# 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:** $X \sim \sum_{k=1}^{r} \pi_k \mathrm{Pois}(\boldsymbol{\lambda}_k)$
- **Unknowns:** Rate matrix $\mathbf{A} = [\boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_r]$ 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 $\hat{\mathbf{U}}_l$ share a **coherent basis**.

**Recovery Steps:**

1. **RIMs:** Solve $\hat{\mathbf{U}}_{l,\uparrow} \boldsymbol{\Psi}_l \approx \hat{\mathbf{U}}_{l,\downarrow}$ for shifts.
2. **Joint Diagonalization:** Find $\mathbf{T}$ to diagonalize all $\boldsymbol{\Psi}_l$ simultaneously.
3. **Extraction:** Eigenvalues $\rightarrow$ Rates $\boldsymbol{\lambda}_k$. Least Squares $\rightarrow$ Probs $\boldsymbol{\pi}$.

# Results: Robustness & Phase Unwrapping

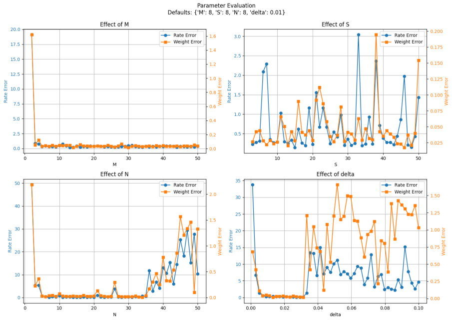


Figure: Error without Unwrapping

**Finding 1: Phase Unwrapping**

- Standard ESPRIT suggests unwrapping.
- **Result:** Unwrapping *increases* error in this PGF context.
- Best performance at optimal scale $\Delta \approx 1/\max(A)$.

# Results: Sample Complexity

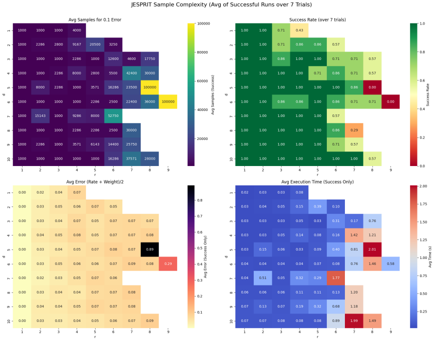


Figure: Success Rate (Large Range)

**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 $d$.