

# Pole Projet P15 - Sketching

subtitle

Team Name

January 23, 2026

## Section 1

# JESPRIT for Parameters Estimation of Mixed Poisson distribution

# Problem Formulation: Mixed Poisson Recovery

**Goal:** Recover latent parameters of a Mixed Poisson distribution from count data.

- **Model:**  $X \sim \sum_{k=1}^r \pi_k \text{Pois}(\lambda_k)$
- **Unknowns:** Rate matrix  $\mathbf{A} = [\lambda_1, \dots, \lambda_r]$  and weights  $\pi$ .

**Challenge:** High-dimensional estimation from unlabeled samples.

## Sketching Approach:

- 1 Compute **Empirical PGF** at specific points:  $\mathbf{t}(\mathbf{u}, n) = \mathbf{1} + j\Delta n \mathbf{u}$ .
- 2 Map PGF samples to **Harmonic Retrieval** model.
- 3 Apply **JESPRIT** to recover parameters.

# Sampling Strategy: The Empirical PGF

## Why Empirical PGF?

- We lack the true distribution, but we have samples  $\mathbf{x}^{(j)}$ .
- PGF form:  $\hat{G}_{\mathbf{x}}(\mathbf{t}) = \frac{1}{N_s} \sum_{j=1}^{N_s} e^{\langle \mathbf{x}^{(j)}, \ln(\mathbf{t}) \rangle}$ .

## Line Sampling for Shift Invariance

- Single points are not enough. We need a **sequence** to capture "rotations".
- For each direction vector  $\mathbf{u}_l$ , we sample along a line in the complex domain:

$$\mathbf{t}(\mathbf{u}_l, n) = \mathbf{1} + j\Delta n \mathbf{u}_l, \quad n = 0, 1, \dots, 2N - 1$$

## Resulting Signal Sequence:

$$y_l[n] = \hat{G}_{\mathbf{x}}(\mathbf{t}(\mathbf{u}_l, n)) \approx \sum_{k=1}^r \pi_k(e^{j\Delta \langle \lambda_k, \mathbf{u}_l \rangle})^n$$

This maps the Poisson recovery problem to Harmonic Retrieval.

# The JESPRIT Algorithm

## Key Innovation: Global Subspace Estimation

- Instead of processing each direction independently, we **stack** Hankel matrices from all directions into  $\mathbf{X}_{\text{glob}}$ .
- **Global SVD:**  $\mathbf{X}_{\text{glob}} \approx \mathbf{U}_{\text{glob}} \mathbf{\Sigma} \mathbf{V}^H$ .
- Ensures all directional subspaces  $\hat{\mathbf{U}}_l$  share a **coherent basis**.

## Recovery Steps:

- 1 **RIMs:** Solve  $\hat{\mathbf{U}}_{l,\uparrow} \mathbf{\Psi}_l \approx \hat{\mathbf{U}}_{l,\downarrow}$  for shifts.
- 2 **Joint Diagonalization:** Find  $\mathbf{T}$  such that for all  $l$ :

$$\mathbf{\Psi}_l \approx \mathbf{T} \mathbf{\Phi}_l \mathbf{T}^{-1}, \quad \mathbf{\Phi}_l = \text{diag}(\phi_{1,l}, \dots, \phi_{r,l})$$

- 3 **Extraction:** Eigenvalues  $\phi_{k,l}$  give Rates  $\lambda_k$  via:

$$\phi_{k,l} = e^{j\Delta \langle \lambda_k, \mathbf{u}_l \rangle}$$

- 4 **Probabilities ( $\pi$ ):** Solve linear system  $\mathbf{y} \approx \mathbf{E}\pi$  via Least Squares:

$$\min_{\pi} \|\mathbf{y} - \mathbf{E}\pi\|_2^2 \quad \text{where } E_{l,k} = e^{j\Delta \langle \hat{\lambda}_k, \mathbf{u}_l \rangle}$$

# Results: Parameter Analysis & Phase Unwrapping

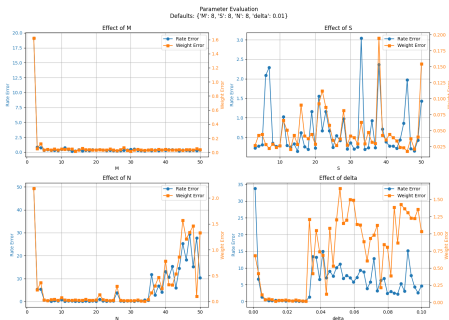


Figure: Error without Unwrapping

## 1. Phase Unwrapping:

- Unwrapping *increases* error here.
- Optimal scale:  $\Delta \approx 1/\max(A)$ .

## 2. Parameter Sensitivity:

- **Directions ( $M$ ) & Snapshots ( $S$ ):** Robust to over-sampling.
- **Samples ( $N$ ):** Sensitive. Large  $N$  causes aliasing/wrapping issues.

# Results: Sample Complexity

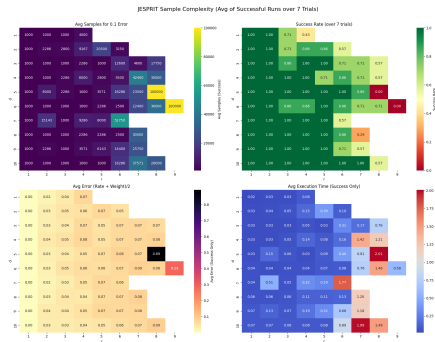


Figure: Success Rate (Large Range)

## Experiment Setup:

- $A = \text{random}(\text{range}, d, r)$ .
- Increase number of samples until  $\text{MRE} < 10\%$ .
- Increase dimension  $r$  while success rate  $> 75\%$ .

## Finding 2: Dynamic Range

- A larger rate range in  $A$  improves recovery.

## Scalability:

- Algorithm performance depends on Rank  $r$ , not Dimension  $d$ .