

# Pole Projet P15 - Sketching

## subtitle

Team Name

January 21, 2026

# 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  $\boldsymbol{\pi}$ .

**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.

# 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:**  $\mathbf{X}_{\text{glob}} \approx \mathbf{U}_{\text{glob}} \boldsymbol{\Sigma} \mathbf{V}^H$ .
- Ensures all directional subspaces  $\hat{\mathbf{U}}_I$  share a **coherent basis**.

## Recovery Steps:

- ① **RIMs:** Solve  $\hat{\mathbf{U}}_{I,\uparrow} \Psi_I \approx \hat{\mathbf{U}}_{I,\downarrow}$  for shifts.
- ② **Joint Diagonalization:** Find  $\mathbf{T}$  to diagonalize all  $\Psi_I$  simultaneously.
- ③ **Extraction:** Eigenvalues  $\rightarrow$  Rates  $\lambda_k$ . Least Squares  $\rightarrow$  Probs  $\pi$ .

# Results: Parameter Analysis & Phase Unwrapping

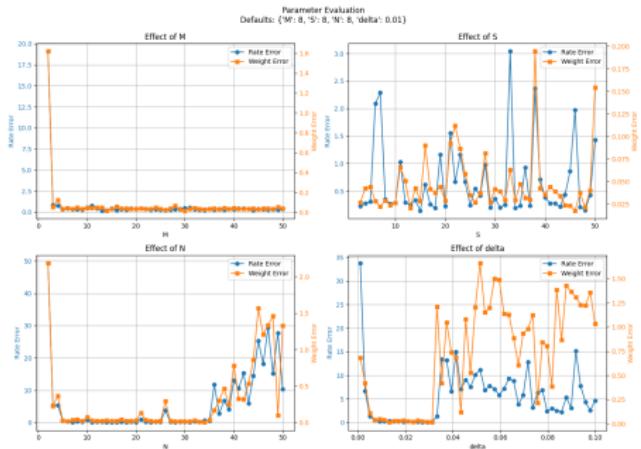


Figure: Error without Unwrapping

## 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

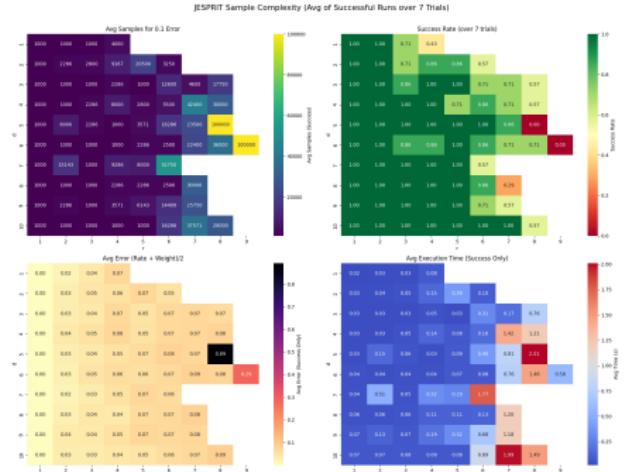


Figure: Success Rate (Large Range)

## Finding 2: Dynamic Range

- Larger rate range  $[0, 10^4]$  improves recovery.
- Distinct "directions" in count space act as higher SNR.

## Scalability:

- Sample complexity depends on Rank  $r$ , not Dimension  $d$ .