Title

Zitong Zhang
May 28, 2020

Networks

- Pictures of social networks
- Pictures of neural networks
- Pictures of static and dynamic networks

Dynamic networks

Pictures of:

- Networks with slowly-changing communities & connecting probabilities
  - Include related literatures
- <u>Growing networks</u> in nervous system
  - However, not many literatures

Growing networks

- Picture of real growing neural network (Cell paper)
- Pictures of their results
  - Prior expectation: roles of nodes
- Weakness of their method

Outline

- Model growing networks via stochastic block model
- Inference procedure
- Simulation and application to real data

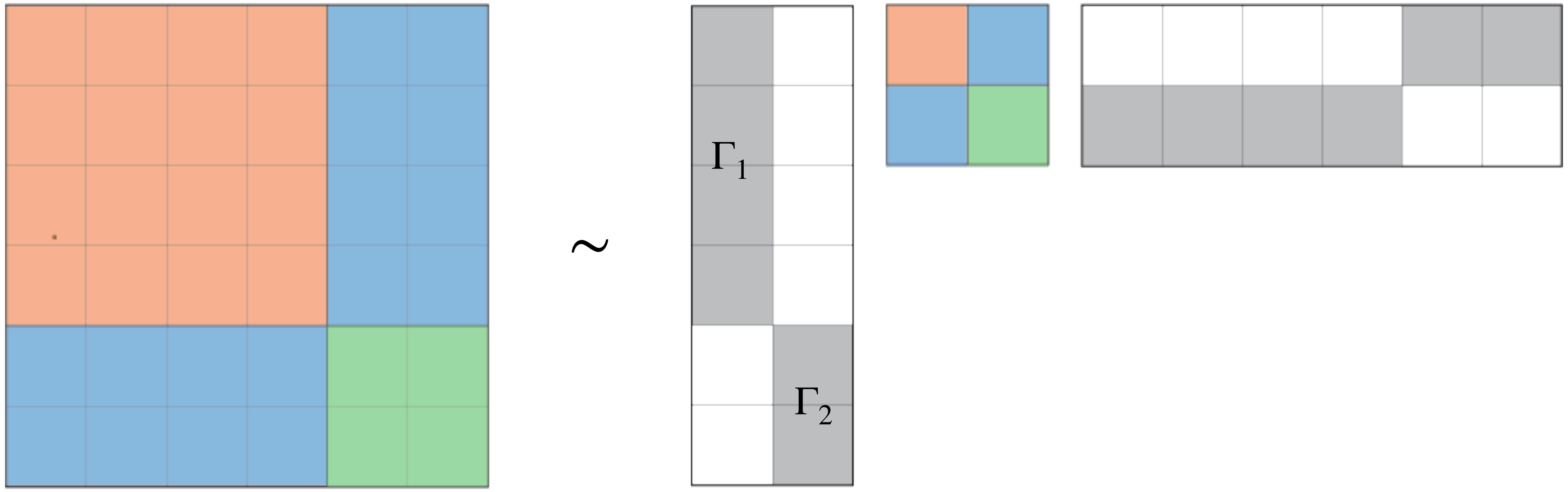
Modeling the growing networks

Introduce terms and notations through pictures:

- Observed neural activity: N_{i,j}
- Roles of nodes: z_i
- “Bridge” between multiple networks and common cell types — — connecting pattern: \lambda_{q,l}

Graphical representation of the model

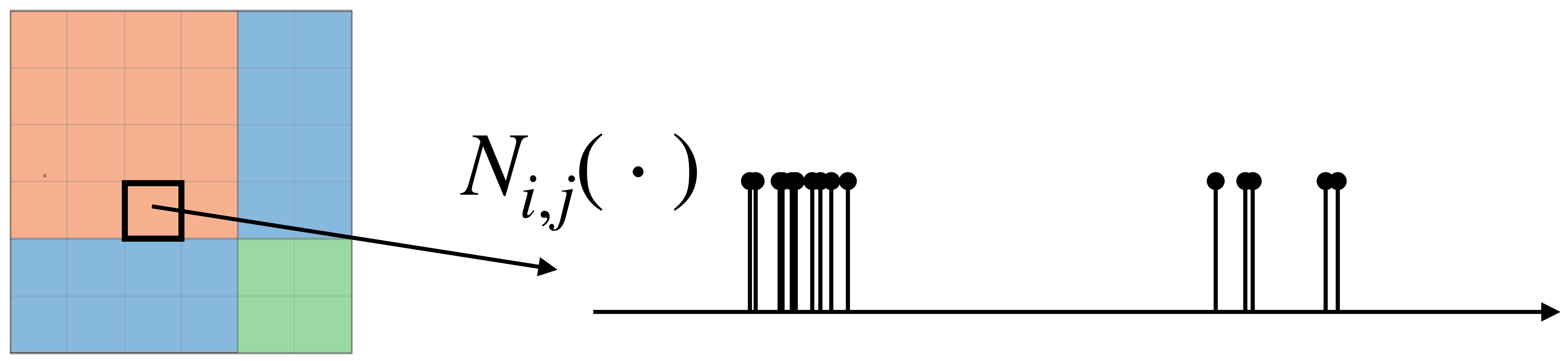
Observed $N_{i,j}(\cdot),\ i, j \in [n]$ **Membership** z_i **Connecting pattern** $\lambda_{q,l}(\cdot)$

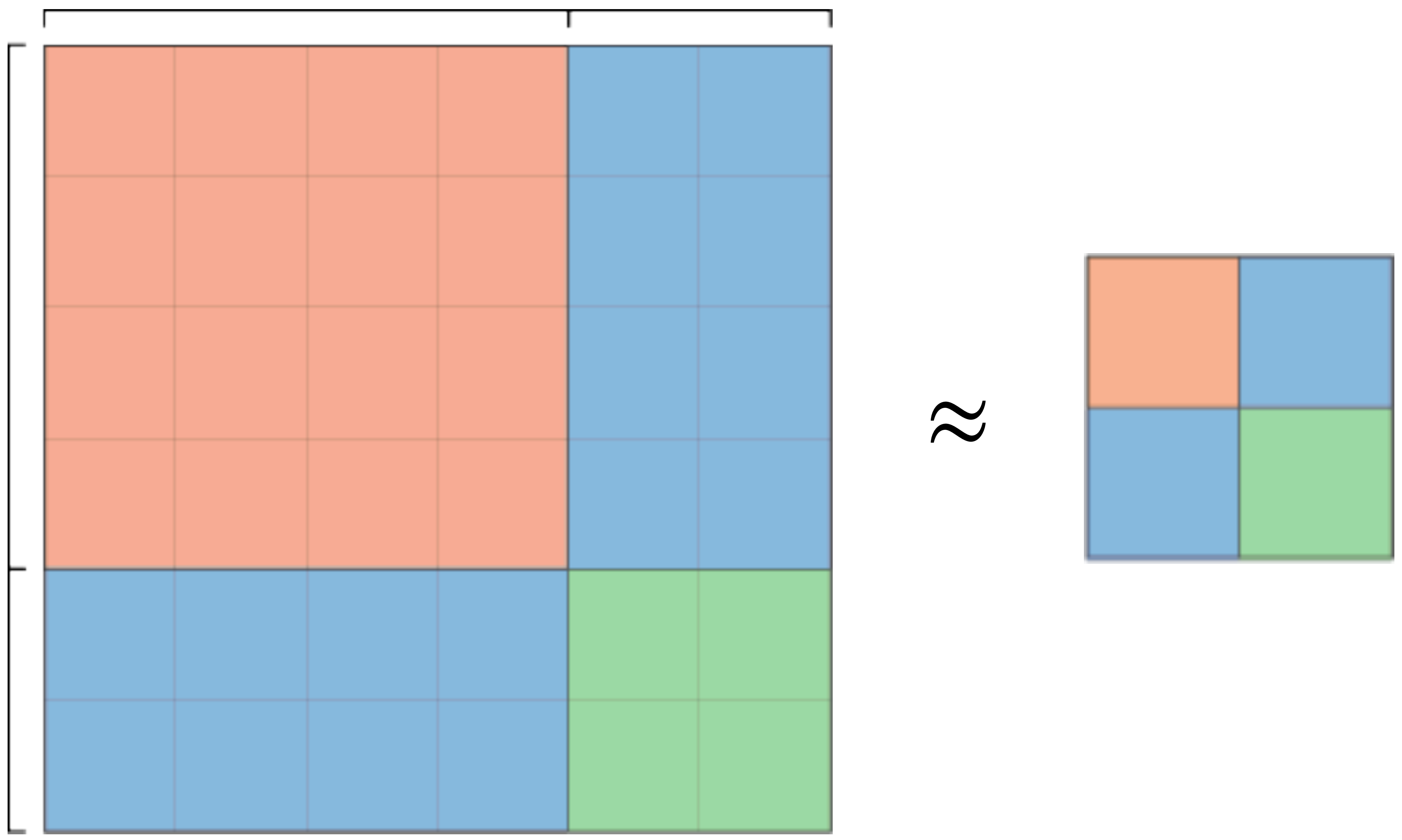


The conditional density of $N_{i,j}(\cdot)$ **given** $z_i = q$ **and** $z_j = l$ **is** $\lambda_{q,l}(\cdot)$

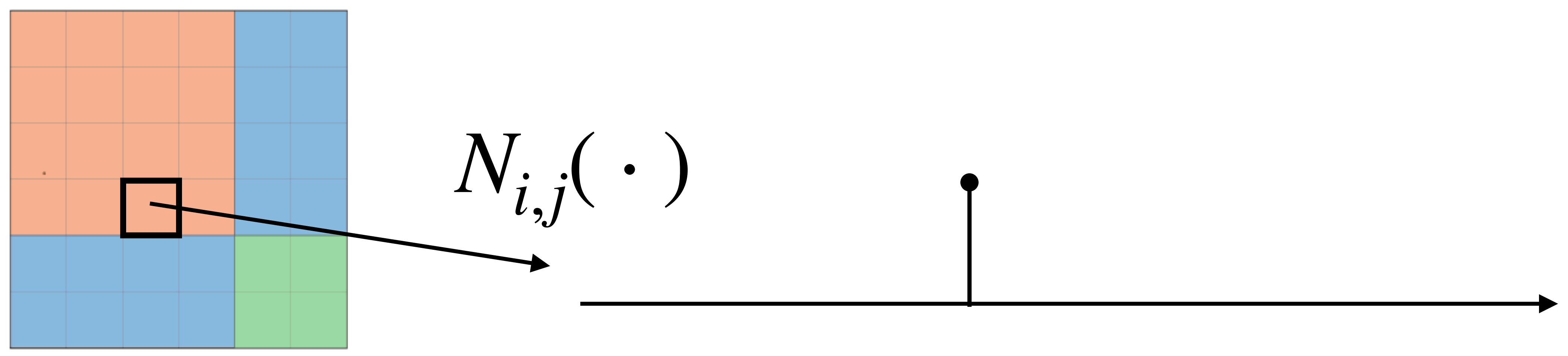
(Matias et al, 2018)

There is at most one event in each point process

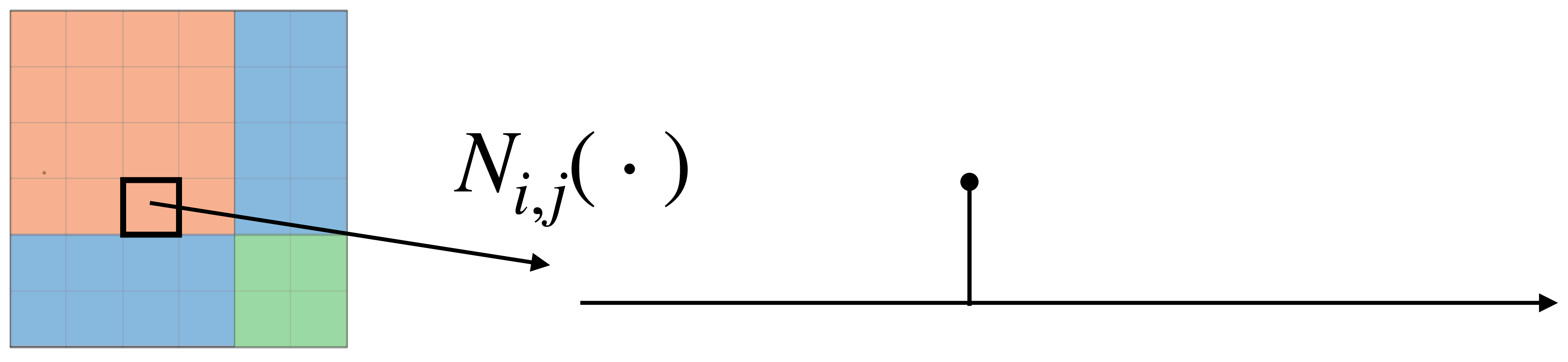


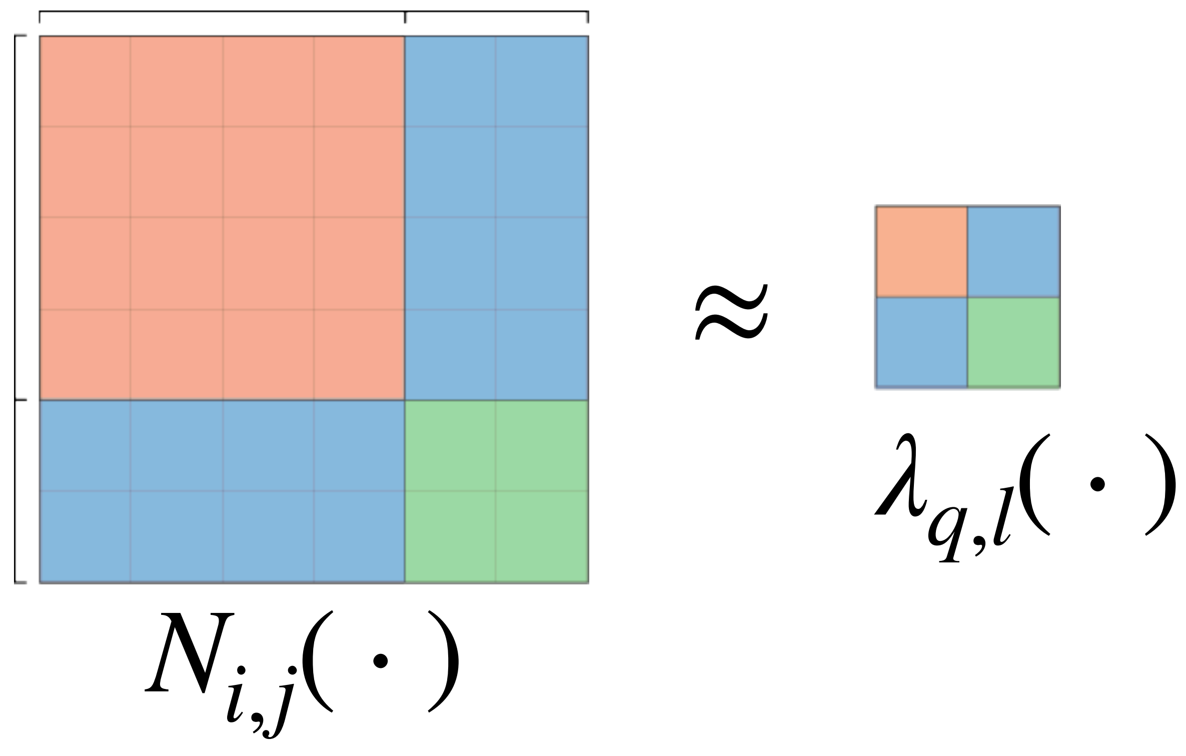


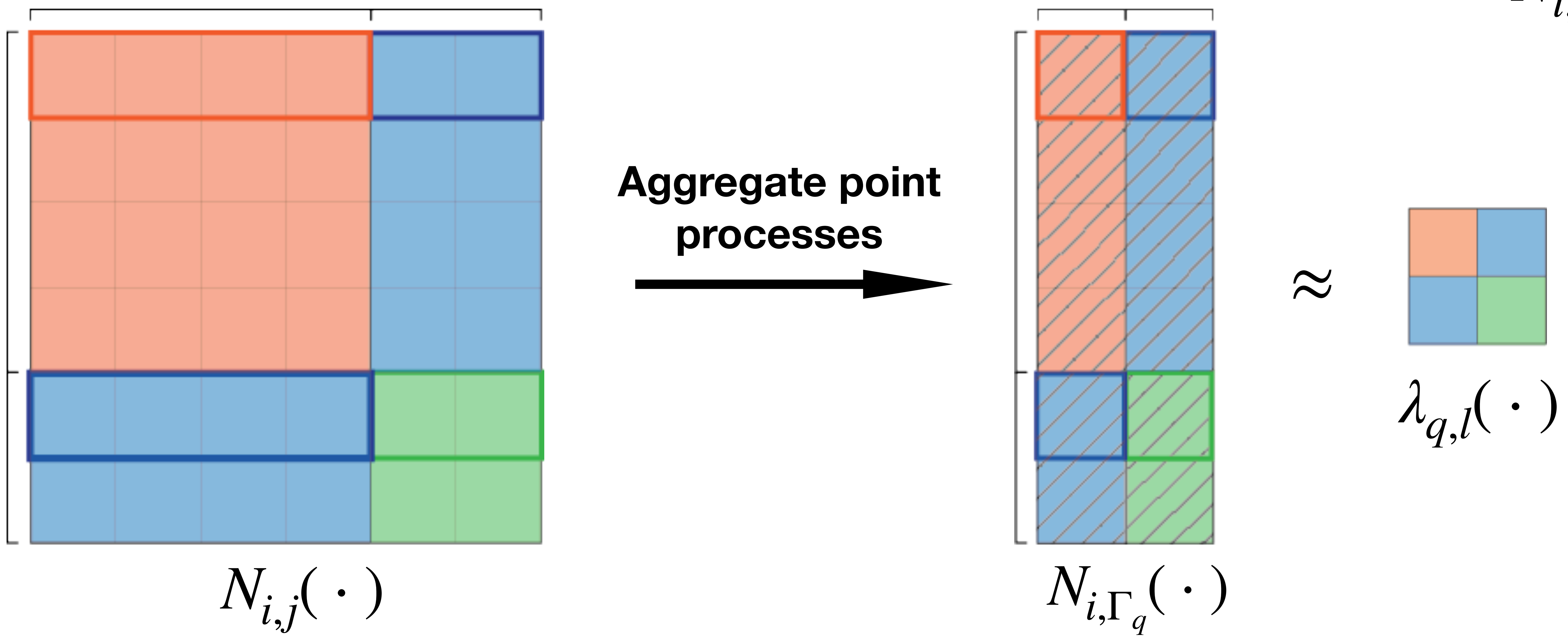
There is at most one event in each point process



There is at most one event in each point process



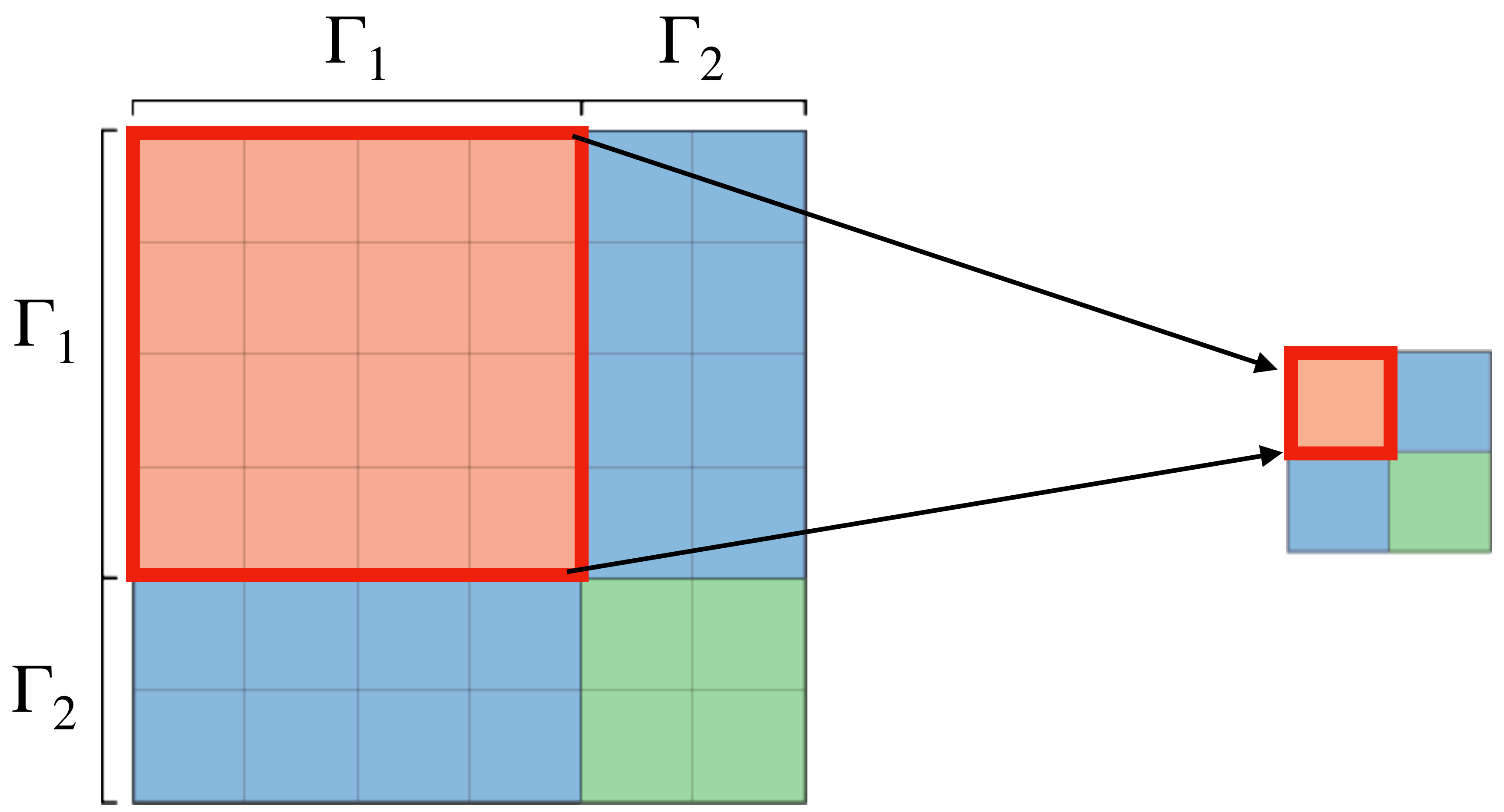




Inference

Approximate k-means method

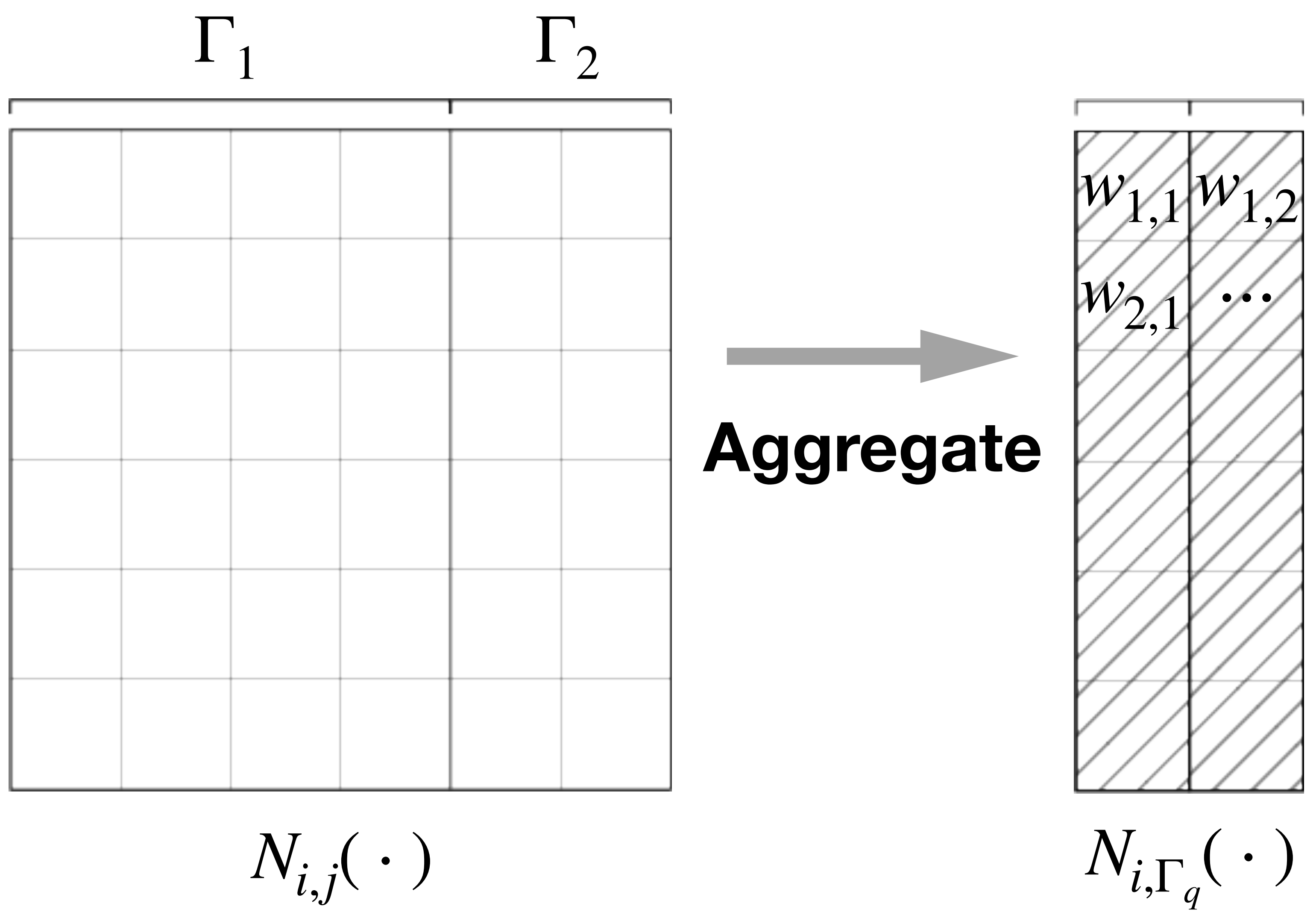
Re-center based on current clustering



Kernel density estimation

$$\hat{f}_h(x) = \frac{1}{nh}\sum_{i=1}^{n} K(\frac{x - x_i}{h})$$

Re-cluster based on estimated connecting patterns and current clustering

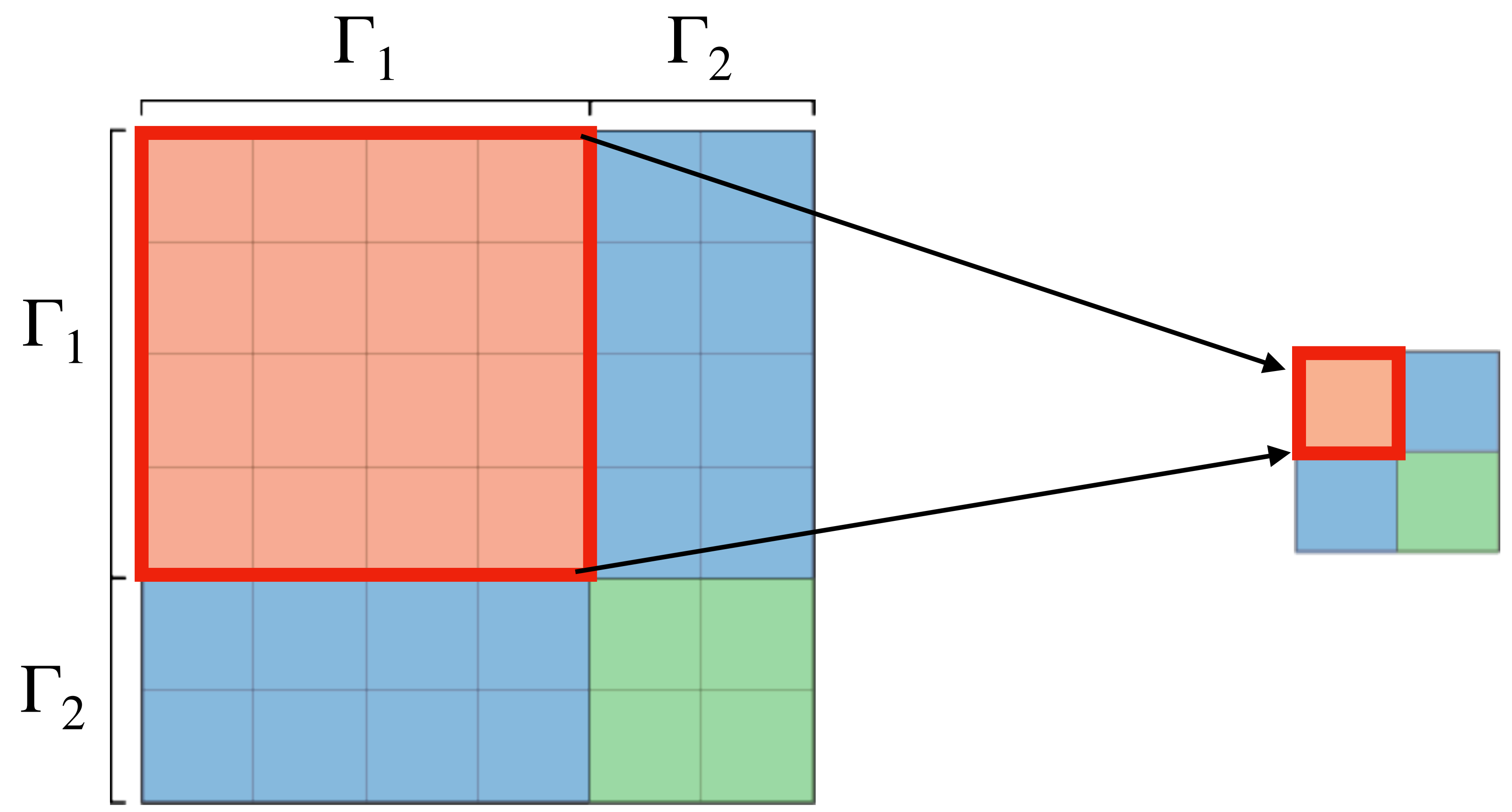


$$\sum_q w_{1,q} = 1$$

Inference

Approximate k-means method

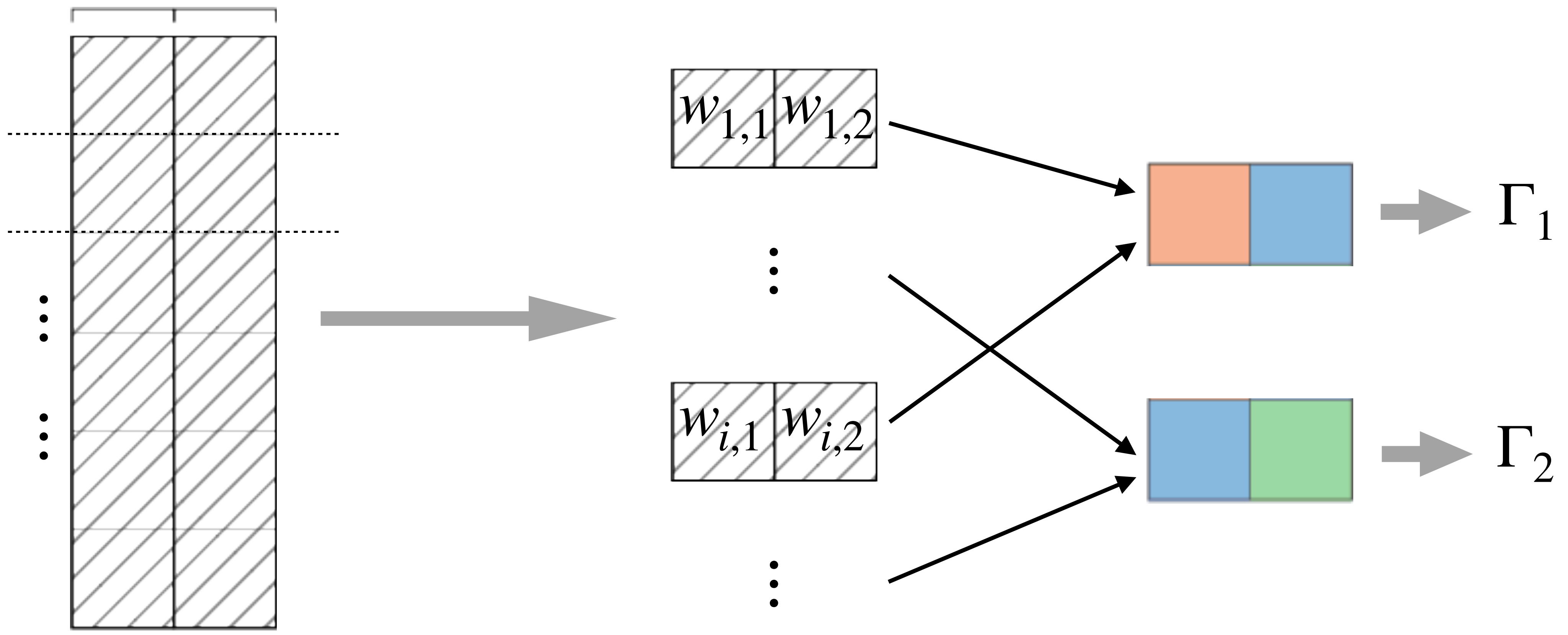
Re-center based on current clustering



Kernel density estimation

$$\hat{f}_h(x) = \frac{1}{nh}\sum_{i=1}^{n} K(\frac{x - x_i}{h})$$

Re-cluster based on estimated connecting patterns and current clustering

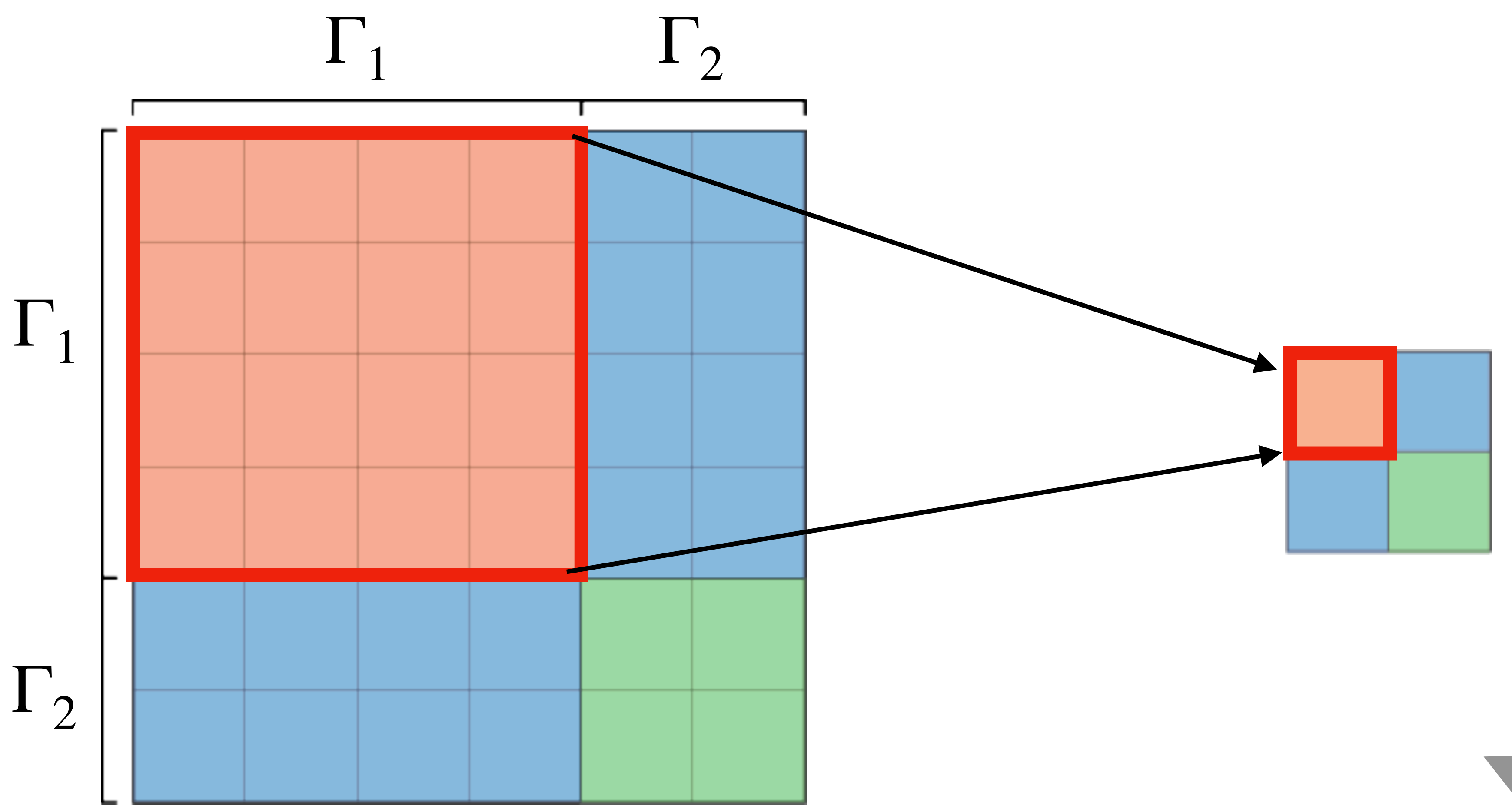


$$dist(i, \Gamma_q) = \sum_l w_{i,l} \cdot d(N_{i,\Gamma_l}, \lambda_{q,l}) = \sum_l w_{i,l} \cdot \|f_{N_{i,\Gamma_l}} - \lambda_{q,l}\|_2$$

Inference

Approximate k-means method

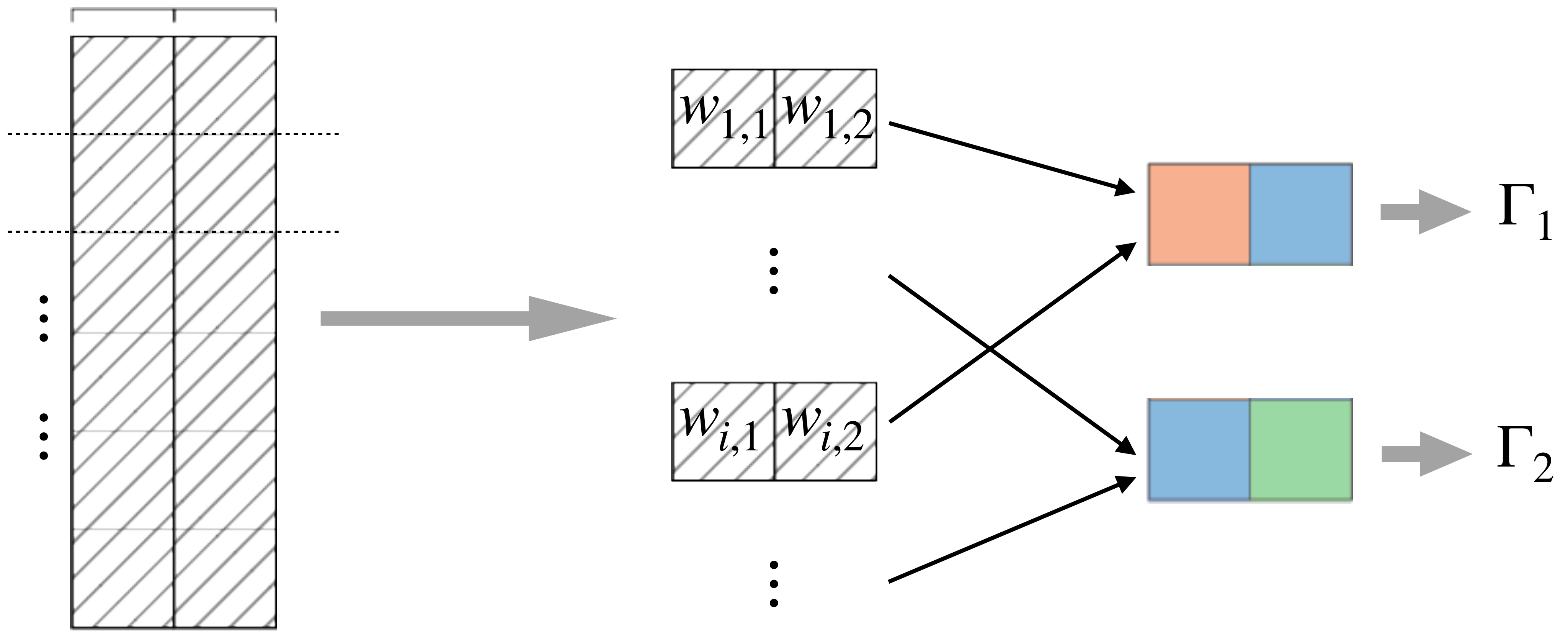
Re-center based on current clustering



Kernel density estimation

$$\hat{f}_h(x) = \frac{1}{nh}\sum_{i=1}^{n} K(\frac{x - x_i}{h})$$

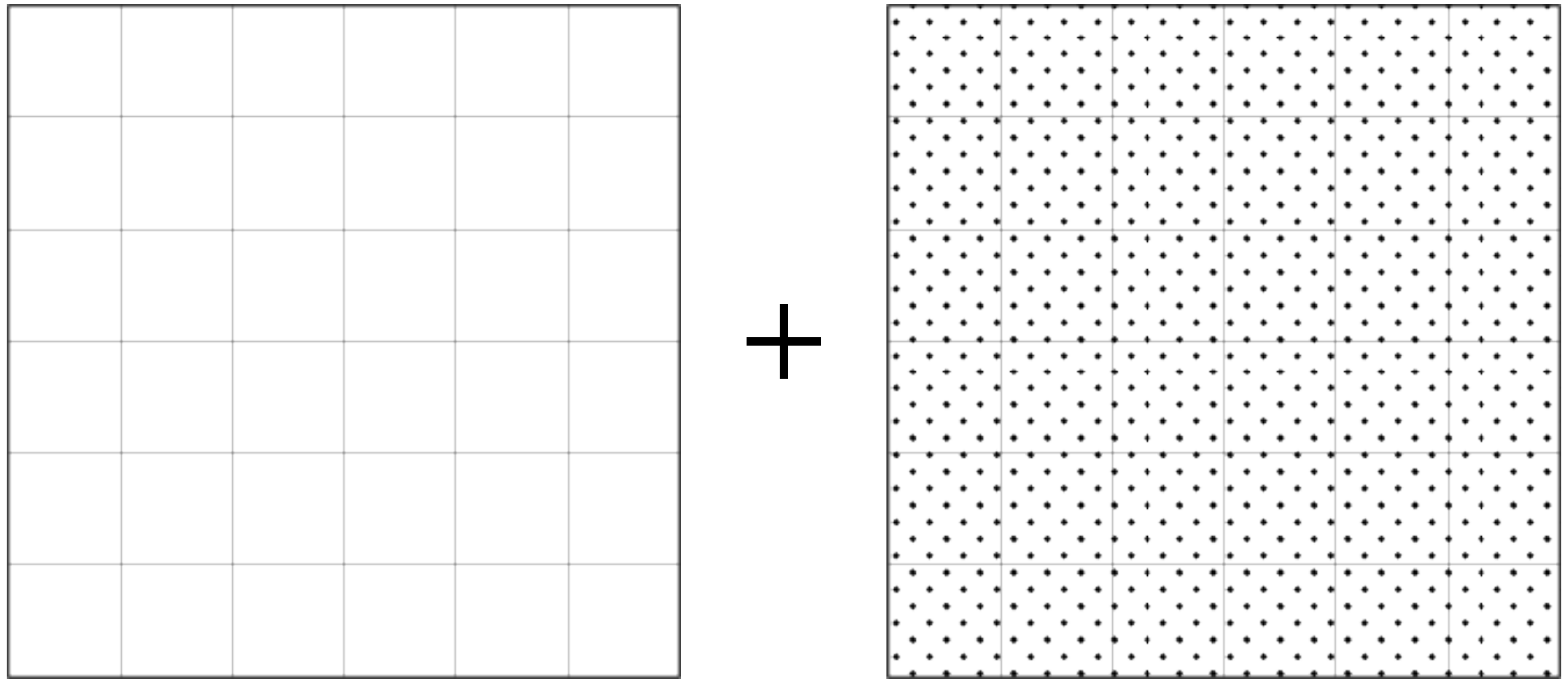
Re-cluster based on estimated connecting patterns and current clustering



$$dist(i, \Gamma_q) = \sum_l w_{i,l} \cdot d(N_{i,\Gamma_l}, \lambda_{q,l}) = \sum_l w_{i,l} \cdot \|f_{N_{i,\Gamma_l}} - \lambda_{q,l}\|_2$$

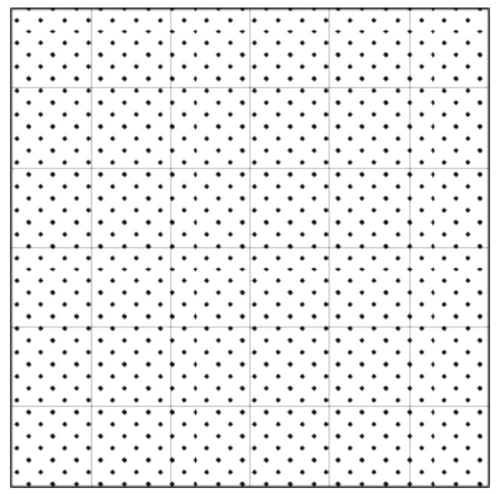
Time shifts associated with each edge

- Pictures from real data showing the existence of time shifts



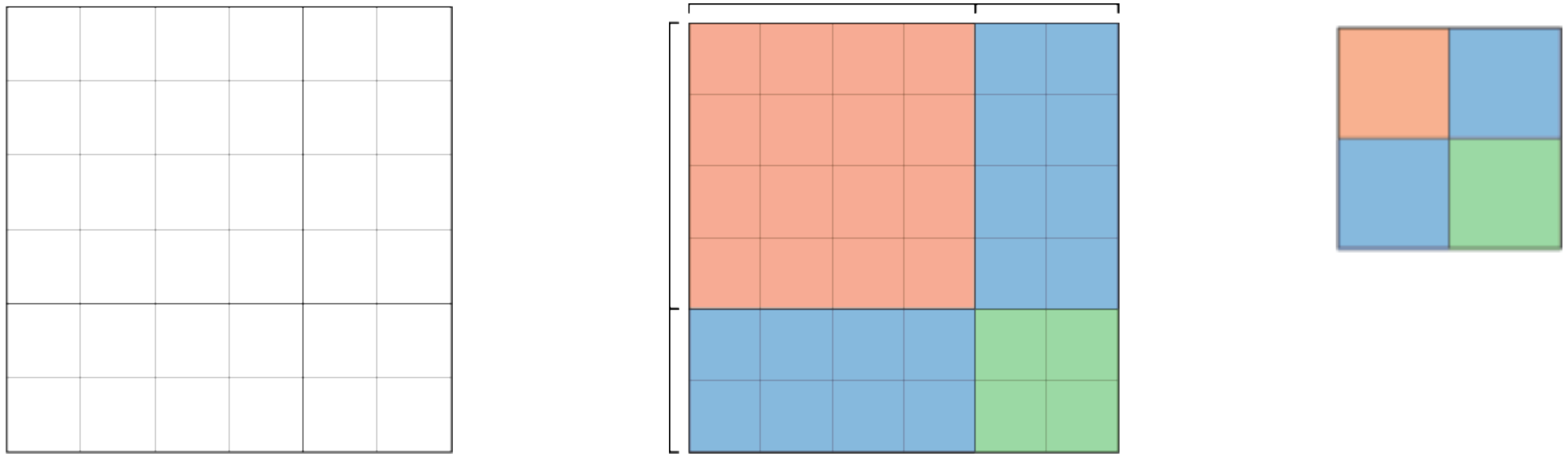
$$N_{i,j}(\,\cdot - \tau_{i,j}),\ i,j \in [n]$$

Problem



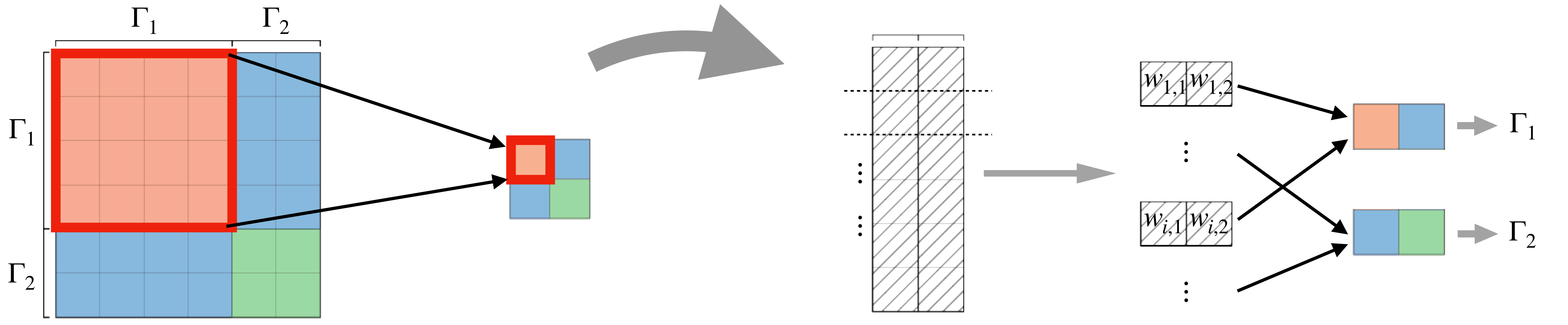
$$N_{i,j}(\cdot - \tau_{i,j}),\ i, j \in [n]$$

Align back **Find clusters** **Connecting patterns**



Incorporating time shifts

Given aligned point processes:

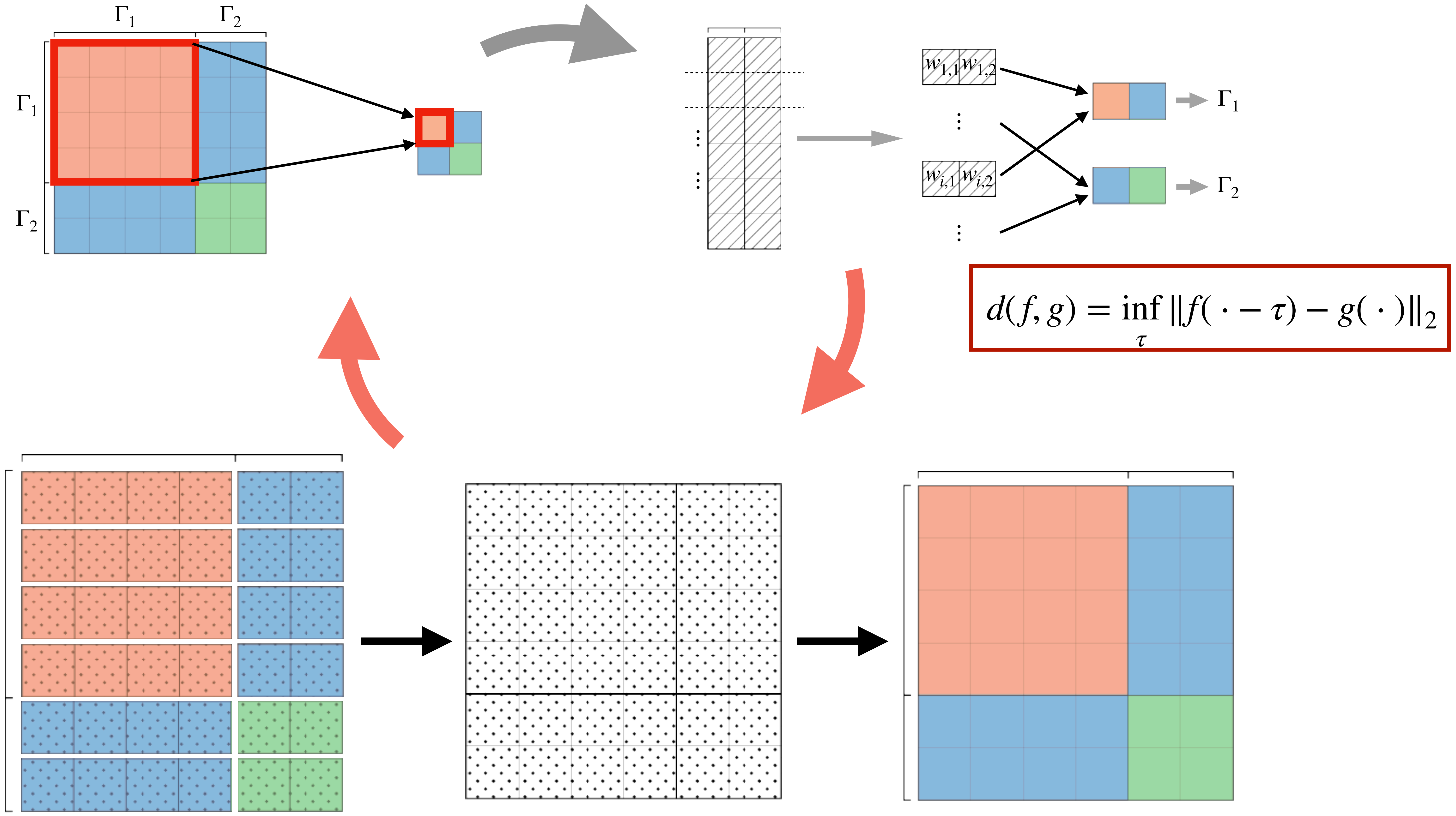


$$\boxed{d(f, g) = \inf_{\tau} \|f(\cdot - \tau) - g(\cdot)\|_2}$$

Incorporating time shifts

Given aligned point processes:

Re-align point processes based on current clustering

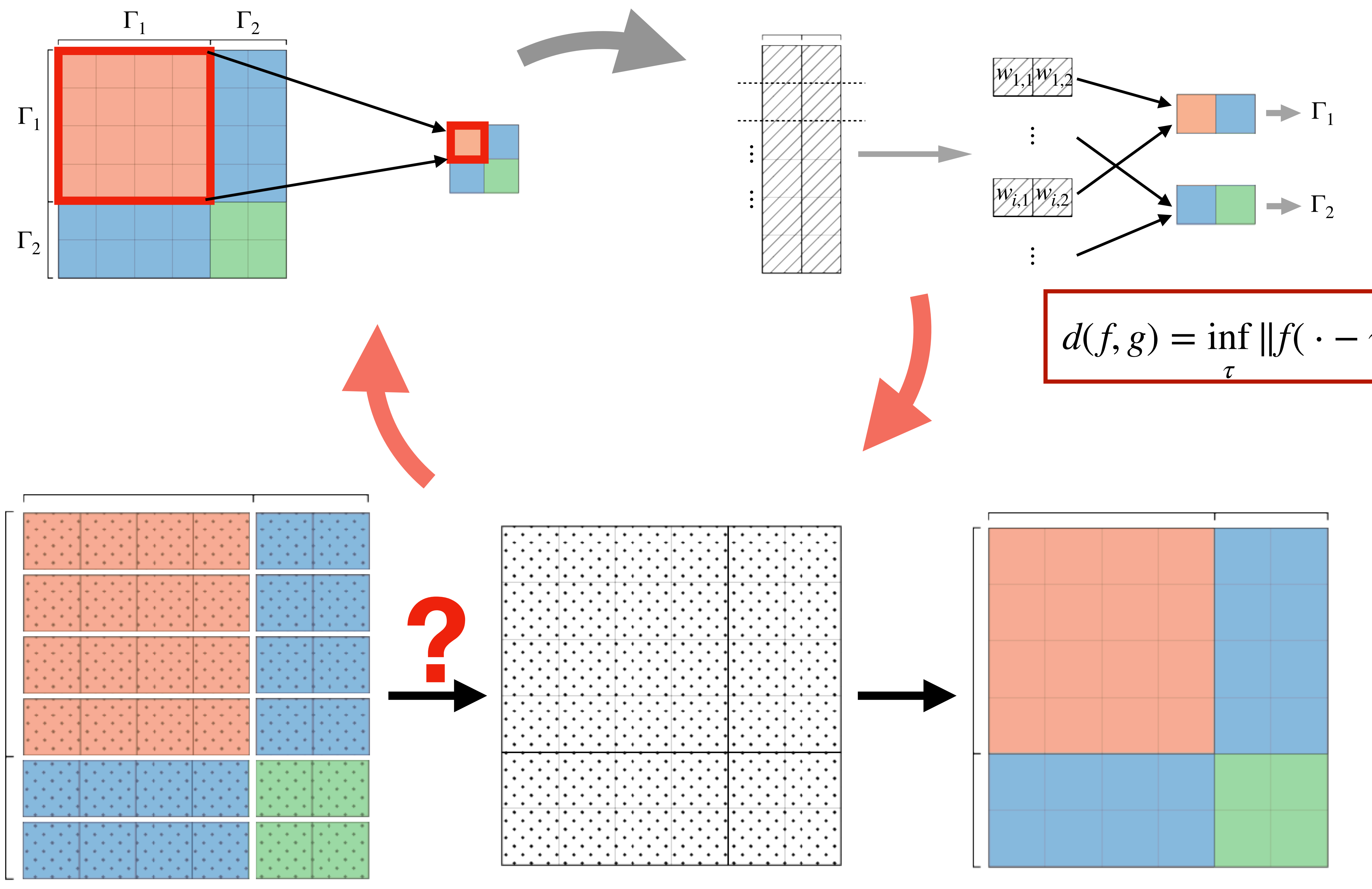


$$d(f,g) = \inf_{\tau} \|f(\cdot - \tau) - g(\cdot)\|_2$$

Incorporating time shifts

Given aligned point processes:

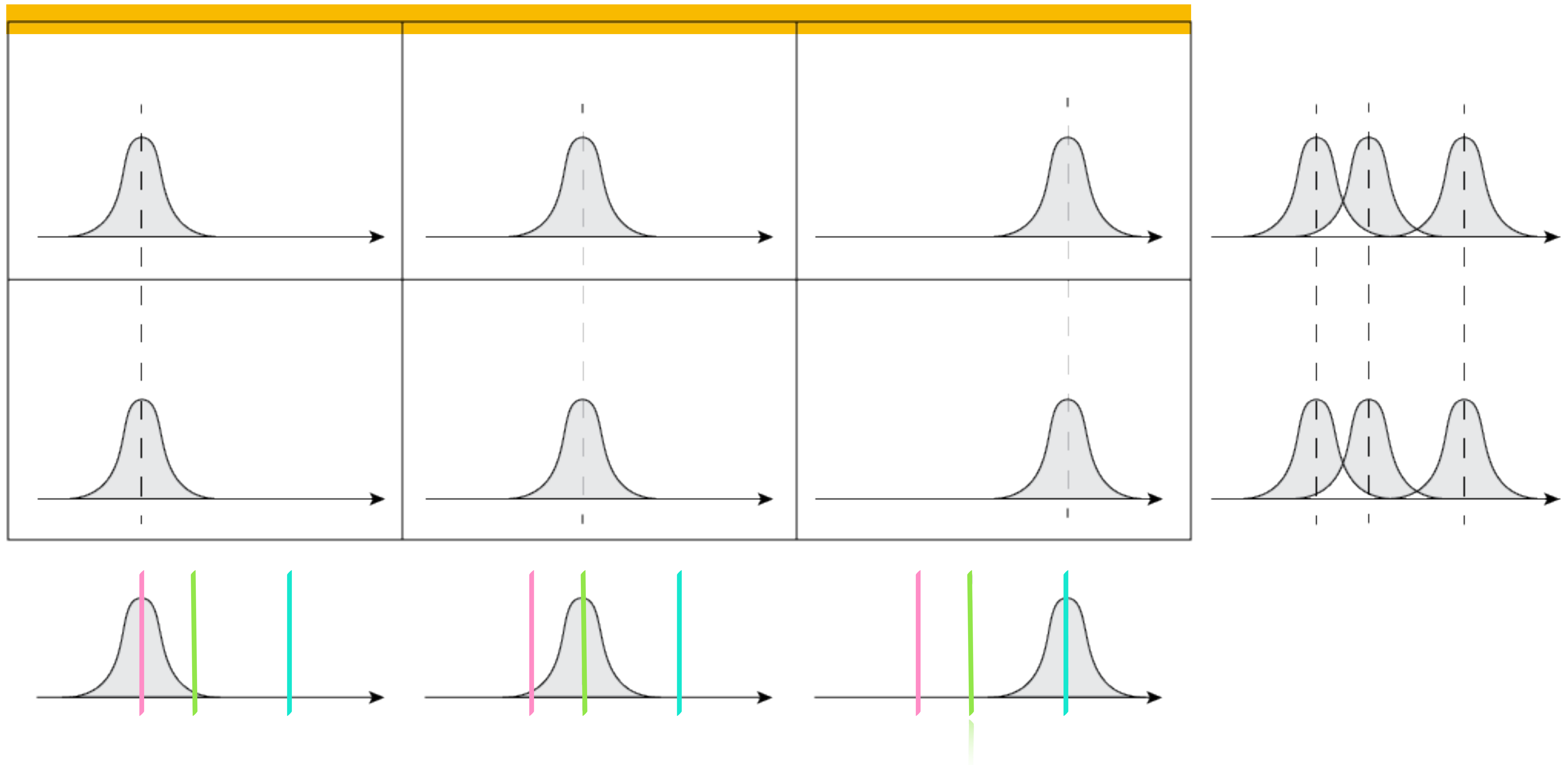
Re-align point processes based on current clustering



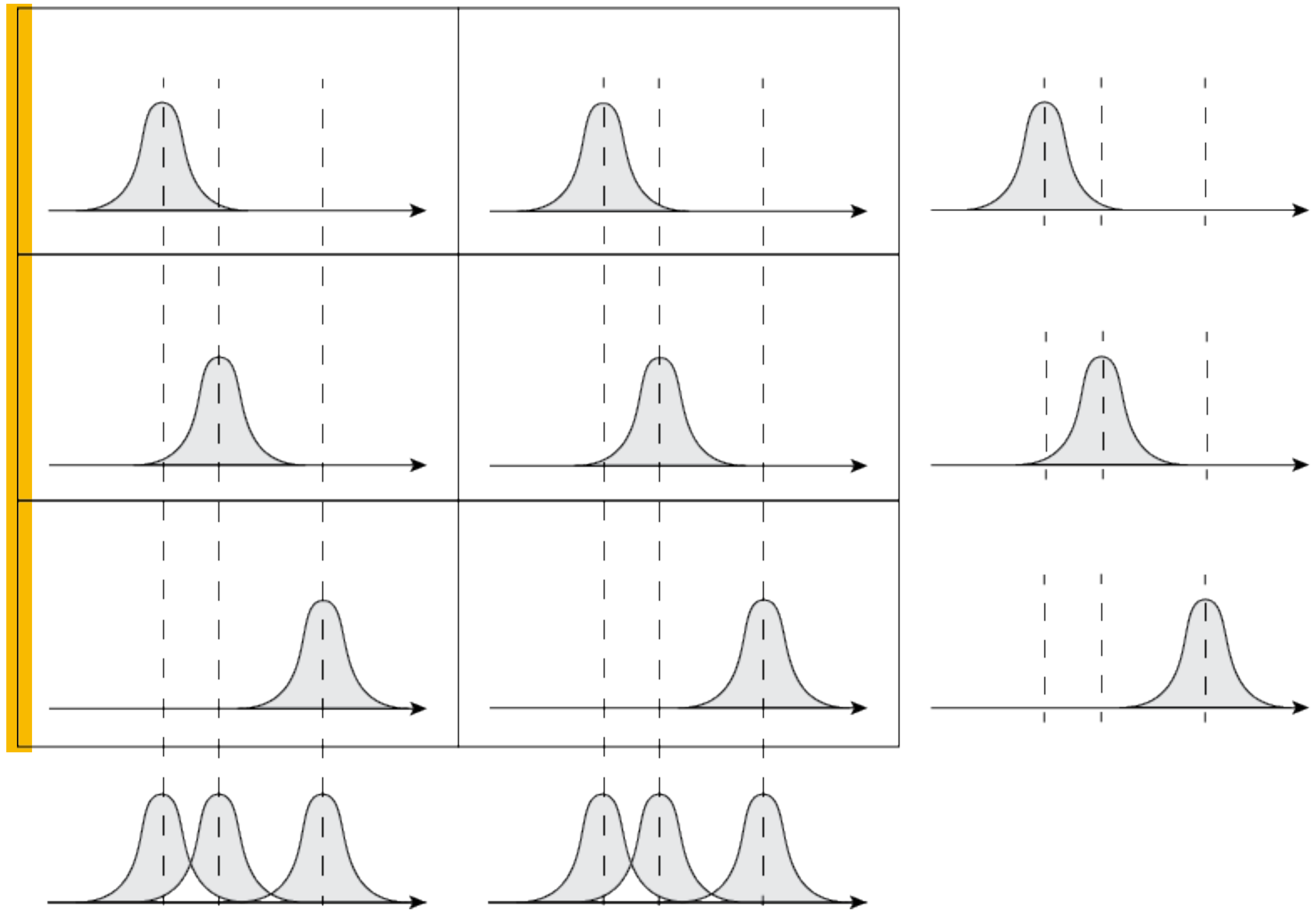
$$d(f, g) = \inf_{\tau} \|f(\cdot - \tau) - g(\cdot)\|_2$$

Assumption on time shifts

Align columns

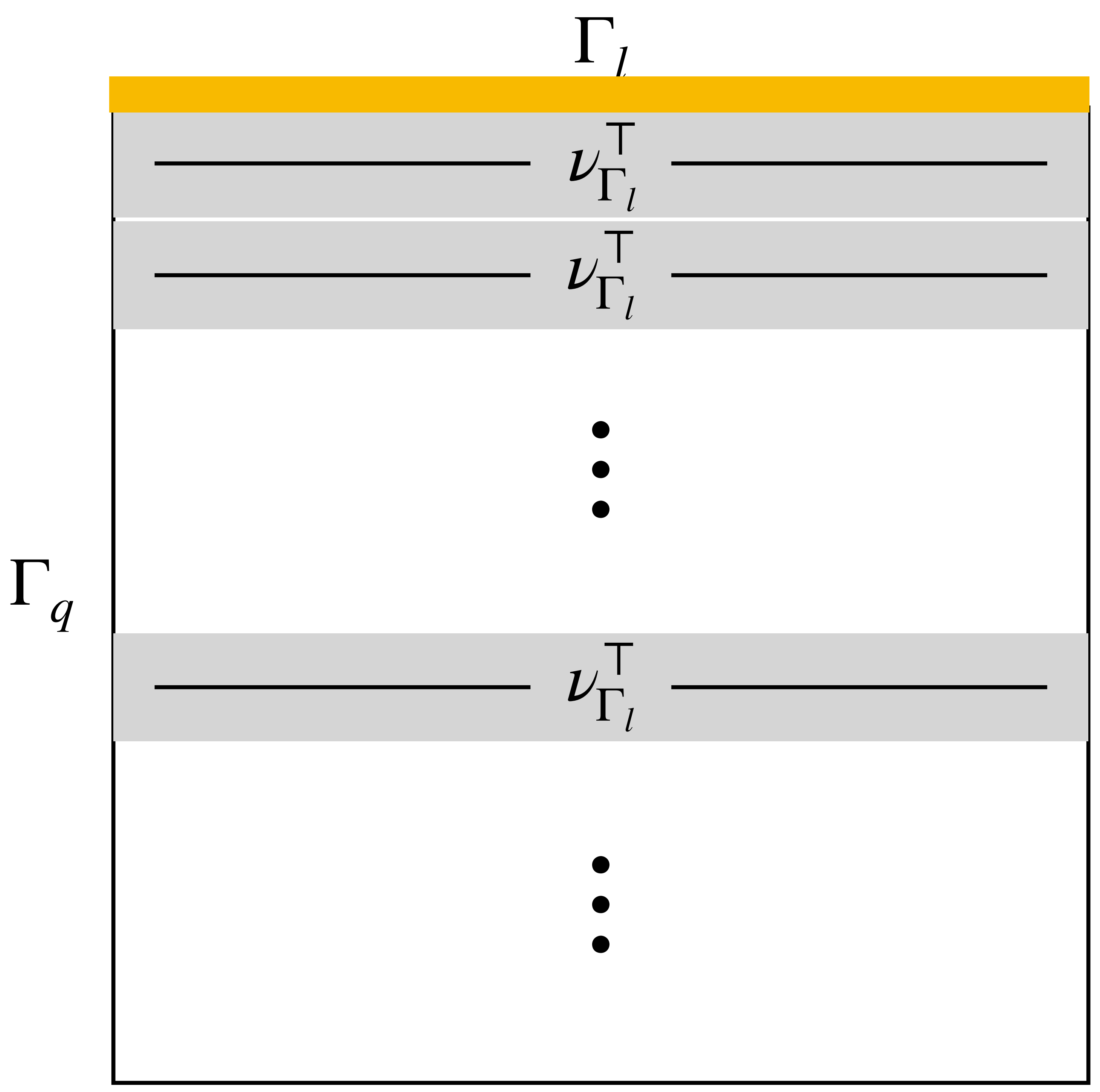


Align rows



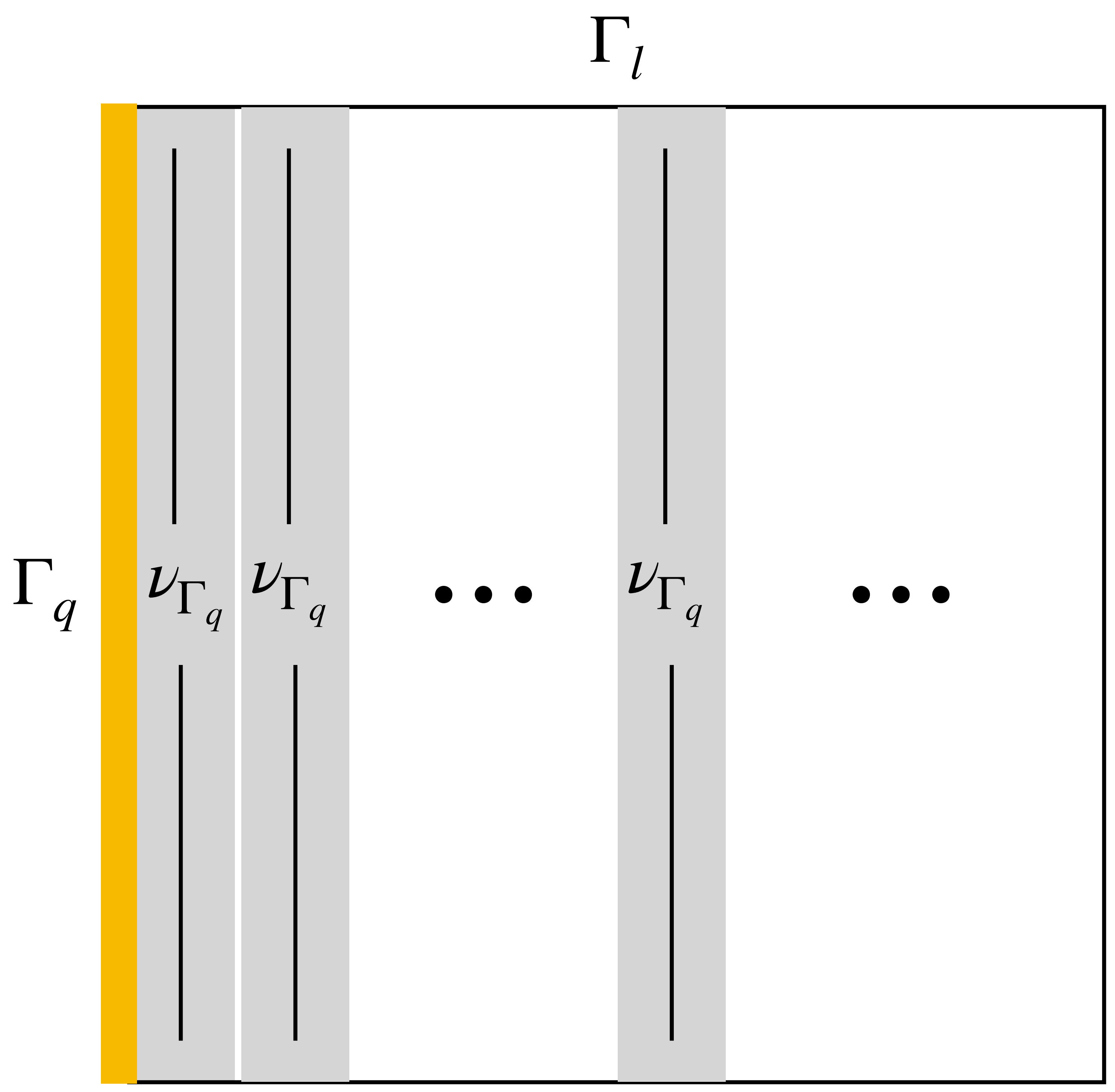
Assumption on time shifts

Align columns



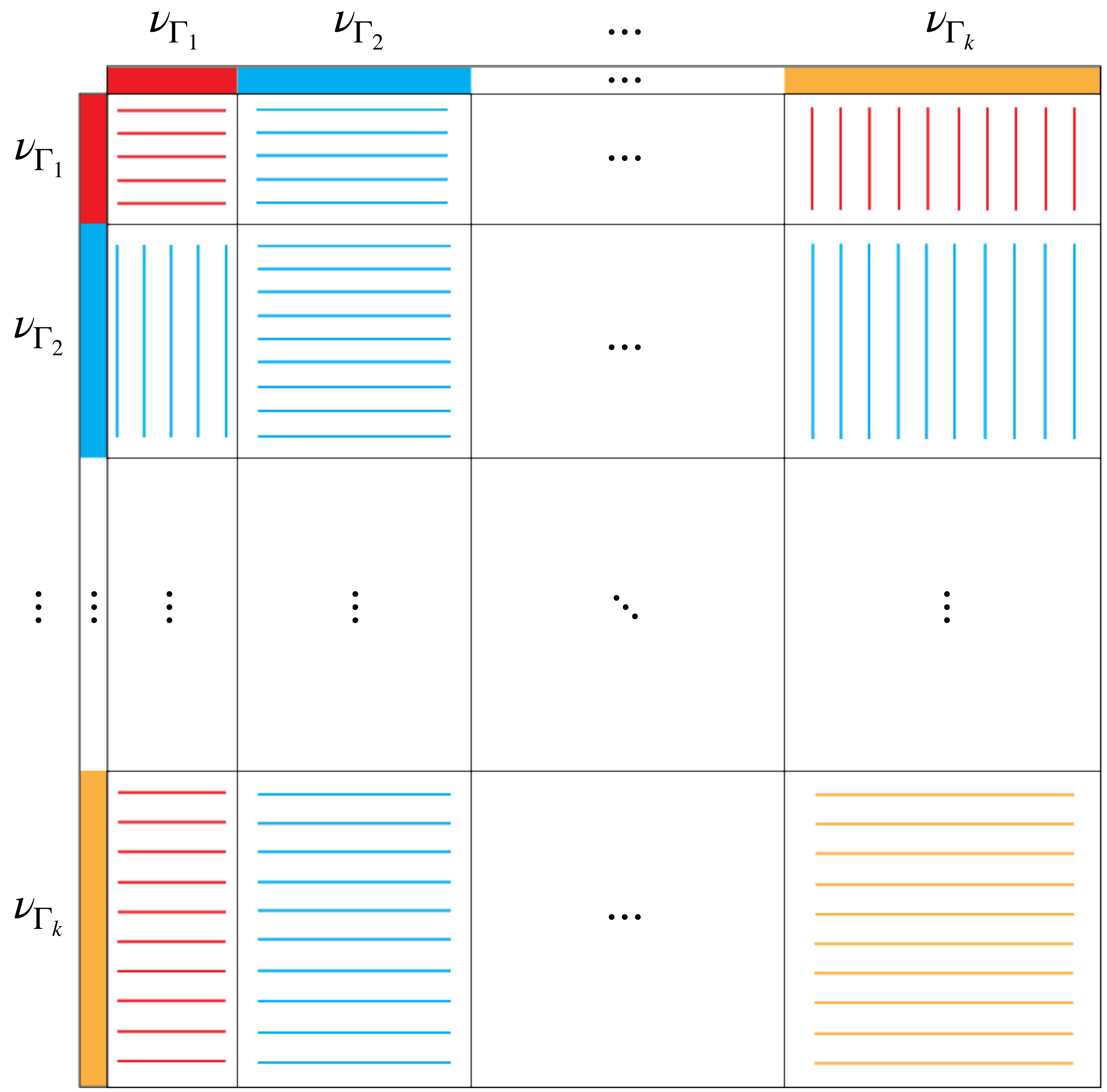
$$\tau_{i,\cdot} - \tau_{j,\cdot} = 0$$

Align rows



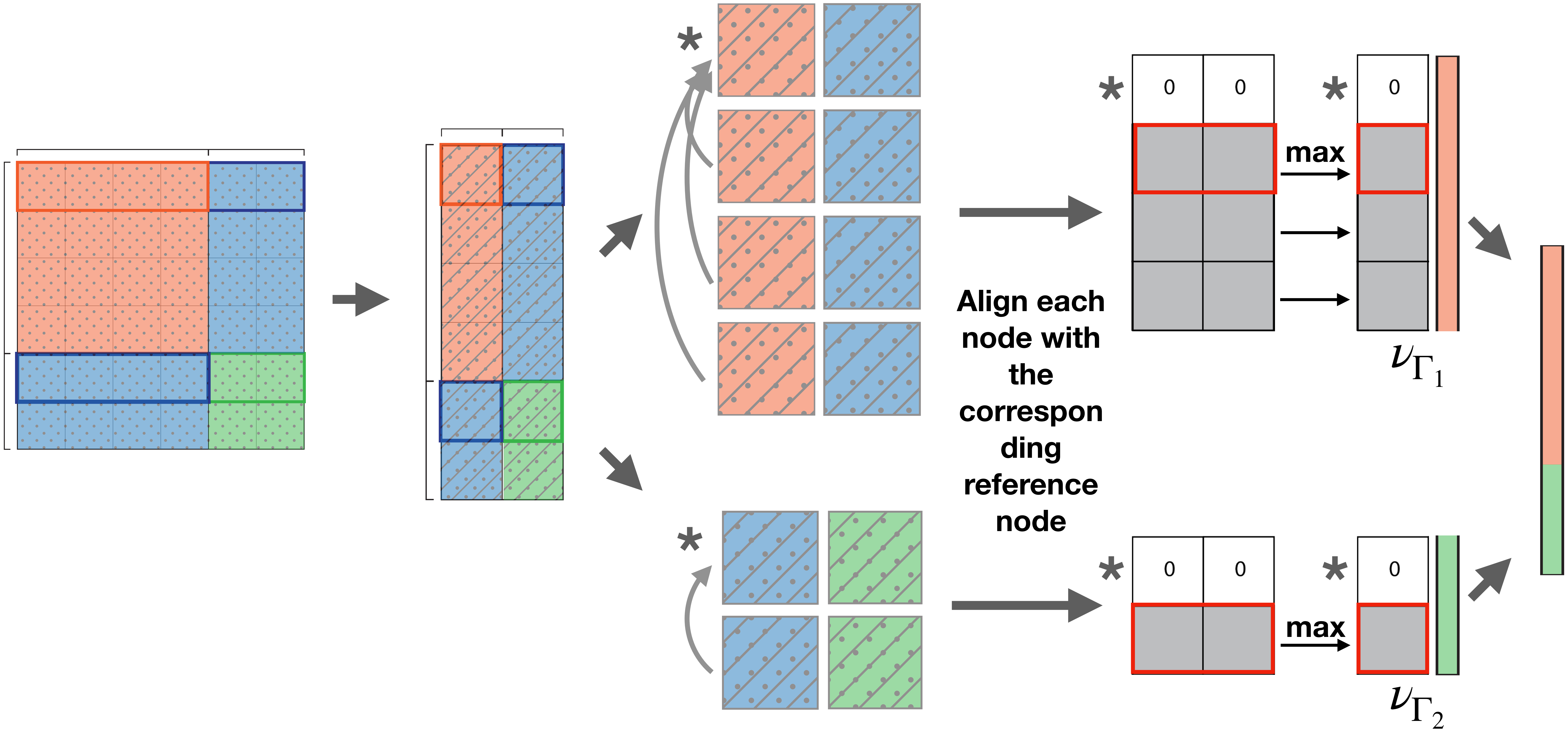
$$\tau_{i,\cdot} - \tau_{j,\cdot} = \nu_i - \nu_j$$

Assumption on time shifts

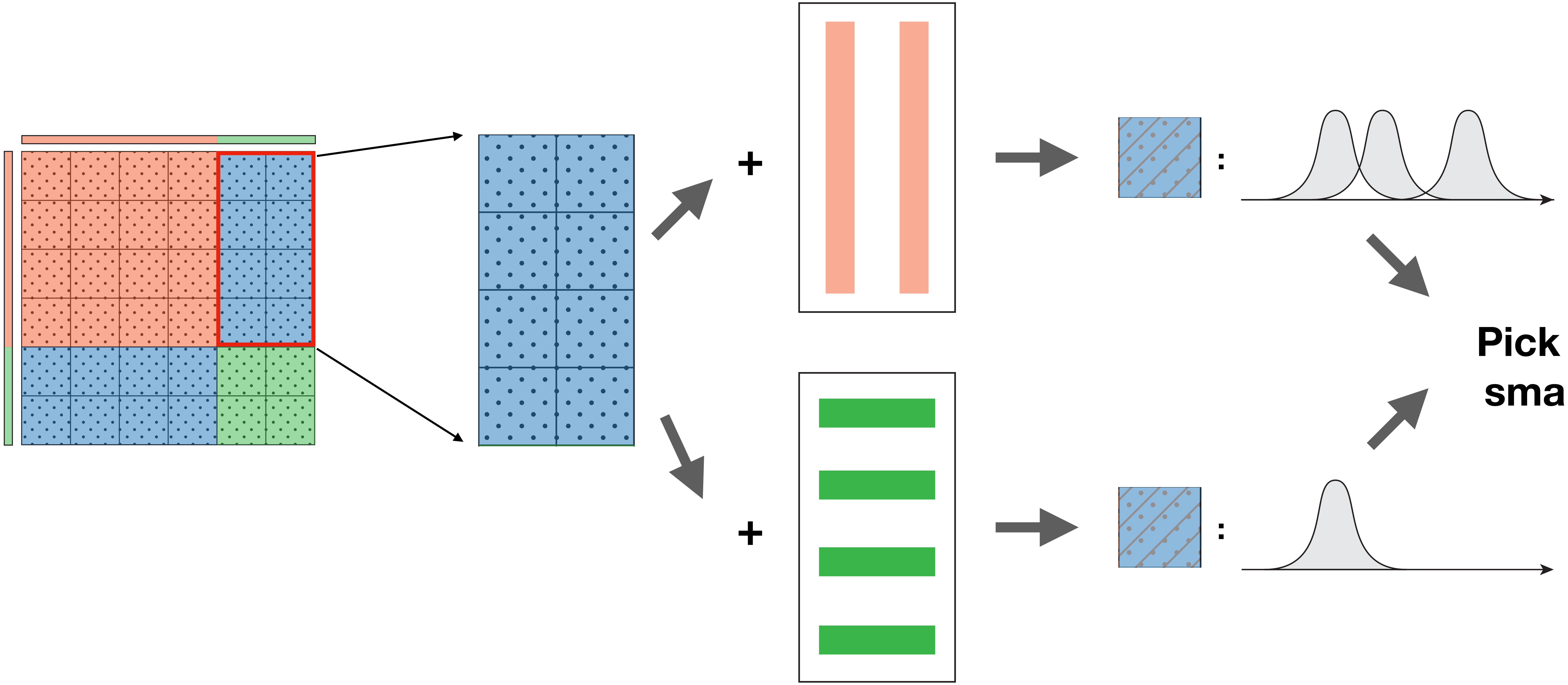


Matrix of time shifts

Estimate time shifts

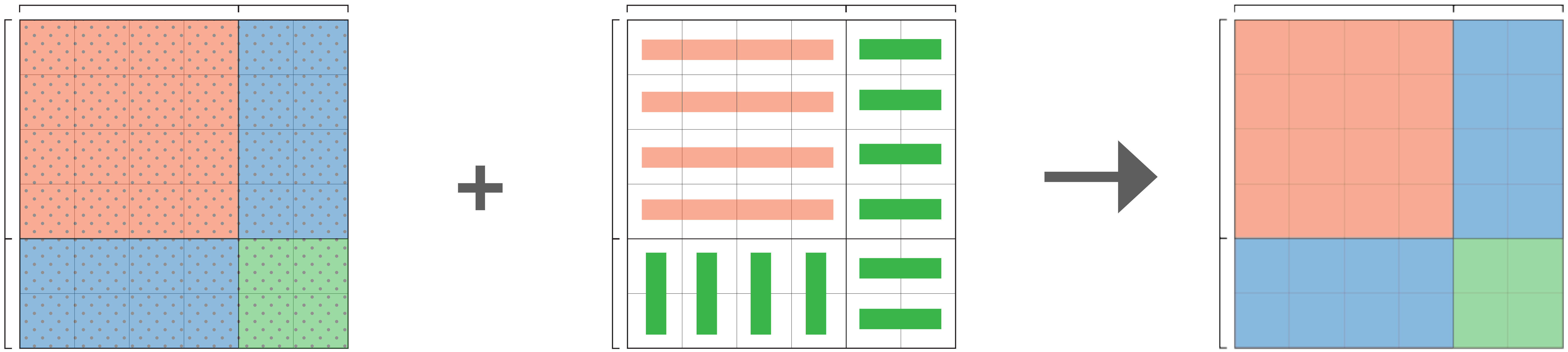


Estimate time shifts

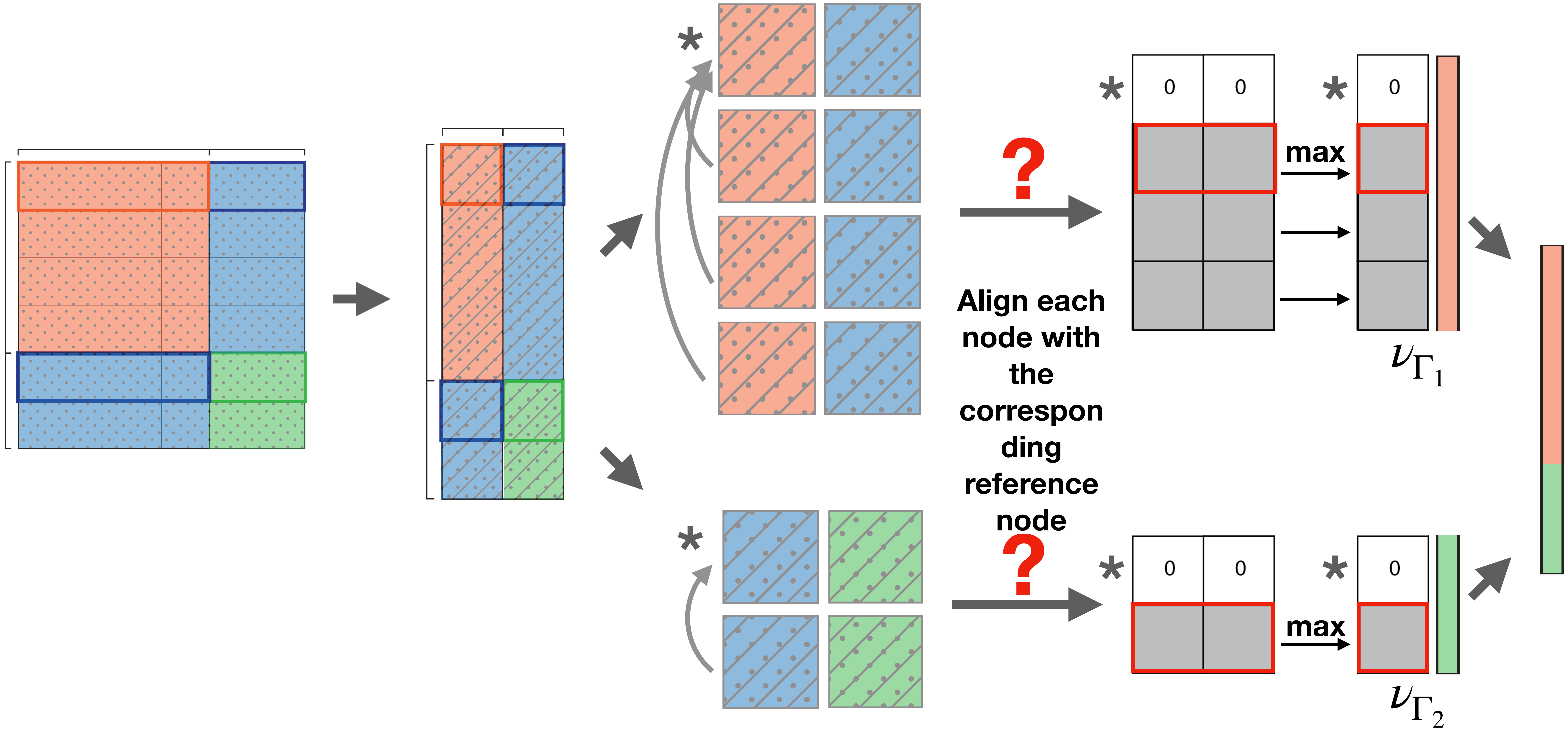


Pick the one with smaller variance

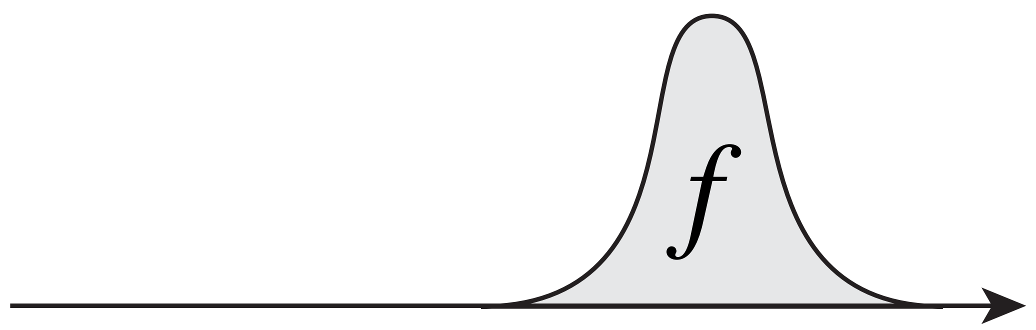
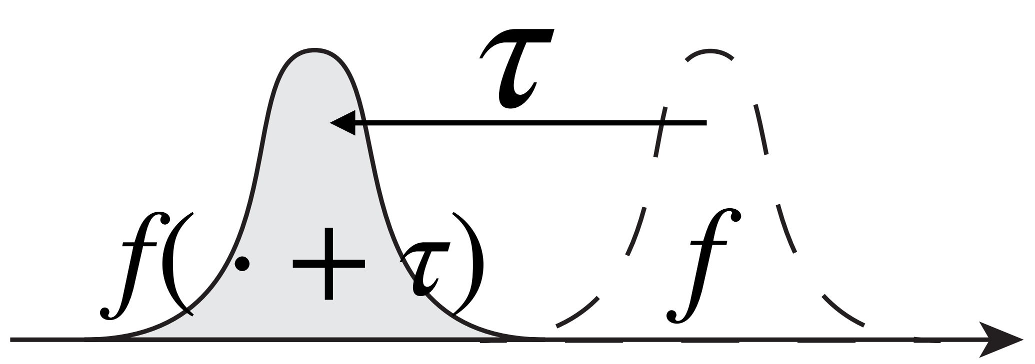
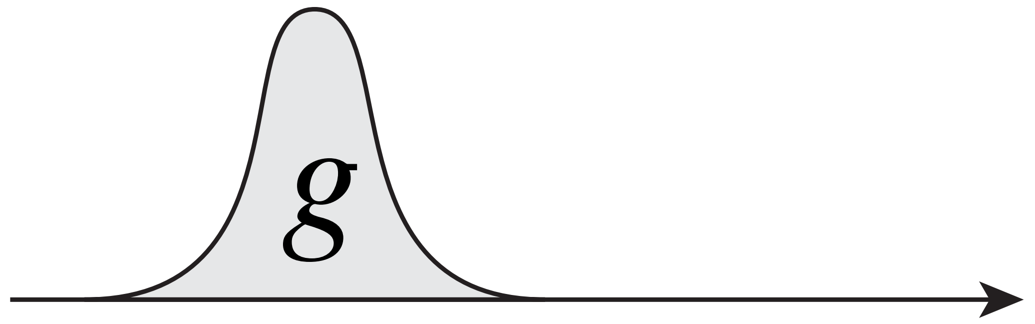
Re-align point processes



Estimate time shifts



Aligning curves and evaluating their distance

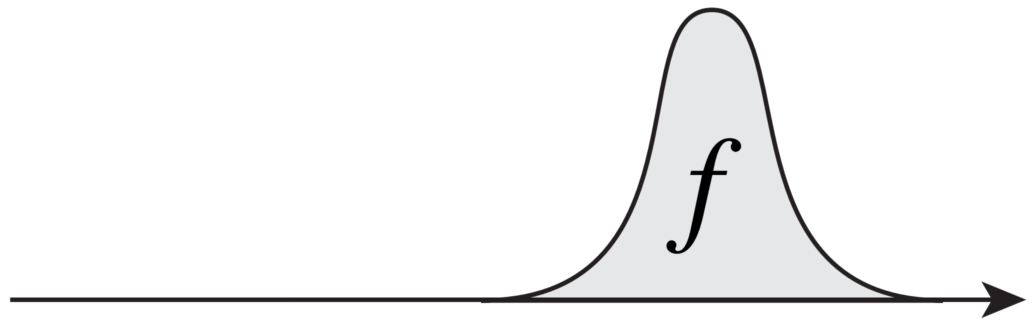
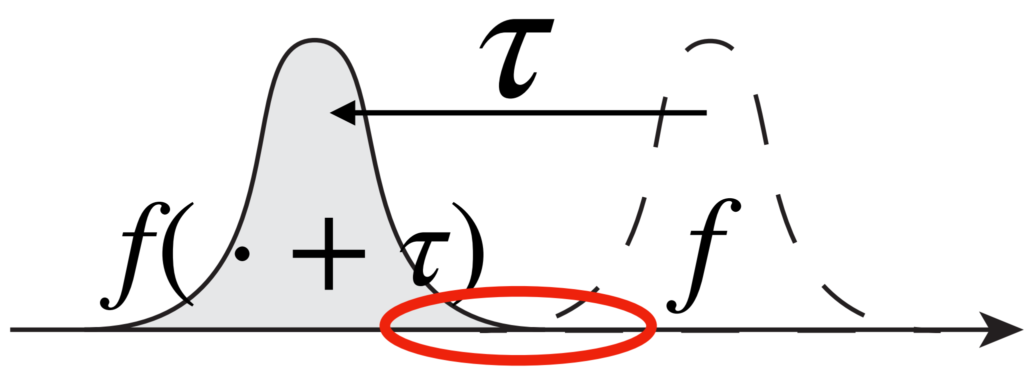
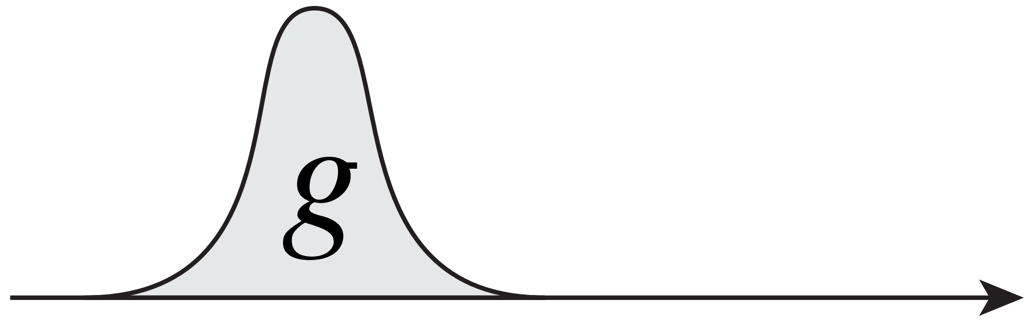
Curves	Fourier coefficients
	$\theta_j \quad {\scriptstyle j=-\frac{N-1}{2},\dots,\frac{N-1}{2}}$
	$\theta'_j = \theta_j e^{i2\pi j(n_0/N)} \quad {\scriptstyle n_0=\frac{\tau}{T}N}$
	$\gamma_j \quad {\scriptstyle j=-\frac{N-1}{2},\dots,\frac{N-1}{2}}$

$$d(f,g) = \min_{\tau} \|f-g\|_2$$

$$= \min_{n_0} \left[\frac{T}{N^2} \boxed{\sum_j |\theta'_j - \gamma_j|^2} \right]^{1/2}$$

Apply gradient descent algorithm

Aligning curves and evaluating their distance

Curves	Fourier coefficients
	$\theta_j \quad {}_{j=-\frac{N-1}{2},\dots,\frac{N-1}{2}}$
	$\theta'_j = \theta_j e^{i2\pi j(n_0/N)} \quad n_0 = \frac{\tau}{T}N$
	$\gamma_j \quad {}_{j=-\frac{N-1}{2},\dots,\frac{N-1}{2}}$

$$d(f,g) = \min_{\tau} \|f-g\|_2$$

$$= \min_{n_0} \left[\frac{T}{N^2} \boxed{\sum_j |\theta'_j - \gamma_j|^2} \right]^{1/2}$$

Apply gradient descent algorithm

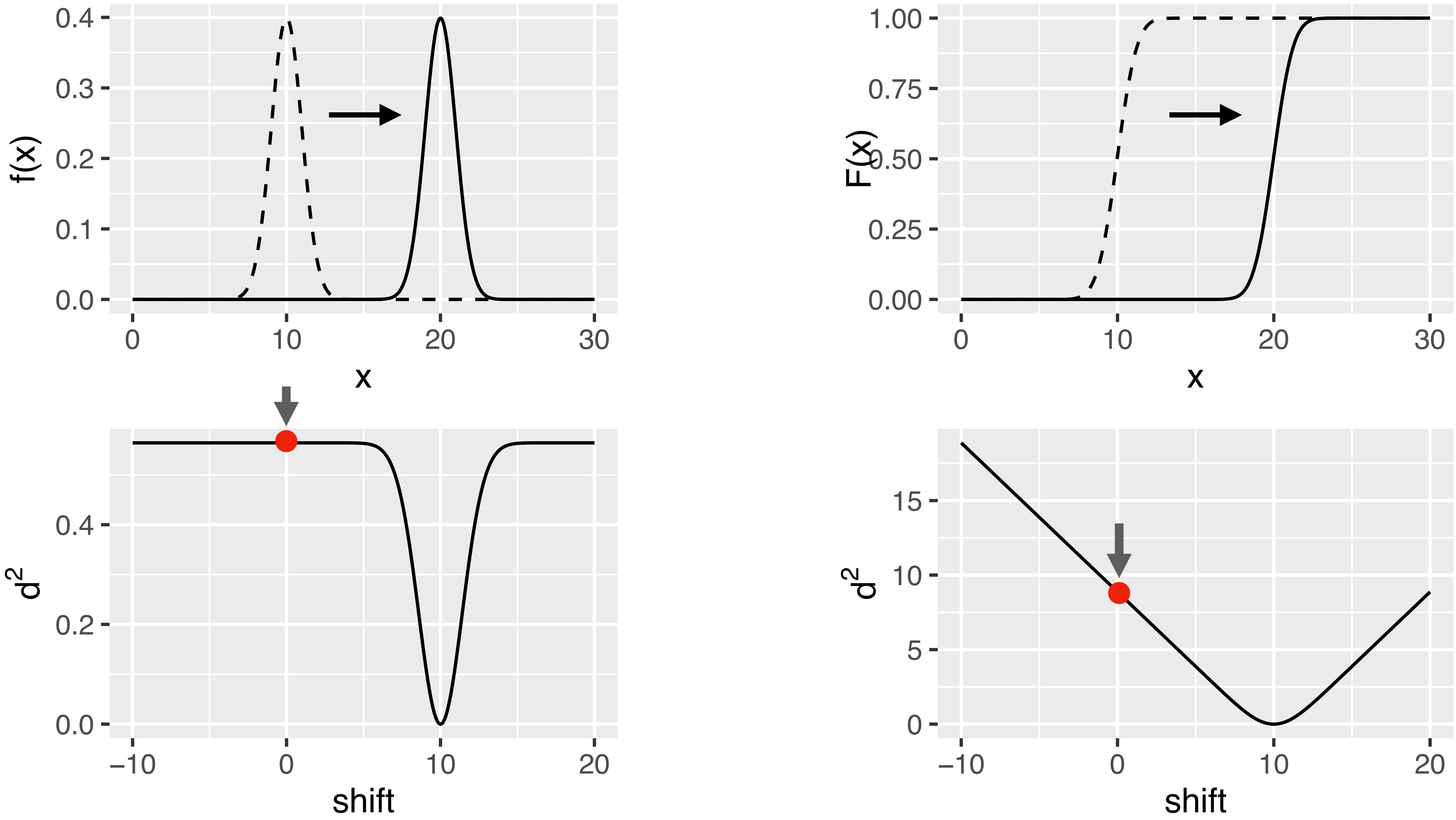
Good initialization?

Initializing gradient descent by aligning c.d.f.

Align p.d.f. | **Align c.d.f.**

Curves

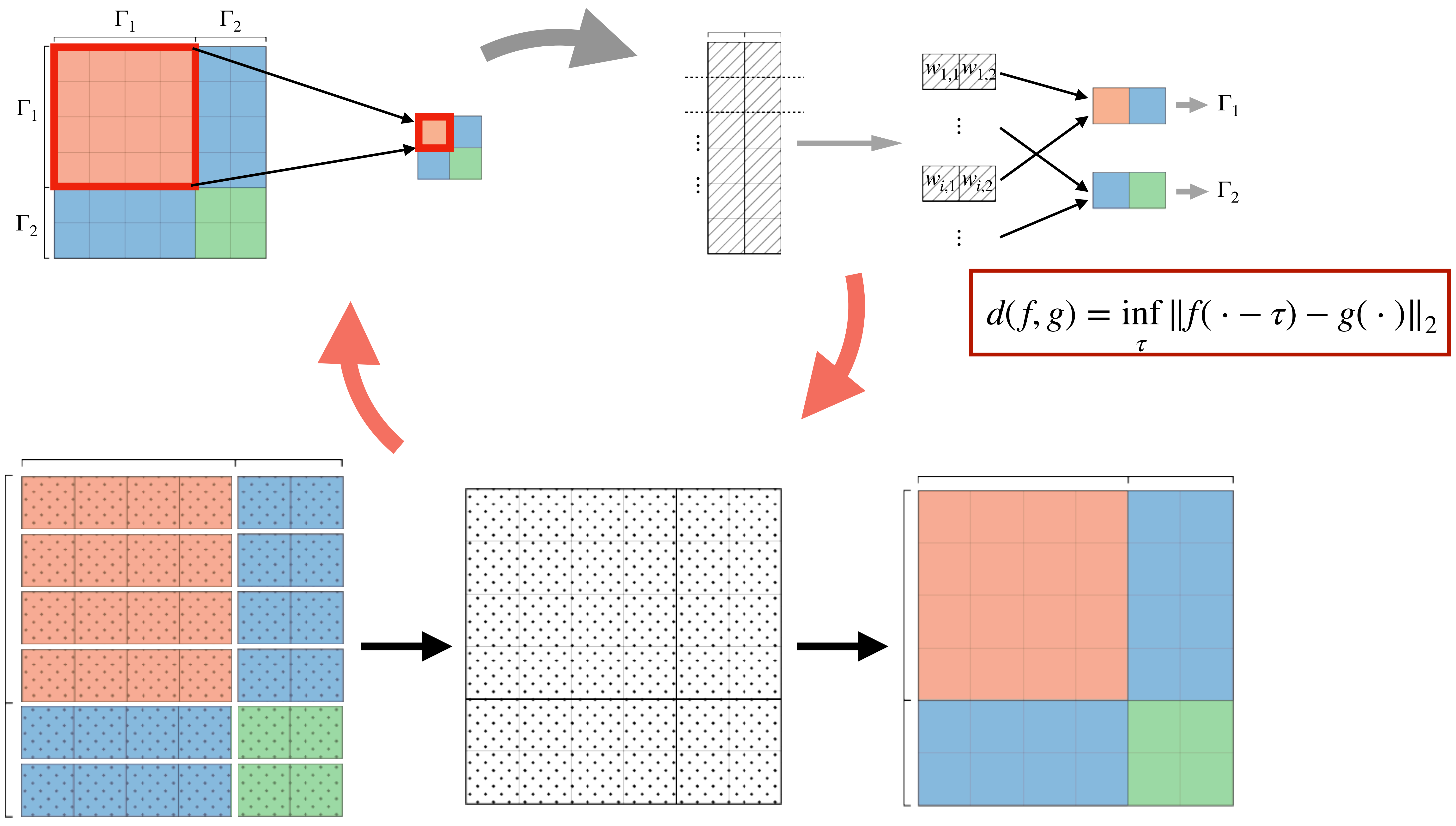
Loss function



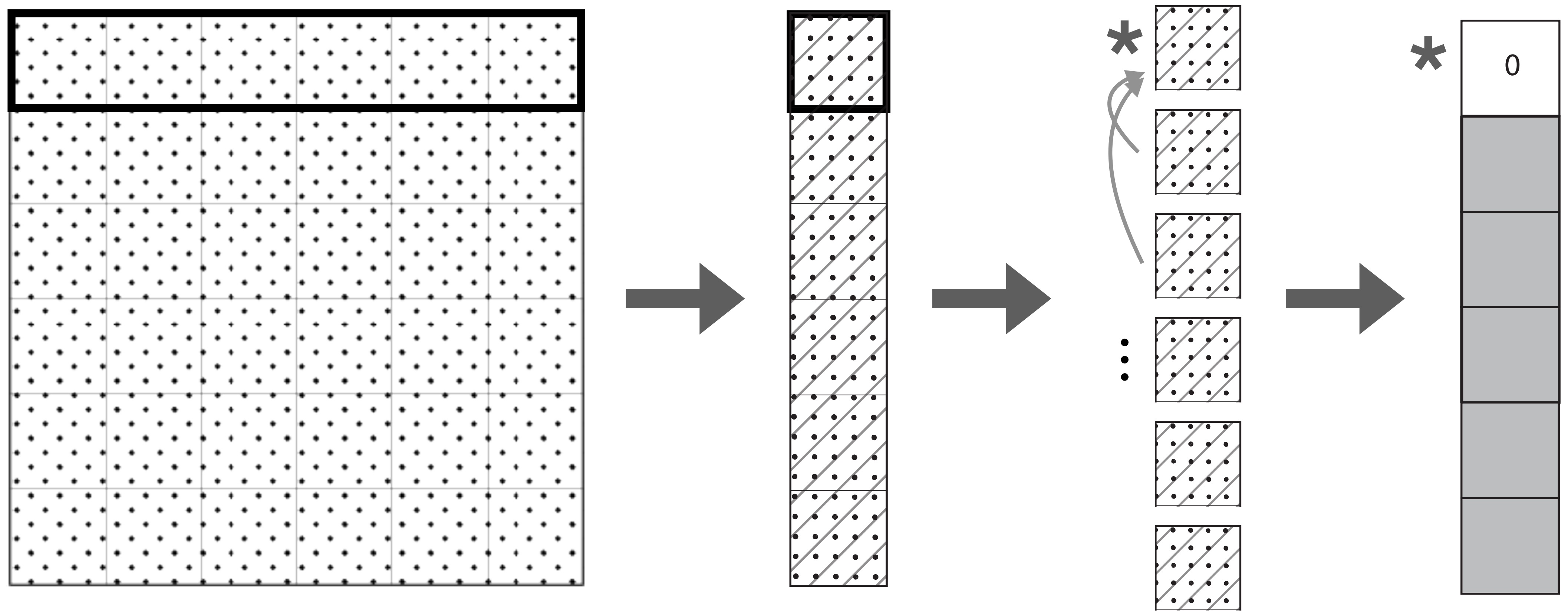
Algorithm

Given aligned point processes:

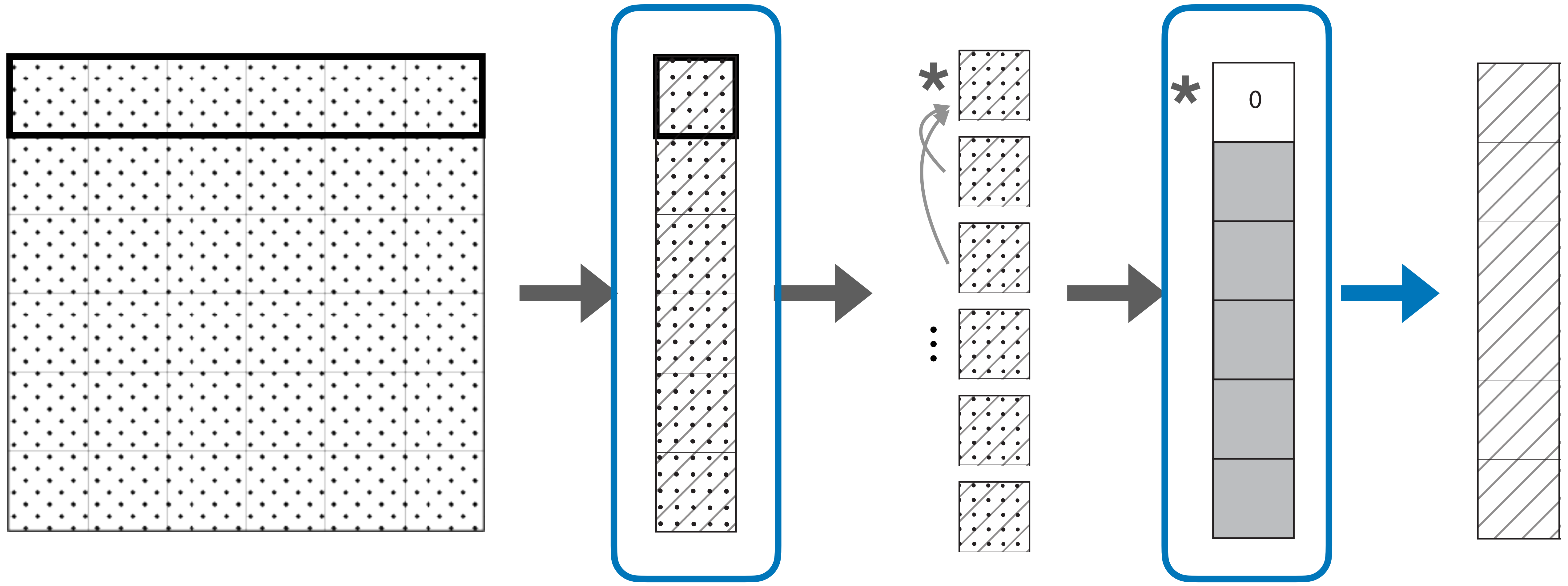
Re-align point processes based on current clustering



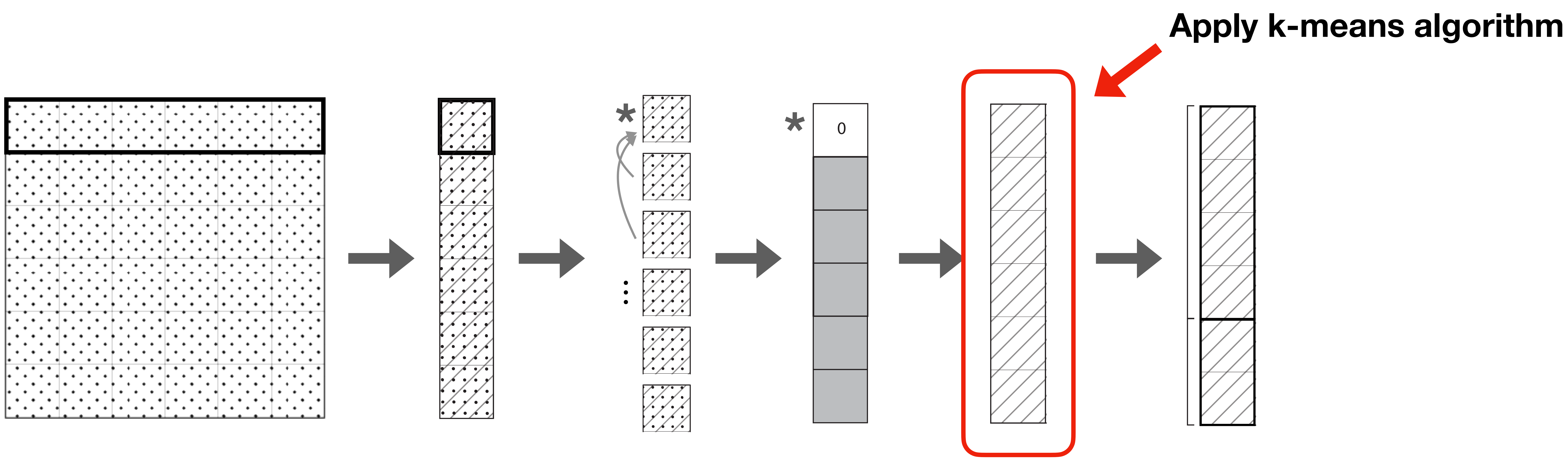
Initialization of clustering and time shifts



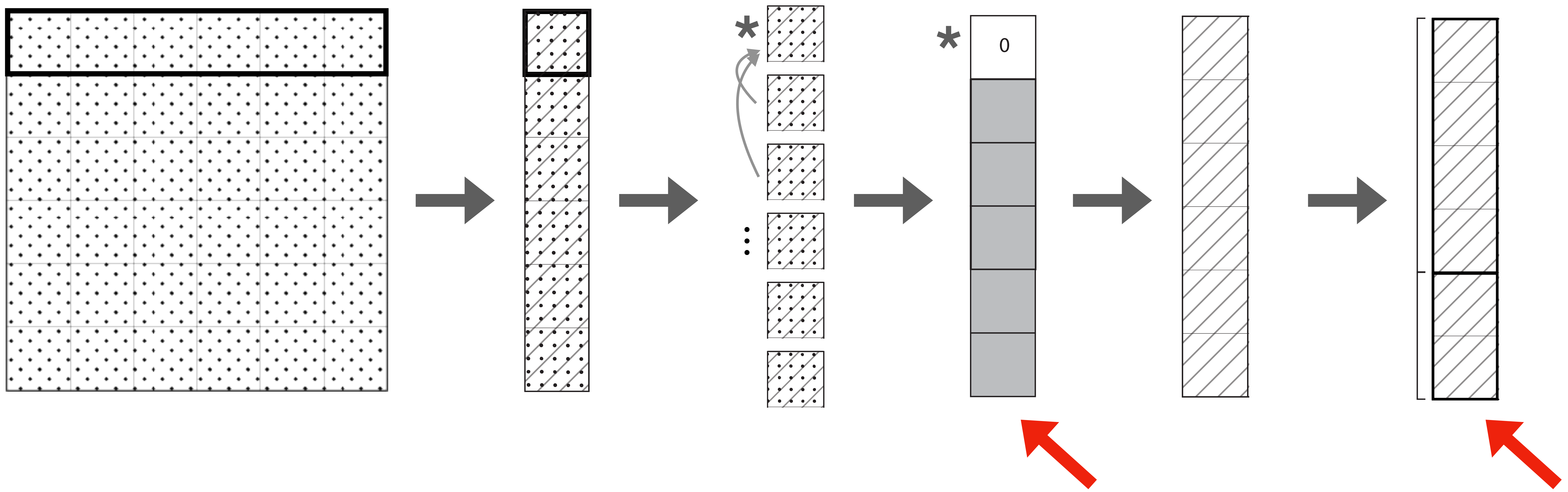
Initialization of clustering and time shifts



Initialization of clustering and time shifts



Initialization of clustering and time shifts



Simulation

Over-clustering can help to find the correct cluster numbers

Real data

Future work

Thank you

Appendix

- Derivation of aligning pdf and cdf
- Details about simulation set up, tuning parameters in algorithm
- Details about dealing with negative event time