A Data-Driven Prism: Multi-View Source Separation with Diffusion Model Priors

EUCAIF 2025

Sebastian Wagner-Carena

Aizhan Akhmetzhanova

Sydney Erickson

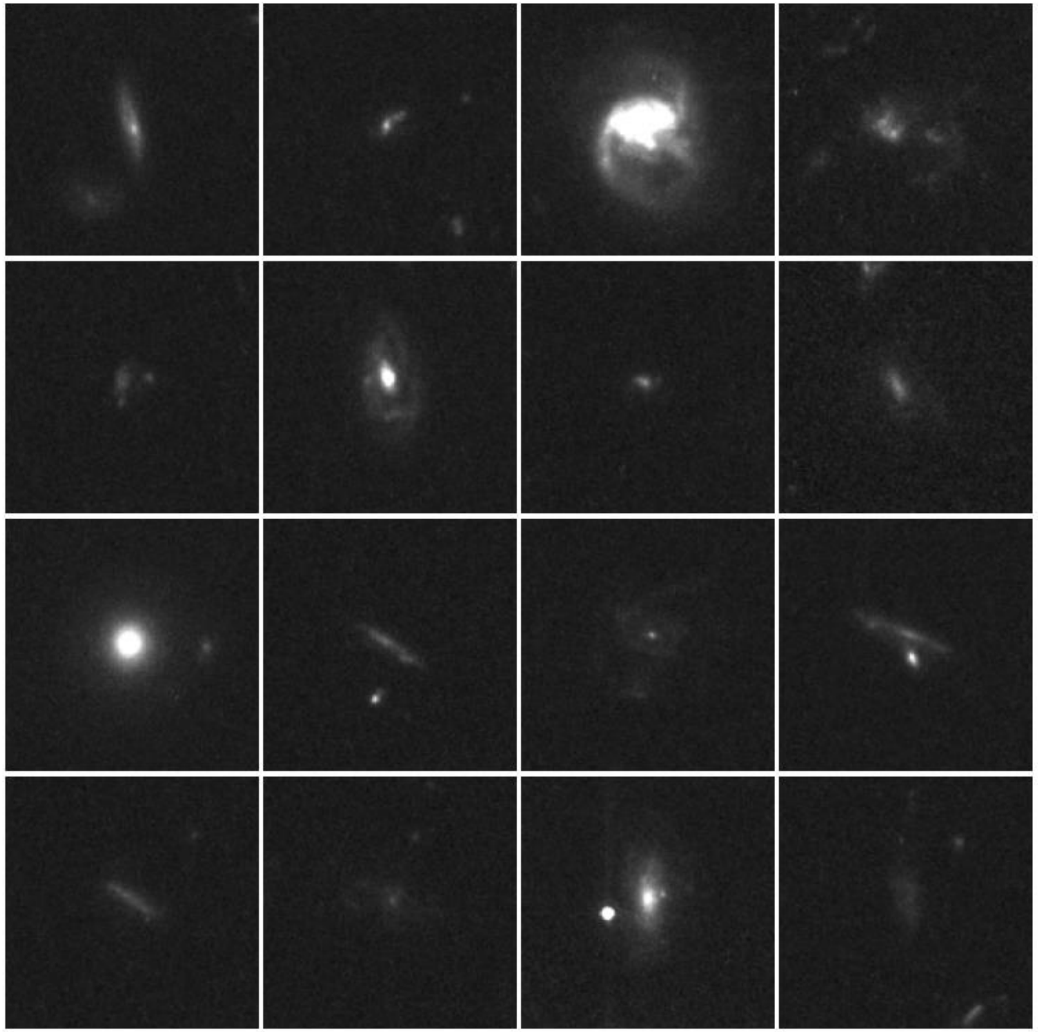
Motivating Examples: Galaxies

- What does a galaxy **look like**? **How** do we know?

Koekemoer et al., 2007

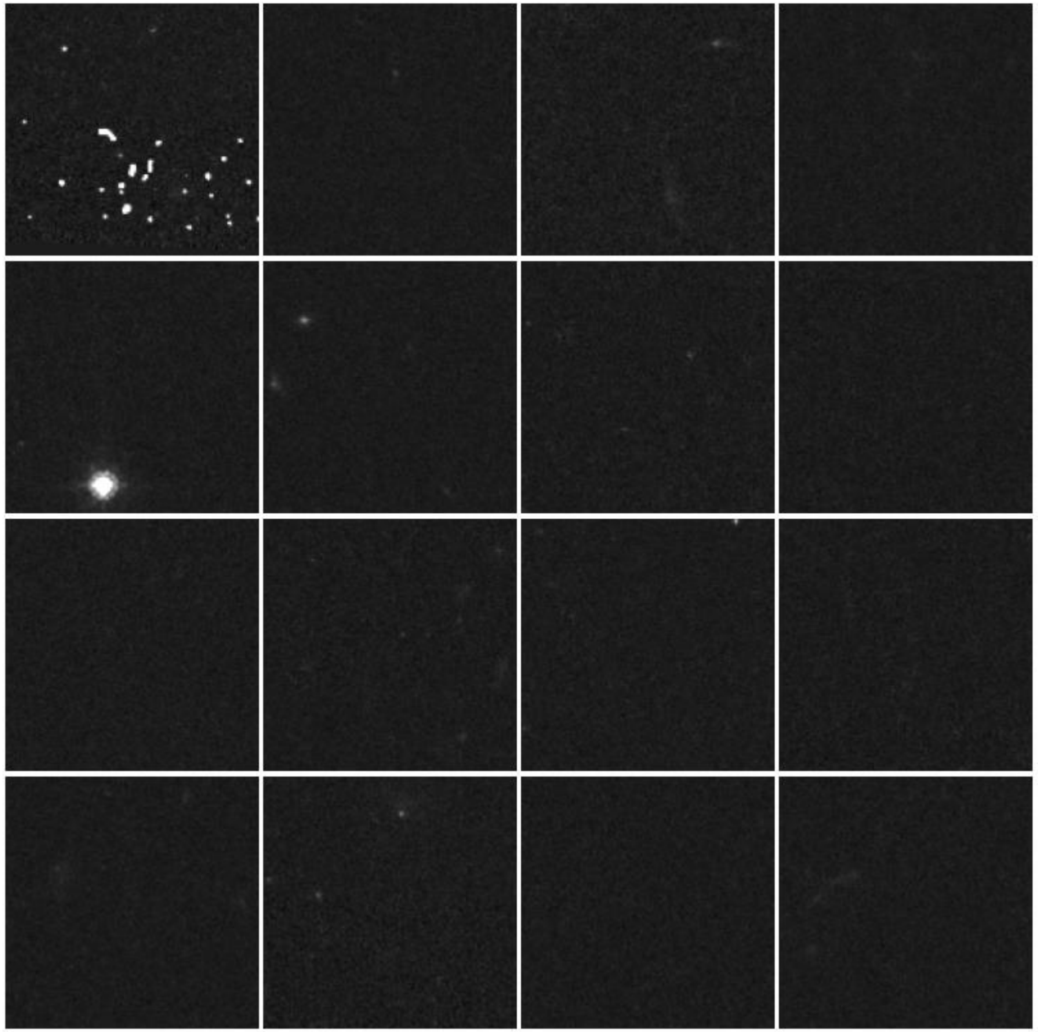
Motivating Examples: Galaxies

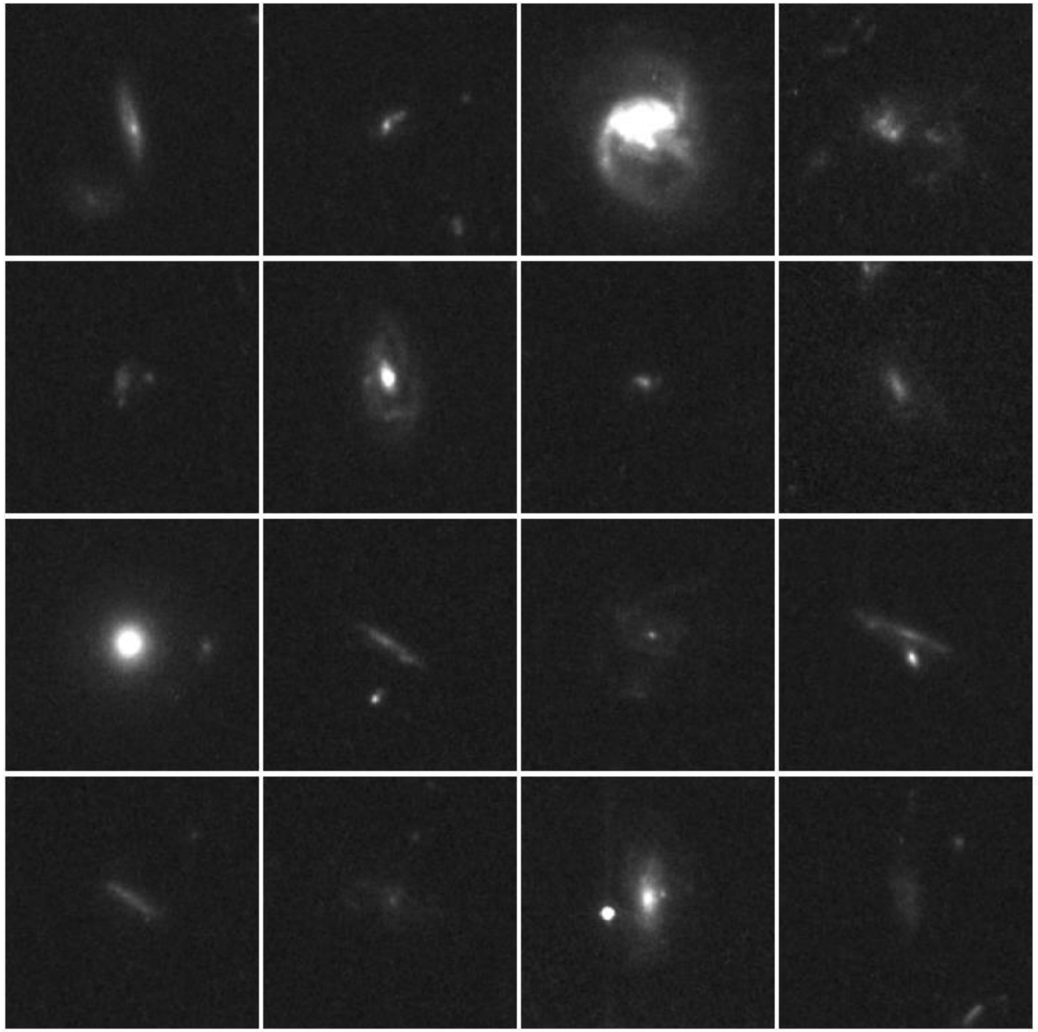
- What does a galaxy **look like**? **How** do we know?
- What part of the light comes from the **central galaxy**? **How** do we know?



Koekemoer et al., 2007

Motivating Examples: Galaxies

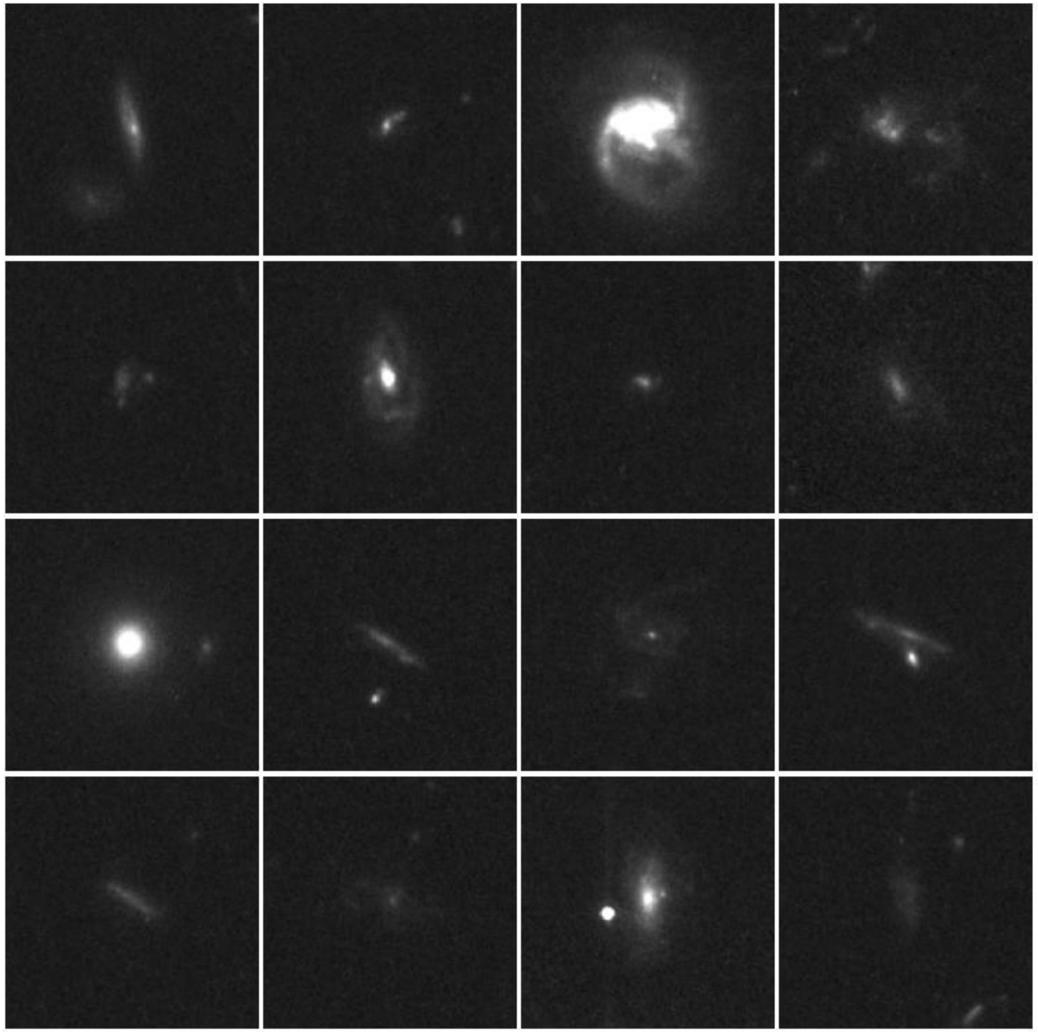




Koekemoer et al., 2007

Motivating Examples: Galaxies

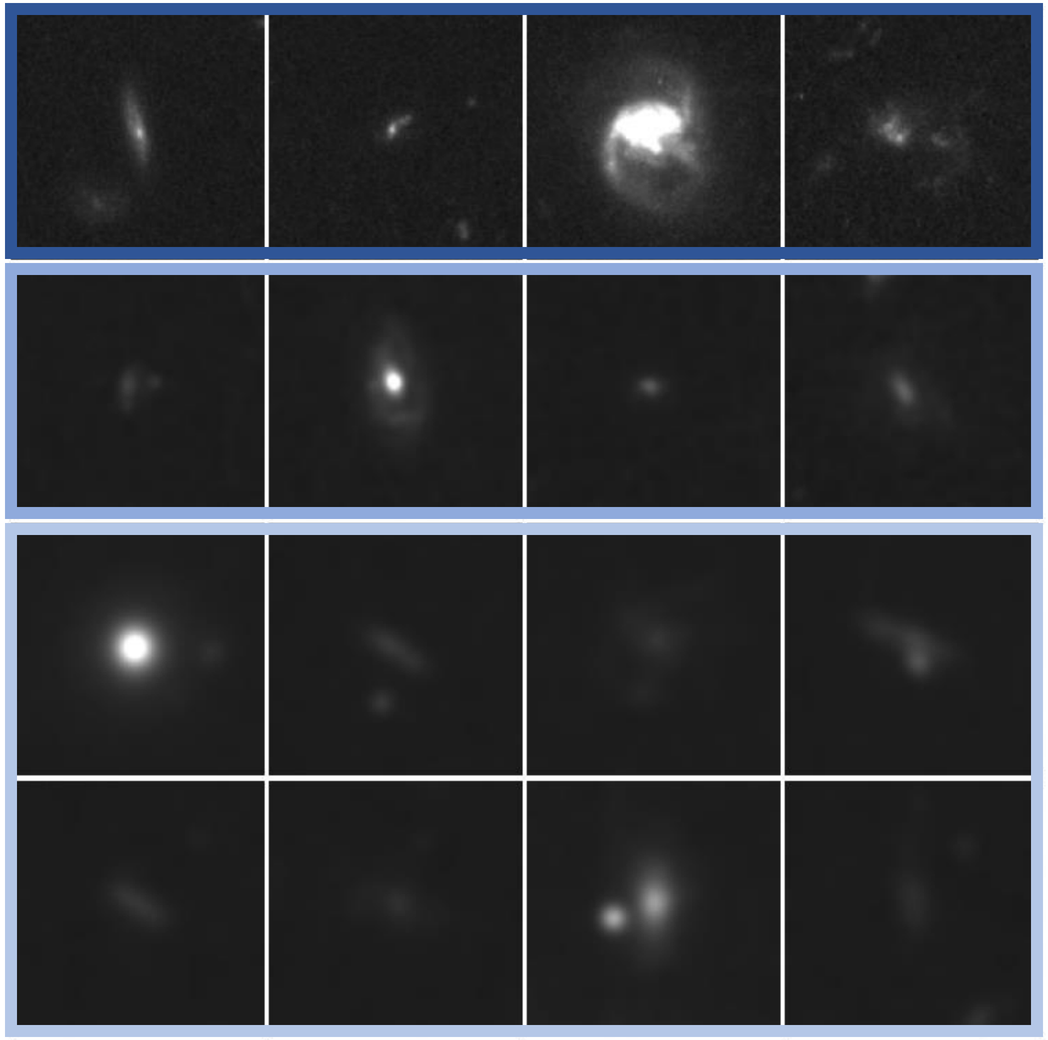
- What does a galaxy **look like**? **How** do we know?
- What part of the light comes from the **central galaxy**? **How** do we know?
- We want:
  - A **prior** for galaxies: $p(\mathbf{x}_{\text{gal}})$
  - **Posterior** sampling of just the galaxy light: $p(\mathbf{x}_{\text{gal}}|\mathbf{y}_{\text{obs}})$
- We have: line-of-sight, noise



Koekemoer et al., 2007

Motivating Examples: Galaxies

- What does a galaxy **look like**? **How** do we know?
- What part of the light comes from the **central galaxy**? **How** do we know?
- We want:
  - A **prior** for galaxies: $p(\mathbf{x}_{\text{gal}})$
  - **Posterior** sampling of just the galaxy light: $p(\mathbf{x}_{\text{gal}}|\mathbf{y}_{\text{obs}})$
- We have: line-of-sight, noise, and **multi-resolution**



Multi-View Source Separation

$$\mathbf{y}_{i_\alpha}^{\alpha} = \left(\sum_{\beta=1}^{N_s} \mathbf{A}_{i_\alpha}^{\alpha\beta} \mathbf{x}_{i_\alpha}^{\beta} \right) + \eta_{i_\alpha}^{\alpha}$$

$$\alpha \in \{1, \ldots, N_{\text{views}}\}$$

$$\beta \in \{1, \ldots, N_s\}$$

Multi-View Source Separation

- Consider:
    - A **noisy observation** (of a specific **view** – i.e. galaxies or randoms)…

$$\mathbf{y}^{\alpha}_{i_\alpha} = \left(\sum_{\beta=1}^{N_s} \mathbf{A}^{\alpha\beta}_{i_\alpha} \mathbf{x}^{\beta}_{i_\alpha} \right) + \eta^{\alpha}_{i_\alpha}$$

$$\alpha \in \{1, \ldots, N_{\text{views}}\}$$

$$\beta \in \{1, \ldots, N_s\}$$

Multi-View Source Separation

- Consider:
  - A **noisy observation** (of a specific **view** – i.e. galaxies or randoms)…
  - composed of a linear mixture of **independent sources**.

$$\mathbf{y}^{\alpha}_{i_\alpha} = \left(\sum_{\beta=1}^{N_s} \mathbf{A}^{\alpha\beta}_{i_\alpha} \mathbf{x}^{\beta}_{i_\alpha}\right) + \eta^{\alpha}_{i_\alpha}$$

$$\alpha \in \{1, \ldots, N_{\text{views}}\}$$

$$\beta \in \{1, \ldots, N_s\}$$

Multi-View Source Separation

- Consider:
  - A **noisy observation** (of a specific **view** – i.e. galaxies or randoms)…
  - composed of a linear mixture of **independent sources**.
  - The mixture of each source is given by a **matrix** that depends on the **view**, the **source**, and the **specific sample**.

$$\mathbf{y}^{\alpha}_{i_\alpha} = \left(\sum_{\beta=1}^{N_s} \mathbf{A}^{\alpha\beta}_{i_\alpha} \mathbf{x}^{\beta}_{i_\alpha} \right) + \eta^{\alpha}_{i_\alpha}$$

$$\alpha \in \{1, \ldots, N_{\text{views}}\}$$

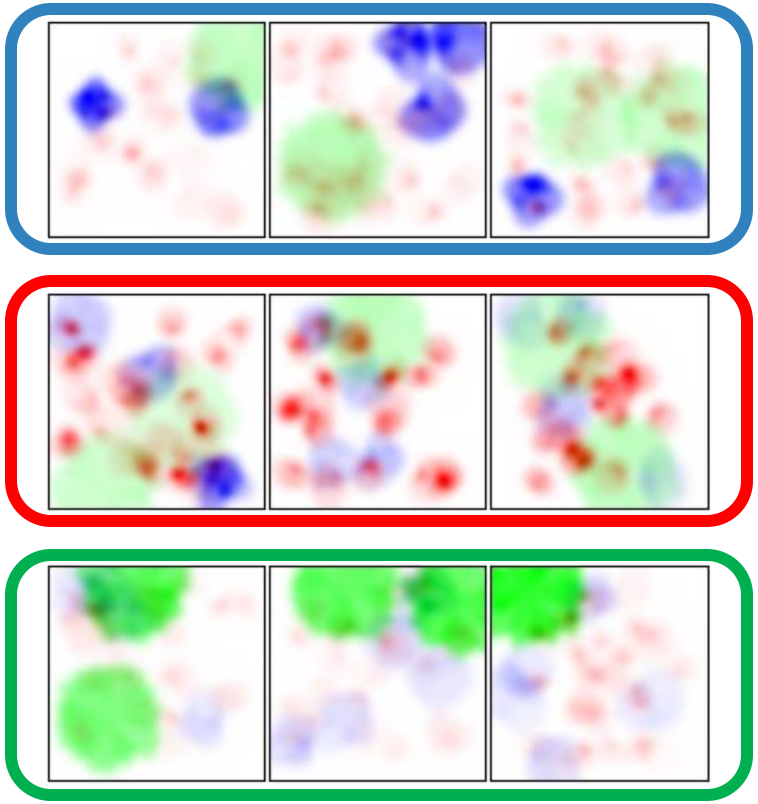
$$\beta \in \{1, \ldots, N_s\}$$

Multi-View Source Separation

$\alpha = 1$

$\alpha = 2$

$\alpha = 3$

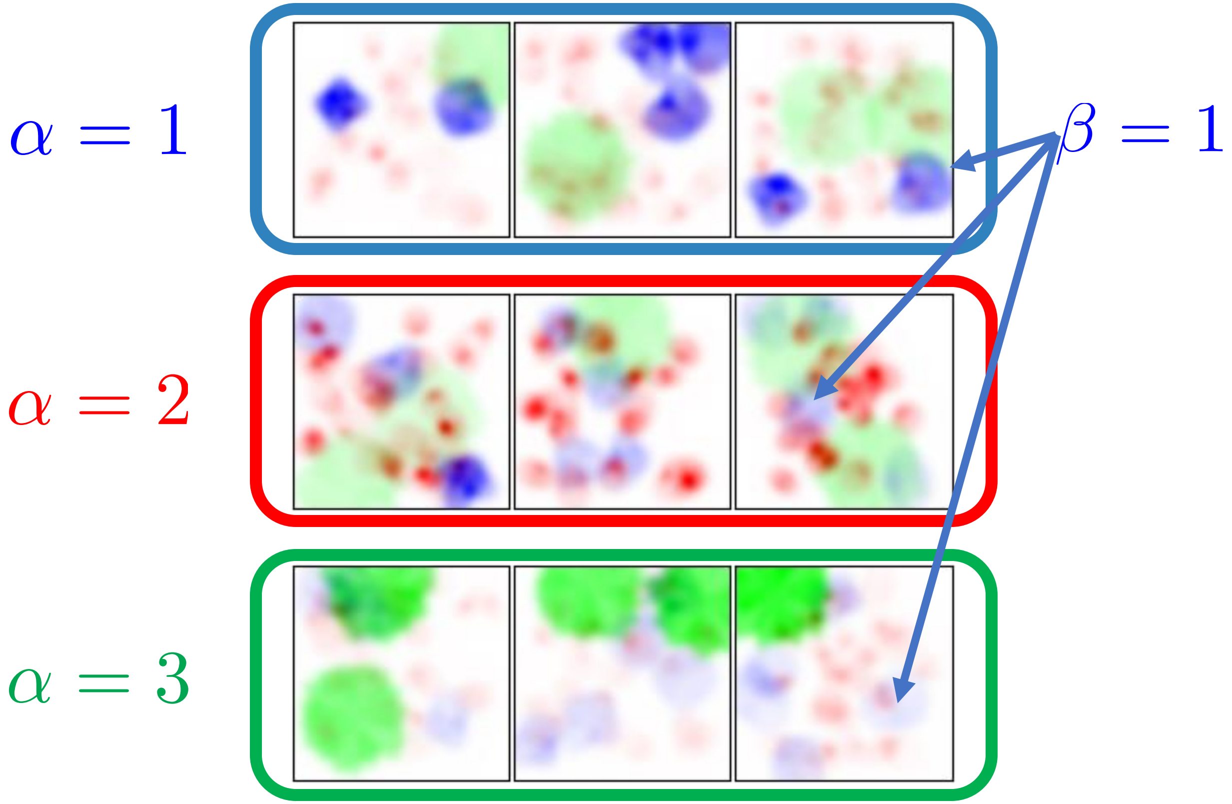


$$\mathbf{y}^{\alpha}_{i_\alpha} = \left(\sum_{\beta=1}^{N_s} \mathbf{A}^{\alpha\beta}_{i_\alpha} \mathbf{x}^{\beta}_{i_\alpha} \right) + \eta^{\alpha}_{i_\alpha}$$

$$\alpha \in \{1, \ldots, N_{\text{views}}\}$$

$$\beta \in \{1, \ldots, N_s\}$$

Multi-View Source Separation

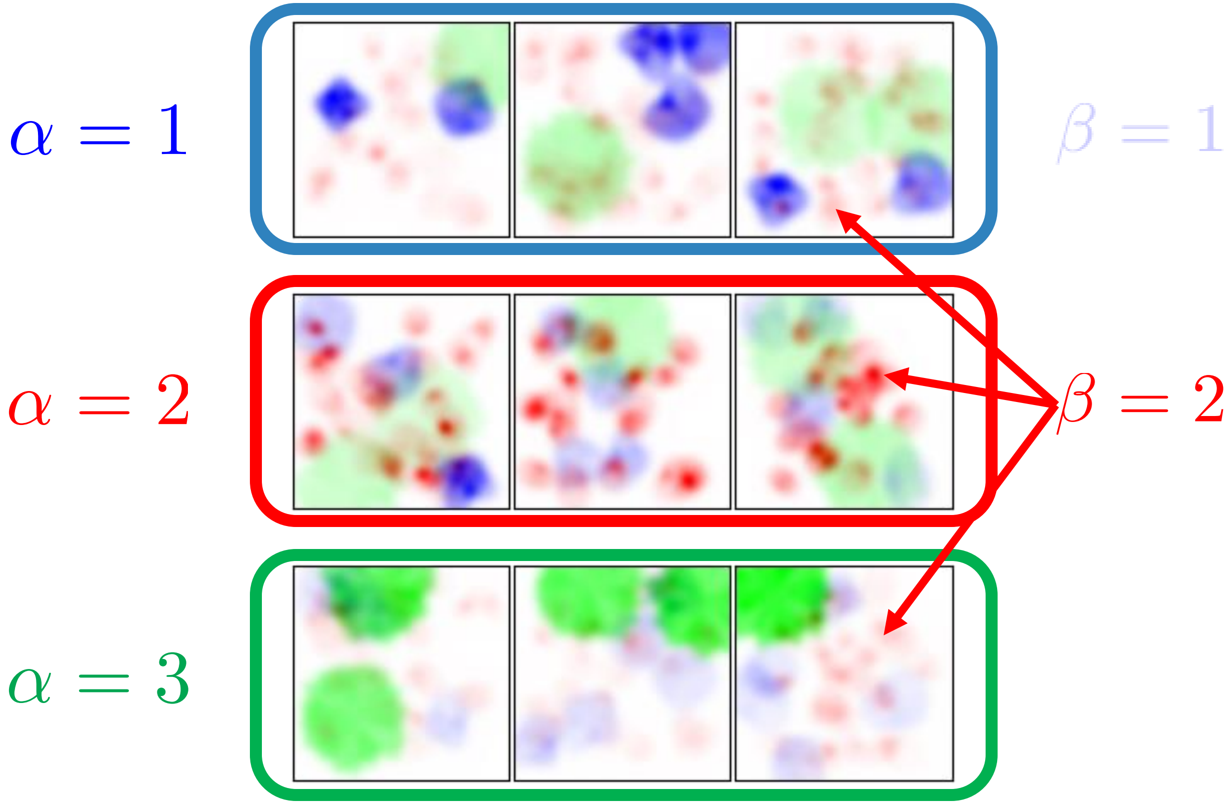


$$\mathbf{y}^{\alpha}_{i_\alpha} = \left(\sum_{\beta=1}^{N_s} \mathbf{A}^{\alpha\beta}_{i_\alpha} \mathbf{x}^{\beta}_{i_\alpha} \right) + \eta^{\alpha}_{i_\alpha}$$

$$\alpha \in \{1, \ldots, N_{\text{views}}\}$$

$$\beta \in \{1, \ldots, N_s\}$$

Multi-View Source Separation

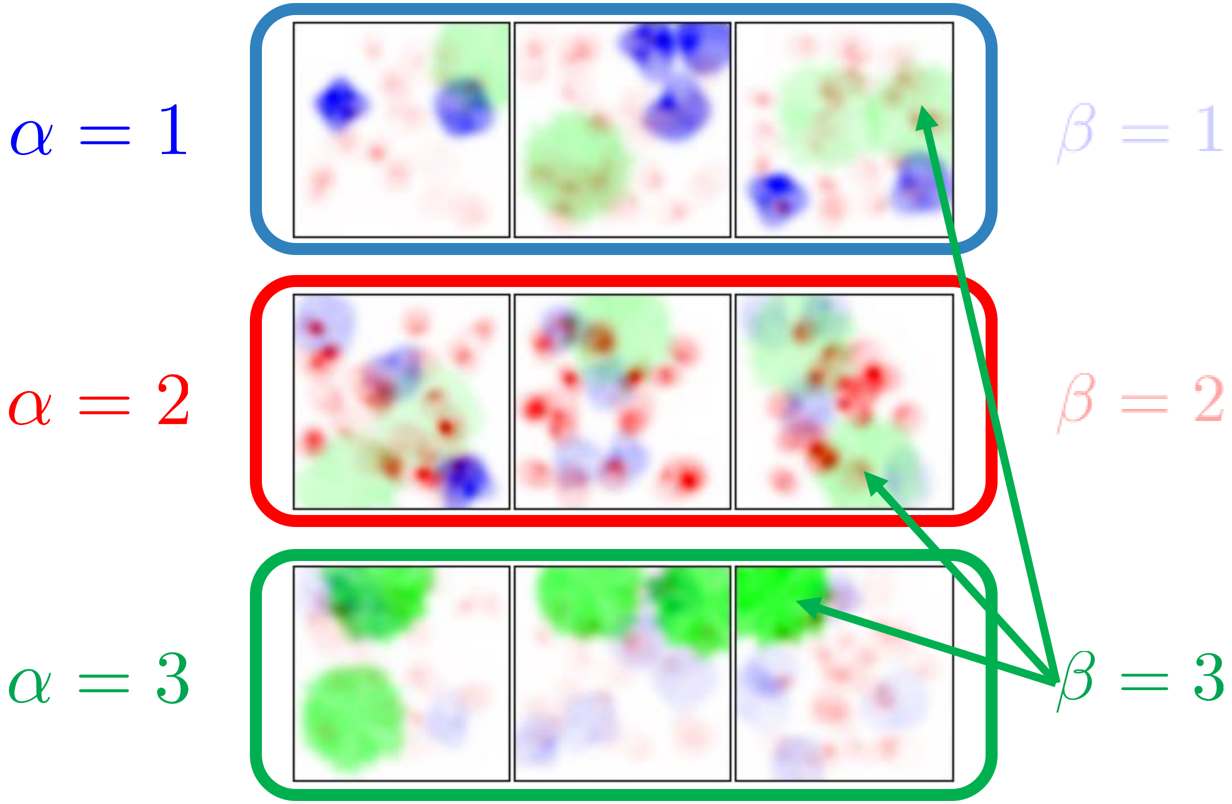


$$\mathbf{y}_{i_\alpha}^{\alpha} = \left(\sum_{\beta=1}^{N_s} \mathbf{A}_{i_\alpha}^{\alpha\beta} \mathbf{x}_{i_\alpha}^{\beta} \right) + \eta_{i_\alpha}^{\alpha}$$

$$\alpha \in \{1, \ldots, N_{\text{views}}\}$$

$$\beta \in \{1, \ldots, N_s\}$$

Multi-View Source Separation



$$\mathbf{y}^{\alpha}_{i_\alpha} = \left(\sum_{\beta=1}^{N_s} \mathbf{A}^{\alpha\beta}_{i_\alpha} \mathbf{x}^{\beta}_{i_\alpha} \right) + \eta^{\alpha}_{i_\alpha}$$

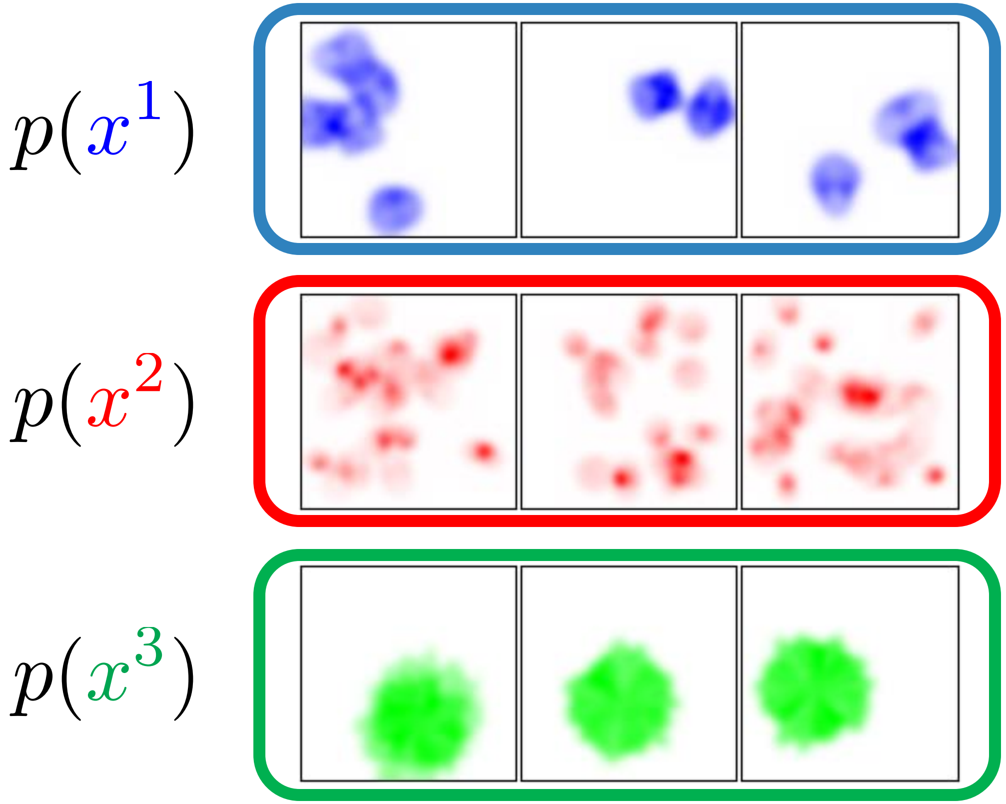
$$\alpha \in \{1, \ldots, N_{\text{views}}\}$$

$$\beta \in \{1, \ldots, N_s\}$$

Multi-View Source Separation

- **Goals**:
  - We aim to infer the individual prior distributions of each source.

$$p(\mathbf{x}^{\beta})$$



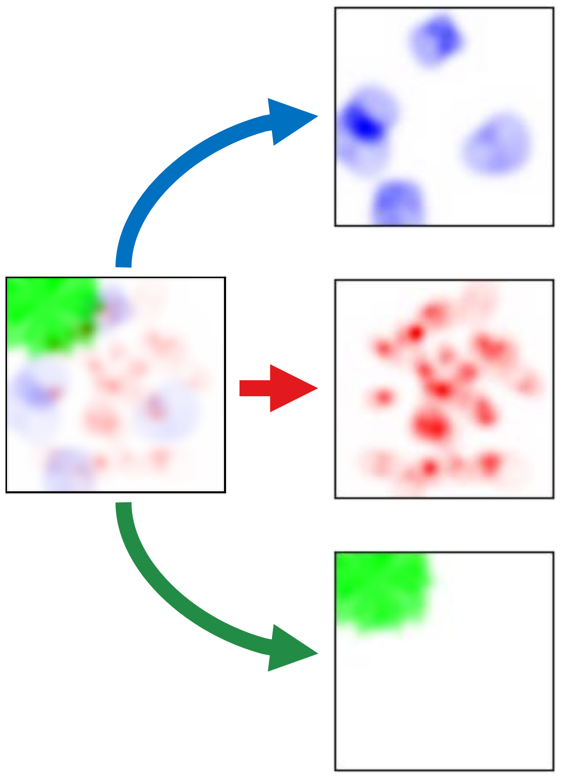
Multi-View Source Separation

- **Goals**:
  - We aim to infer the individual prior distributions of each source.

$$p(\mathbf{x}^{\beta})$$

  - We then want to separate the sources by sampling from the joint posterior:

$$p(\{\mathbf{x}^{\beta}\}|\mathbf{y}_i^{\alpha}, \mathbf{A}_i^{\alpha\beta})$$



Ingredients

1. Flexible Bayesian prior that can capture our unknown source distributions
2. Ability to train that prior given incomplete, noisy data
3. Ability to disentangle multiple sources from the same observation

Song et al., 2021

Review: Diffusion Models

- Start with observed samples and perturb them through a diffusion process given by the SDE:

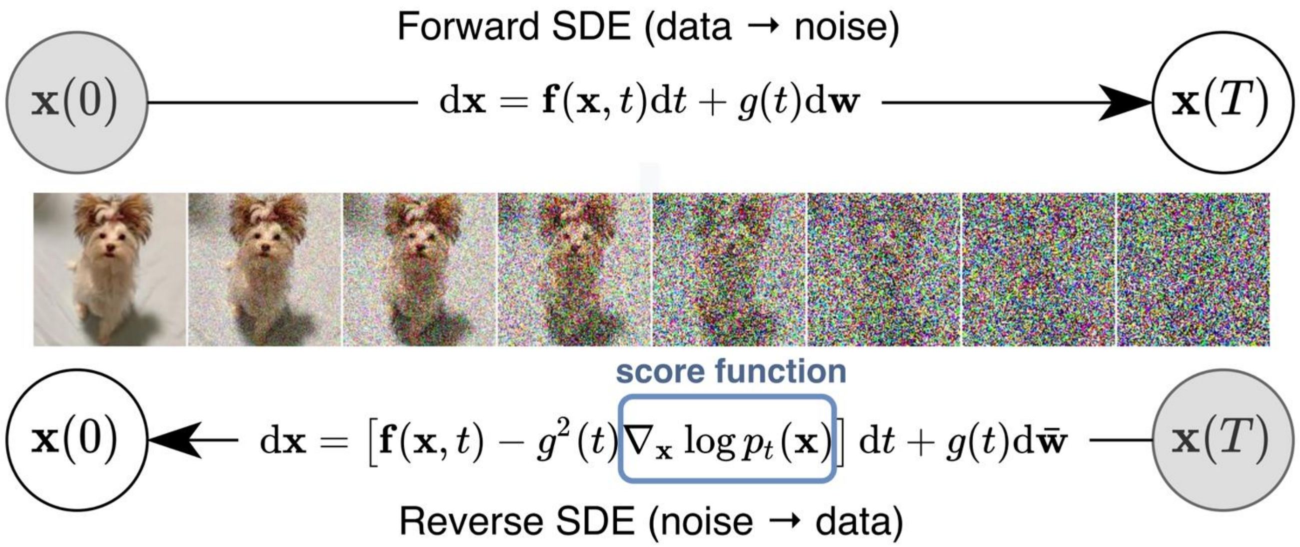
$$\mathrm{d}x_t = f_t x_t \mathrm{d}t + g_t \mathrm{d}w_t$$

- The reverse SDE is given by:

$$\mathrm{d}x_t = \left[f_t x_t - g(t)^2 \boxed{\nabla_{x_t} \log p(x_t)}\right] \mathrm{d}t + g(t)\mathrm{d}\bar{w}_t$$

Song et al., 2021

Review: Diffusion Models



- With enough perturbation, final timestep is equivalent to noise.
- Integrating backwards with reverse SDE gives a sample from the original distribution. We just need the **score.** Train a model to estimate it.

Song et al., 2021

Review: Diffusion Models as Priors

- What if we want to condition on an observation? We need the **score of the posterior**:

$$\mathrm{d}x_t = \left[f_t x_t - g(t)^2 \boxed{\nabla_{x_t} \log p(x_t|y)}\right] \mathrm{d}t + g(t)\mathrm{d}\bar{w}_t$$

- From Bayes rule:

$$\nabla_{x_t} \log p(x_t|y) = \nabla_{x_t} \log p(x_t) + \nabla_{x_t} \log p(y|x_t).$$

- First term we already have, second term comes from integral:

$$p(y|x_t) = \int p(y|x_0)p(x_0|x_t)\mathrm{d}x_0$$

Ingredients

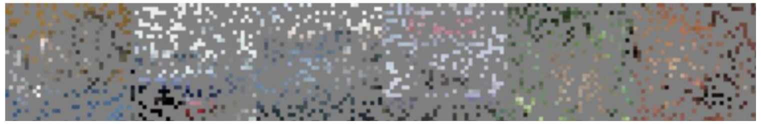
1. ~~Flexible Bayesian prior that can capture our unknown source distributions~~
2. Ability to train that prior given incomplete, noisy data
3. Ability to disentangle multiple sources from the same observation

Rozet et al., 2024

Review: Diffusion Priors from Observations

- In MVSS, we don’t know the prior, so we still can’t sample from the likelihood.

y

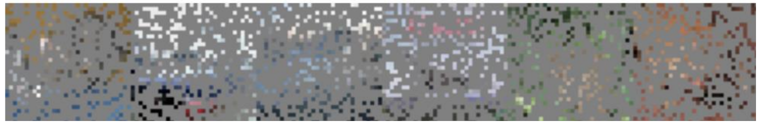


x

Review: Diffusion Priors from Observations

- In MVSS, we don't know the prior, so we still can't sample from the likelihood. The solution is **Expectation-Maximization**:
    1. Start from an initial diffusion model

y

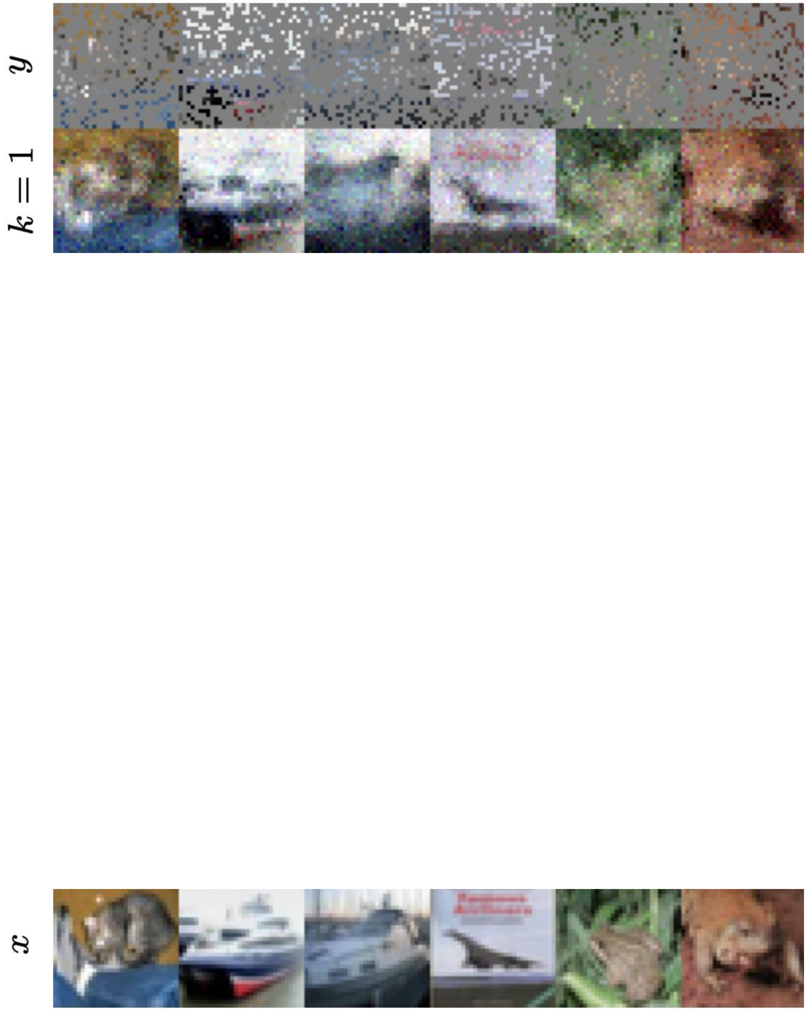


x

Rozet et al., 2024

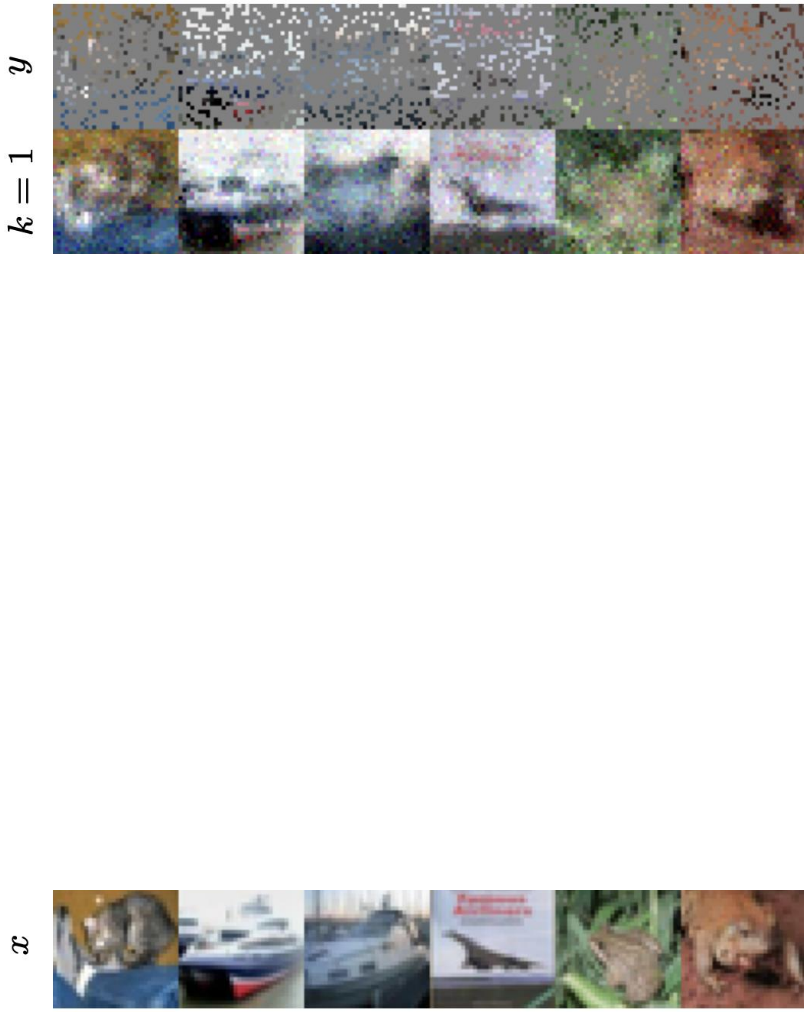
Review: Diffusion Priors from Observations

- In MVSS, we don’t know the prior, so we still can’t sample from the likelihood. The solution is **Expectation-Maximization**:
  1. Start from an initial diffusion model
  2. **Expectation:** Sample from the posterior of your observations given current diffusion model



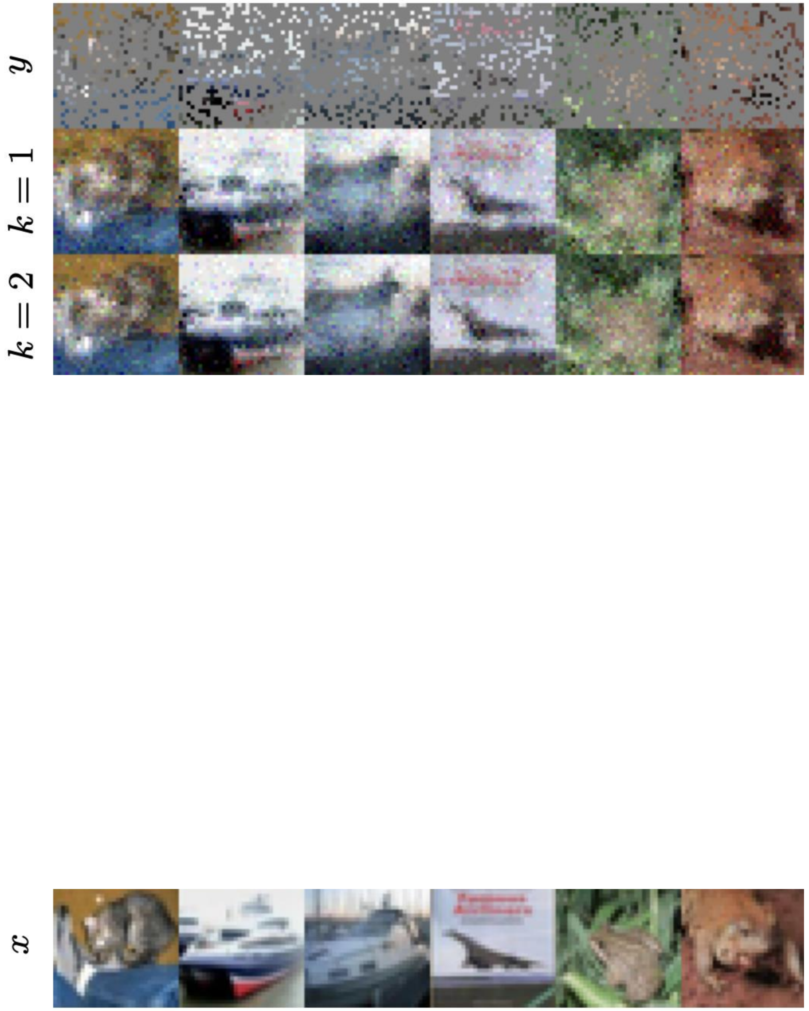
Review: Diffusion Priors from Observations

- In MVSS, we don't know the prior, so we still can't sample from the likelihood. The solution is **Expectation-Maximization**:
    1. Start from an initial diffusion model
    2. **Expectation:** Sample from the posterior of your observations given current diffusion model
    3. **Maximization:** Fit the diffusion model to posterior samples



Review: Diffusion Priors from Observations

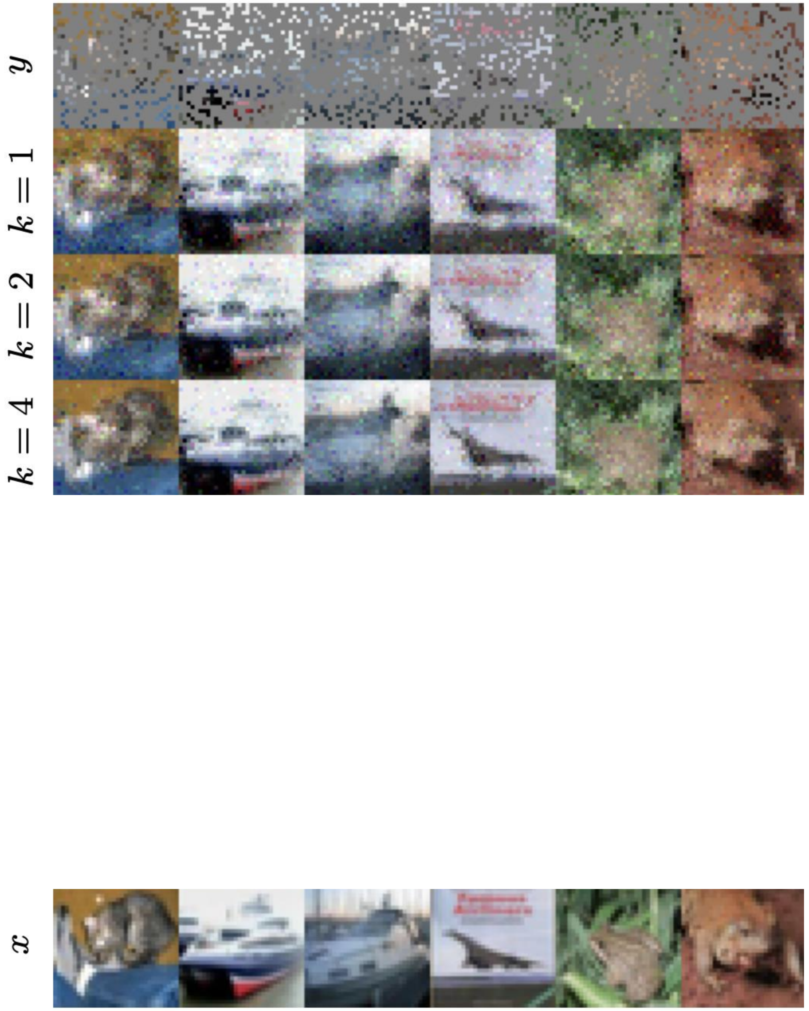
- In MVSS, we don't know the prior, so we still can't sample from the likelihood. The solution is **Expectation-Maximization**:
    1. Start from an initial diffusion model
    2. **Expectation:** Sample from the posterior of your observations given current diffusion model
    3. **Maximization:** Fit the diffusion model to posterior samples
    4. Return to step 2 until converged



Rozet et al., 2024

Review: Diffusion Priors from Observations

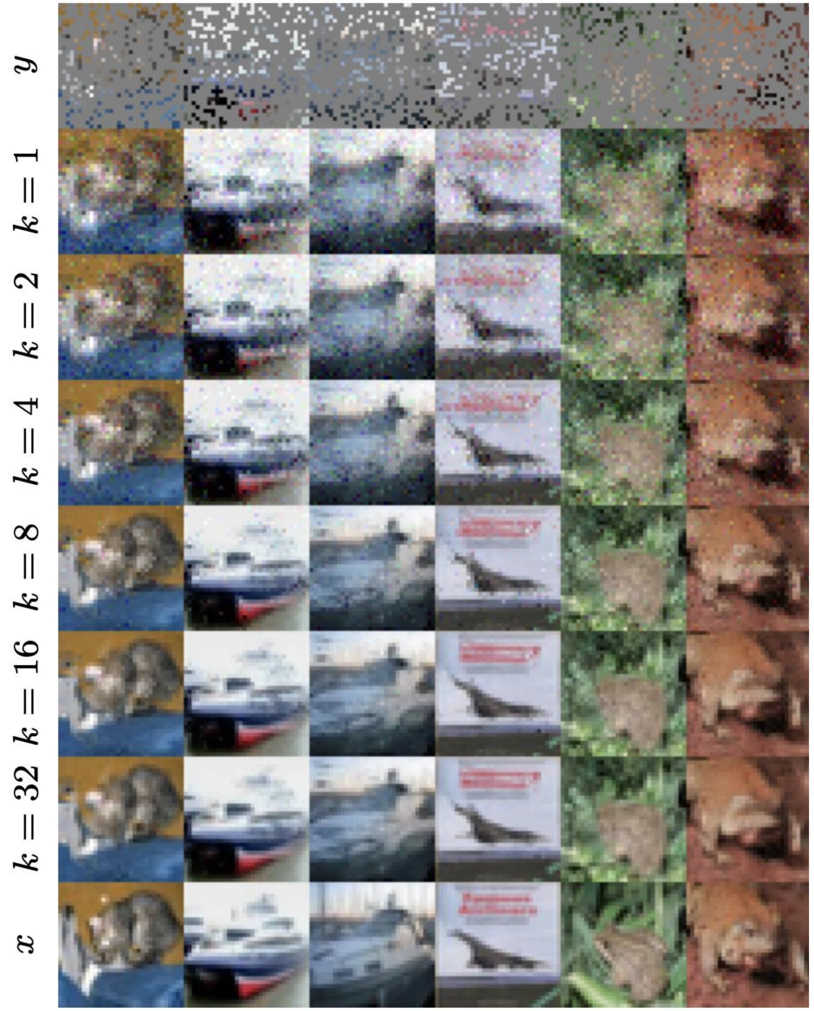
- In MVSS, we don't know the prior, so we still can't sample from the likelihood. The solution is **Expectation-Maximization**:
    1. Start from an initial diffusion model
    2. **Expectation:** Sample from the posterior of your observations given current diffusion model
    3. **Maximization:** Fit the diffusion model to posterior samples
    4. Return to step 2 until converged



Rozet et al., 2024

Review: Diffusion Priors from Observations

- In MVSS, we don't know the prior, so we still can't sample from the likelihood. The solution is **Expectation-Maximization**:
  1. Start from an initial diffusion model
  2. **Expectation:** Sample from the posterior of your observations given current diffusion model
  3. **Maximization:** Fit the diffusion model to posterior samples
  4. Return to step 2 until converged

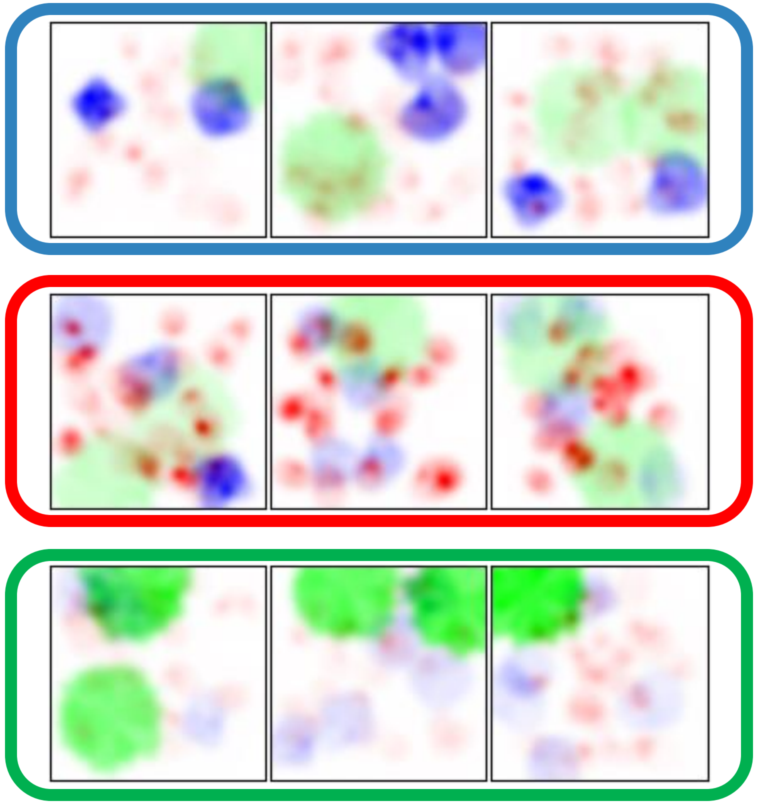


Ingredients

1. ~~Flexible Bayesian prior that can capture our unknown source distributions~~
2. ~~Ability to train that prior given incomplete, noisy data~~
3. Ability to disentangle multiple sources from the same observation

Joint Posterior Sampling with Diffusion Priors

Observations

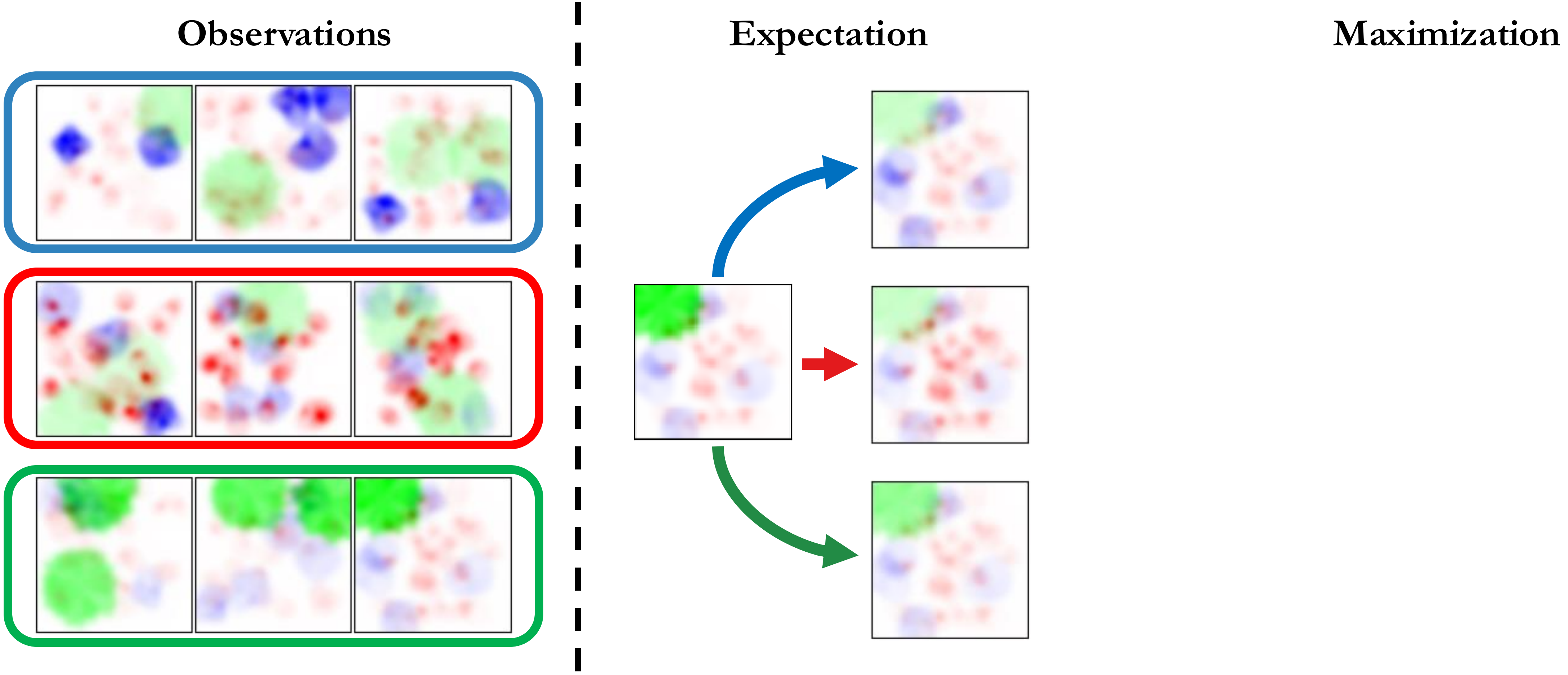


Expectation

Maximization

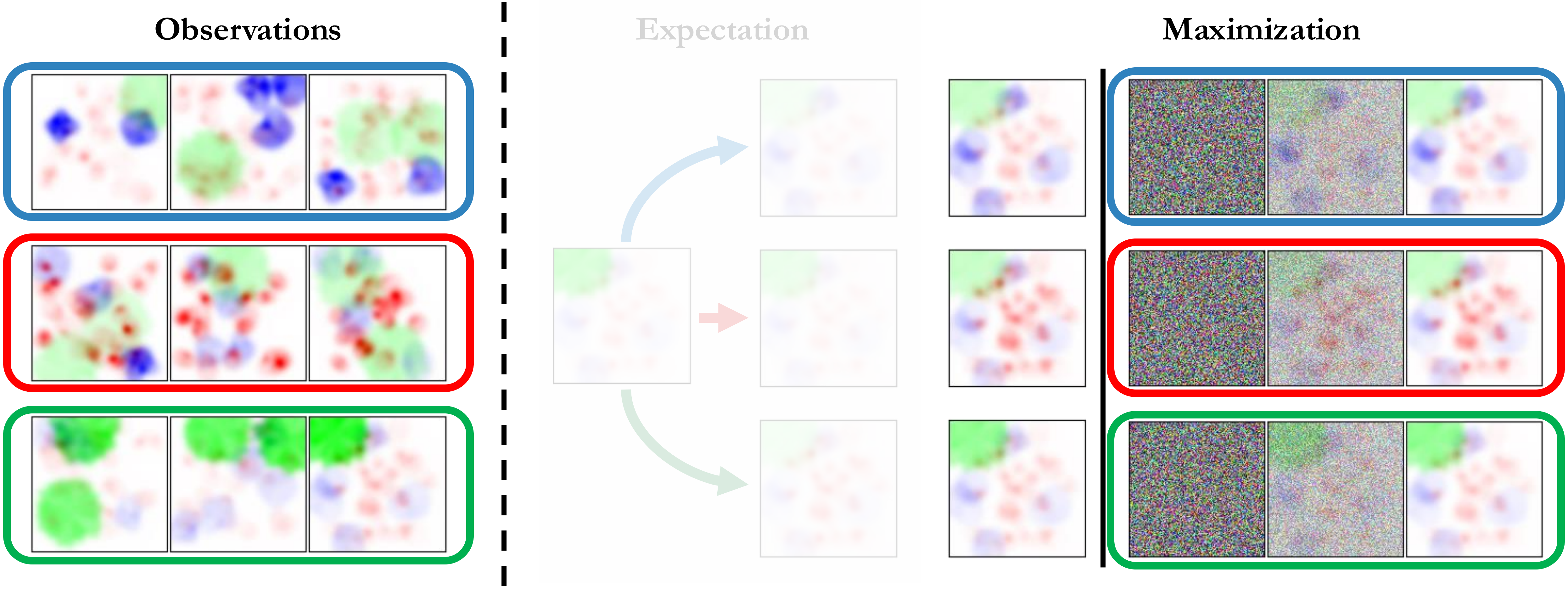
Wagner-Carena, Akhmetzhanova, et al., submitted.

Joint Posterior Sampling with Diffusion Priors



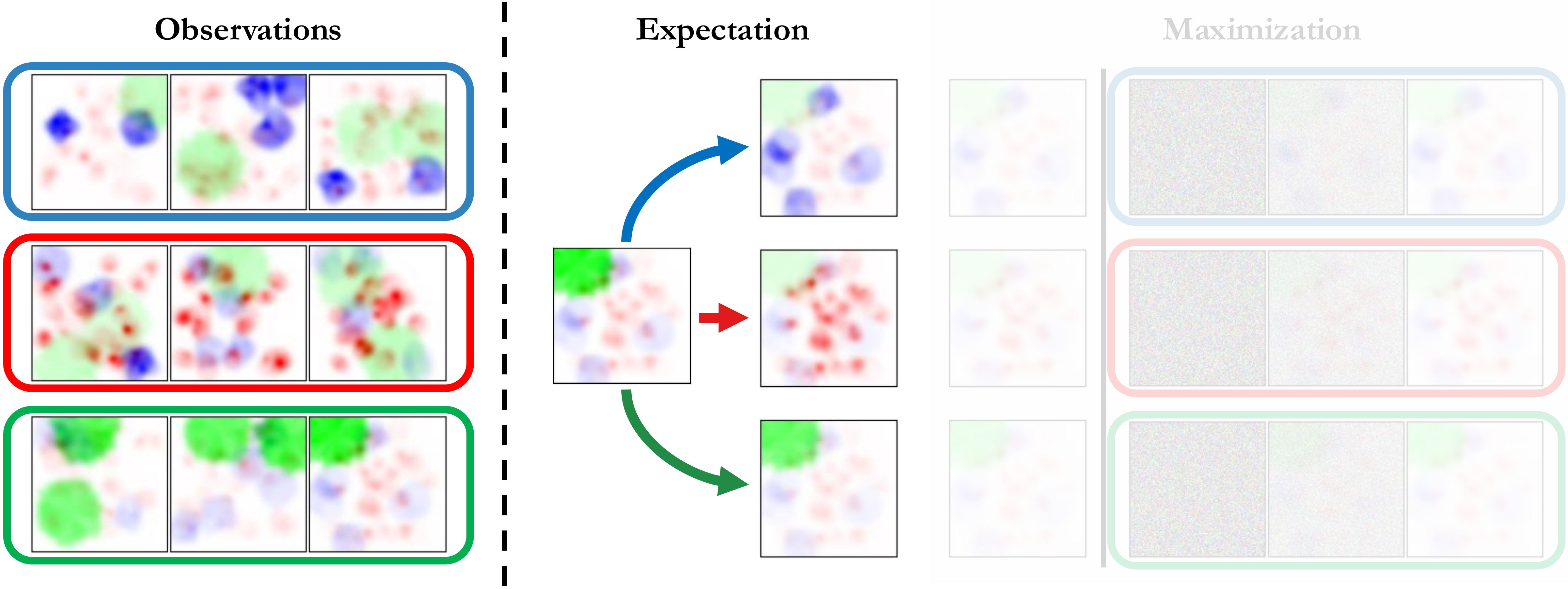
Wagner-Carena, Akhmetzhanova, et al., submitted.

Joint Posterior Sampling with Diffusion Priors

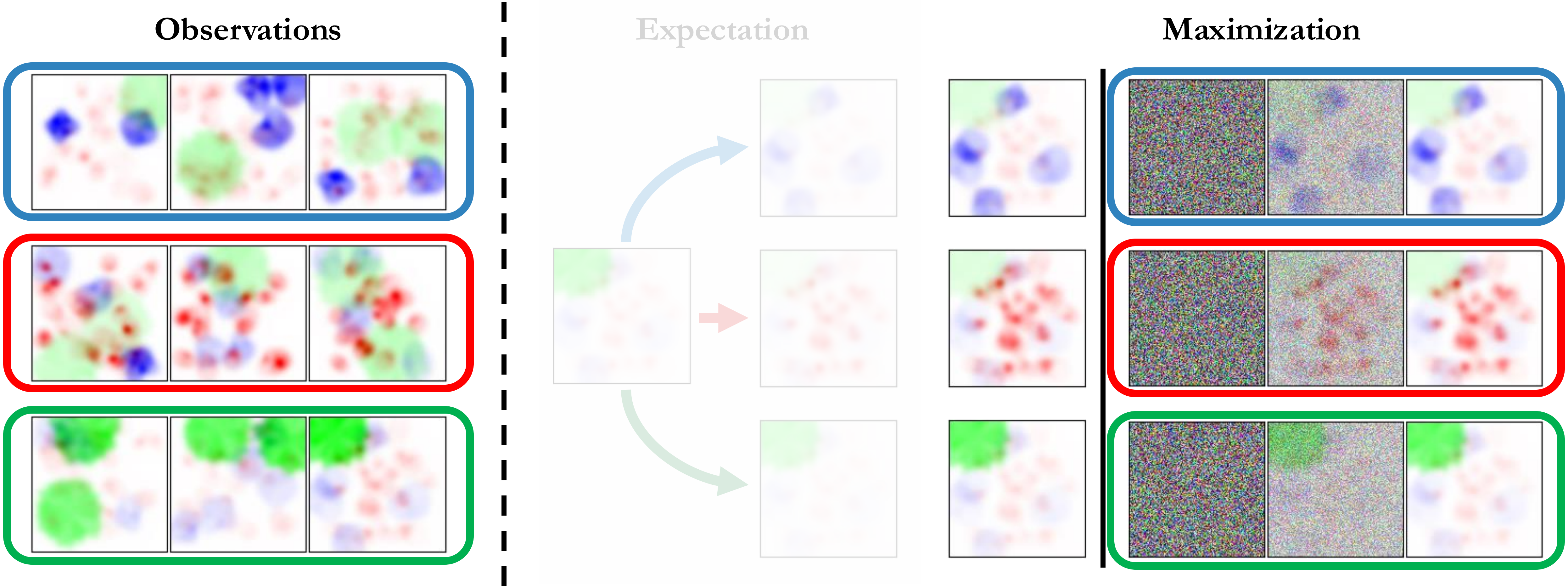


Wagner-Carena, Akhmetzhanova, et al., submitted.

Joint Posterior Sampling with Diffusion Priors

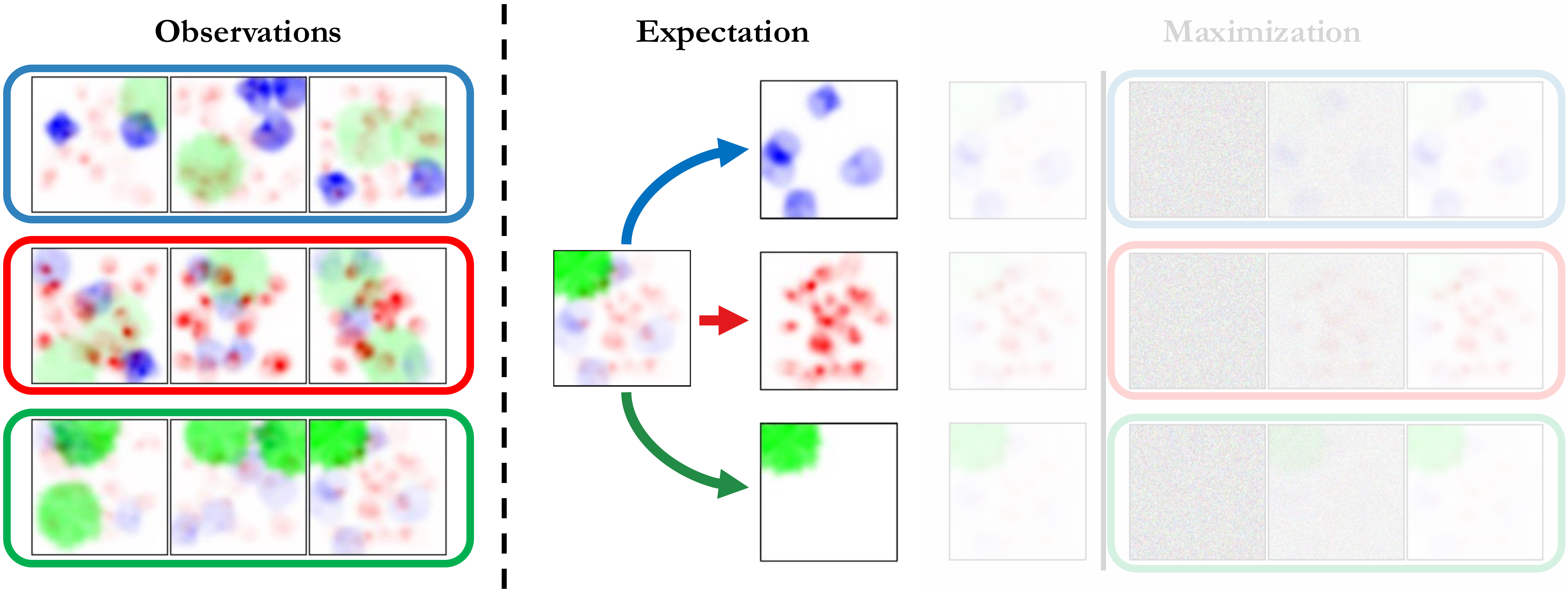


Joint Posterior Sampling with Diffusion Priors



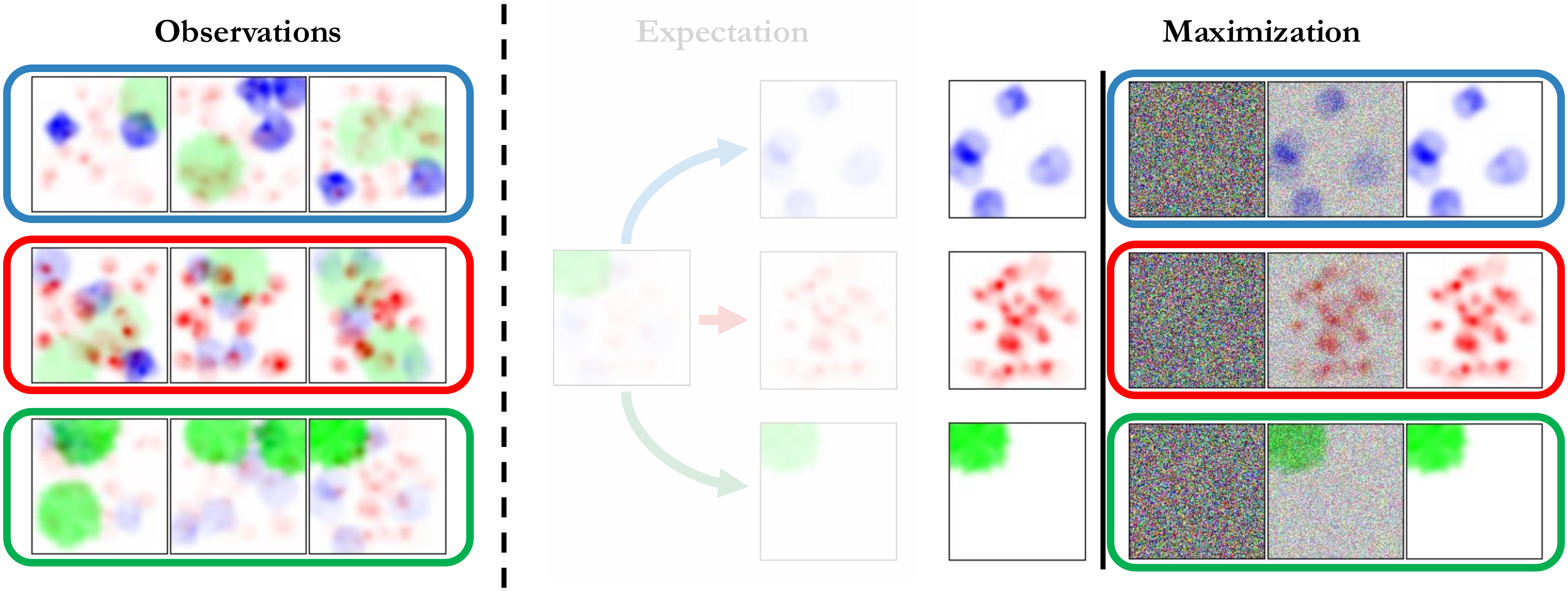
Wagner-Carena, Akhmetzhanova, et al., submitted.

Joint Posterior Sampling with Diffusion Priors



Wagner-Carena, Akhmetzhanova, et al., submitted.

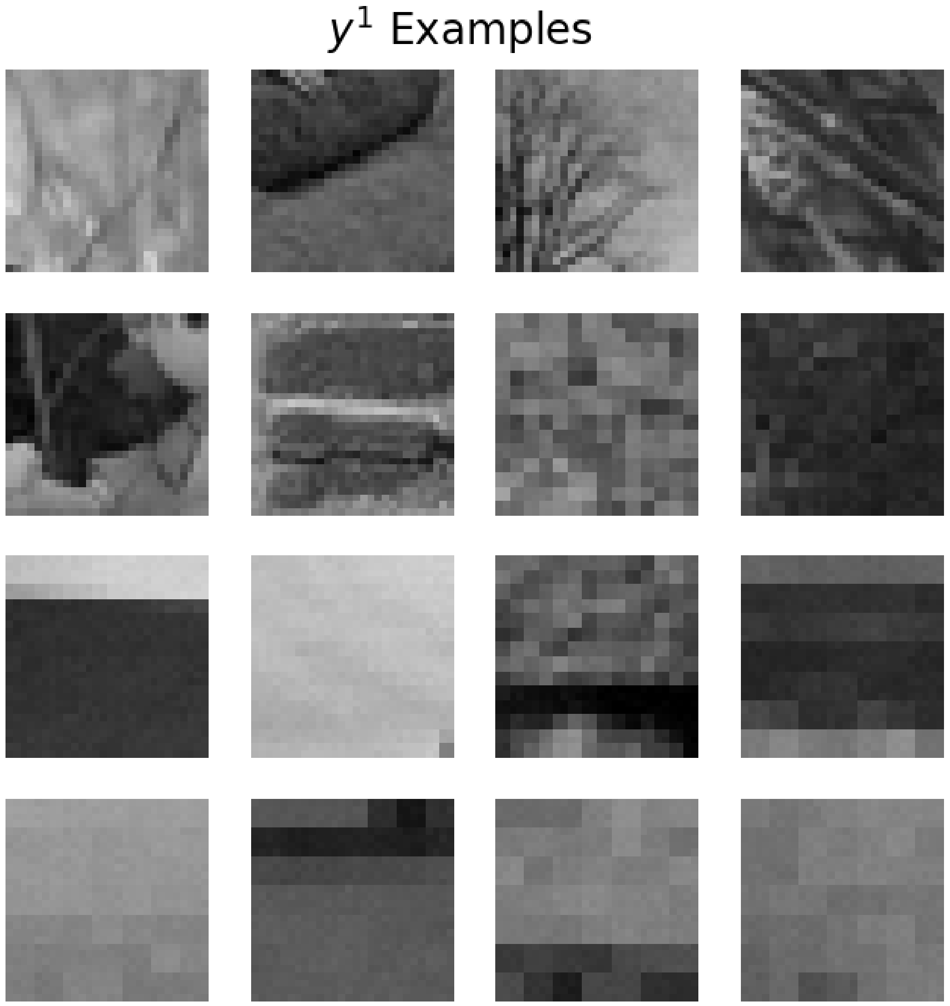
Joint Posterior Sampling with Diffusion Priors



Ingredients

1. ~~Flexible Bayesian prior that can capture our unknown source distributions~~
2. ~~Ability to train that prior given incomplete, noisy data~~
3. ~~Ability to disentangle multiple sources from the same observation~~

Image Data: Contrastive



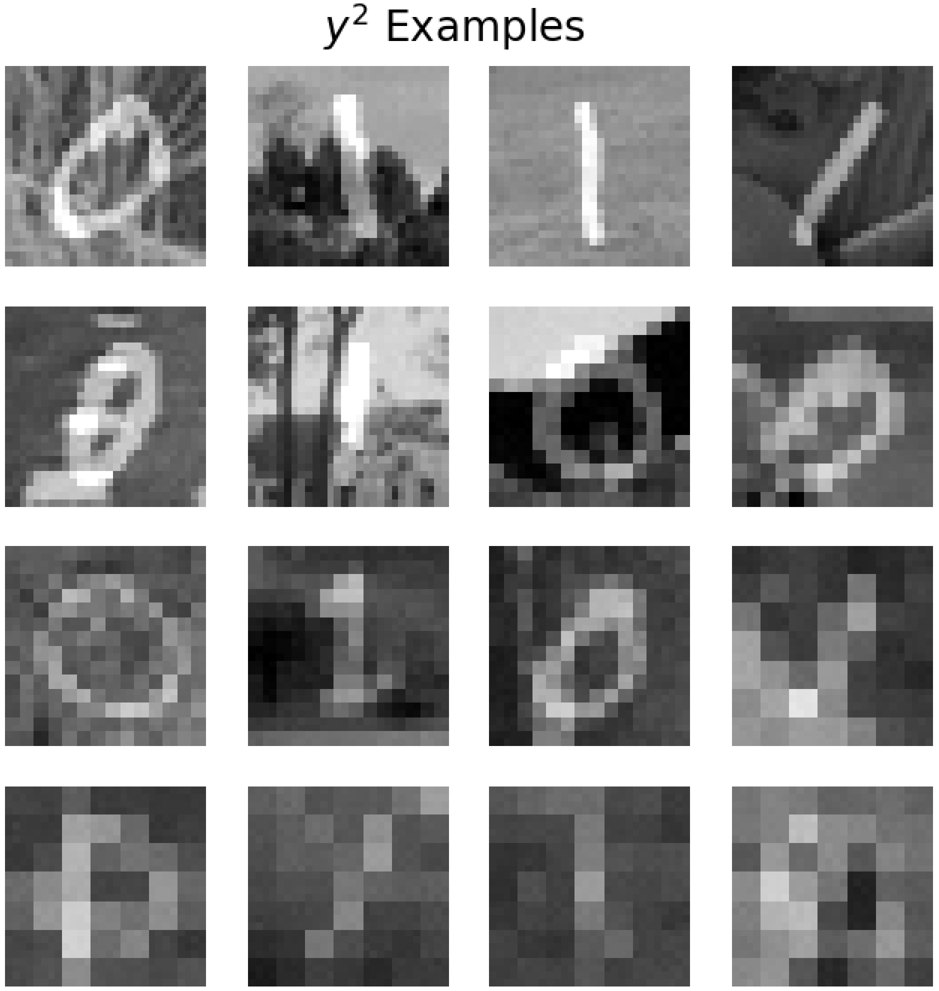


Image Data: Contrastive

- Two views:
    1. **Pure grass** background
    2. **Mixture** of digits and grass background
- Assume we have **noisy** and **incomplete** data
- **Goal**: separate two source distributions
    - Start by training initial grass distribution from first view, then train MNIST distribution using second view

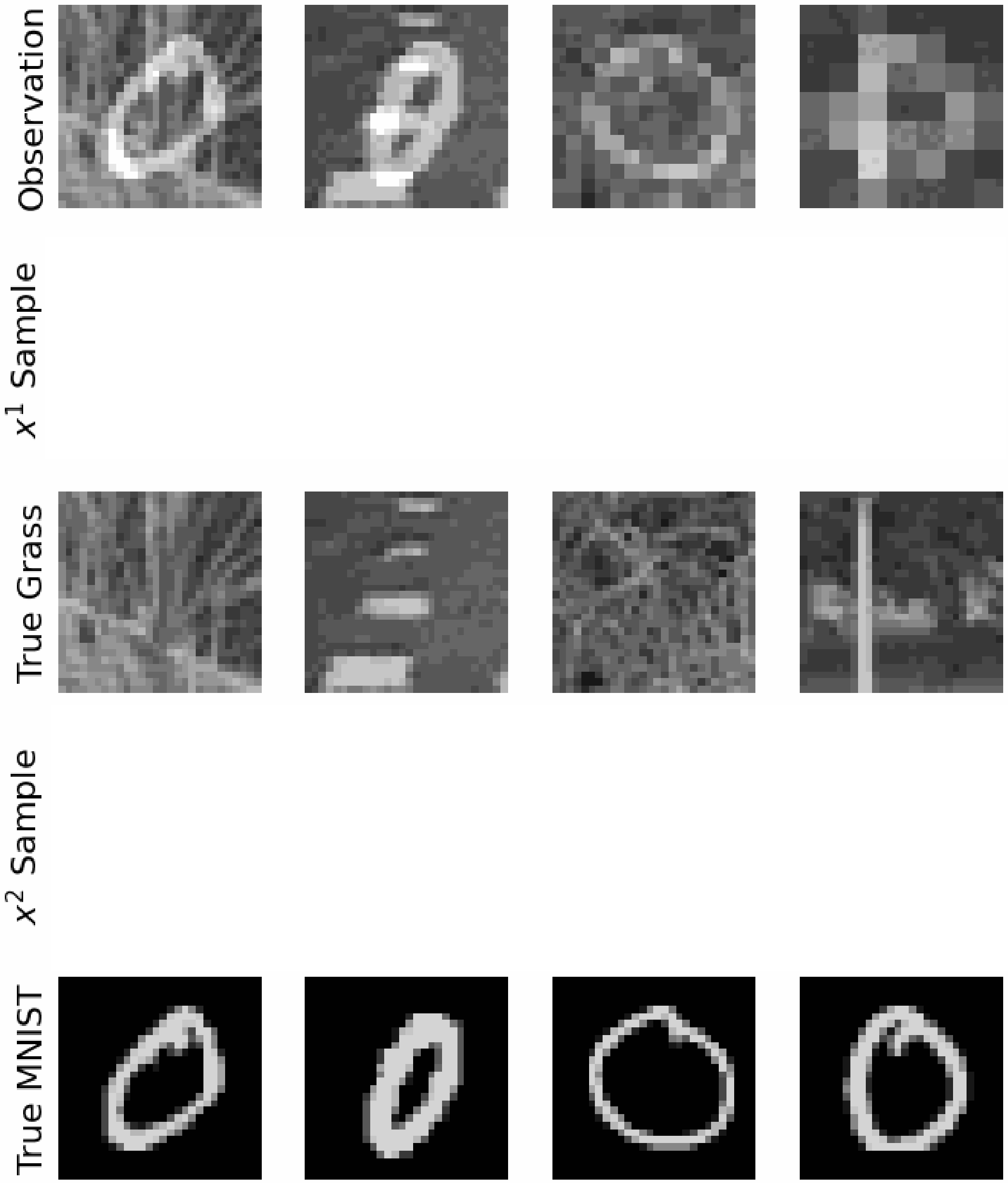
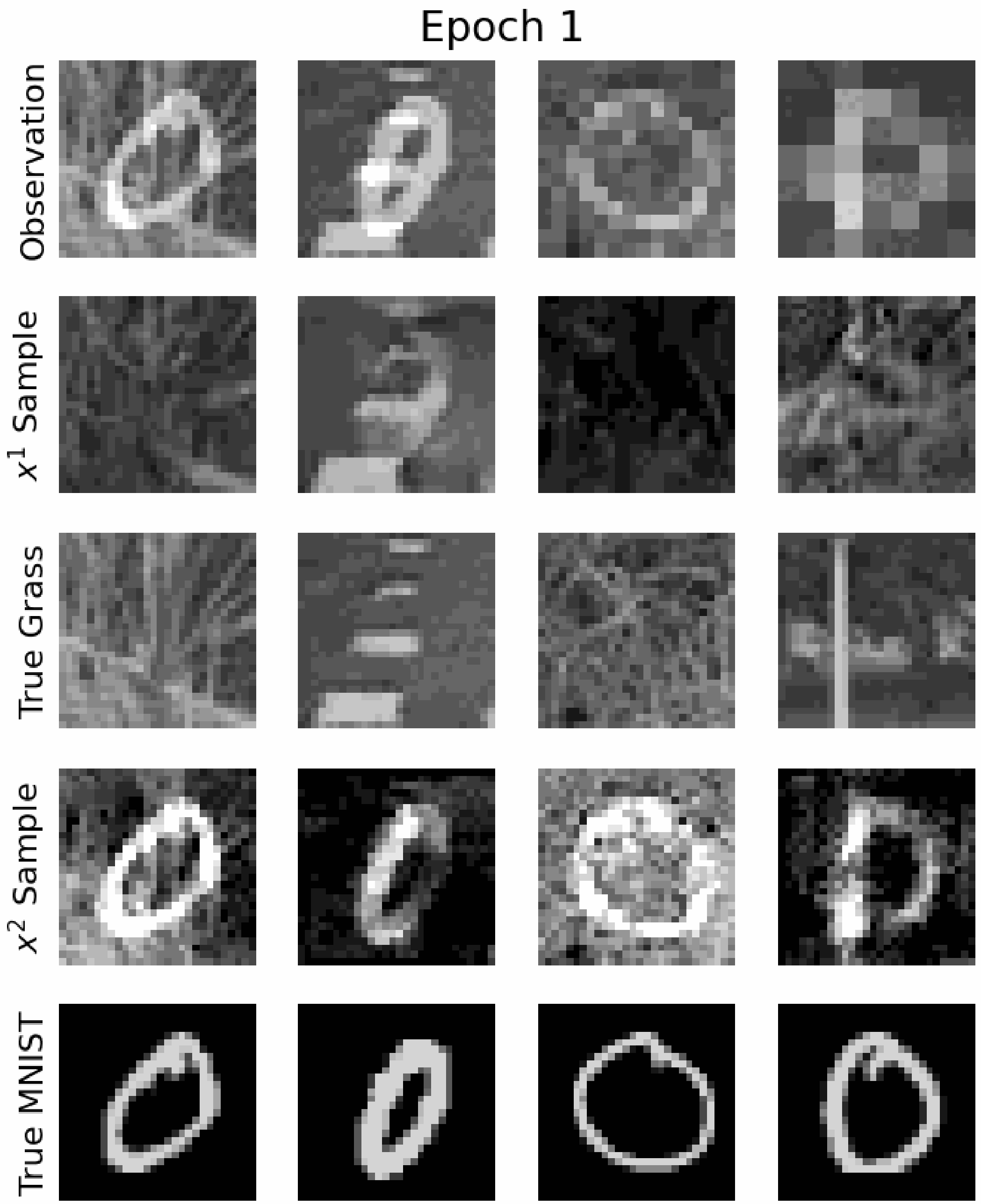


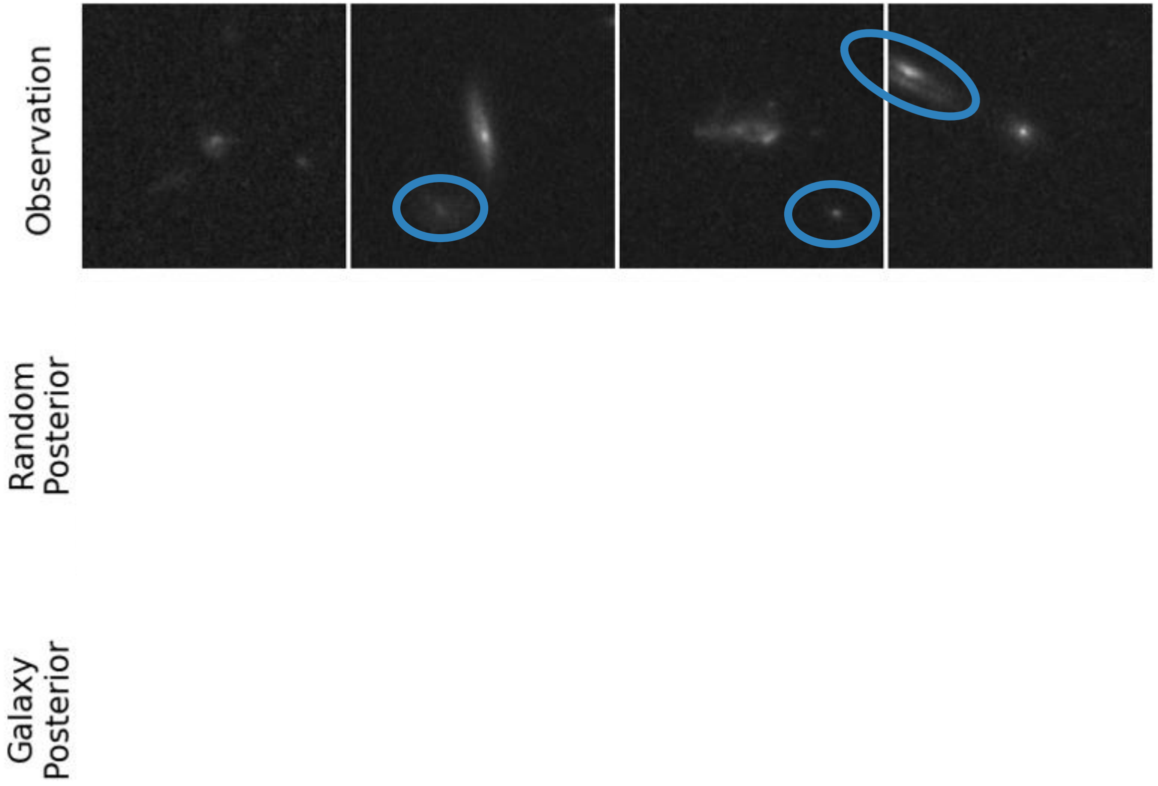
Image Data: Contrastive

- Two views:
  1. **Pure grass** background
  2. **Mixture** of digits and grass background
- Assume we have **noisy** and **incomplete** data
- **Goal**: separate two source distributions
  - Start by training initial grass distribution from first view, then train MNIST distribution using second view
- Successfully **disentangled sources!**



Galaxies: Contrastive

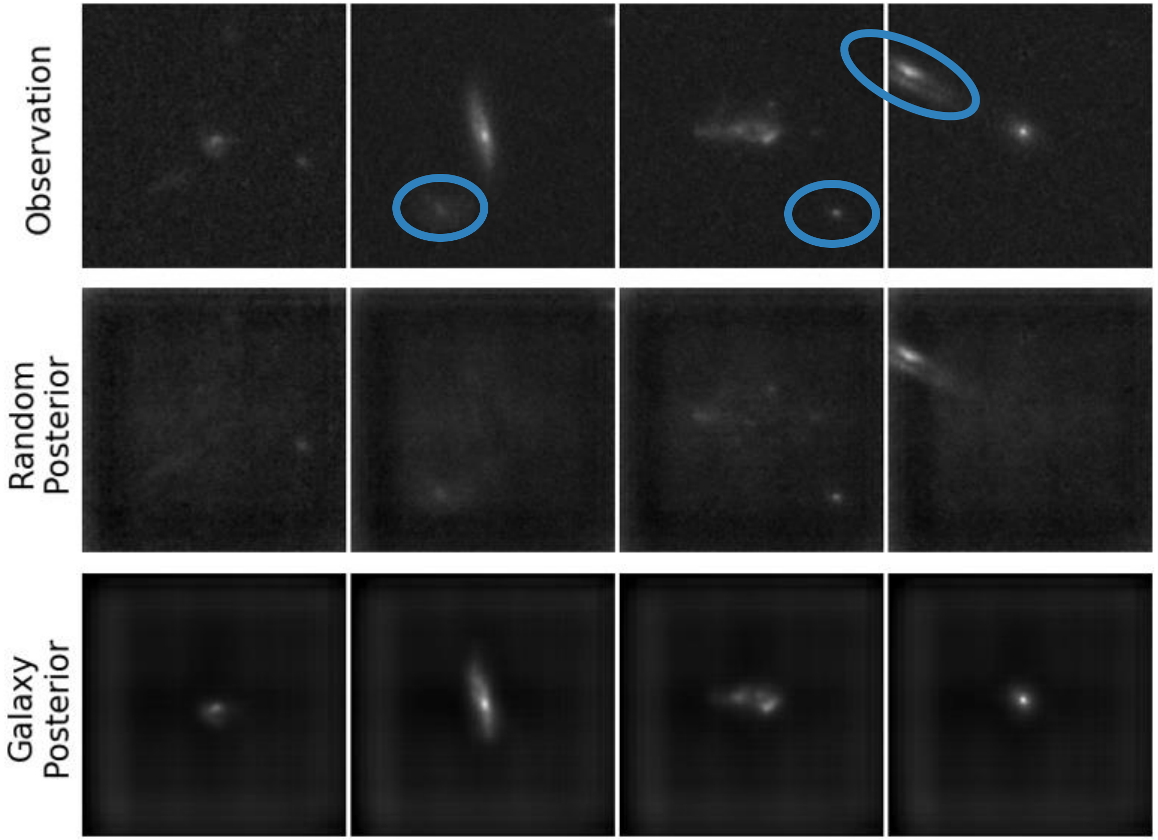
- Two views:
    1. **Random** foreground / background
    2. **Galaxies** with the foreground / background
- **Goal**: separate two source distributions (galaxies and randoms)



Wagner-Carena, Akhmetzhanova, et al., submitted.

Galaxies: Contrastive

- Two views:
    1. **Random** foreground / background
    2. **Galaxies** with the foreground / background
- **Goal**: separate two source distributions (galaxies and randoms)
- Convincing **separation**!
    - Not perfect:
        - Artifacts from random model
        - Metrics to evaluate performance



Conclusions

- For astrophysics, pristine, isolated observations are rare; leveraging these datasets requires solving a **source separation problem**
- We've presented a method that is **effective** even when:
  - **No source is ever individually observed** (mixed sources)
  - Observations are **noisy and incomplete**
  - **Resolution and mixing varies** (i.e. different observatories)
- Diffusion models enable us to directly probe what our sources look like
  - **Data-driven priors** form our next-generation sky surveys
  - **Statistically principled and interpretable posteriors**

Extra Slides

1D Manifolds: Mixed

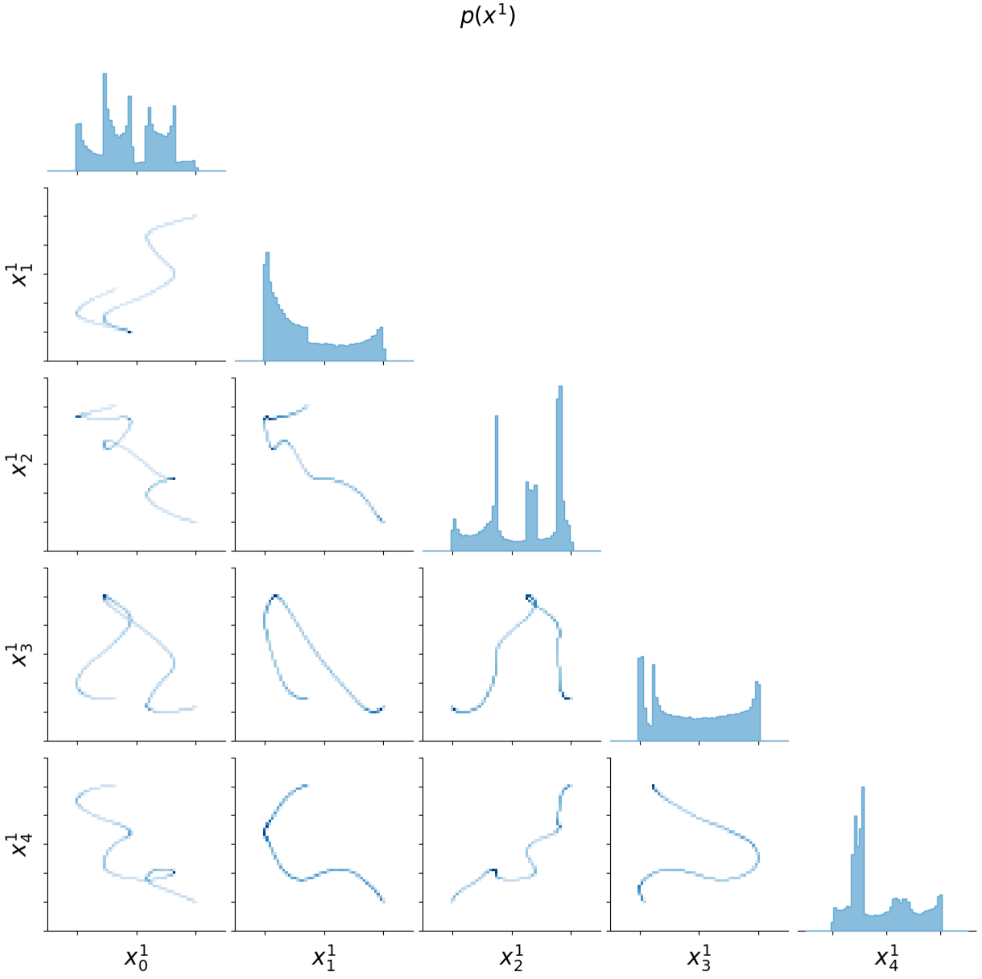
- Sources are two distinct 1D manifolds embedded in a 5D space
- Observations are 2D random mixture of the manifolds
- Want to disentangle both manifolds

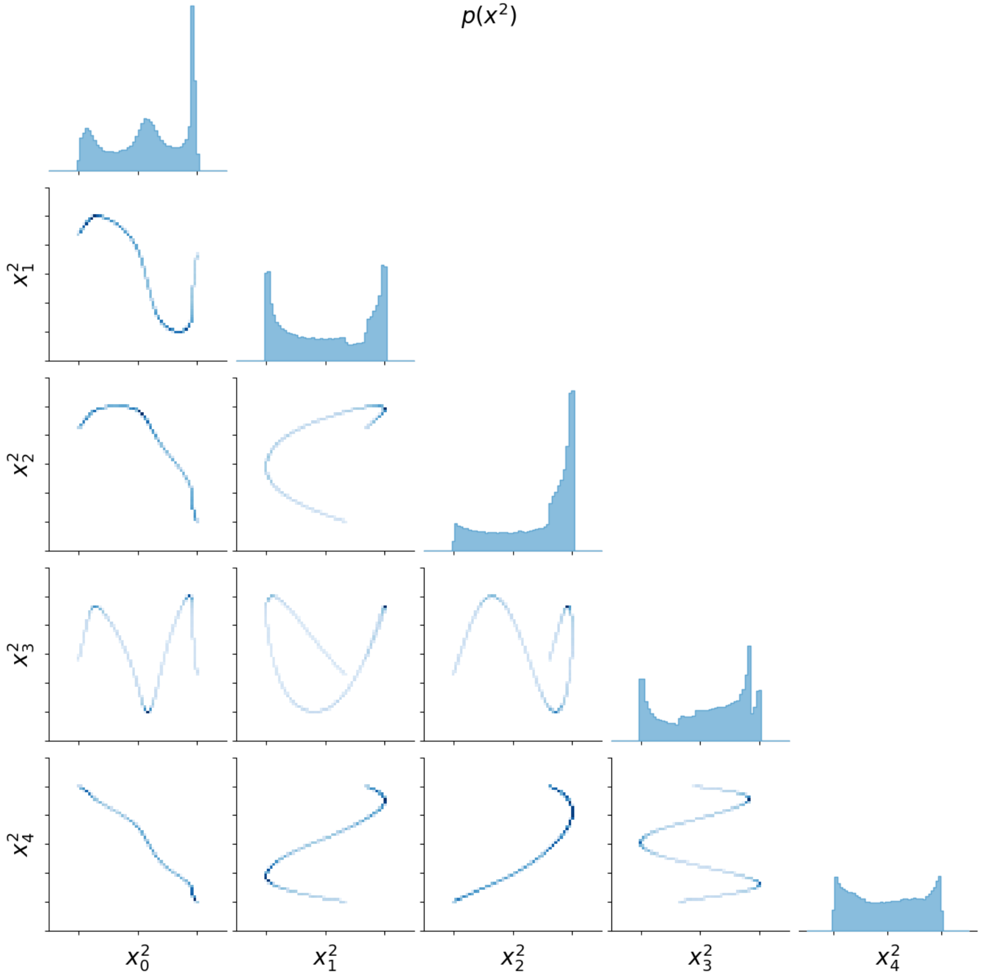
$$\mathbf{y}^{\alpha}_{i_\alpha} = \mathbf{A}_{i_\alpha} \left(\sum_{\beta=1}^{N_s} c^{\alpha\beta} \mathbf{x}^{\beta}_{i_\alpha} \right) + \eta^{\alpha}_{i_\alpha}$$

$$c^{\alpha\beta} = \begin{cases} 1 & \text{if } \alpha = \beta \\ f_{\text{mix}} & \text{if } \alpha \neq \beta \end{cases}$$

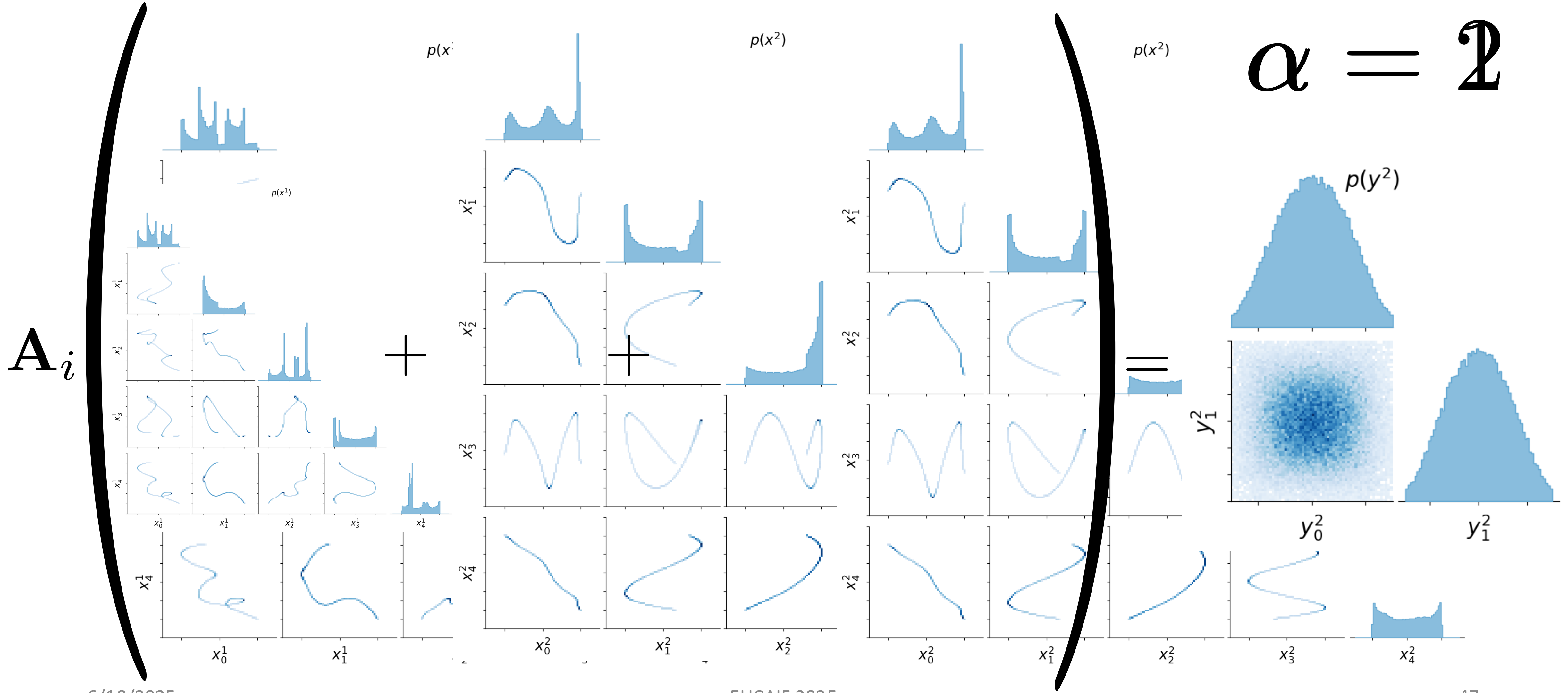
$$N_{\text{view}} \geq N_s$$

1D Manifolds: Mixed





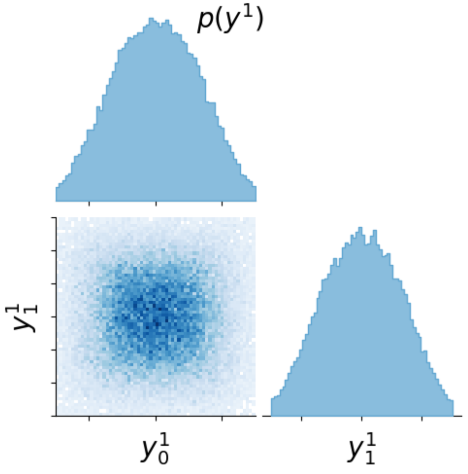
1D Manifolds: Mixed



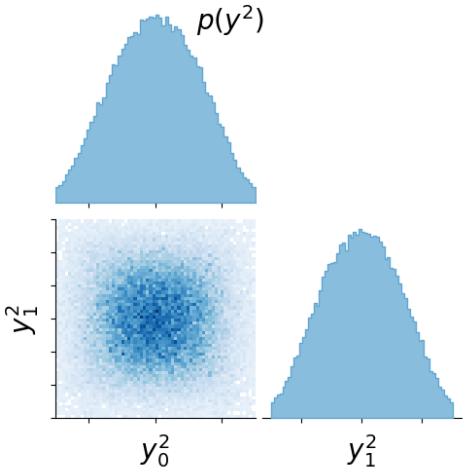
Wagner-Carena et al., in prep.

1D Manifolds: Mixed

- We have **samples from the two views** and their corresponding **mixing matrices**

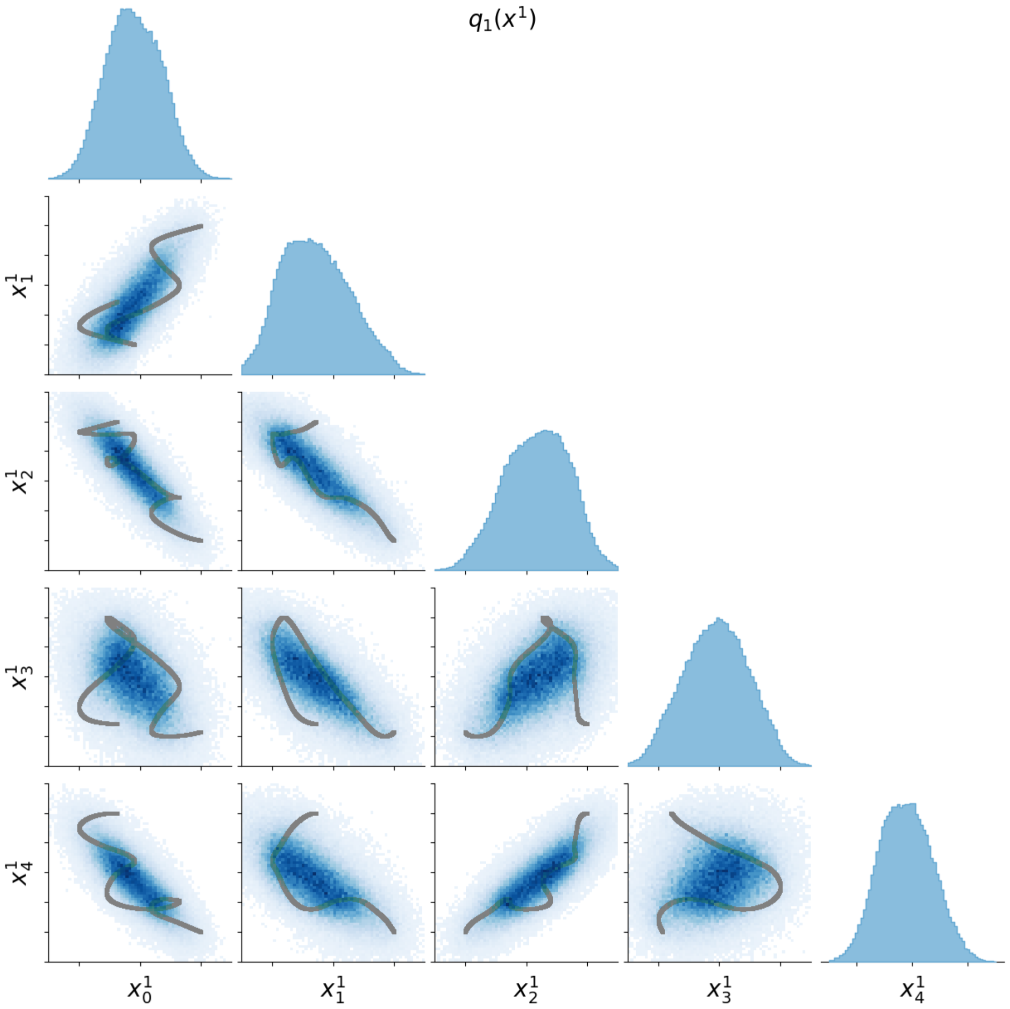


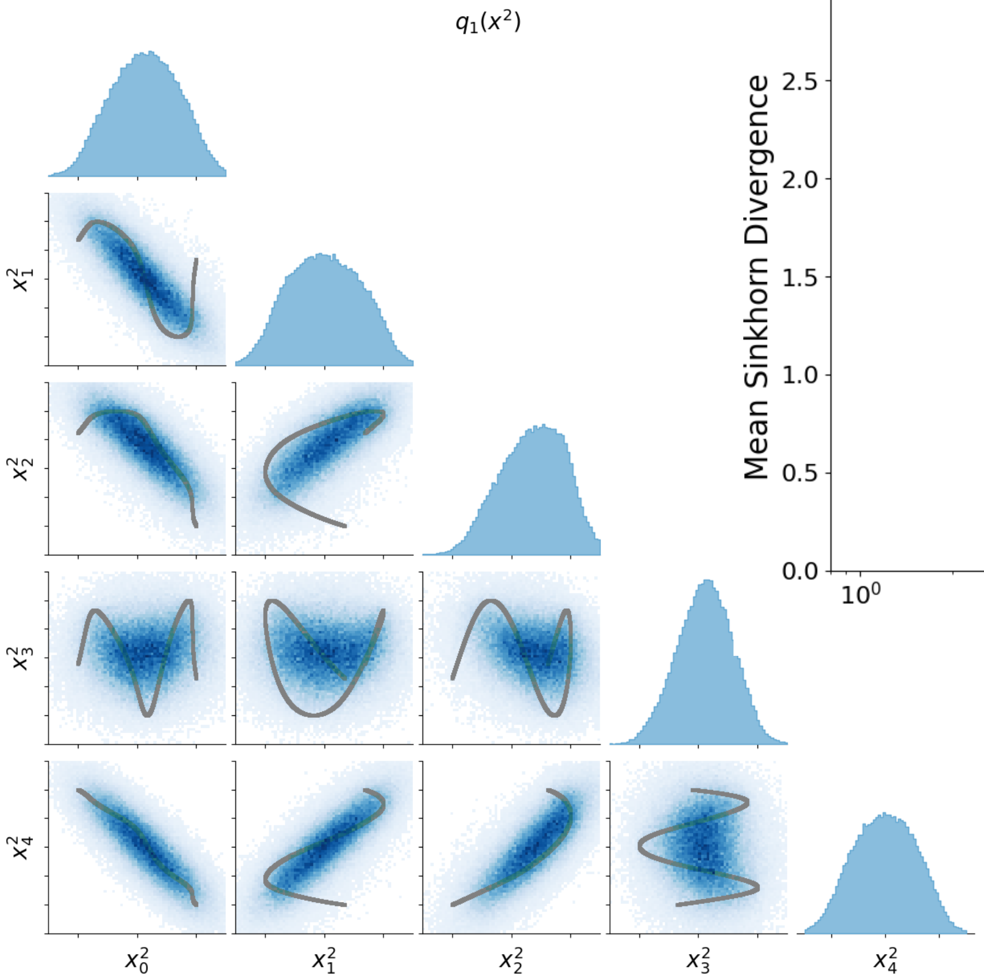
$$+\mathbf{A}_i$$

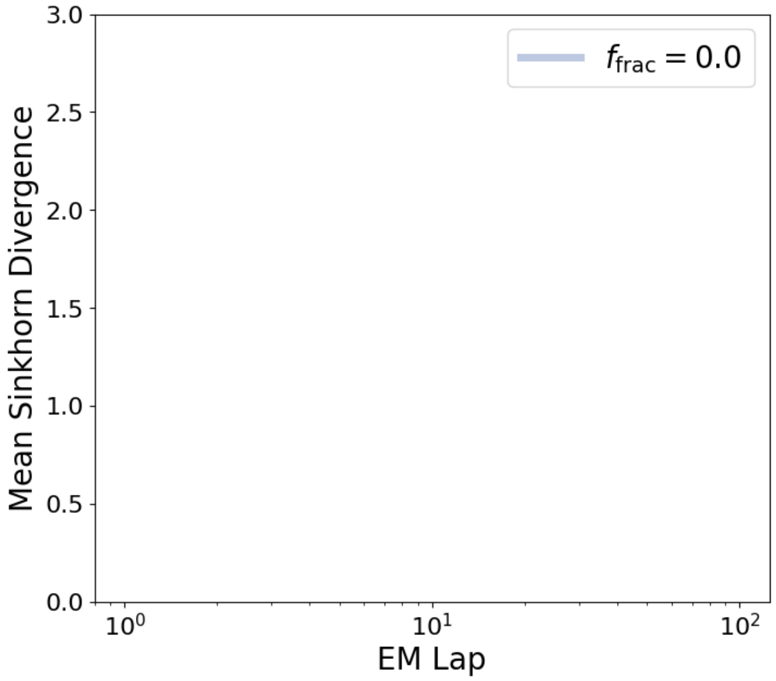


$$+\mathbf{A}_i$$

1D Manifolds: Mixed

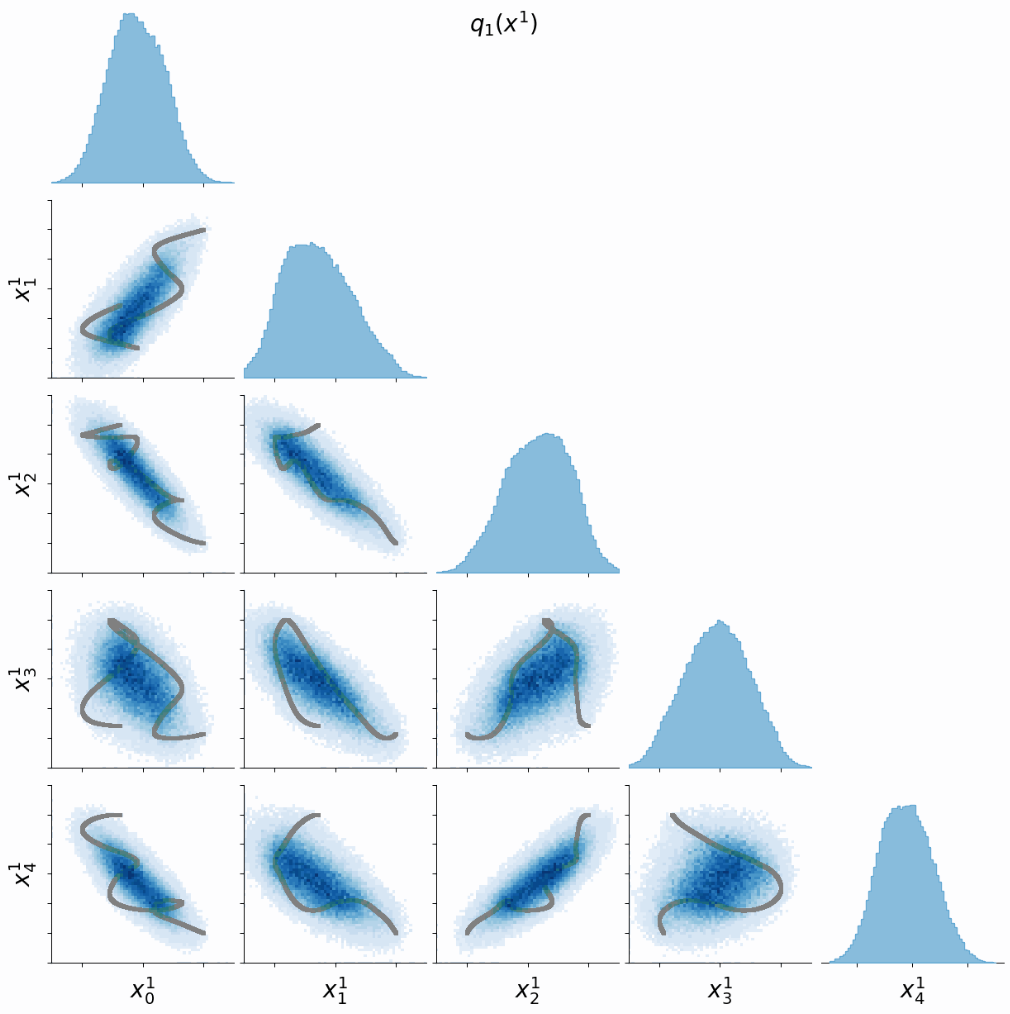


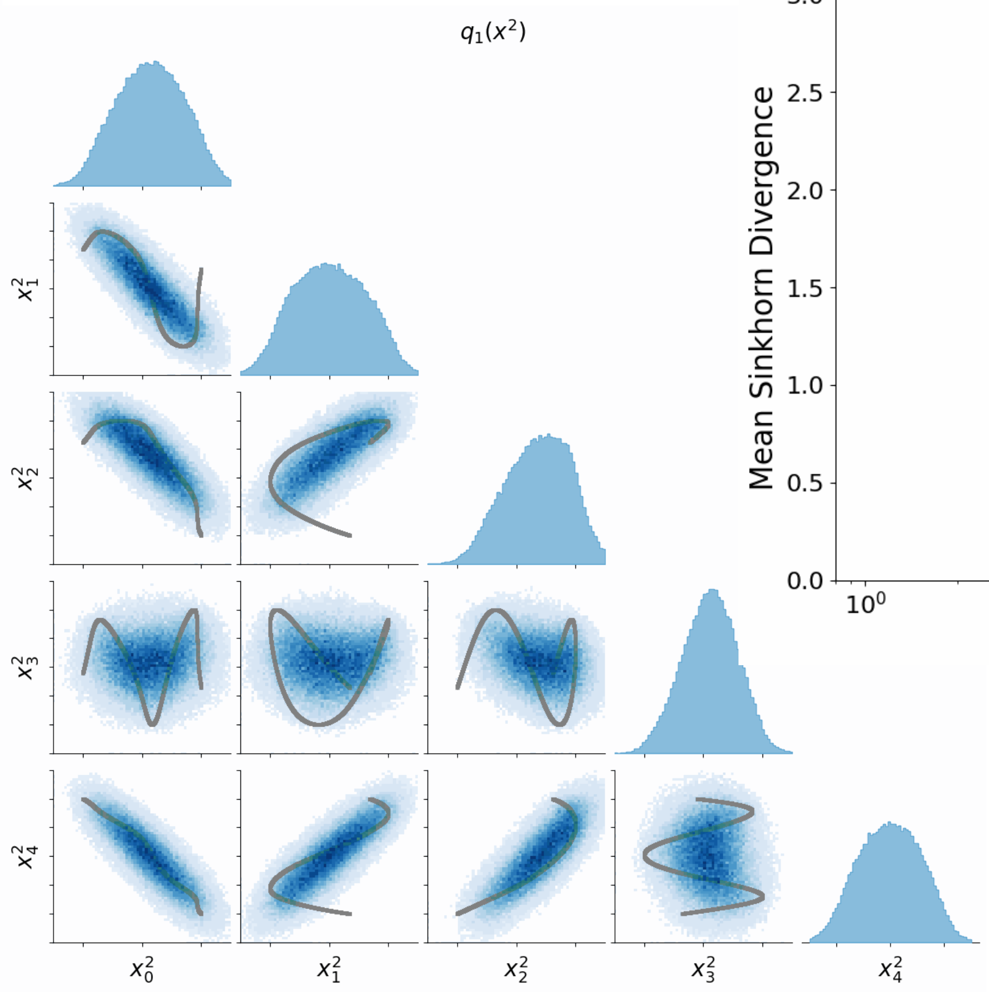


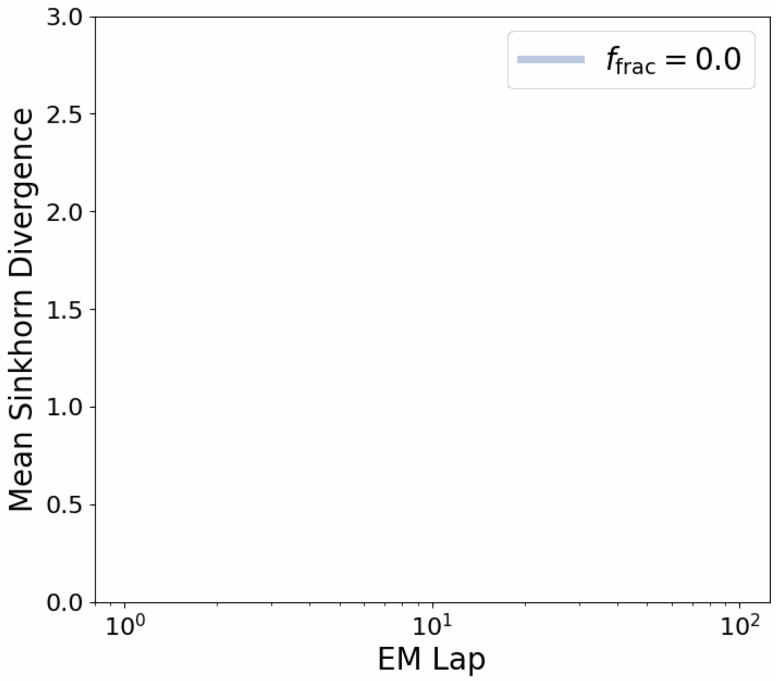


Wagner-Carena et al., in prep.

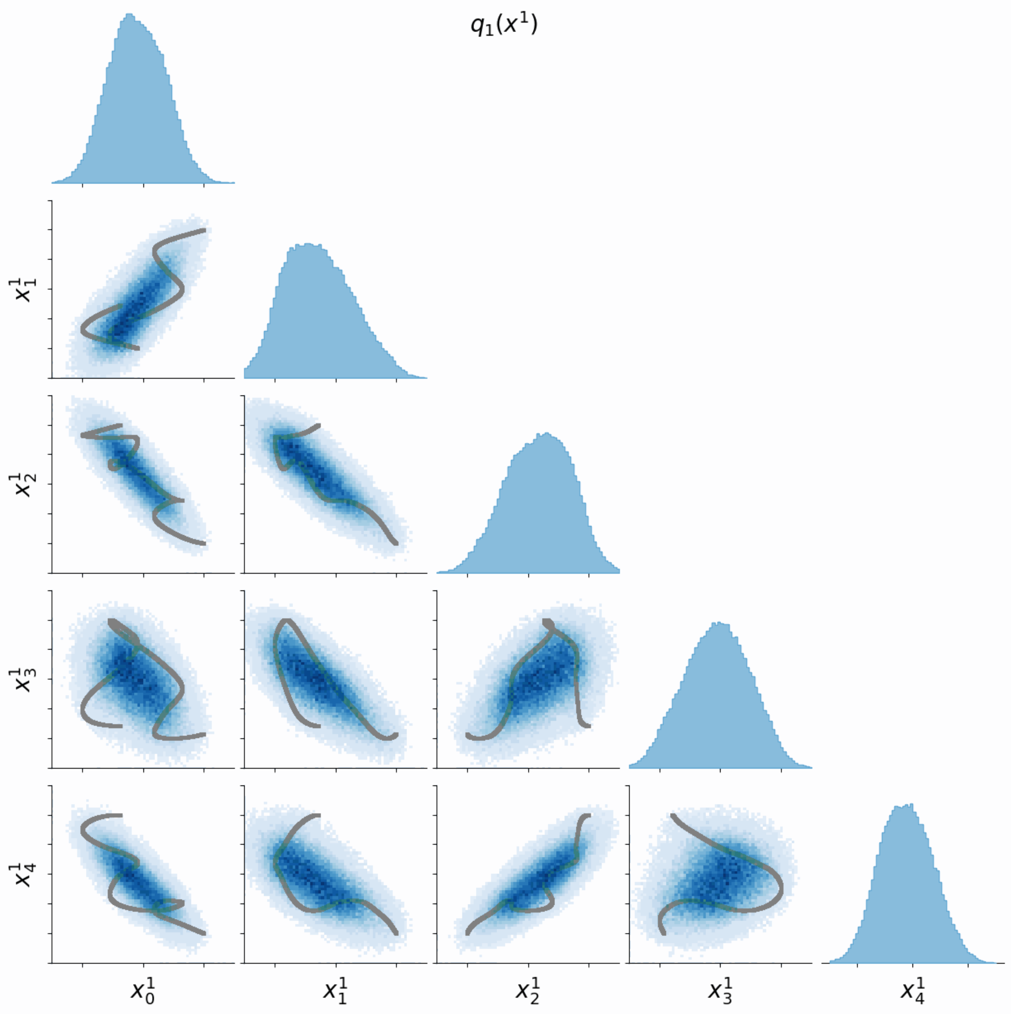
1D Manifolds: Mixed

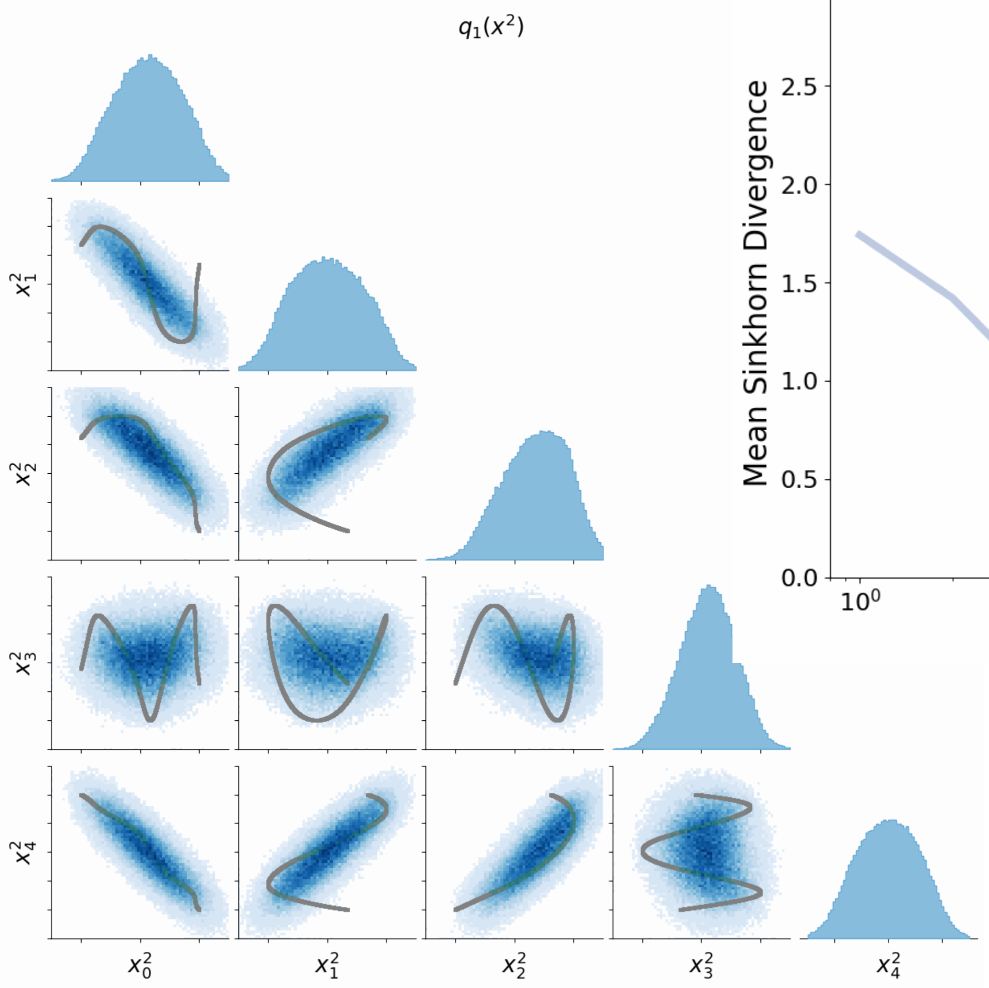


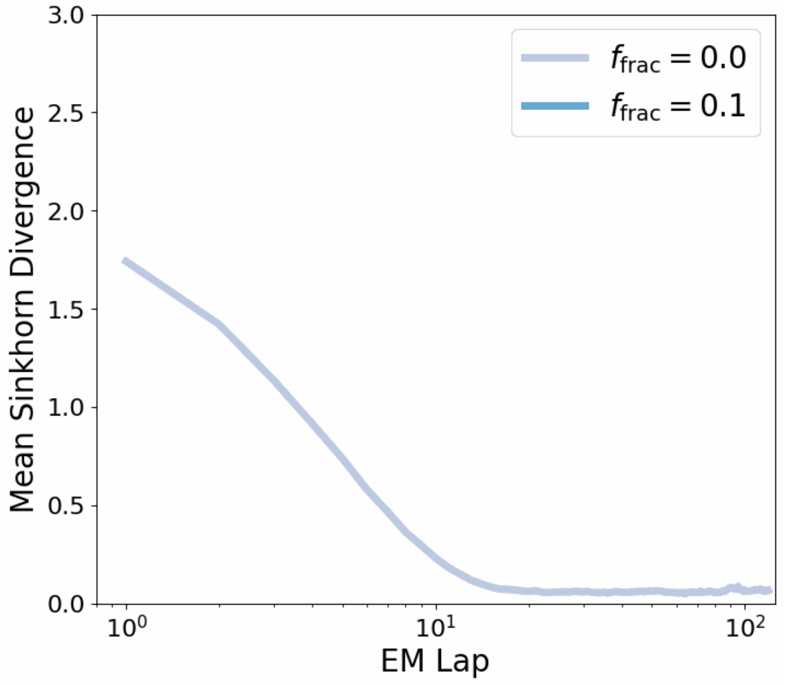




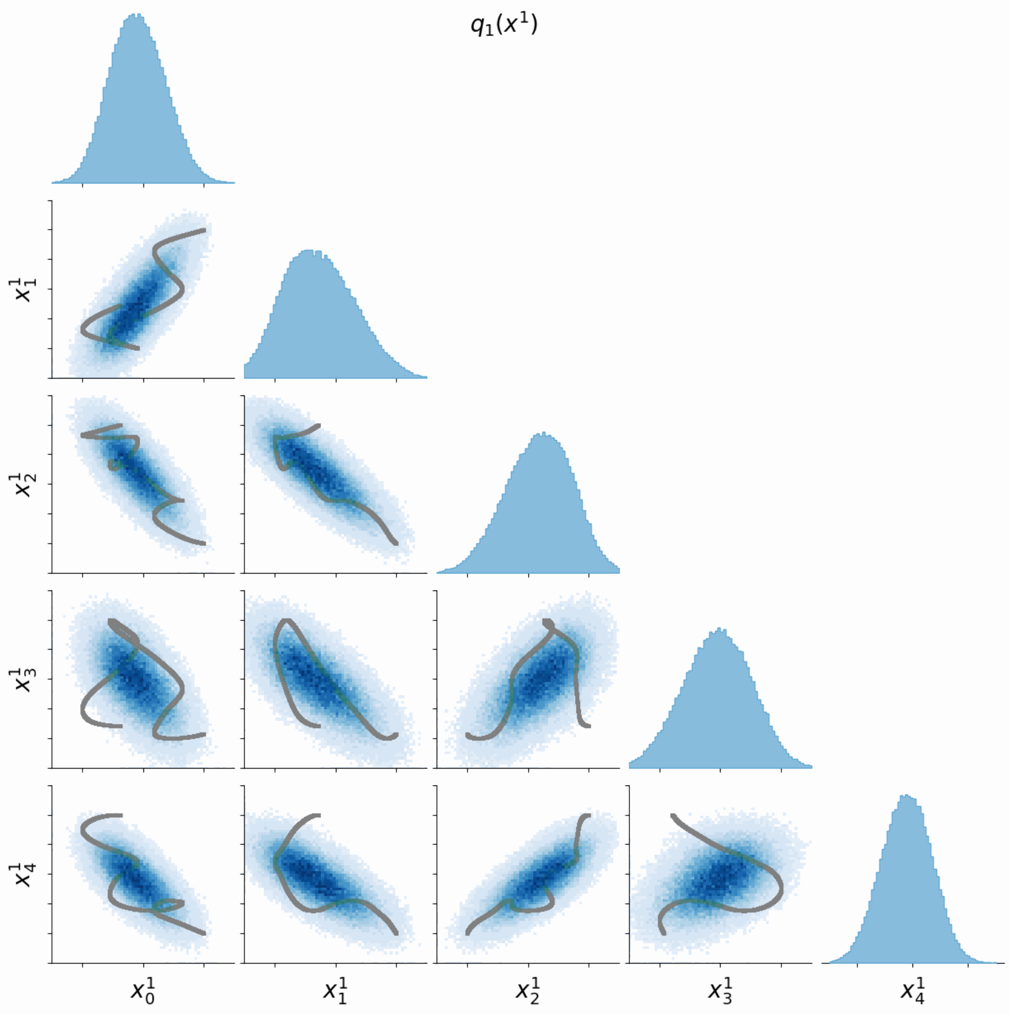
1D Manifolds: Mixed

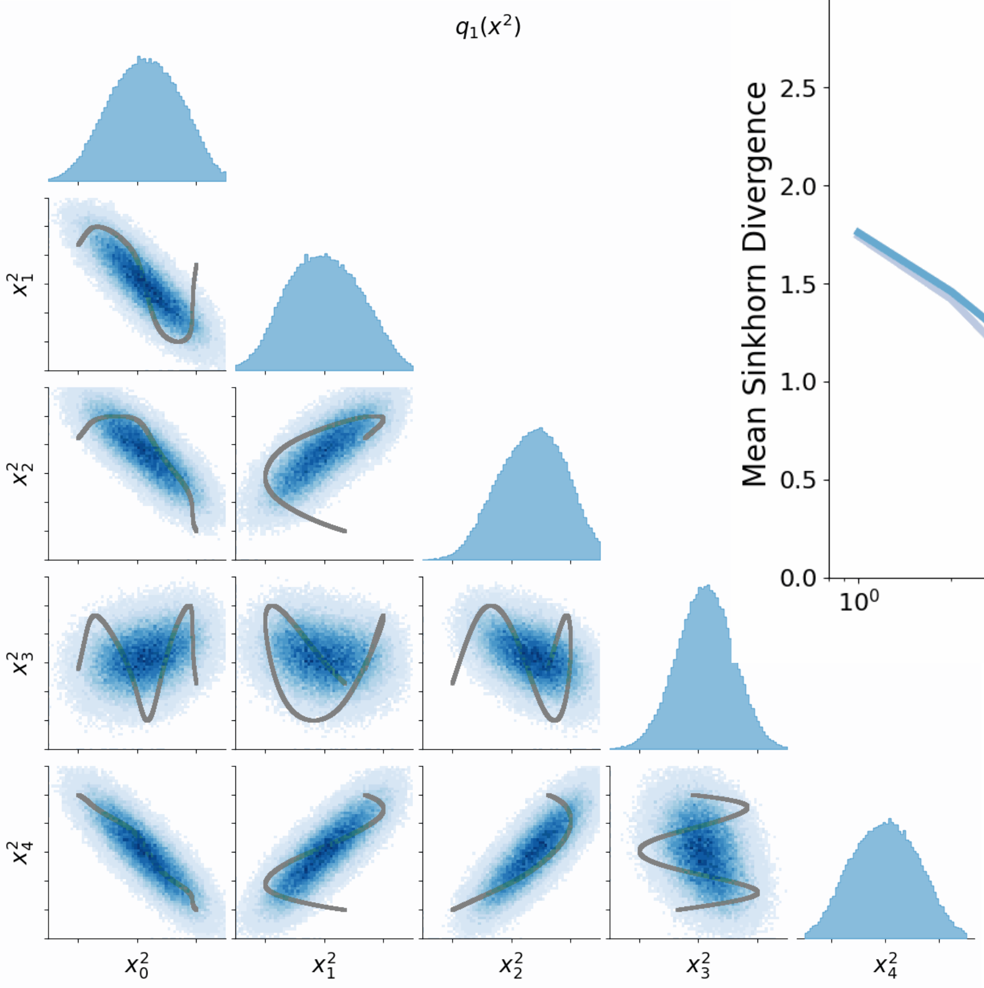


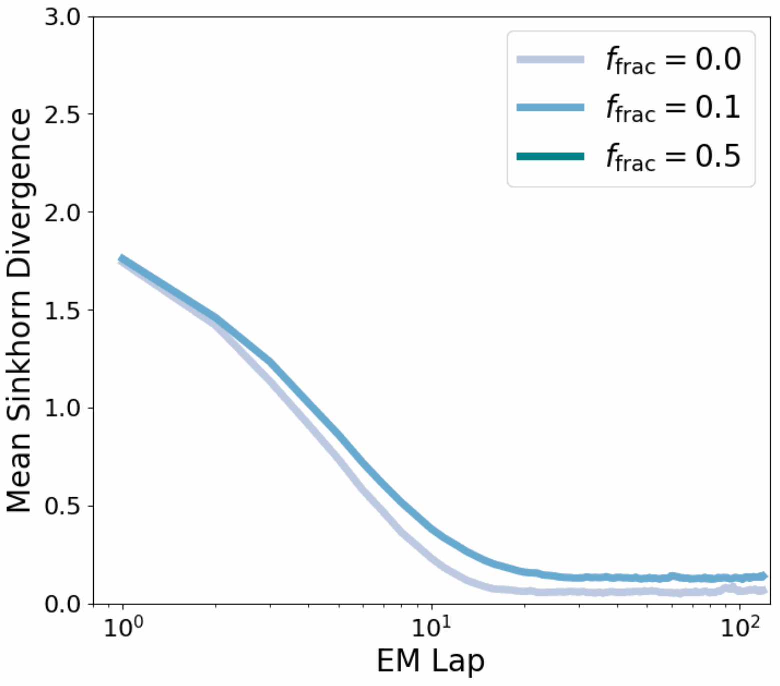




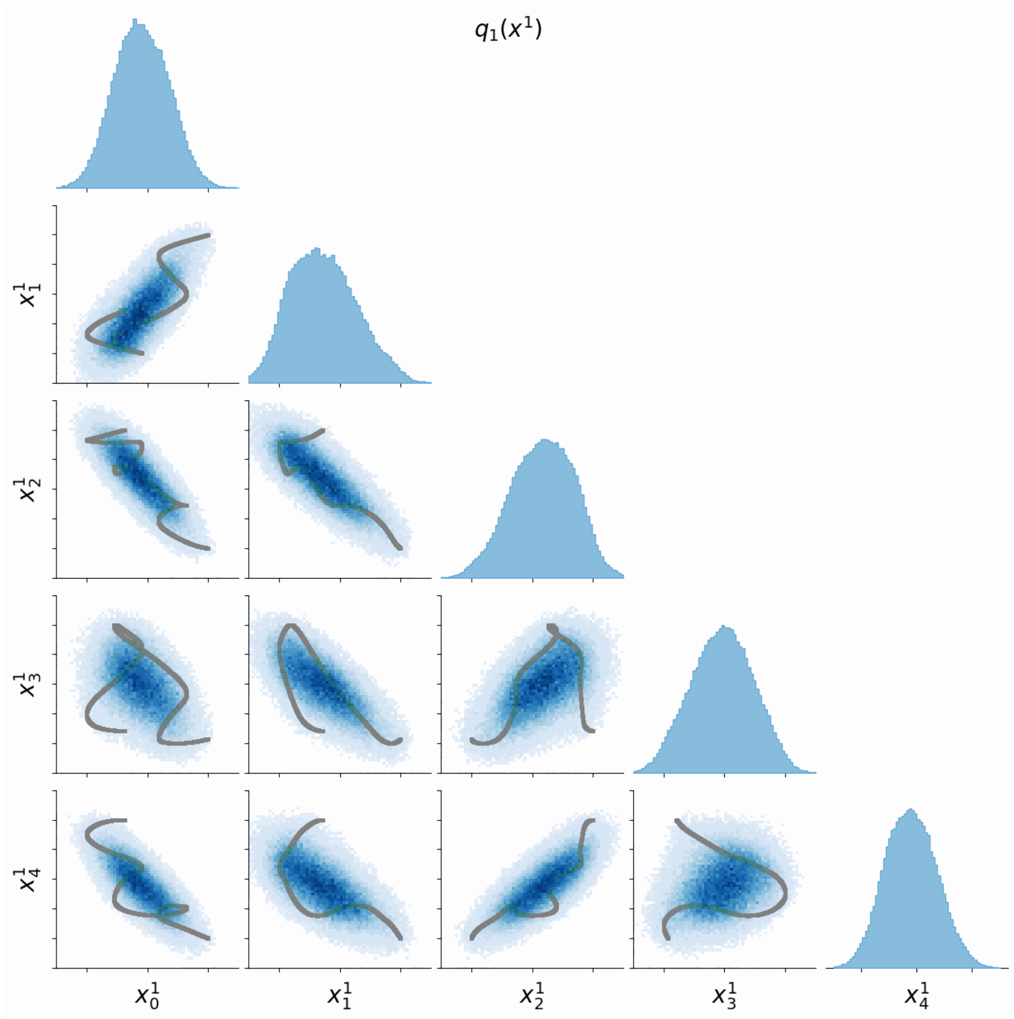
1D Manifolds: Mixed

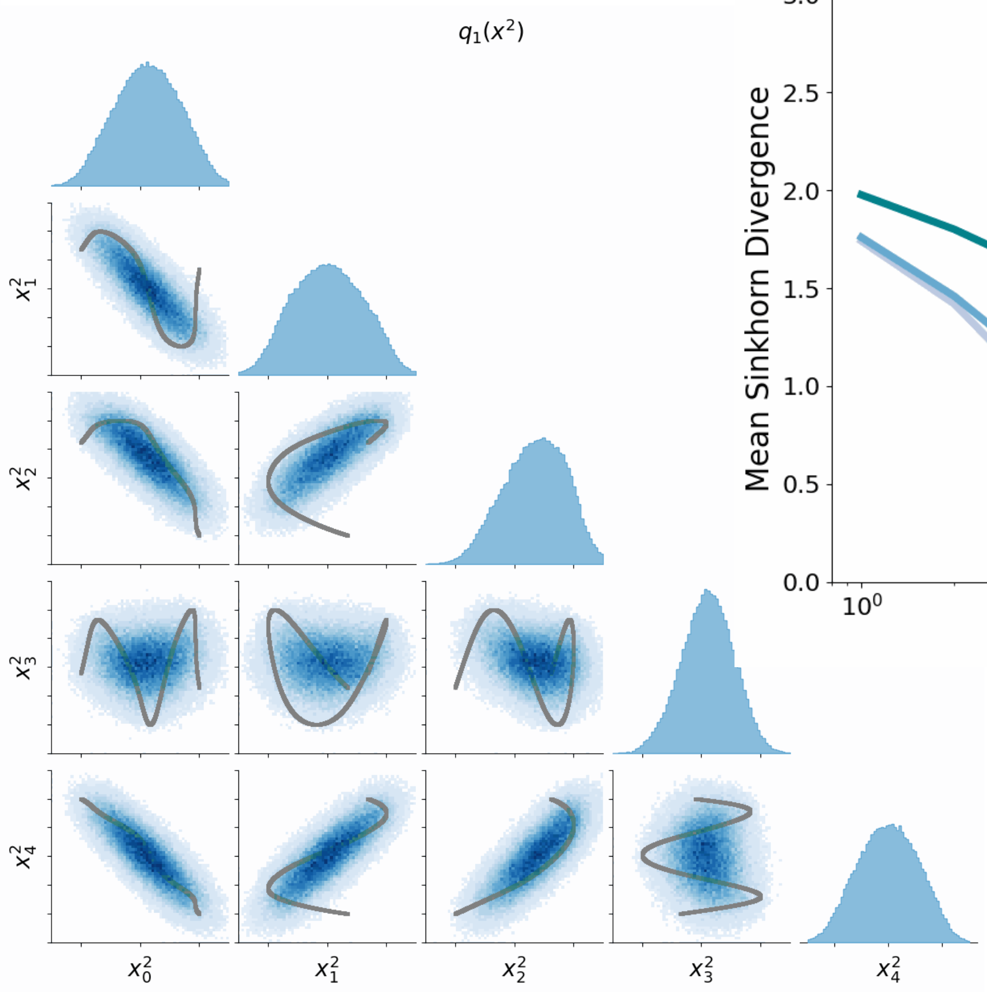


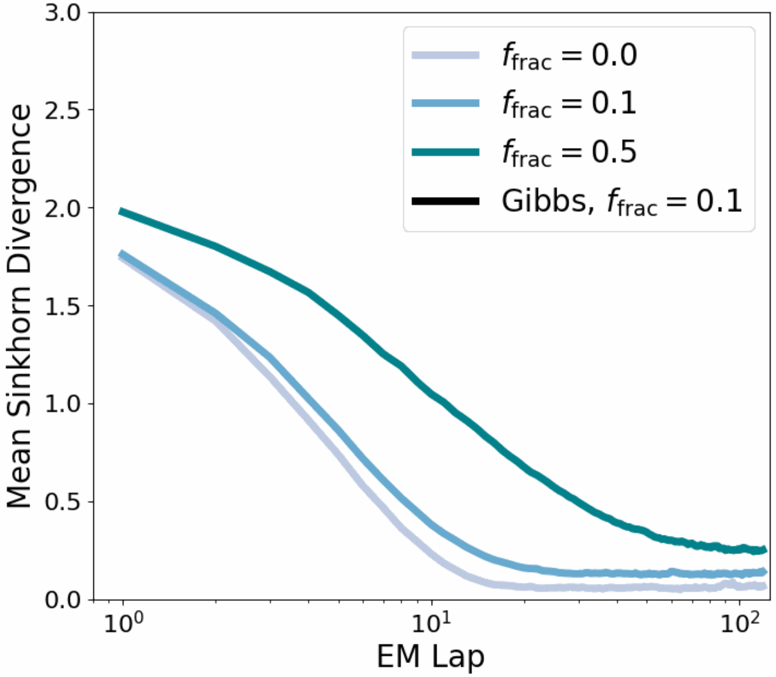




1D Manifolds: Mixed







1D Manifolds: Mixed

- Can disentangle the sources, even when **no source is seen on its own**.
- **Outperforms** baseline Gibbs sampling approach, even for much larger mixing fractions
- Works with **significant information loss** on individual observations.