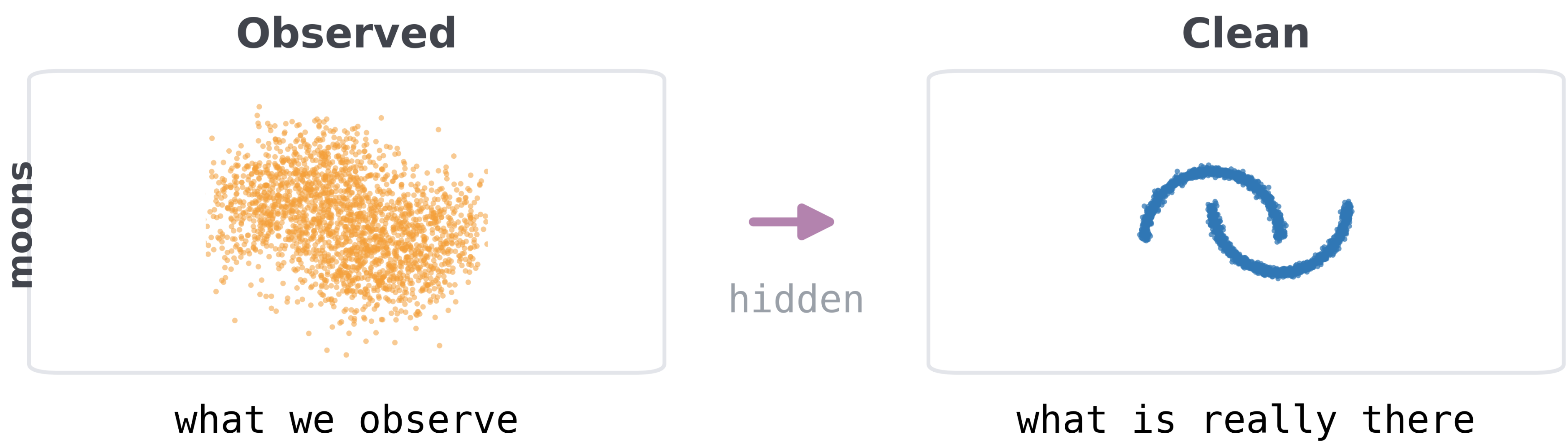
Convolutional Maximum Mean Discrepancy for Inference in Noisy Data

Ritwik Vashistha (https://ritwikvashistha.github.io/)

Joint work with Jeff Phillips, Abhra Sarkar and Arya Farahi

The Problem

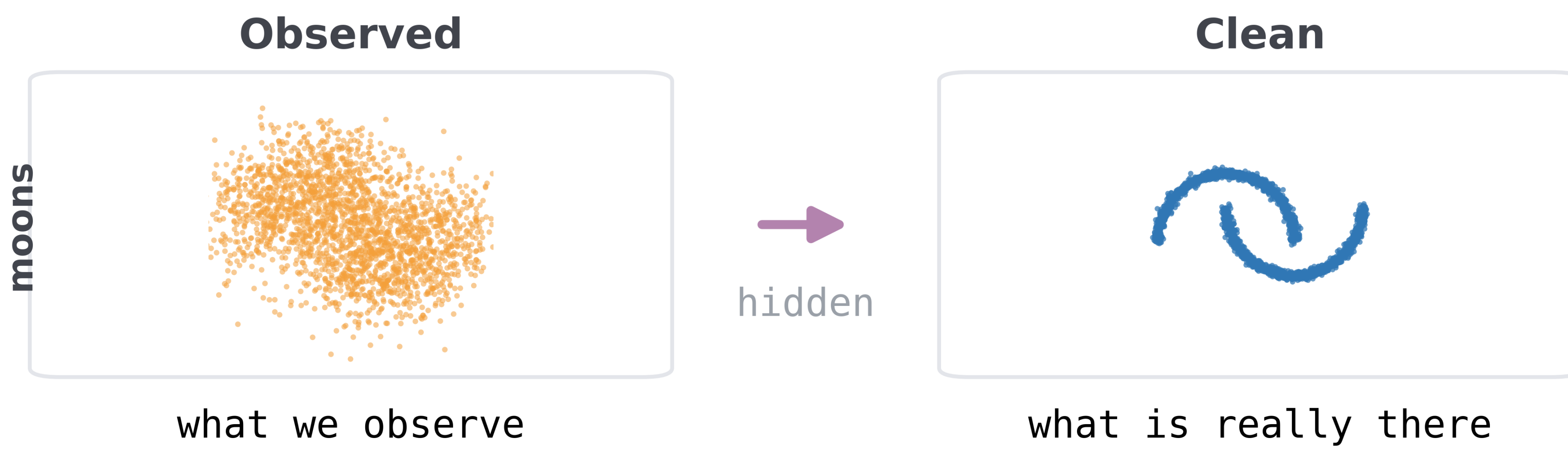
Noise hides the signal science cares about.



We observe $\tilde{W} = W + U$ but we wish to recover the latent W.

Likelihood based methods are optimal but make strong assumptions: known likelihood, correct noise, no outliers etc.

The Problem



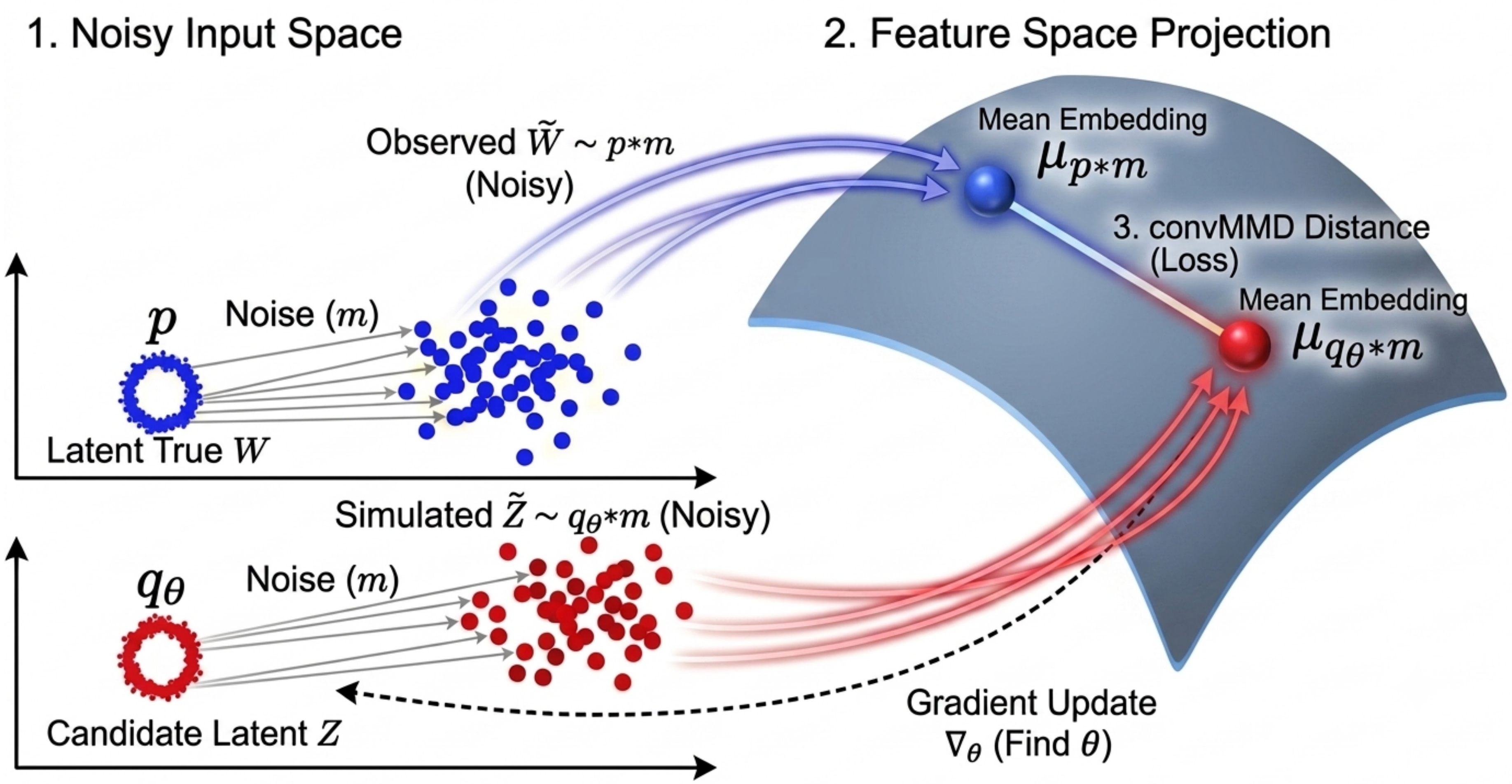
We are interested in two problems:

Deconvolution	Denoising
Recover distribution of W	Recover each W_i from $\tilde{W}_i$

No parametric form is assumed for distribution of W.

Idea

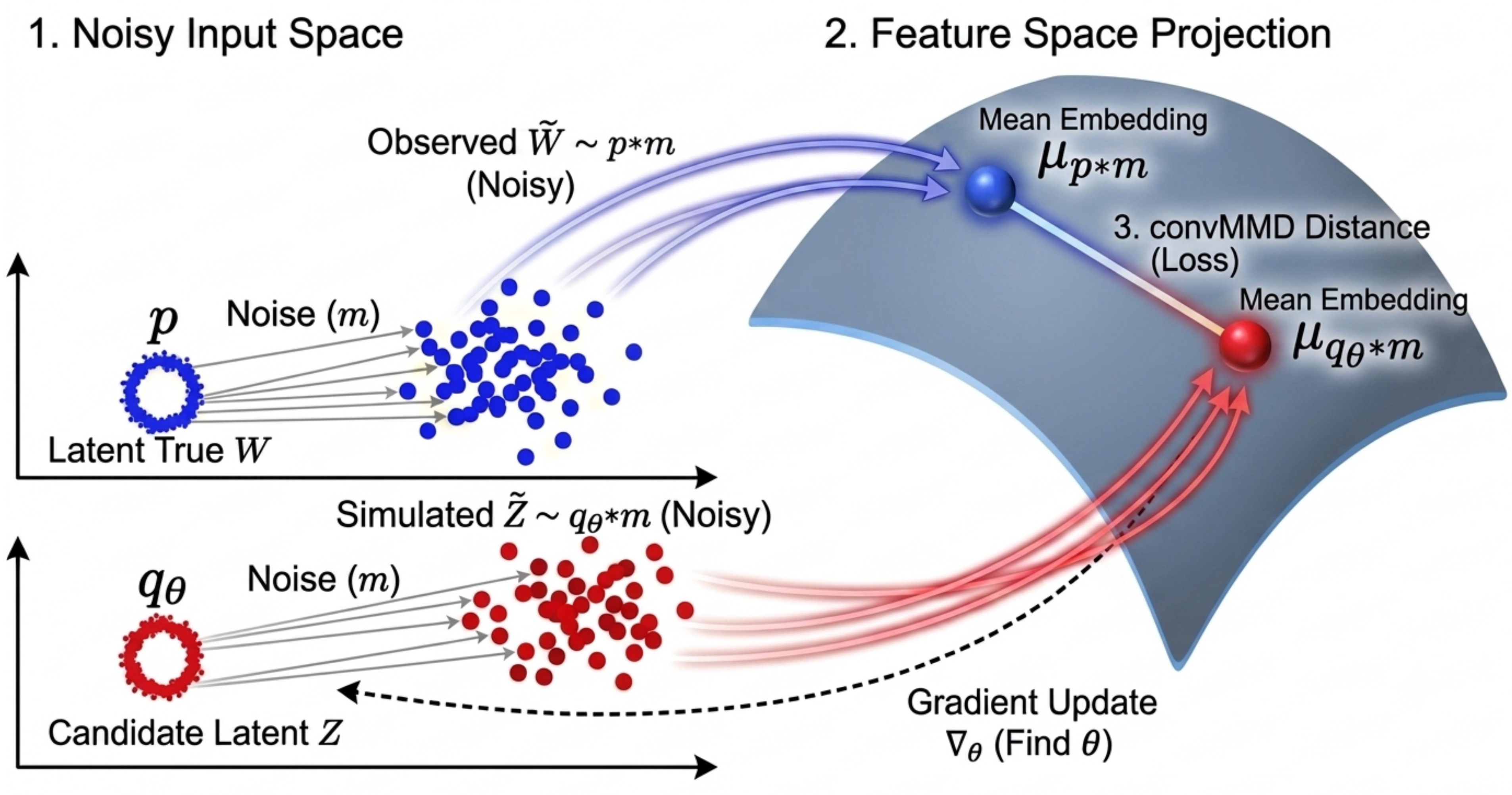
Stage 1: Deconvolution



Compare convoluted (conv) noisy samples using Maximum Mean Discrepancy (MMD) distance.

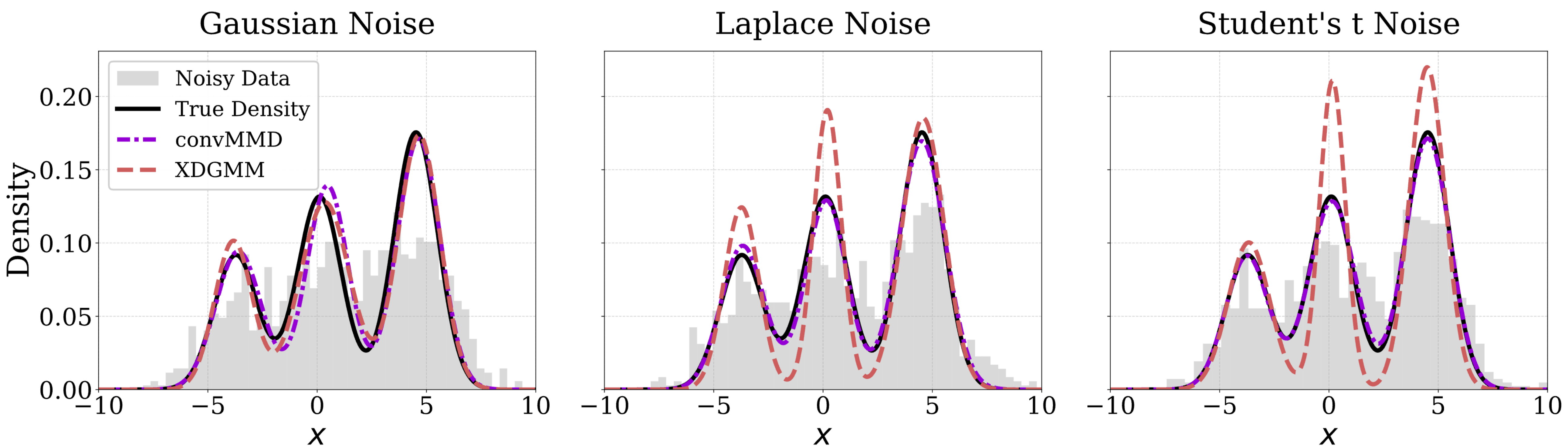
Idea

Stage 1: Deconvolution



We learn $q_{\hat{\theta}}$ by minimizing convMMD(p, q_θ, m) where, convMMD (p, q_θ, m)= $\|\mu_{p*m} - \mu_{q_\theta*m}\|_{\mathscr{H}}$

Deconvolution Illustration



Estimated density using different methods across three different noise distributions: Gaussian (left), Laplace (middle), Student's t (right).

Idea

Stage 2: Denoising

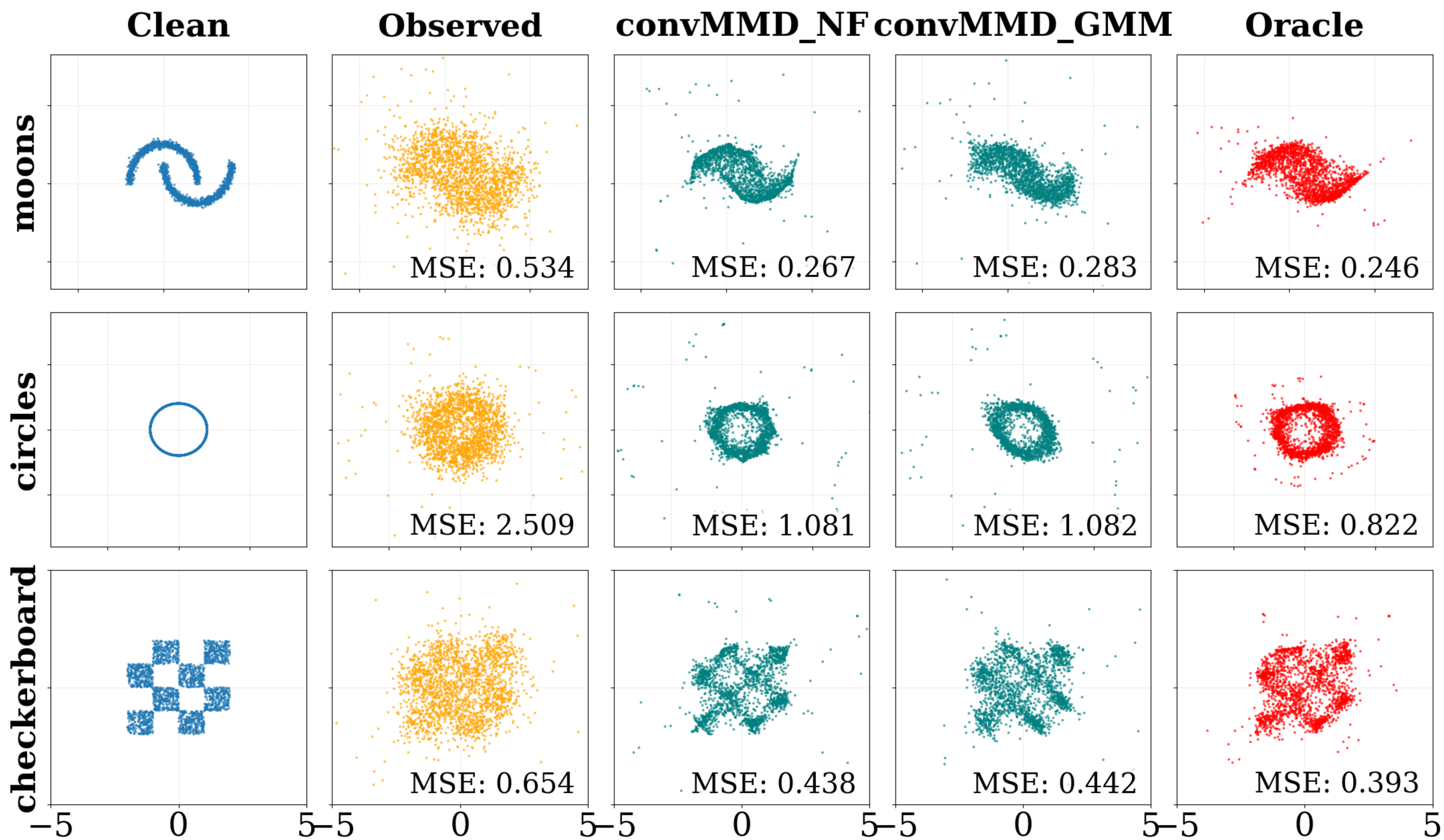
We now use $q_{\hat{\theta}}$ learnt in Stage 1 to perform denoising using Empirical Bayes. We want $p(\mathbf{W} \mid \tilde{\mathbf{W}})$, which can be obtained as follows:

$$\boxed{p(\mathbf{W} \mid \tilde{\mathbf{W}}) \propto \mathbf{p}(\tilde{\mathbf{W}} \mid \mathbf{W}) \cdot \mathbf{q}_{\hat{\theta}}(\mathbf{W})}$$

Per Point Denoised Observation

$$\boxed{\hat{\mathbf{W}} = \mathbb{E}(\mathbf{W} \mid \hat{\mathbf{W}})}$$

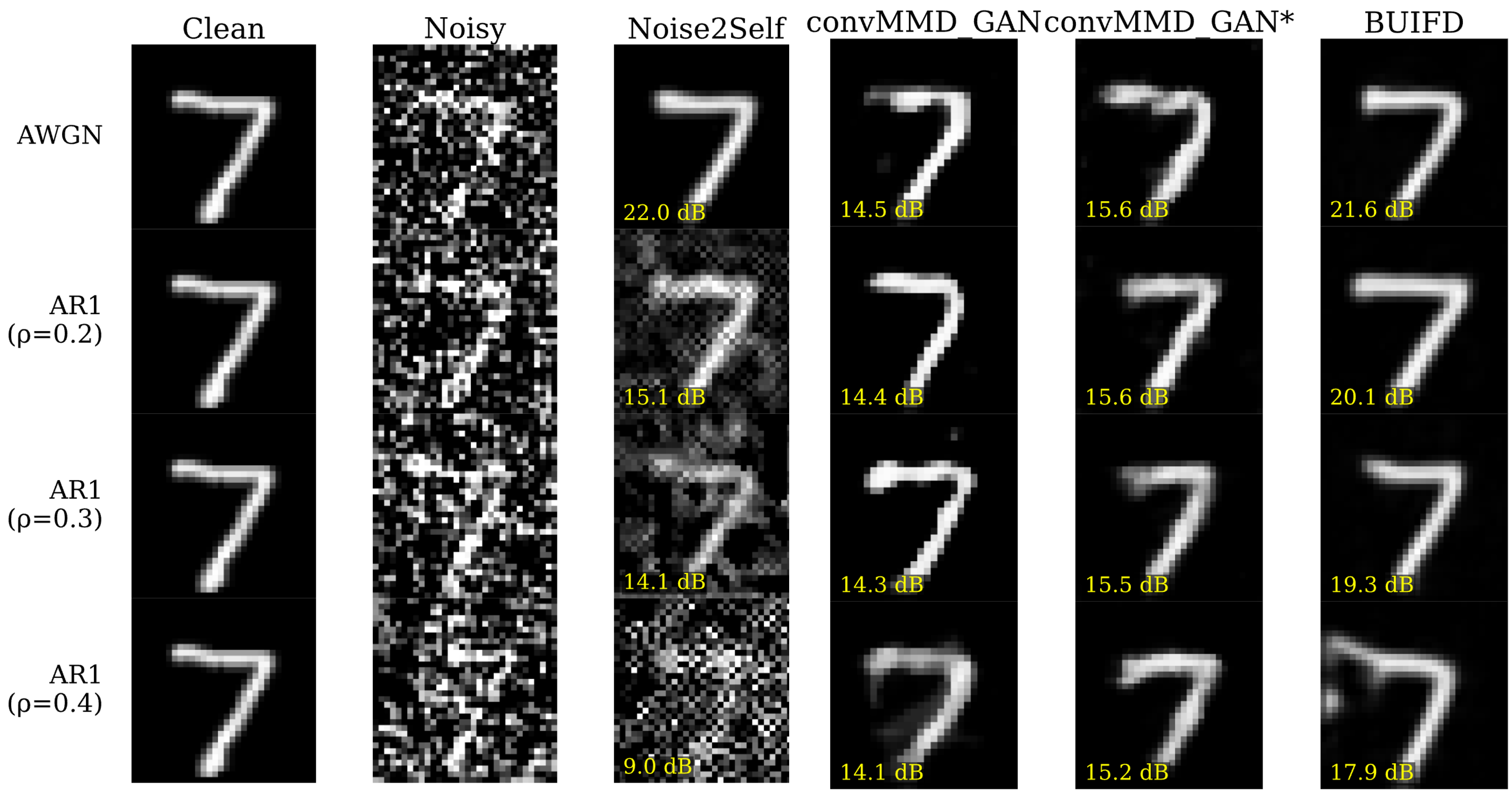
Denoising Illustration



Two Moons, Circles and Checkerboard data corrupted by heteroscedastic noise and outliers.

Vashistha, Sarkar & Farahi (arXiv: 2606.21907)

Scalability to High Dimensions



Denoising under pixel-independent Gaussian Noise (AWGN) and spatially-correlated AR(1) noise.

Vashistha, Sarkar & Farahi (arXiv: 2606.21907)

Can we also solve regression problems?

Consider the following data generating process:

$$Y = \alpha + \beta \cdot X + \epsilon;$$

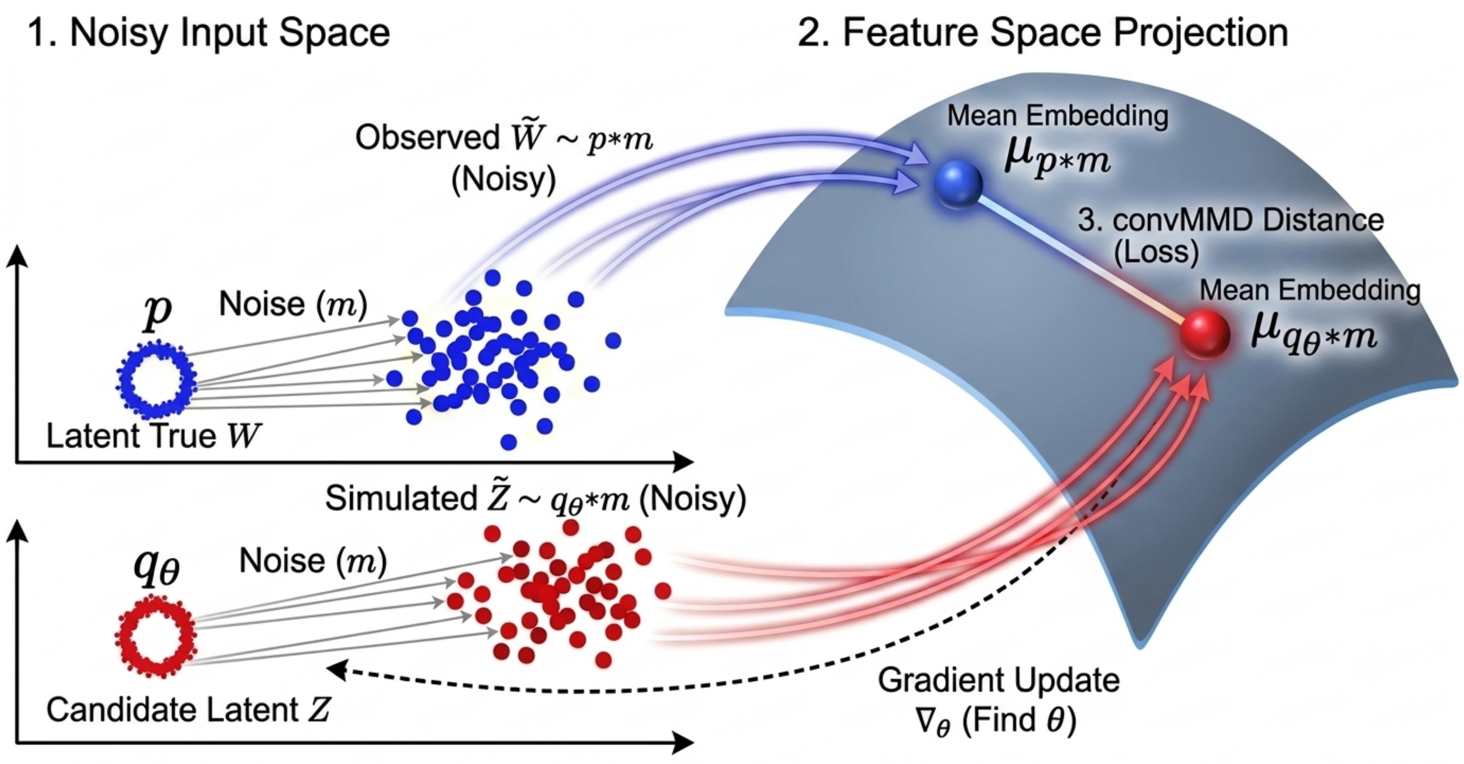
$$\tilde{X} = X + U_X;$$

$$\tilde{Y} = Y + U_Y.$$

We only observe $(\tilde{X}, \tilde{Y})$, but we wish to estimate α, β.

Can we also solve regression problems?

Lets go back to the idea for deconvolution.



Here, $q_{\boldsymbol{\theta}}(X, Y) = \underbrace{q\left(Y \mid X, \boldsymbol{\theta}_{\text{reg}}\right)}_{\text{Regression Model}} \cdot \underbrace{q\left(X \mid \boldsymbol{\theta}_X\right)}_{\text{Latent Covariate Model}}$

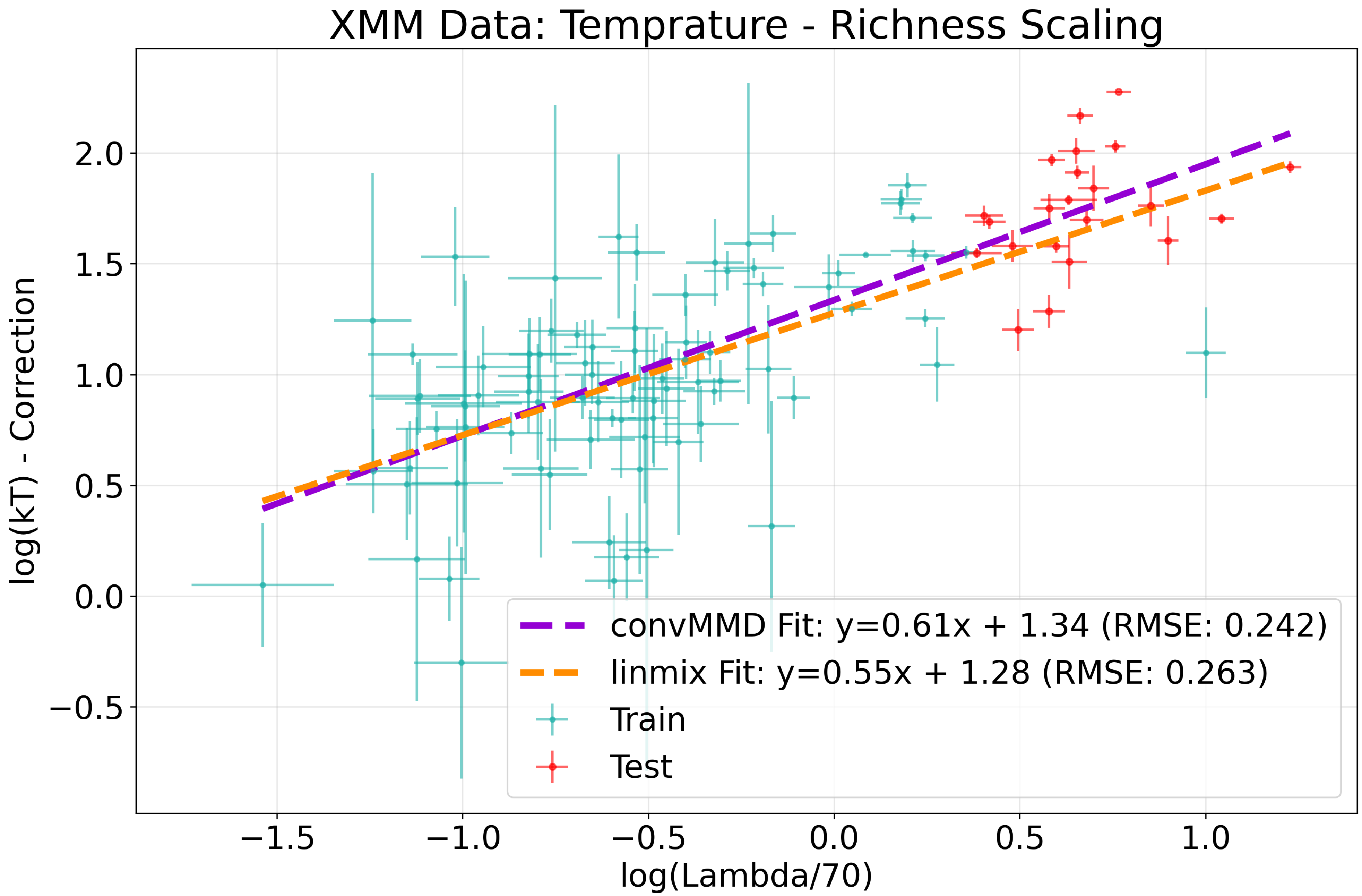
Can we also solve regression problems?

Here, $q_{\boldsymbol{\theta}}(X, Y) = \underbrace{q\left(Y \mid X, \boldsymbol{\theta}_{\text{reg}}\right)}_{\text{Regression Model}} \cdot \underbrace{q\left(X \mid \boldsymbol{\theta}_X\right)}_{\text{Latent Covariate Model}}$

Let $\tilde{W} = \{(\tilde{X}, \tilde{Y})^{\mathrm{T}}\}$ and $\tilde{Z} = \{(\tilde{X}_{\text{sim}}, \tilde{Y}_{\text{sim}})^{\mathrm{T}}\}$

We learn $q_{\hat{\theta}}$ by minimizing convMMD$(\boldsymbol{p}, \boldsymbol{q}_{\theta}, \boldsymbol{m})$ where,
convMMD $(\boldsymbol{p}, \boldsymbol{q}_{\theta}, \boldsymbol{m})$= $\|\mu_{p*m} - \mu_{q_\theta * m}\|_{\mathscr{H}}$

Regression Illustration



See, Vashistha et al. (arXiv:2604.12022)

Final Thoughts

- A single framework for solving multiple inverse problems: deconvolution, denoising, and error in variables regression.
- Scalable to high dimensions and robust to presence of outliers.
- Ongoing work: noise misspecification, improving optimization
- Looking for new applications!

Paper 1 (arXiv:2604.12022)

Paper 2 (arXiv: 2606.21907)