **Monotone VI** which has solution true parameters $\mathbf{B} = [\text{vec}(\boldsymbol{\Theta}_i)]_{i=1}^N$

$$\mathscr{A}_t(\mathbf{B}) := \text{vec}([\mathbf{R}_i \xi_{i,t}]_{i=1}^N) : \mathbb{R}^{m \times N} \to \mathbb{R}^{CN}$$

$$\mathscr{A}_t^*(\mathbf{y}) := [\text{vec}([x_{i,c} \boldsymbol{\Theta}_i^T \mathbf{e}_c])_{c=1}^C)]_{i=1}^N : \mathbb{R}^{CN} \to \mathbb{R}^{(C^2 d + C) \times N}$$

$$\mathbf{y}_t := \text{vec}([\mathbf{x}_{i,t}]_{i=1}^N) \in \mathbb{R}^{CN}$$

Consider now matrix-space vector field $\Psi : \mathbb{R}^{(C^2 d + C) \times N} \to \mathbb{R}^{(C^2 d + C) \times N}$

$$\boxed{\Psi(\mathbf{B}) = \mathbb{E}_t[\mathscr{A}_t^*(\phi(\mathscr{A}_t(\mathbf{B})) - \mathbf{y}_t)]}$$

Claim: If the observed data obey $\mathbb{E}[\mathbf{x}_{i,t} \,|\, \mathscr{H}_{i,t}] = \eta(\boldsymbol{\Theta}_i \xi_{i,t})$, then true parameters are a zero of Ψ.

Monotone VI for Parameter Recovery

Goal: Construct a **Monotone VI** which has solution true parameters $\mathbf{B} = [\text{vec}(\mathbf{\Theta}_i)]_{i=1}^N$

$$\mathscr{A}_t(\mathbf{B}) := \text{vec}([\mathbf{R}_i \xi_{i,t}]_{i=1}^N) : \mathbb{R}^{m \times N} \to \mathbb{R}^{CN}$$

$$\mathscr{A}_t^*(\mathbf{y}) := [\text{vec}([x_{i,c} \mathbf{\Theta}_i^T \mathbf{e}_c])_{c=1}^C)]_{i=1}^N : \mathbb{R}^{CN} \to \mathbb{R}^{(C^2 d + C) \times N}$$

$$\mathbf{y}_t := \text{vec}([\mathbf{x}_{i,t}]_{i=1}^N) \in \mathbb{R}^{CN}$$

Consider now matrix-space vector field $\Psi : \mathbb{R}^{(C^2 d + C) \times N} \to \mathbb{R}^{(C^2 d + C) \times N}$

$$\boxed{\Psi(\mathbf{B}) = \mathbb{E}_t[\mathscr{A}_t^*(\phi(\mathscr{A}_t(\mathbf{B})) - \mathbf{y}_t)]}$$

Claim: If the observed data obey $\mathbb{E}[\mathbf{x}_{i,t} \,|\, \mathscr{H}_{i,t}] = \eta(\mathbf{\Theta}_i \xi_{i,t})$, then true parameters are a zero of Ψ.

Special Case: $\phi = \mathbf{Id} \implies$ VI is **least squares gradient field** $\Psi(\mathbf{X}) = \nabla_{\mathbf{X}}[\ell(\mathbf{X})]$.

Monotone VI for Parameter Recovery

Goal: Construct a **Monotone VI** which has solution true parameters $\mathbf{B} = [\mathrm{vec}(\mathbf{\Theta}_i)]_{i=1}^N$

$$\mathscr{A}_t(\mathbf{B}) := \mathrm{vec}([\mathbf{R}_i \xi_{i,t}]_{i=1}^N) : \mathbb{R}^{m \times N} \to \mathbb{R}^{CN}$$

$$\mathscr{A}_t^*(\mathbf{y}) := [\mathrm{vec}([x_{i,c} \mathbf{\Theta}_i^T \mathbf{e}_c])_{c=1}^C)]_{i=1}^N : \mathbb{R}^{CN} \to \mathbb{R}^{(C^2 d + C) \times N}$$

$$\mathbf{y}_t := \mathrm{vec}([\mathbf{x}_{i,t}]_{i=1}^N) \in \mathbb{R}^{CN}$$

Solve $(\Psi, \{\|\cdot\|_* \leq \lambda\})$
via first order methods

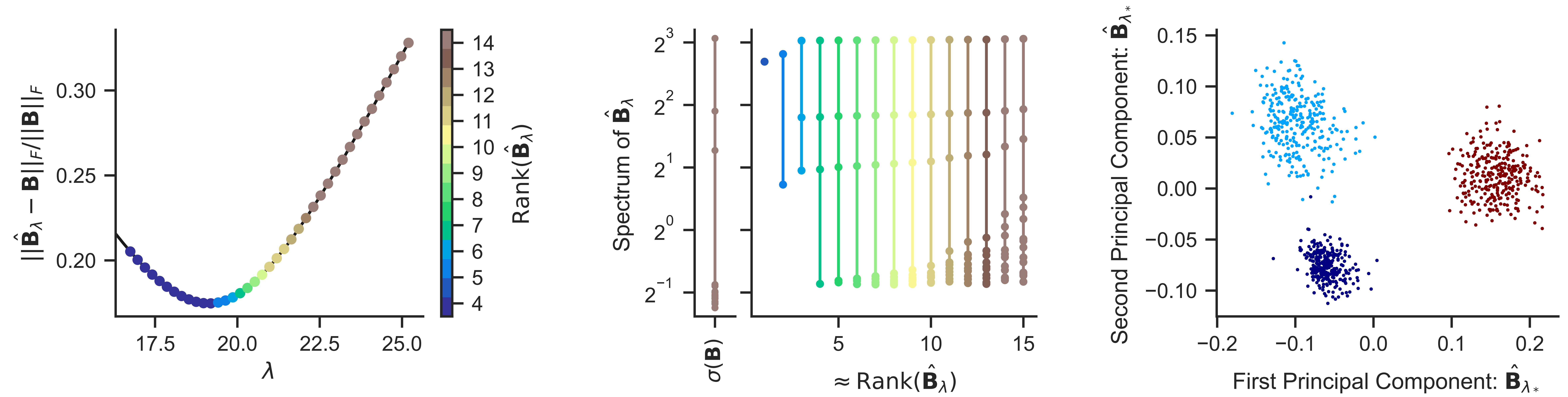
Consider now matrix-space vector field $\Psi : \mathbb{R}^{(C^2 d + C) \times N} \to \mathbb{R}^{(C^2 d + C) \times N}$

$$\boxed{\Psi(\mathbf{B}) = \mathbb{E}_t[\mathscr{A}_t^*(\phi(\mathscr{A}_t(\mathbf{B})) - \mathbf{y}_t)]}$$

Claim: If the observed data obey $\mathbb{E}[\mathbf{x}_{i,t} \,|\, \mathscr{H}_{i,t}] = \eta(\mathbf{\Theta}_i \xi_{i,t})$, then true parameters are a zero of Ψ.

Special Case: $\phi = \mathbf{Id} \implies$ VI is **least squares gradient field** $\Psi(\mathbf{X}) = \nabla_{\mathbf{X}}[\ell(\mathbf{X})]$.

Synthetic Sequences



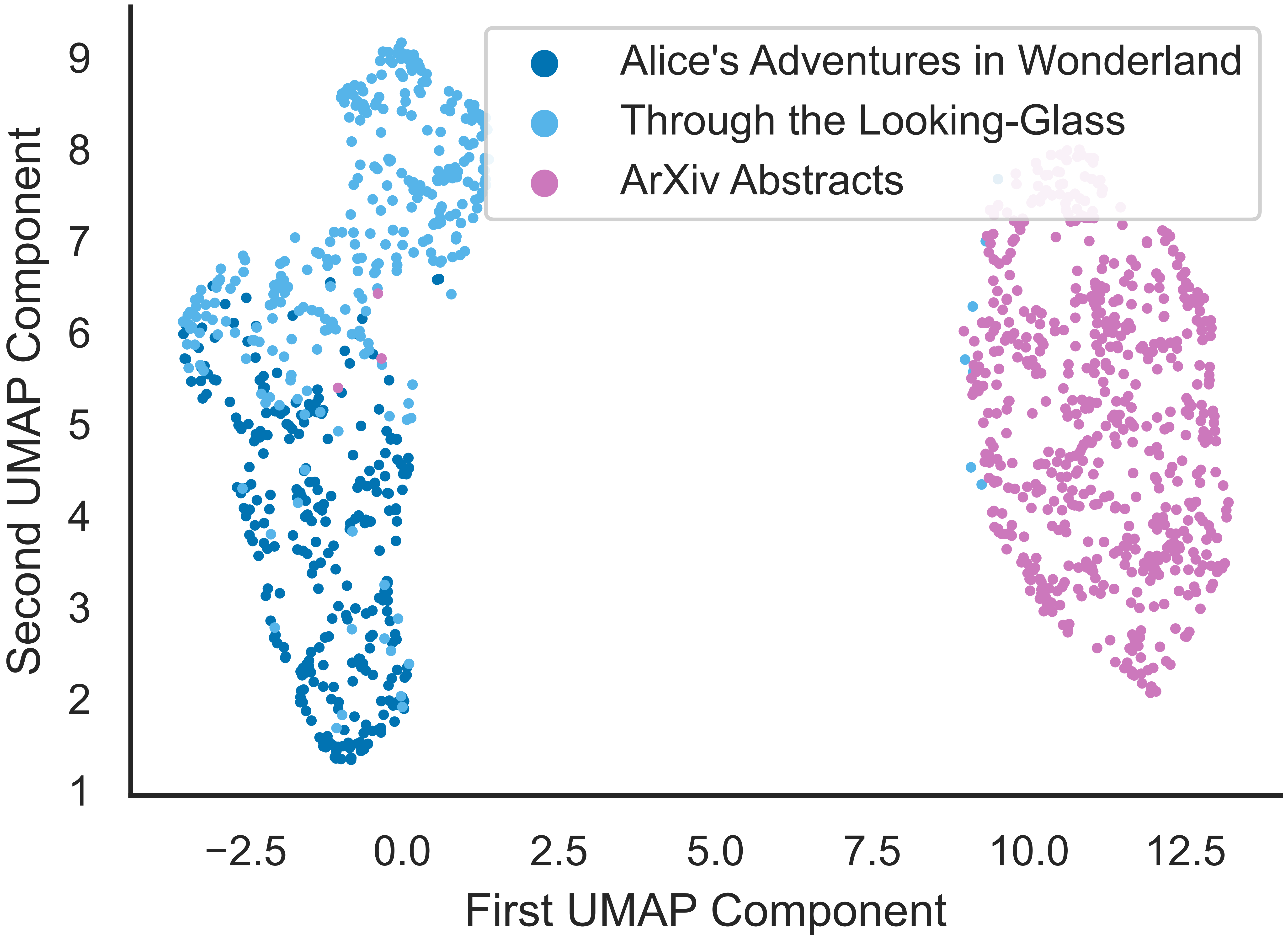
Parameter recovery for univariate time series from $k = 3$ classes

Left: Reconstruction error and rank across nuclear constraint λ

Center: Singular values of true and recovered parameter matrices

Right: First two principal components of the best recovery, with original class labels

Symbolic Sequences



Autoregressive language modeling of paragraphs (≈ 500 chars) from

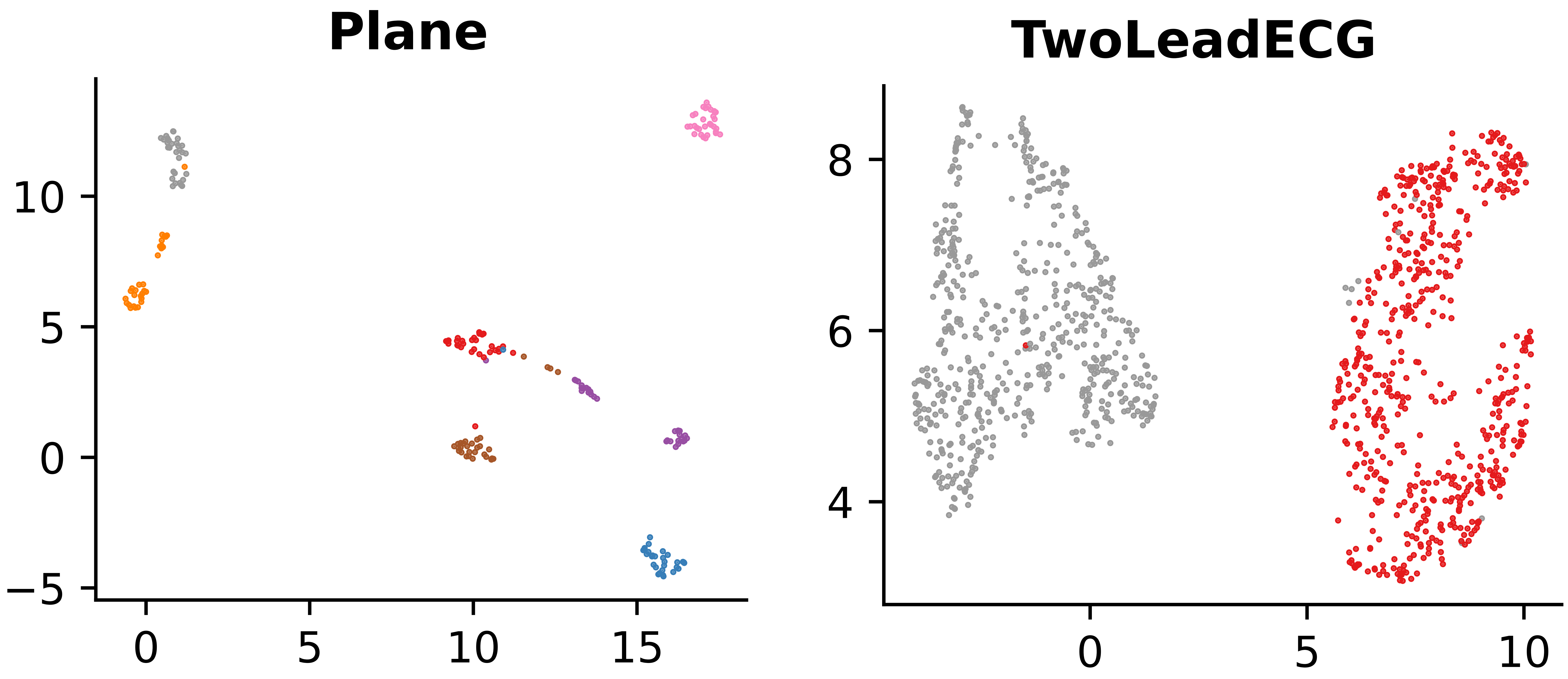
A. *Alice's Adventures in Wonderland* $(n = 228)$

B. *Through the Looking Glass* $(n = 316)$

C. ML abstracts from *ArXiv* $(n = 600)$

Two distinct clusters. Paragraph embeddings of fictional works not clearly separable ($d = 75$, $r = 3$)

Real Time Series



Low rank representation (projected into 2D) for UCR time series data with $d = 20$

Plane: Sensor readings from different aircraft | *TwoLeadECG:* Electrocardiograms

Nonlinear Sequence Data Embedding by Monotone Variational Inequality

Jonathan Y. Zhou, Yao Xie
School of Industrial and Systems Engineering
Georgia Institute to Technology

ICLR 2025 | Apr. 24-28 | Singapore EXPO

Georgia Tech College of Engineering
H. Milton Stewart School of
Industrial and Systems Engineering