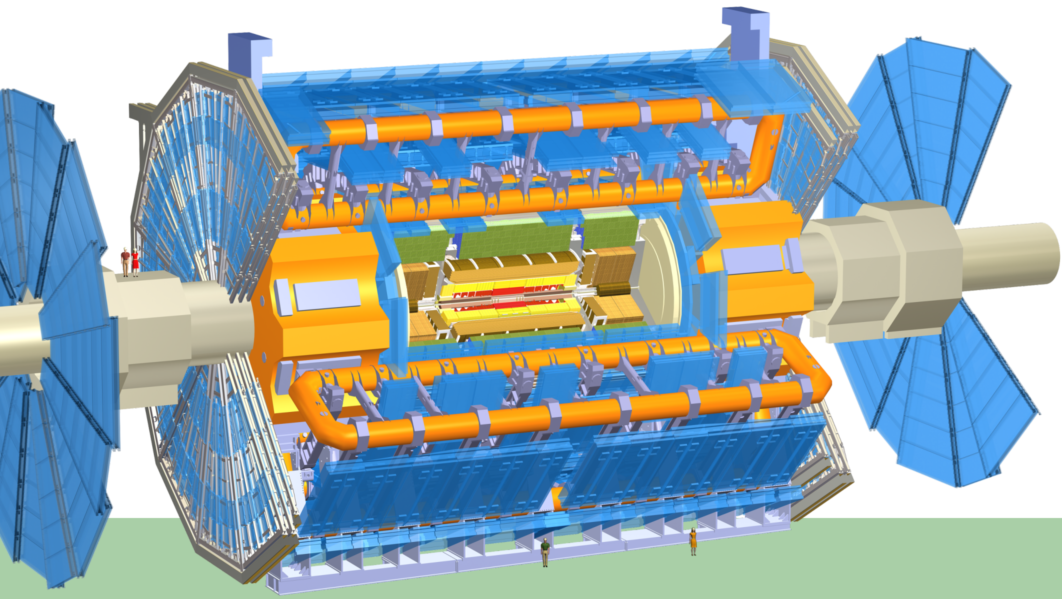


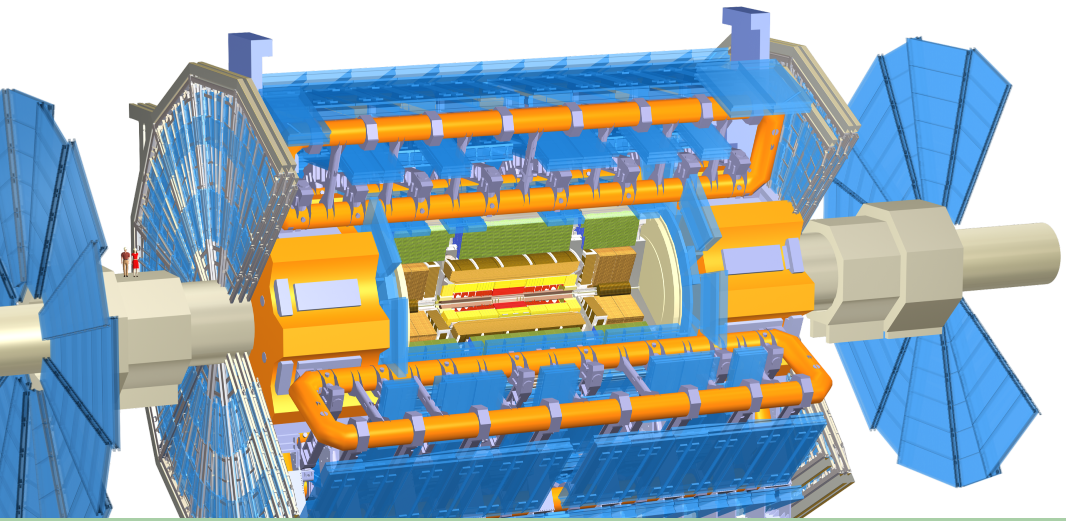
Denoising with a Wasserstein Loss

arXiv:2603.20903

Katy Craig
University of California, Santa Barbara

Joint with Benjamin Faktor (UCLA Math) and Benjamin Nachman (Stanford Physics)

Optimal Transport Reading Group
Carnegie Mellon University
May 18th, 2026



Unfolding with a Wasserstein Loss

arXiv:2603.20903

Katy Craig
University of California, Santa Barbara

Joint with Benjamin Faktor (UCLA Math) and Benjamin Nachman (Stanford Physics)

Optimal Transport Reading Group
Carnegie Mellon University
May 18th, 2026

Plan

Plan

1. Motivation

Plan

1. Motivation

2. Previous Work

Plan

1. Motivation

2. Previous Work

3. Well-posedness vs Ill-posedness

Plan

1. Motivation

2. Previous Work

3. Well-posedness vs Ill-posedness

4. Numerical Method

Plan

1. Motivation
2. Previous Work
3. Well-posedness vs Ill-posedness
4. Numerical Method
5. Numerical Results

Classical Unfolding

Classical Unfolding

Setup:

Classical Unfolding

Setup:

Metric space of parameters (Ω, d_Ω)

Classical Unfolding

Setup:

Metric space of parameters (Ω, d_Ω)

Parametric statistical/noise model: $\{\nu_\sigma : \sigma \in \Omega\} \subseteq \mathscr{P}(\mathscr{Y})$

Classical Unfolding

Setup:

Metric space of parameters (Ω, d_Ω)

Parametric statistical/noise model: $\{\nu_\sigma : \sigma \in \Omega\} \subseteq \mathscr{P}(\mathscr{Y})$

Measured data: $\nu \in \mathscr{P}(\mathscr{Y})$

Classical Unfolding

Setup:

Metric space of parameters (Ω, d_Ω)

Parametric statistical/noise model: $\{\nu_\sigma : \sigma \in \Omega\} \subseteq \mathscr{P}(\mathscr{Y})$

Measured data: $\nu \in \mathscr{P}(\mathscr{Y})$

Classical Unfolding

Setup:

Metric space of parameters (Ω, d_Ω)

Parametric statistical/noise model: $\{\nu_\sigma : \sigma \in \Omega\} \subseteq \mathscr{P}(\mathscr{Y})$

Measured data: $\nu \in \mathscr{P}(\mathscr{Y})$

Goal: Find data $\sigma \in \Omega$ so that the noisy data $\nu_\sigma \in \mathscr{P}(\mathscr{Y})$ satisfies

$$\nu_\sigma \approx \nu.$$

Classical Unfolding

Setup:

Metric space of parameters (Ω, d_Ω)

Parametric statistical/noise model: $\{\nu_\sigma : \sigma \in \Omega\} \subseteq \mathscr{P}(\mathscr{Y})$

Measured data: $\nu \in \mathscr{P}(\mathscr{Y})$

Goal: Find data $\sigma \in \Omega$ so that the noisy data $\nu_\sigma \in \mathscr{P}(\mathscr{Y})$ satisfies

$$\nu_\sigma \approx \nu.$$

Classical Unfolding

Setup:

Metric space of parameters (Ω, d_Ω)

Parametric statistical/noise model: $\{\nu_\sigma : \sigma \in \Omega\} \subseteq \mathscr{P}(\mathscr{Y})$

Measured data: $\nu \in \mathscr{P}(\mathscr{Y})$

Goal: Find data $\sigma \in \Omega$ so that the noisy data $\nu_\sigma \in \mathscr{P}(\mathscr{Y})$ satisfies

$$\nu_\sigma \approx \nu.$$

Common additional hypotheses:

Classical Unfolding

Setup:

Metric space of parameters (Ω, d_Ω)

Parametric statistical/noise model: $\{\nu_\sigma : \sigma \in \Omega\} \subseteq \mathscr{P}(\mathscr{Y})$

Measured data: $\nu \in \mathscr{P}(\mathscr{Y})$

Goal: Find data $\sigma \in \Omega$ so that the noisy data $\nu_\sigma \in \mathscr{P}(\mathscr{Y})$ satisfies

$$\nu_\sigma \approx \nu.$$

Common additional hypotheses:

Well-specified: $\nu = \nu_{\sigma_*}$ for some $\sigma_* \in \Omega$

Classical Unfolding

Setup:

Metric space of parameters (Ω, d_Ω)

Parametric statistical/noise model: $\{\nu_\sigma : \sigma \in \Omega\} \subseteq \mathscr{P}(\mathscr{Y})$

Measured data: $\nu \in \mathscr{P}(\mathscr{Y})$

Goal: Find data $\sigma \in \Omega$ so that the noisy data $\nu_\sigma \in \mathscr{P}(\mathscr{Y})$ satisfies

$$\nu_\sigma \approx \nu.$$

Common additional hypotheses:

Well-specified: $\nu = \nu_{\sigma_*}$ for some $\sigma_* \in \Omega$

Continuity of inverse: $\nu_{\sigma_n} \to \nu_\sigma \implies \sigma_n \to \sigma$

Classical Unfolding

Setup:

Metric space of parameters (Ω, d_Ω)

Parametric statistical/noise model: $\{\nu_\sigma : \sigma \in \Omega\} \subseteq \mathscr{P}(\mathscr{Y})$

Measured data: $\nu \in \mathscr{P}(\mathscr{Y})$

Goal: Find data $\sigma \in \Omega$ so that the noisy data $\nu_\sigma \in \mathscr{P}(\mathscr{Y})$ satisfies

$$\nu_\sigma \approx \nu.$$

Common additional hypotheses:

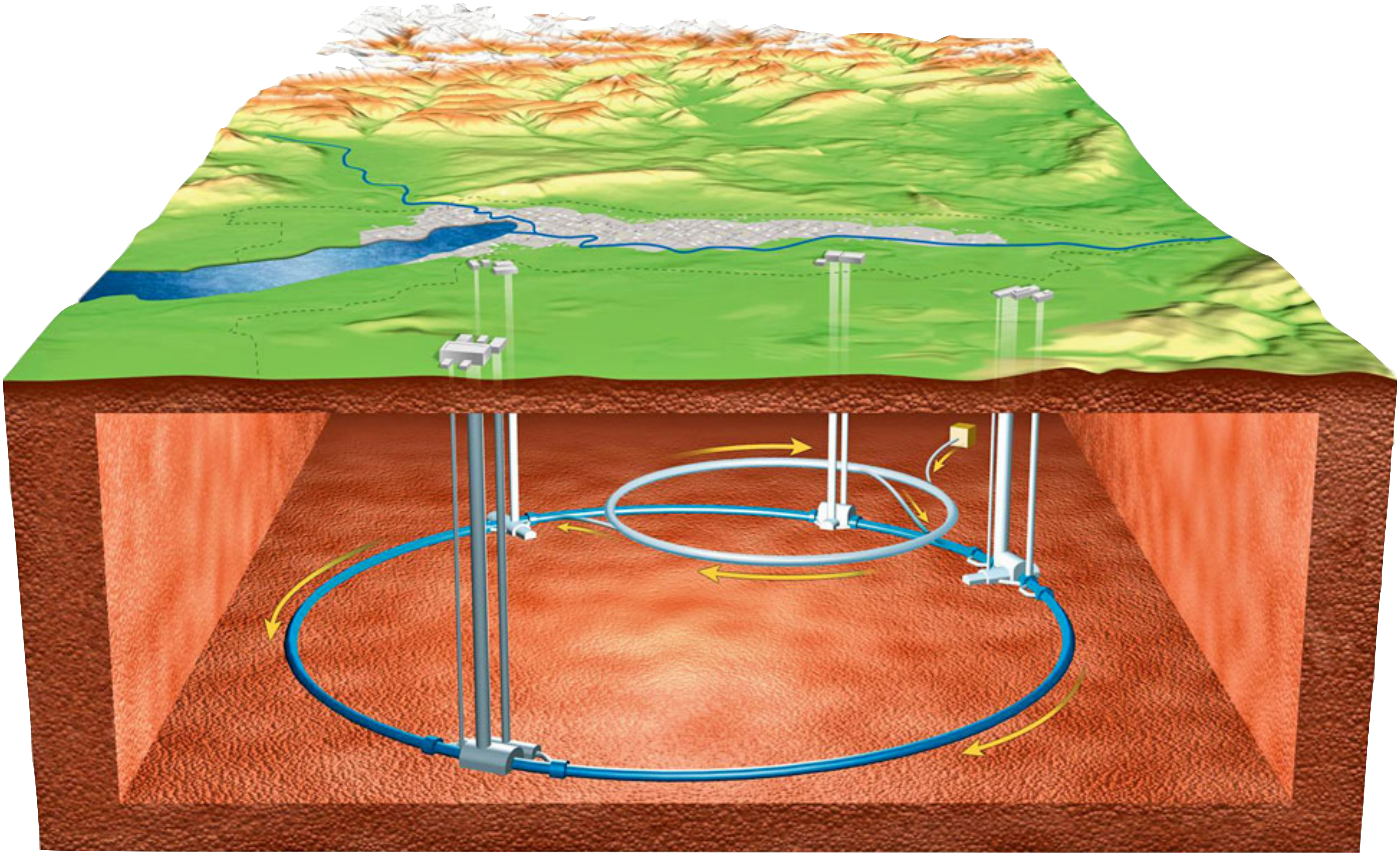
Well-specified: $\nu = \nu_{\sigma_*}$ for some $\sigma_* \in \Omega$

Continuity of inverse: $\nu_{\sigma_n} \to \nu_\sigma \implies \sigma_n \to \sigma$

Equivalent Goal: Find estimator $\sigma \in \Omega$ so that

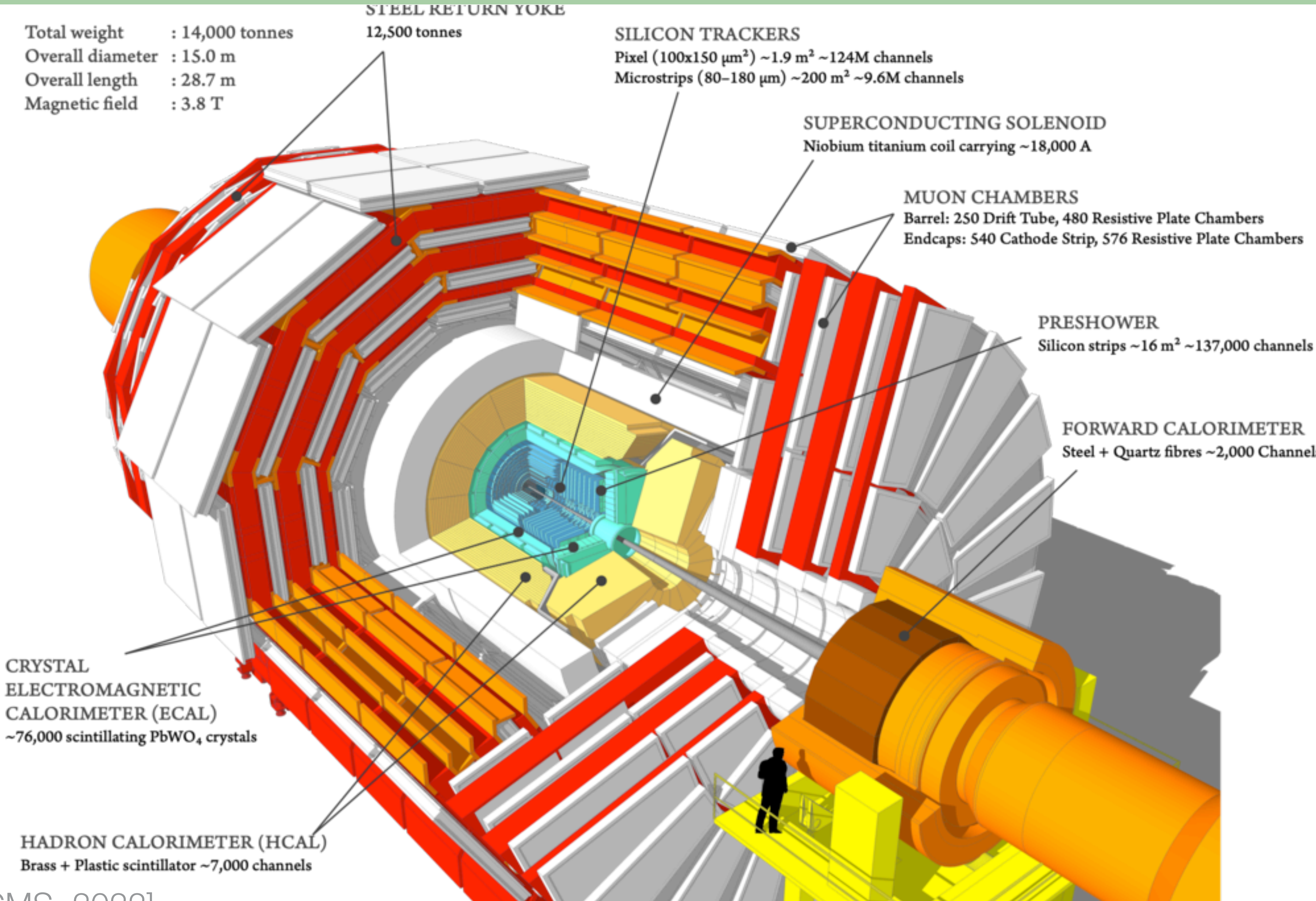
$$\sigma \approx \sigma_*.$$

Motivation



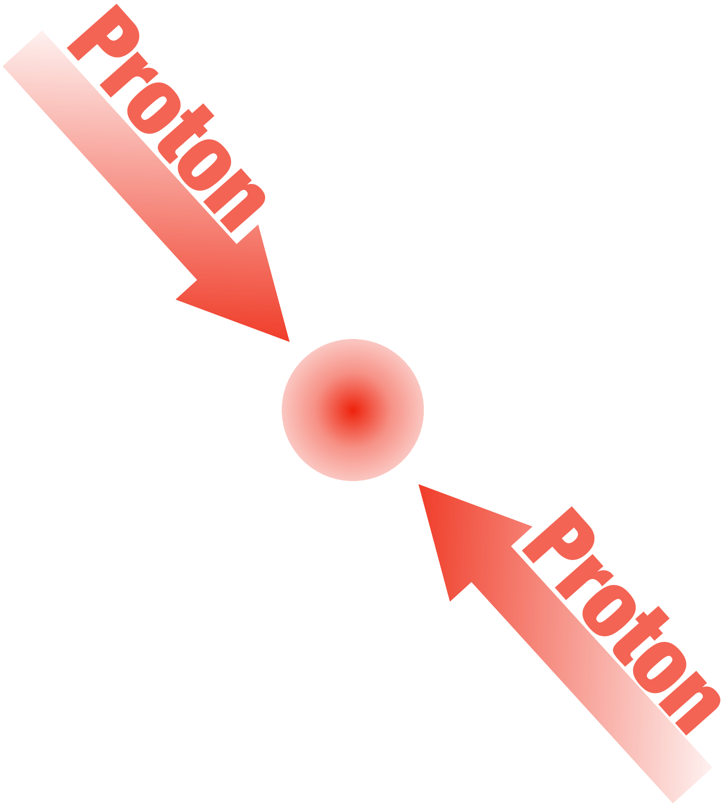
[CERN]

CMS Detector

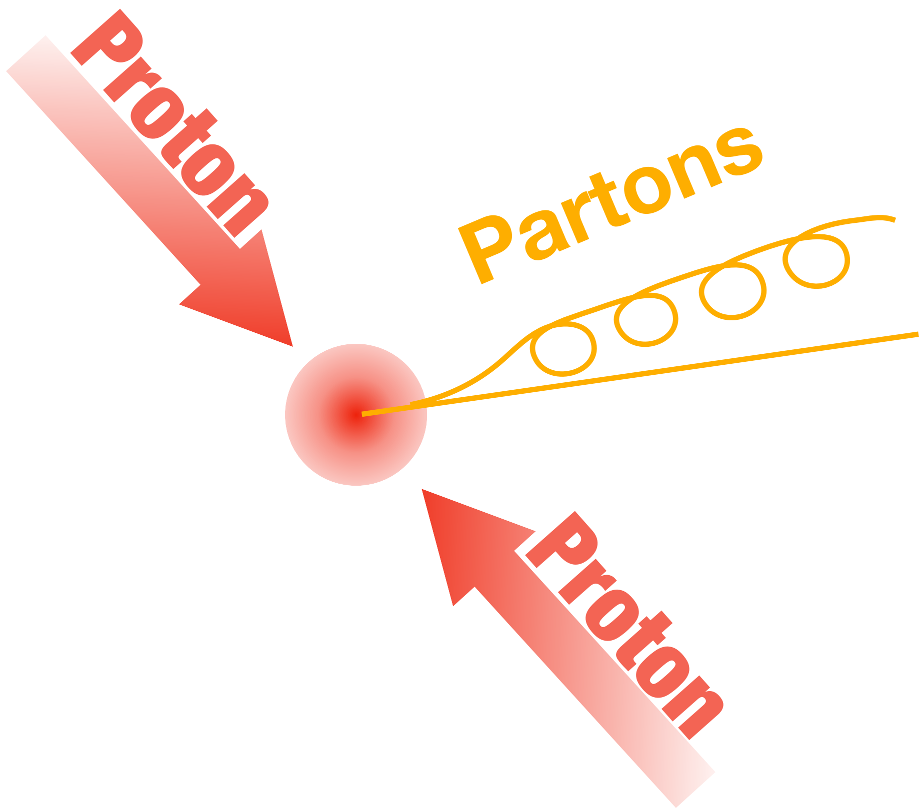


[CMS, 2022]

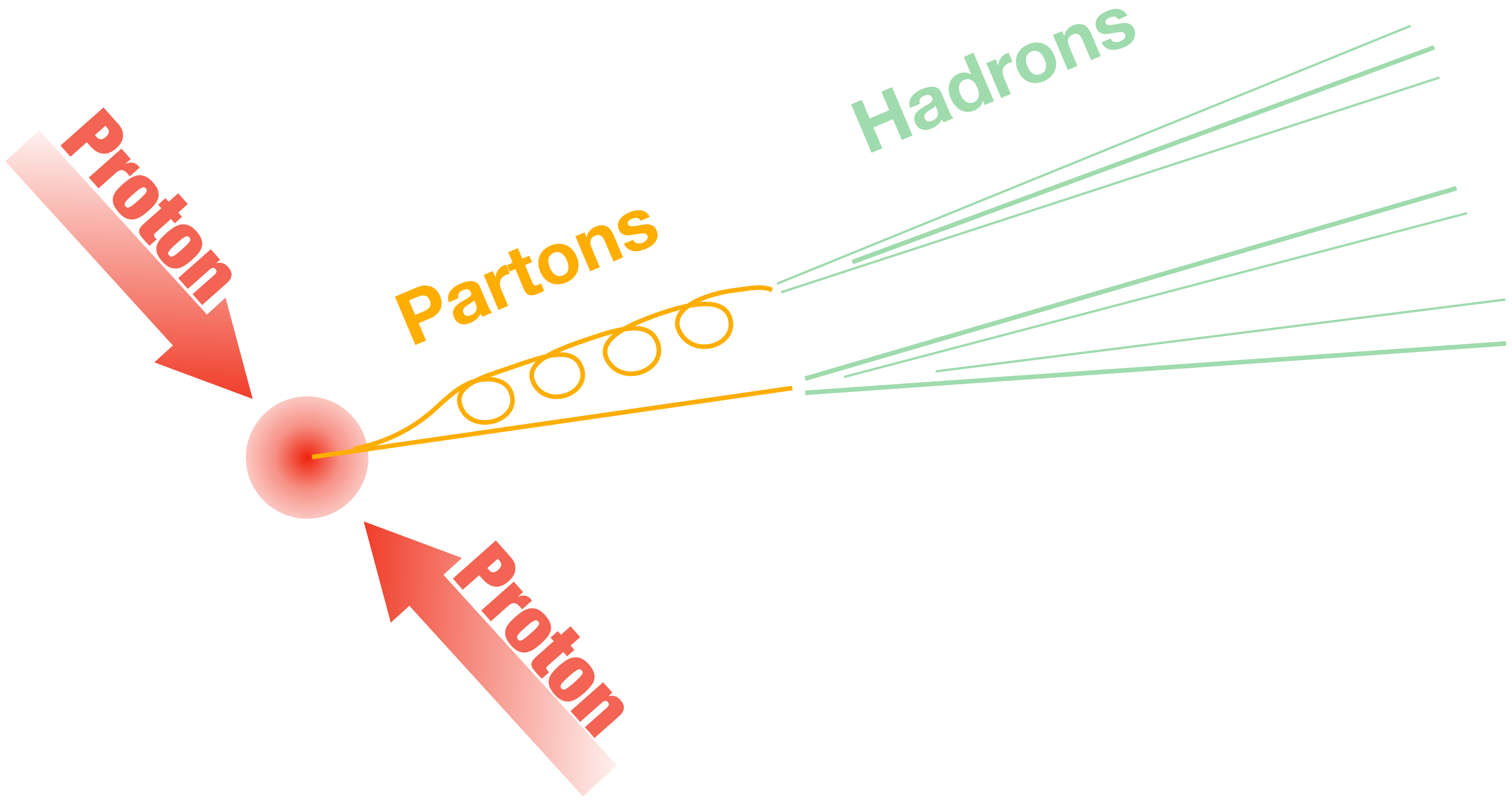
From protons to probability measures



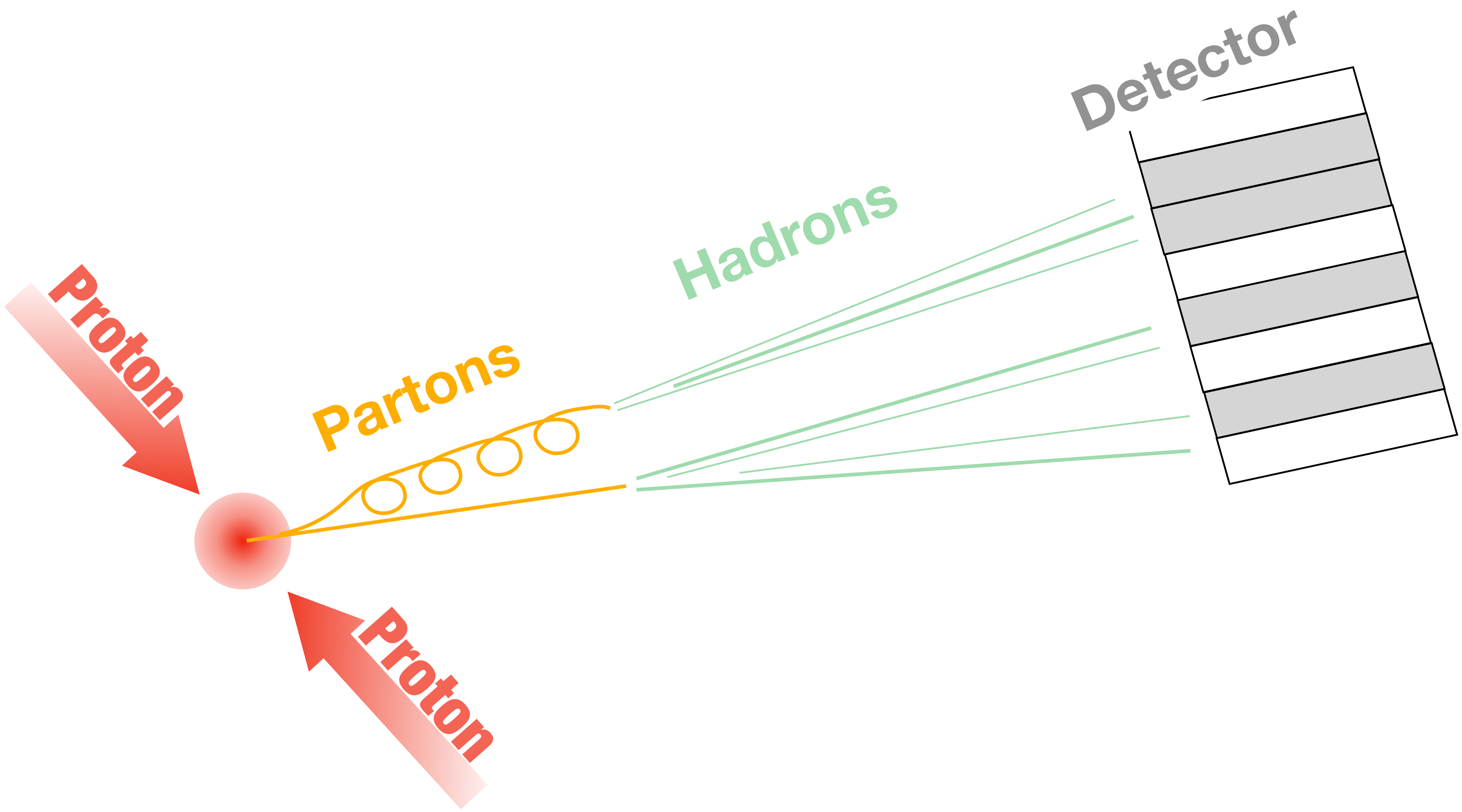
From protons to probability measures



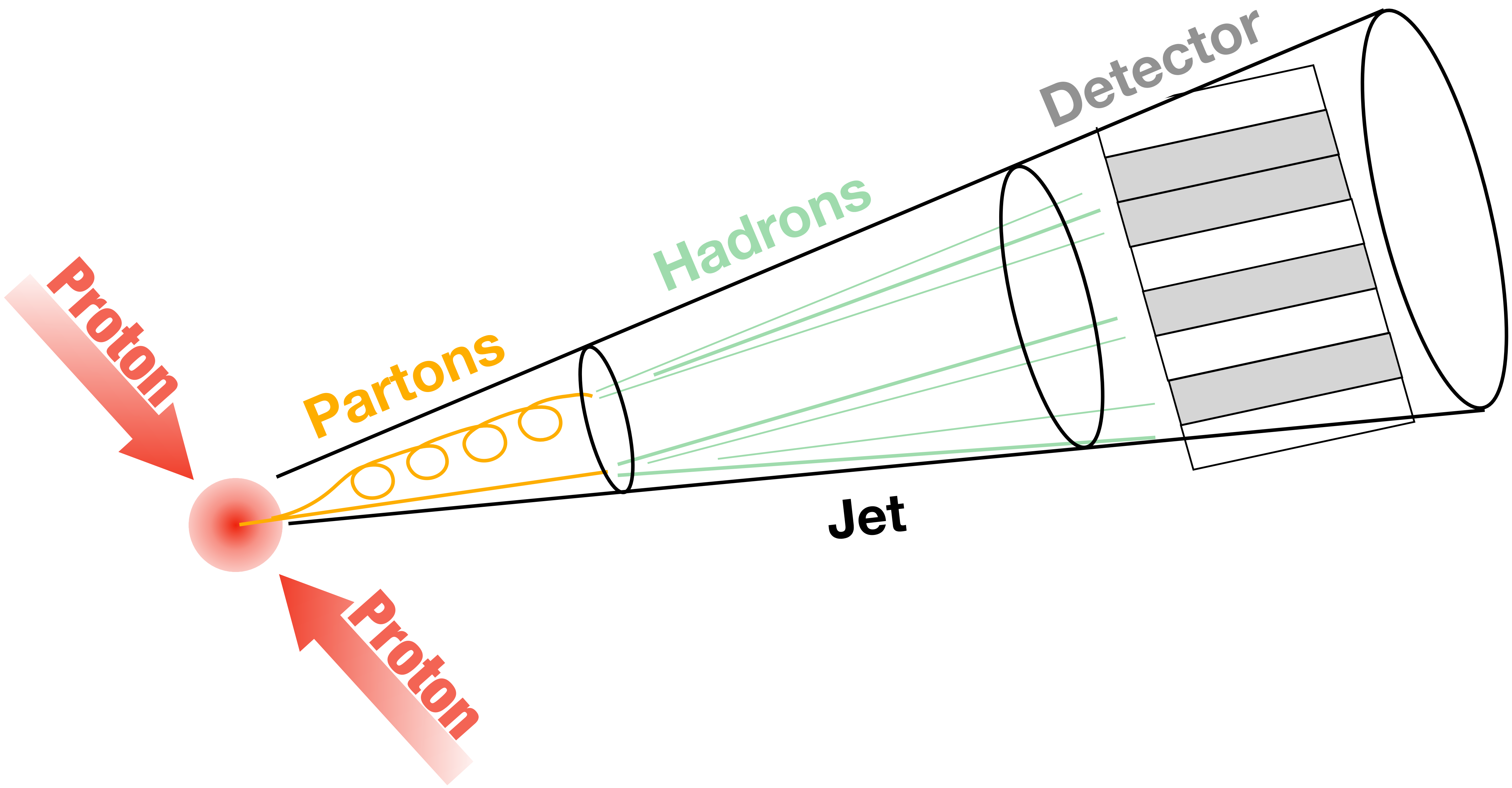
From protons to probability measures



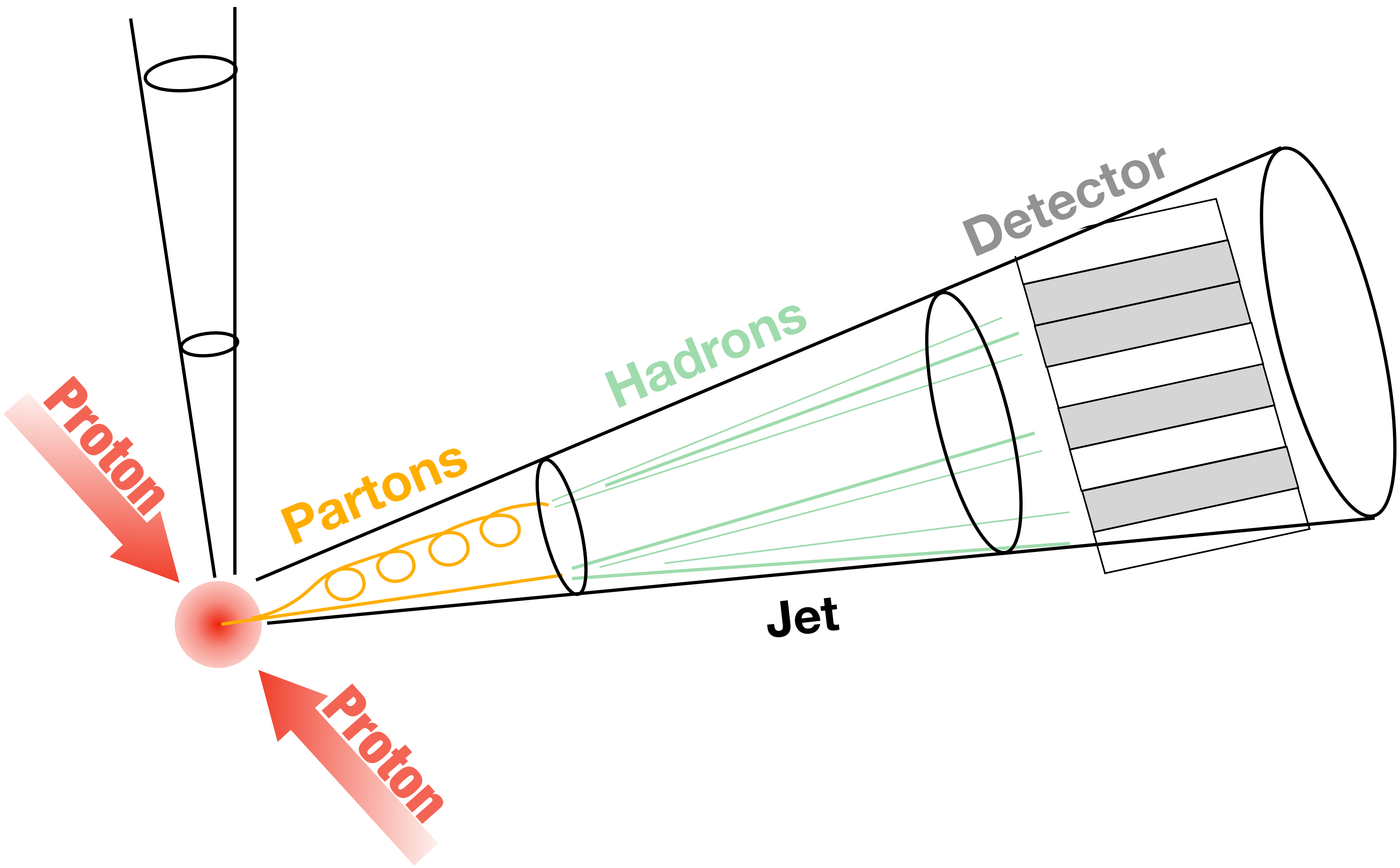
From protons to probability measures



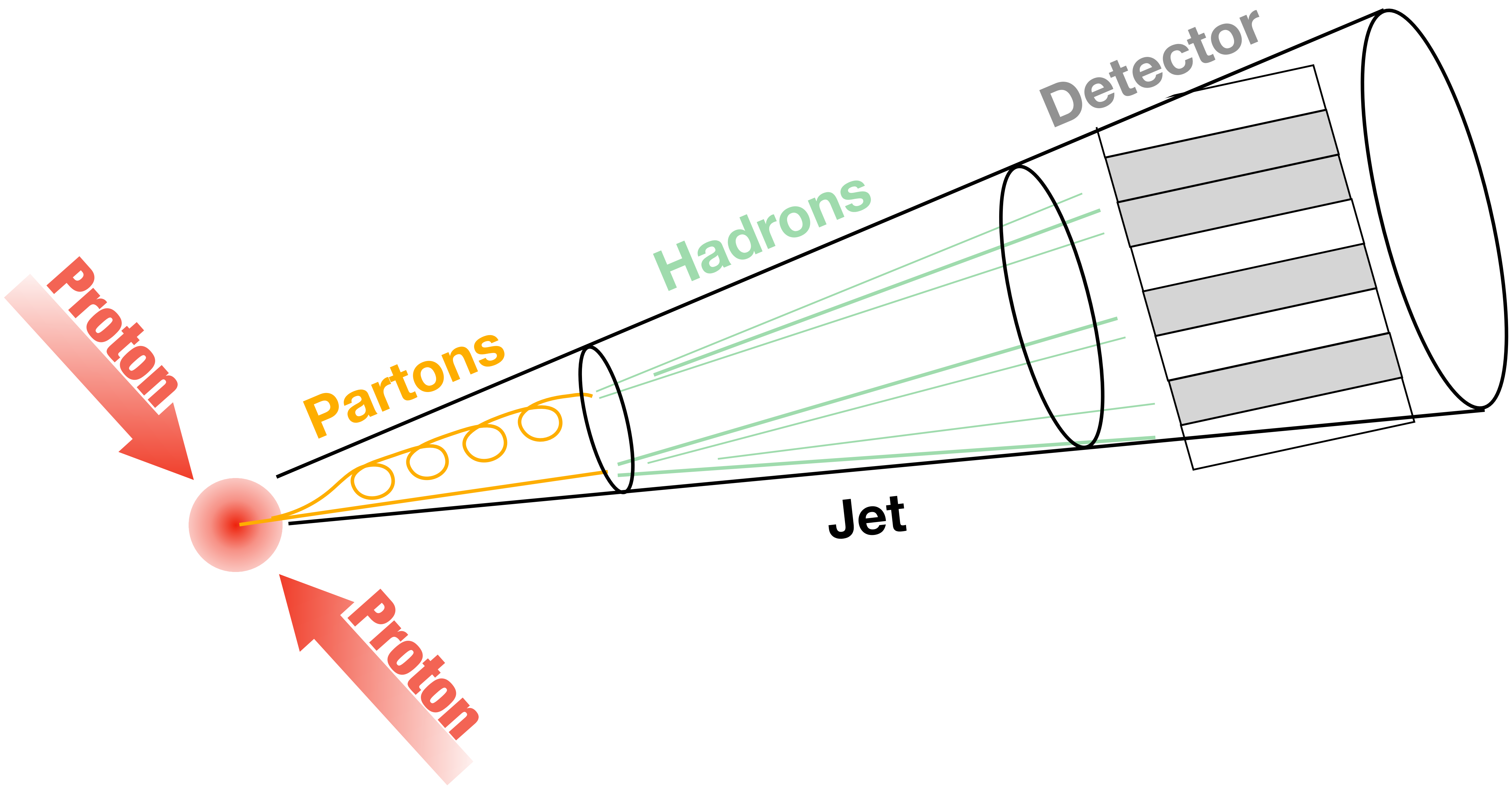
From protons to probability measures



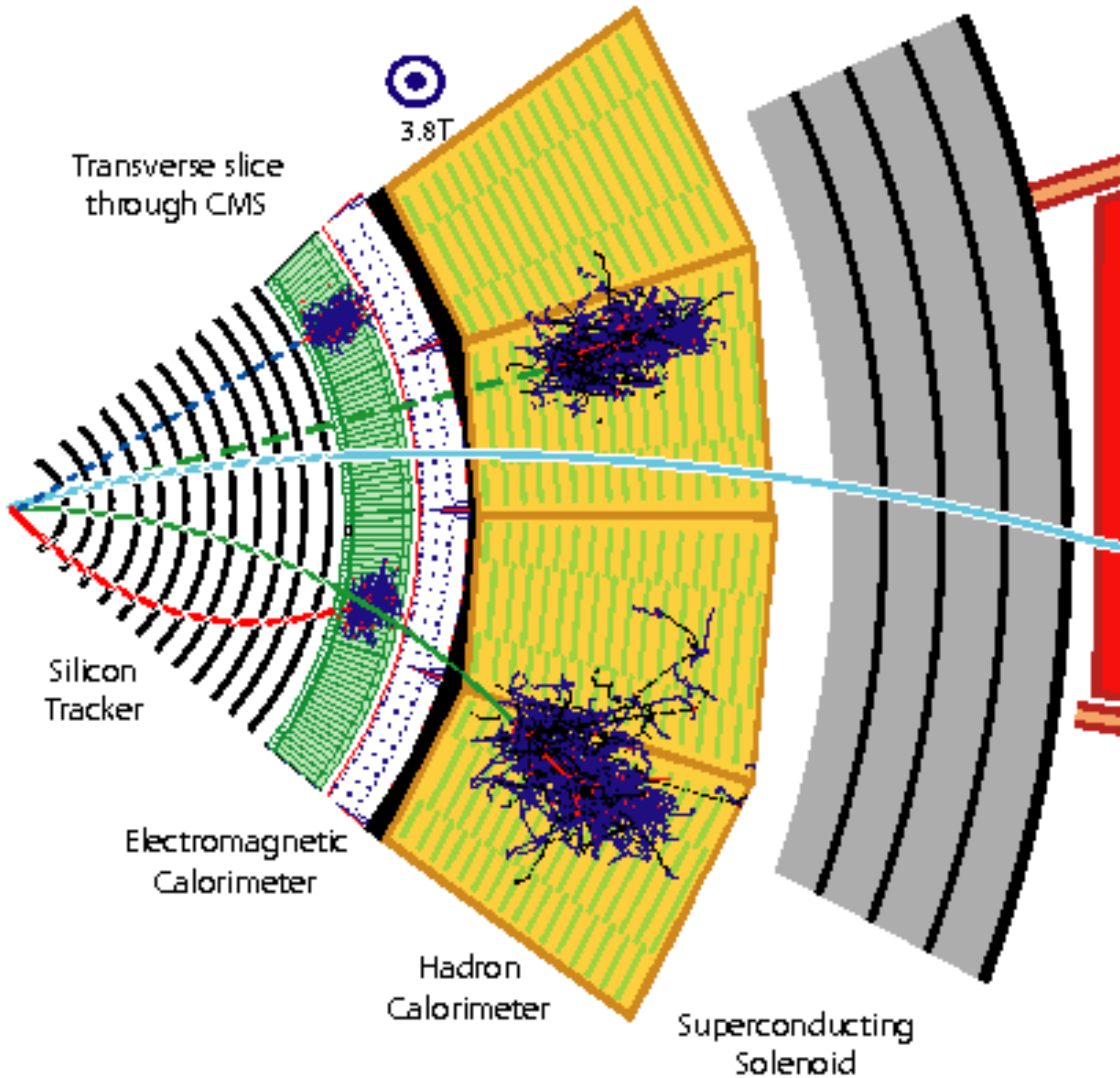
From protons to probability measures



From protons to probability measures



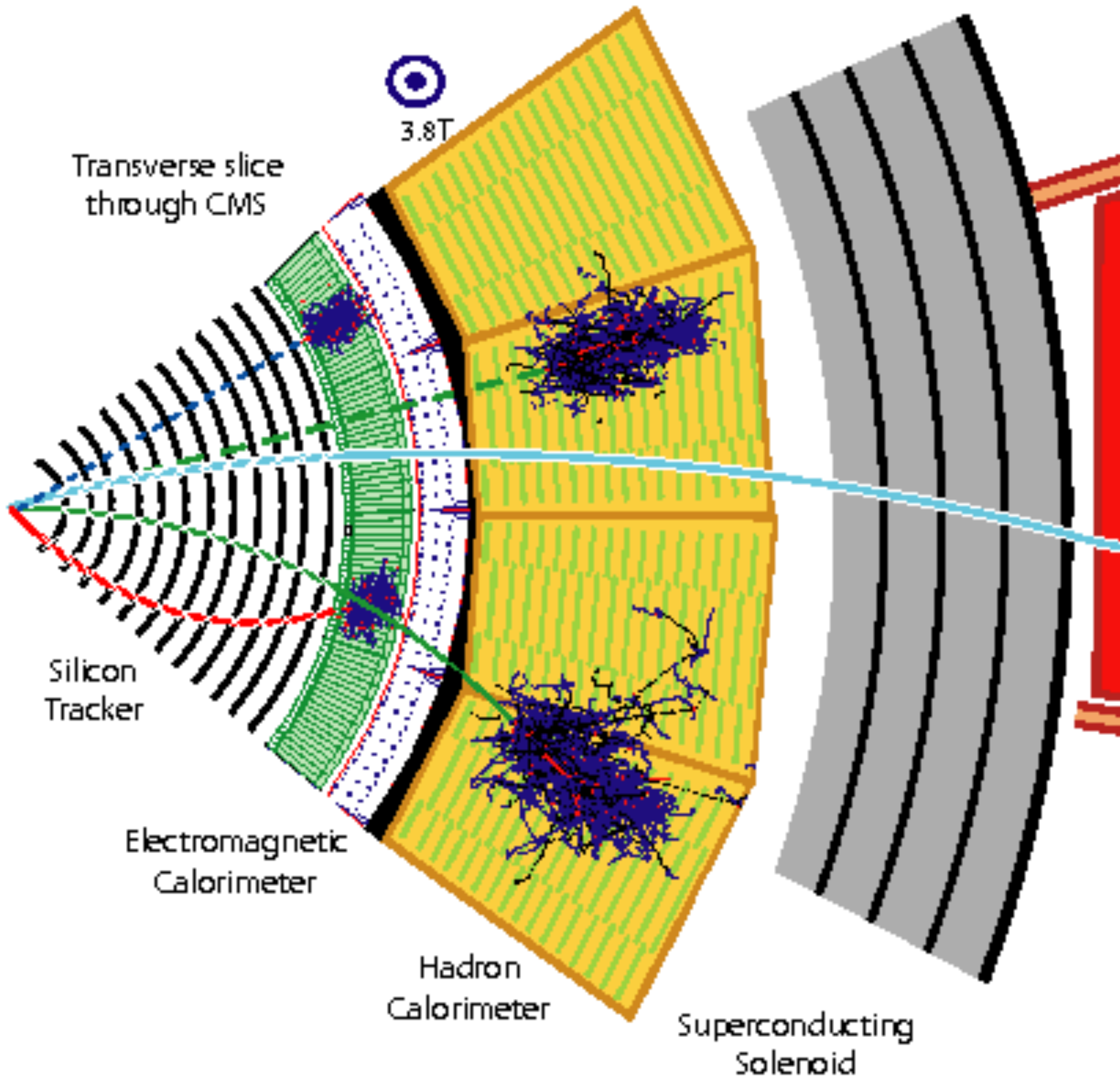
Detector Effects



Detector Effects

Distortions from detector:

- miss particles
- incorrectly measure intensity/detector location
- add background events from other processes…



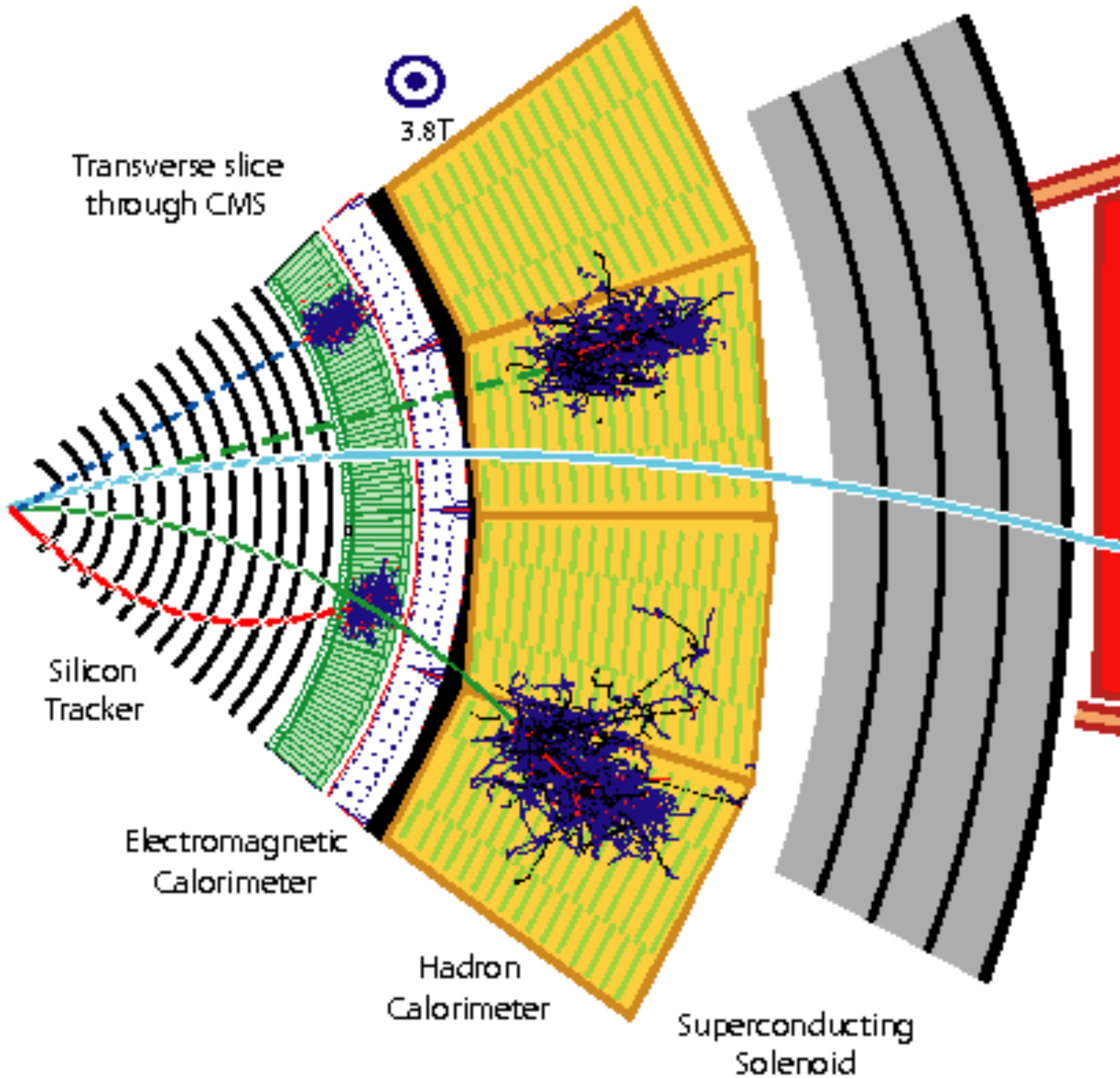
Detector Effects

Distortions from detector:

- miss particles
- incorrectly measure intensity/detector location
- add background events from other processes…

Noise model, $\sigma \mapsto \nu_\sigma$

- Detailed microscopic models of detector effects: Geant4, ATLAS FullSim, CMS FullSim, …
- Faster approximations: Delphes



Classical Approach: MLE

Classical Approach: MLE

Goal: Find data $\sigma \in \Omega$ so that the noisy data $\nu_\sigma \in \mathscr{P}(\mathscr{Y})$ satisfies

Classical Approach: MLE

Goal: Find data $\sigma \in \Omega$ so that the noisy data $\nu_\sigma \in \mathscr{P}(\mathscr{Y})$ satisfies

$$\sigma_* = \mathrm{argmin}_{\sigma \in \Omega} \mathrm{KL}(\nu \,|\, \nu_\sigma)$$

Classical Approach: MLE

Goal: Find data $\sigma \in \Omega$ so that the noisy data $\nu_\sigma \in \mathscr{P}(\mathscr{Y})$ satisfies

$$\sigma_* = \operatorname{argmin}_{\sigma \in \Omega} \mathrm{KL}(\nu \,|\, \nu_\sigma)$$

Common additional hypotheses:

Absolute cty: $\exists\ m \in \mathscr{P}(\mathscr{Y})$ s.t. $\nu \ll \nu_\sigma \ll m$ and $\mathrm{KL}(\nu \,|\, m) < +\infty$

Classical Approach: MLE

Goal: Find data $\sigma \in \Omega$ so that the noisy data $\nu_\sigma \in \mathscr{P}(\mathscr{Y})$ satisfies

$$\sigma_* = \operatorname{argmin}_{\sigma \in \Omega} \mathrm{KL}(\nu \,|\, \nu_\sigma)$$

Common additional hypotheses:

Absolute cty: $\exists\ m \in \mathscr{P}(\mathscr{Y})$ s.t. $\nu \ll \nu_\sigma \ll m$ and $\mathrm{KL}(\nu \,|\, m) < +\infty$

Equivalent Goal: Find data $\sigma \in \Omega$ so that

Classical Approach: MLE

Goal: Find data $\sigma \in \Omega$ so that the noisy data $\nu_\sigma \in \mathscr{P}(\mathscr{Y})$ satisfies

$$\sigma_* = \text{argmin}_{\sigma \in \Omega} \text{KL}(\nu \,|\, \nu_\sigma)$$

Common additional hypotheses:

Absolute cty: $\exists\ m \in \mathscr{P}(\mathscr{Y})$ s.t. $\nu \ll \nu_\sigma \ll m$ and $\text{KL}(\nu \,|\, m) < +\infty$

Equivalent Goal: Find data $\sigma \in \Omega$ so that

$$\sigma_* = \text{argmax}_{\sigma \in \Omega} \int \log\left(\frac{d\nu_\sigma}{dm}\right) d\nu$$

Classical Approach: MLE

Goal: Find data $\boldsymbol{\sigma} \in \Omega$ so that the noisy data $\nu_{\boldsymbol{\sigma}} \in \mathscr{P}(\mathscr{Y})$ satisfies

$$\sigma_* = \operatorname{argmin}_{\sigma \in \Omega} \mathrm{KL}(\nu \,|\, \nu_\sigma)$$

Common additional hypotheses:

Absolute cty: $\exists\ m \in \mathscr{P}(\mathscr{Y})$ s.t. $\nu \ll \nu_\sigma \ll m$ and $\mathrm{KL}(\nu \,|\, m) < +\infty$

Equivalent Goal: Find data $\boldsymbol{\sigma} \in \Omega$ so that

$$\sigma_* = \operatorname{argmax}_{\sigma \in \Omega} \int \log\left(\frac{d\nu_\sigma}{dm}\right) d\nu$$

Empirical: $m = \frac{1}{N}\sum_{i=1}^{N} \delta_{y_i} \implies \nu = \sum_{i=1}^{n} \delta_{y_i} \boldsymbol{\nu_i}, \quad \nu_{\boldsymbol{\sigma}} = \sum_{i=1}^{n} \delta_{y_i} (\boldsymbol{\nu_\sigma})_i$

Classical Approach: MLE

Goal: Find data $\boldsymbol{\sigma} \in \Omega$ so that the noisy data $\nu_{\boldsymbol{\sigma}} \in \mathscr{P}(\mathscr{Y})$ satisfies

$$\sigma_* = \operatorname{argmin}_{\sigma \in \Omega} \mathrm{KL}(\nu \,|\, \nu_\sigma)$$

Common additional hypotheses:

Absolute cty: $\exists\ m \in \mathscr{P}(\mathscr{Y})$ s.t. $\nu \ll \nu_\sigma \ll m$ and $\mathrm{KL}(\nu \,|\, m) < +\infty$

Equivalent Goal: Find data $\boldsymbol{\sigma} \in \Omega$ so that

$$\sigma_* = \operatorname{argmax}_{\sigma \in \Omega} \int \log\left(\frac{d\nu_\sigma}{dm}\right) d\nu$$

Empirical: $m = \dfrac{1}{N} \sum_{i=1}^{N} \delta_{y_i} \implies \nu = \sum_{i=1}^{n} \delta_{y_i} \boldsymbol{\nu_i}, \quad \nu_\sigma = \sum_{i=1}^{n} \delta_{y_i} (\boldsymbol{\nu_\sigma})_i$

Find the MLE: $\sigma_* = \operatorname{argmax}_\sigma \sum_{i=1}^{N} \log\left((\boldsymbol{\nu}_\sigma)_i\right) \boldsymbol{\nu}_i$

Classical Approach: MLE

Goal: Find data $\boldsymbol{\sigma} \in \Omega$ so that the noisy data $\boldsymbol{\nu}_{\boldsymbol{\sigma}} \in \mathscr{P}(\mathscr{Y})$ satisfies

$$\sigma_* = \text{argmin}_{\sigma \in \Omega} \text{KL}(\nu \,|\, \nu_\sigma)$$

Common additional hypotheses:

Absolute cty: $\exists\ m \in \mathscr{P}(\mathscr{Y})$ s.t. $\nu \ll \nu_\sigma \ll m$ and $\text{KL}(\nu \,|\, m) < +\infty$

Equivalent Goal: Find data $\boldsymbol{\sigma} \in \Omega$ so that

$$\sigma_* = \text{argmax}_{\sigma \in \Omega} \int \log\left(\frac{d\nu_\sigma}{dm}\right) d\nu$$

Empirical: $m = \frac{1}{N}\sum_{i=1}^{N} \delta_{y_i} \implies \nu = \sum_{i=1}^{n} \delta_{y_i} \boldsymbol{\nu_i}, \quad \nu_\sigma = \sum_{i=1}^{n} \delta_{y_i} (\boldsymbol{\nu_\sigma})_i$

Find the MLE: $\sigma_* = \text{argmax}_\sigma \sum_{i=1}^{N} \log\left((\boldsymbol{\nu}_\sigma)_i\right) \boldsymbol{\nu}_i$

Limitations of MLE approach

Limitations of MLE approach

Absolute cty: $\exists\ m \in \mathscr{P}(\mathscr{Y})$ s.t. $\nu \ll \nu_\sigma \ll m$ and $\mathrm{KL}(\nu\,|\,m) < +\infty$

Limitations of MLE approach

Absolute cty: $\exists\ m \in \mathscr{P}(\mathscr{Y})$ s.t. $\nu \ll \nu_\sigma \ll m$ and $\mathrm{KL}(\nu\,|\,m) < +\infty$

In LHC applications, ν_σ is empirical and known only through simulation.

Limitations of MLE approach

Absolute cty: $\exists\ m \in \mathscr{P}(\mathscr{Y})$ s.t. $\nu \ll \nu_\sigma \ll m$ and $\mathrm{KL}(\nu \,|\, m) < +\infty$

In LHC applications, ν_σ is empirical and known only through simulation.

To enforce the absolute cty hypothesis, reference measure m is taken to be empirical, uniformly distributed on a grid; both ν and ν_σ are replaced with approximations on the same grid.

Limitations of MLE approach

Absolute cty: $\exists\ m \in \mathcal{P}(\mathcal{Y})$ s.t. $\nu \ll \nu_\sigma \ll m$ and $\mathrm{KL}(\nu\,|\,m) < +\infty$

In LHC applications, ν_σ is empirical and known only through simulation.

To enforce the absolute cty hypothesis, reference measure m is taken to be empirical, uniformly distributed on a grid; both ν and ν_σ are replaced with approximations on the same grid.

Limitations of MLE approach

Absolute cty: $\exists\ m \in \mathscr{P}(\mathscr{Y})$ s.t. $\nu \ll \nu_\sigma \ll m$ and $\mathrm{KL}(\nu \,|\, m) < +\infty$

In LHC applications, ν_σ is empirical and known only through simulation.

To enforce the absolute cty hypothesis, reference measure m is taken to be empirical, uniformly distributed on a grid; both ν and ν_σ are replaced with approximations on the same grid.

Limitations:

Limitations of MLE approach

Absolute cty: $\exists\ m \in \mathscr{P}(\mathscr{Y})$ s.t. $\nu \ll \nu_\sigma \ll m$ and $\mathrm{KL}(\nu\,|\,m) < +\infty$

In LHC applications, ν_σ is empirical and known only through simulation.

To enforce the absolute cty hypothesis, reference measure m is taken to be empirical, uniformly distributed on a grid; both ν and ν_σ are replaced with approximations on the same grid.

Limitations:

- Experiments must coordinate bin centers ahead of time

Limitations of MLE approach

Absolute cty: $\exists\ m \in \mathcal{P}(\mathcal{Y})$ s.t. $\nu \ll \nu_\sigma \ll m$ and $\mathrm{KL}(\nu \,|\, m) < +\infty$

In LHC applications, ν_σ is empirical and known only through simulation.

To enforce the absolute cty hypothesis, reference measure m is taken to be empirical, uniformly distributed on a grid; both ν and ν_σ are replaced with approximations on the same grid.

Limitations:

- Experiments must coordinate bin centers ahead of time
- Optimal σ_* polluted by discretization errors

Limitations of MLE approach

Absolute cty: $\exists\ m \in \mathscr{P}(\mathscr{Y})$ s.t. $\nu \ll \nu_\sigma \ll m$ and $\mathrm{KL}(\nu \,|\, m) < +\infty$

In LHC applications, ν_σ is empirical and known only through simulation.

To enforce the absolute cty hypothesis, reference measure m is taken to be empirical, uniformly distributed on a grid; both ν and ν_σ are replaced with approximations on the same grid.

Limitations:

- Experiments must coordinate bin centers ahead of time
- Optimal σ_* polluted by discretization errors
- Computational complexity of grid-based methods scales exponentially with dimension $\Longrightarrow$ ν and ν_σ are typically replaced with lower dimensional projections, throwing away information

Unbinned Unfolding

Unbinned Unfolding

Motivated by the need for an unbinned unfolding method, which admits rigorous guarantees and a provably convergent numerical method…

Unbinned Unfolding

Motivated by the need for an unbinned unfolding method, which admits rigorous guarantees and a provably convergent numerical method…

Goal: Find data $\sigma \in \Omega$ so that the noisy data $\nu_\sigma \in \mathscr{P}(\mathscr{Y})$ satisfies

Unbinned Unfolding

Motivated by the need for an unbinned unfolding method, which admits rigorous guarantees and a provably convergent numerical method…

Goal: Find data $\sigma \in \Omega$ so that the noisy data $\nu_\sigma \in \mathscr{P}(\mathscr{Y})$ satisfies

$$\sigma_* = \mathrm{argmin}_{\sigma \in \Omega} W_p^p(\nu_\sigma, \nu)$$

Unbinned Unfolding

Motivated by the need for an unbinned unfolding method, which admits rigorous guarantees and a provably convergent numerical method…

Goal: Find data $\sigma \in \Omega$ so that the noisy data $\nu_\sigma \in \mathscr{P}(\mathscr{Y})$ satisfies

$$\sigma_* = \mathrm{argmin}_{\sigma \in \Omega} W_p^p(\nu_\sigma, \nu)$$

We consider a nonparametric formulation:

Unbinned Unfolding

Motivated by the need for an unbinned unfolding method, which admits rigorous guarantees and a provably convergent numerical method…

Goal: Find data $\boldsymbol{\sigma} \in \boldsymbol{\Omega}$ so that the noisy data $\boldsymbol{\nu}_{\boldsymbol{\sigma}} \in \mathscr{P}(\mathscr{Y})$ satisfies

$$\sigma_* = \mathrm{argmin}_{\sigma \in \Omega} W_p^p(\nu_\sigma, \nu)$$

We consider a nonparametric formulation:

- $\boldsymbol{\Omega} = \mathscr{P}(\mathscr{X})$, where $\mathscr{X}$ is a Polish space; $\mathscr{Y} = \mathbb{R}^d$

Unbinned Unfolding

Motivated by the need for an unbinned unfolding method, which admits rigorous guarantees and a provably convergent numerical method…

Goal: Find data $\boldsymbol{\sigma} \in \boldsymbol{\Omega}$ so that the noisy data $\boldsymbol{\nu}_{\sigma} \in \mathscr{P}(\mathscr{Y})$ satisfies

$$\sigma_* = \mathrm{argmin}_{\sigma \in \Omega} W_p^p(\nu_\sigma, \nu)$$

We consider a nonparametric formulation:

- $\boldsymbol{\Omega} = \mathscr{P}(\mathscr{X})$, where $\mathscr{X}$ is a Polish space; $\mathscr{Y} = \mathbb{R}^d$
- Given a Markov kernel $\rho : \mathscr{X} \to \mathscr{P}(\mathbb{R}^d) : x \mapsto \rho_x$, define $\boldsymbol{\sigma} \mapsto \boldsymbol{\nu}_{\sigma}$ by

Unbinned Unfolding

Motivated by the need for an unbinned unfolding method, which admits rigorous guarantees and a provably convergent numerical method…

Goal: Find data $\sigma \in \Omega$ so that the noisy data $\nu_\sigma \in \mathcal{P}(\mathcal{Y})$ satisfies

$$\sigma_* = \mathrm{argmin}_{\sigma \in \Omega} W_p^p(\nu_\sigma, \nu)$$

We consider a nonparametric formulation:

- $\Omega = \mathcal{P}(\mathcal{X})$, where $\mathcal{X}$ is a Polish space; $\mathcal{Y} = \mathbb{R}^d$
- Given a Markov kernel $\rho : \mathcal{X} \to \mathcal{P}(\mathbb{R}^d) : x \mapsto \rho_x$, define $\sigma \mapsto \nu_\sigma$ by

$$\int_{\mathbb{R}^d} f(y) d\nu_\sigma(y) = \int_X \int_{\mathbb{R}^d} f(y) d\rho_x(y) d\sigma(x) \ , \quad \forall f \in C_b(\mathbb{R}^d)$$

Unbinned Unfolding

Motivated by the need for an unbinned unfolding method, which admits rigorous guarantees and a provably convergent numerical method…

Goal: Find data $\sigma \in \Omega$ so that the noisy data $\nu_\sigma \in \mathscr{P}(\mathscr{Y})$ satisfies

$$\sigma_* = \text{argmin}_{\sigma \in \Omega} W_p^p(\nu_\sigma, \nu)$$

Interpretation: true signal at location x is observed on the detector as $\rho_x \in \mathscr{P}(\mathbb{R}^d)$

We consider a nonparametri

- $\Omega = \mathscr{P}(\mathscr{X})$, where $\mathscr{X}$ is
- Given a Markov kernel $\rho : \mathscr{X} \to \mathscr{P}(\mathbb{R}^d) : x \mapsto \rho_x$, define $\sigma \mapsto \nu_\sigma$ by

$$\int_{\mathbb{R}^d} f(y) d\nu_\sigma(y) = \int_X \int_{\mathbb{R}^d} f(y) d\rho_x(y) d\sigma(x) \ , \quad \forall f \in C_b(\mathbb{R}^d)$$

Unbinned Unfolding

Motivated by the need for an unbinned unfolding method, which admits rigorous guarantees and a provably convergent numerical method…

Goal: Find data $\sigma \in \Omega$ so that the noisy data $\nu_\sigma \in \mathscr{P}(\mathscr{Y})$ satisfies

$$\sigma_* = \text{argmin}_{\sigma\in\Omega} W_p^p(\nu_\sigma, \nu)$$

We consider a nonparametri

- $\Omega = \mathscr{P}(\mathscr{X})$, where $\mathscr{X}$ is

<u>Interpretation:</u> true signal at location x is observed on the detector as $\rho_x \in \mathscr{P}(\mathbb{R}^d)$

- Given a Markov kernel $\rho : \mathscr{X} \to \mathscr{P}(\mathbb{R}^d) : x \mapsto \rho_x$, define $\sigma \mapsto \nu_\sigma$ by

$$\int_{\mathbb{R}^d} f(y)d\nu_\sigma(y) = \int_X \int_{\mathbb{R}^d} f(y)d\rho_x(y)d\sigma(x) \ , \quad \forall f \in C_b(\mathbb{R}^d)$$

<u>Examples:</u> *No noise:* $\rho_x = \delta_x \, , \, \nu_\sigma = \sigma$

Unbinned Unfolding

Motivated by the need for an unbinned unfolding method, which admits rigorous guarantees and a provably convergent numerical method…

Goal: Find data $\sigma \in \Omega$ so that the noisy data $\nu_\sigma \in \mathscr{P}(\mathscr{Y})$ satisfies

$$\sigma_* = \operatorname{argmin}_{\sigma \in \Omega} W_p^p(\nu_\sigma, \nu)$$

We consider a nonparametri

- $\Omega = \mathscr{P}(\mathscr{X})$, where $\mathscr{X}$ is

Interpretation: true signal at location x is observed on the detector as $\rho_x \in \mathscr{P}(\mathbb{R}^d)$

- Given a Markov kernel $\rho : \mathscr{X} \to \mathscr{P}(\mathbb{R}^d) : x \mapsto \rho_x$, define $\sigma \mapsto \nu_\sigma$ by

$$\int_{\mathbb{R}^d} f(y) d\nu_\sigma(y) = \int_X \int_{\mathbb{R}^d} f(y) d\rho_x(y) d\sigma(x) \;, \quad \forall f \in C_b(\mathbb{R}^d)$$

Examples: *No noise:* $\rho_x = \delta_x \,,\; \nu_\sigma = \sigma$

Gaussian mixture noise: $\rho_x = \varphi(y - x) d\lambda(y), \;\; \nu_\sigma = \varphi * \sigma \, d\lambda$

Unbinned Unfolding

Motivated by the need for an unbinned unfolding method, which admits rigorous guarantees and a provably convergent numerical method…

Goal: Find data $\sigma \in \Omega$ so that the noisy data $\nu_\sigma \in \mathscr{P}(\mathscr{Y})$ satisfies

$$\sigma_* = \operatorname{argmin}_{\sigma \in \Omega} W_p^p(\nu_\sigma, \nu)$$

Interpretation: true signal at location x is observed on the detector as $\rho_x \in \mathscr{P}(\mathbb{R}^d)$

We consider a nonparametri

- $\Omega = \mathscr{P}(\mathscr{X})$, where $\mathscr{X}$ is
- Given a Markov kernel $\rho : \mathscr{X} \to \mathscr{P}(\mathbb{R}^d) : x \mapsto \rho_x$, define $\sigma \mapsto \nu_\sigma$ by

$$\int_{\mathbb{R}^d} f(y) d\nu_\sigma(y) = \int_X \int_{\mathbb{R}^d} f(y) d\rho_x(y) d\sigma(x) \;, \quad \forall f \in C_b(\mathbb{R}^d)$$

Examples: *No noise:* $\rho_x = \delta_x \,,\, \nu_\sigma = \sigma$

Gaussian mixture noise: $\rho_x = \varphi(y - x) d\lambda(y), \;\; \nu_\sigma = \varphi * \sigma \, d\lambda$

Transport map noise: $\rho_x = \delta_{t(x)} \,,\, \nu_\sigma = t\#\sigma$

Plan

1. Motivation
2. Previous Work
3. Well-posedness vs Ill-posedness
4. Numerical Method
5. Numerical Results

Previous Work

Previous Work

Statistical Estimators (minimum Kantorovich, Wasserstein projection,..)

- [Bassetti, Bodini, and Regazzini '06] consider $\nu = \nu_n$ as an empirical iid approximation of some $\tilde{\nu}_{\text{true}} = \nu_{\sigma_*}$; sufficient conditions for existence of optimum σ_n and convergence of $\sigma_n \to \sigma_*$
- [Bernton, et. al. '19] consider $\Omega = \mathbb{R}^m$, weaken hypothesis on approximation ν_n and that $\tilde{\nu}_{\text{true}} = \nu_{\sigma_*}$; introduce numerical approach for computing minimizer via Monte-Carlo expectation maximization… but only once a method is given to optimize the finite sample problem in σ. Our numerical approach gives such a method.
- [García-Trillos, Jaffe, Sen '25] proved *sensitivity efficiency* of the estimator.

Previous Work

Previous Work

Learning generative models

- [Genevay, Peyré, and Cuturi '18] and [Akyildiz, Girolami, Vadeboncouer, and Stuart '25] consider numerical approach via auto-differentiation for $\nu_\sigma = \varphi * (g\#\sigma)$ for $\varphi,\ g$ smooth

Inverse problems over the space of measures

- [Li, et. al. '24, '25] consider transport map noise, develop sufficient conditions for existence, and raise the open question of sufficient conditions for uniqueness

- [Lasserre '24] considers ν_σ given by a Gaussian mixture where $\operatorname{supp}\ \sigma$ is a compact set of mean and covariance parameters; characterizes the unique minimizer in terms of a moment relaxation problem.

Plan

1. Motivation
2. Previous Work
3. Well-posedness vs Ill-posedness
4. Numerical Method
5. Numerical Results

Well/Ill-posedness of W_p-Unfolding

Well/Ill-posedness of W_p-Unfolding

Assumptions:

Well/Ill-posedness of W_p-Unfolding

Assumptions:

(CTY): $x \mapsto \rho_x$ is continuous

Well/Ill-posedness of W_p-Unfolding

Assumptions:

(CTY): $x \mapsto \rho_x$ is continuous

(CRC): $x \mapsto M_p(\rho_x)$ has totally bounded sublevel sets

Well/Ill-posedness of W_p-Unfolding

Assumptions:

(CTY): $x \mapsto \rho_x$ is continuous

(CRC): $x \mapsto M_p(\rho_x)$ has totally bounded sublevel sets

(INJ): $\sigma \mapsto \nu_\sigma$ is injective on the set of minimizers

Well/Ill-posedness of W_p-Unfolding

Assumptions:

(CTY): $x \mapsto \rho_x$ is continuous

(CRC): $x \mapsto M_p(\rho_x)$ has totally bounded sublevel sets

(INJ): $\sigma \mapsto \nu_\sigma$ is injective on the set of minimizers

Thm [C., Faktor, Nachman '26]:

Well/Ill-posedness of W_p-Unfolding

Assumptions:

(CTY): $x \mapsto \rho_x$ is continuous

(CRC): $x \mapsto M_p(\rho_x)$ has totally bounded sublevel sets

(INJ): $\sigma \mapsto \nu_\sigma$ is injective on the set of minimizers

Thm [C., Faktor, Nachman '26]:

- If (CTY) and (CRC) hold, a solution exists.

Well/Ill-posedness of W_p-Unfolding

Assumptions:

(CTY): $x \mapsto \rho_x$ is continuous

(CRC): $x \mapsto M_p(\rho_x)$ has totally bounded sublevel sets

(INJ): $\sigma \mapsto \nu_\sigma$ is injective on the set of minimizers

Thm [C., Faktor, Nachman ’26]:

- If (CTY) and (CRC) hold, a solution exists.
- If $p > 1$, $\nu \ll \mathscr{L}^d$, and (INJ), then solutions are unique iff the problem is feasible.

Well/Ill-posedness of W_p-Unfolding

Assumptions:

(CTY): $x \mapsto \rho_x$ is continuous

(CRC): $x \mapsto M_p(\rho_x)$ has totally bounded sublevel sets

(INJ): $\sigma \mapsto \nu_\sigma$ is injective on the set of minimizers

Thm [C., Faktor, Nachman '26]:

- If (CTY) and (CRC) hold, a solution exists.
- If $p > 1$, $\nu \ll \mathscr{L}^d$, and (INJ), then solutions are unique iff the problem is feasible.

The hypotheses are **necessary:**

Well/Ill-posedness of W_p-Unfolding

Assumptions:

(CTY): $x \mapsto \rho_x$ is continuous

(CRC): $x \mapsto M_p(\rho_x)$ has totally bounded sublevel sets

(INJ): $\sigma \mapsto \nu_\sigma$ is injective on the set of minimizers

Thm [C., Faktor, Nachman '26]:

- If (CTY) and (CRC) hold, a solution exists.
- If $p > 1$, $\nu \ll \mathscr{L}^d$, and (INJ), then solutions are unique iff the problem is feasible.

The hypotheses are **necessary:**

$\exists$	(CTY)	$\neg$(CTY)
(CRC)	Yes	No
$\neg$(CRC)	No	

$!$	$p > 1$	$p = 1$
$\nu \ll \mathcal{L}^d$	Yes	No
$\neg(\nu \ll \mathcal{L}^d)$	No	

Transport map noise, $\nu_\sigma = t\#\sigma$

Transport map noise, $\nu_\sigma = t\#\sigma$

Assumptions:

Transport map noise, $\nu_\sigma = t\#\sigma$

Assumptions:

(CTY): $t(x)$ is continuous

Transport map noise, $\nu_\sigma = t\#\sigma$

Assumptions:

(CTY): $t(x)$ is continuous

(CRC): $|t(x)|$ has totally bounded sublevel sets

Transport map noise, $\nu_\sigma = t\#\sigma$

Assumptions:

(CTY): $t(x)$ is continuous

(CRC): $|t(x)|$ has totally bounded sublevel sets

(INJ): $t(x)$ is injective

Transport map noise, $\nu_\sigma = t\#\sigma$

Assumptions:

(CTY): $t(x)$ is continuous

(CRC): $|t(x)|$ has totally bounded sublevel sets

(INJ): $t(x)$ is injective

Thm [C., Faktor, Nachman '26]:

Transport map noise, $\nu_\sigma = t\#\sigma$

Assumptions:

(CTY): $t(x)$ is continuous

(CRC): $|t(x)|$ has totally bounded sublevel sets

(INJ): $t(x)$ is injective

Thm [C., Faktor, Nachman '26]:

- If (CTY) and (CRC) hold, a solution exists.

Transport map noise, $\nu_\sigma = t\#\sigma$

Assumptions:

(CTY): $t(x)$ is continuous

(CRC): $|t(x)|$ has totally bounded sublevel sets

(INJ): $t(x)$ is injective

Thm [C., Faktor, Nachman '26]:

- If (CTY) and (CRC) hold, a solution exists.
- If $p > 1$, $\nu \ll \mathscr{L}^d$, and (INJ), then solutions are unique if and only if the problem is feasible.

Transport map noise, $\nu_\sigma = t\#\sigma$

Assumptions:

(CTY): $t(x)$ is continuous

(CRC): $|t(x)|$ has totally bounded sublevel sets

(INJ): $t(x)$ is injective

Thm [C., Faktor, Nachman '26]:

- If (CTY) and (CRC) hold, a solution exists.
- If $p > 1$, $\nu \ll \mathscr{L}^d$, and (INJ), then solutions are unique if and only if the problem is feasible.

Most of the hypotheses are **necessary:**

Transport map noise, $\nu_\sigma = t\#\sigma$

Assumptions:

(CTY): $t(x)$ is continuous

(CRC): $|t(x)|$ has totally bounded sublevel sets

(INJ): $t(x)$ is injective

Thm [C., Faktor, Nachman '26]:

- If (CTY) and (CRC) hold, a solution exists.
- If $p > 1$, $\nu \ll \mathscr{L}^d$, and (INJ), then solutions are unique if and only if the problem is feasible.

Most of the hypotheses are **necessary:**

$\exists$	(CTY)	$\neg$(CTY)
(CRC)	Yes	No
$\neg$(CRC)	No	

!	$p > 1$	$p = 1$
$\nu \ll \mathcal{L}^d$	Yes	Yes $(d = 1)$
$\neg(\nu \ll \mathcal{L}^d)$	No	

Plan

1. Motivation
2. Previous Work
3. Well-posedness vs Ill-posedness
4. Numerical Method
5. Numerical Results

Numerical Regularization

- Measurement $\nu \in \mathscr{P}(Y)$
- Noise model $\rho : X \to \mathscr{P}(Y)$

Goal: Find data $\sigma \in \mathscr{P}(X)$ so that the noisy data $\nu_\sigma \in \mathscr{P}(Y)$ satisfies

$$\sigma_* = \operatorname{argmin}_\sigma \; W_p^p(\nu_\sigma, \nu).$$

Numerical Regularization

- Measurement $\nu \in \mathscr{P}(Y)$
- Noise model $\rho : X \to \mathscr{P}(Y)$

Goal: Find data $\sigma \in \mathscr{P}(X)$ so that the noisy data $\nu_\sigma \in \mathscr{P}(Y)$ satisfies

$$\sigma_* = \operatorname{argmin}_\sigma W_{p,\epsilon}^p(\nu_\sigma, \nu)$$

Numerical Regularization

- Measurement $\nu \in \mathscr{P}(Y)$
- Noise model $\rho : X \to \mathscr{P}(Y)$

Goal: Find data $\sigma \in \mathscr{P}(X)$ so that the noisy data $\nu_\sigma \in \mathscr{P}(Y)$ satisfies

$$\sigma_* = \text{argmin}_\sigma \; W_{p,\epsilon}^p(\nu_\sigma, \nu) + \epsilon \text{KL}(\nu_\sigma \,|\, \hat{m}) + \epsilon KL(\sigma \,|\, \hat{\sigma})$$

Numerical Regularization

- Measurement $\nu \in \mathscr{P}(Y)$
- Noise model $\rho : X \to \mathscr{P}(Y)$

Goal: Find data $\sigma \in \mathscr{P}(X)$ so that the noisy data $\nu_\sigma \in \mathscr{P}(Y)$ satisfies

$$\sigma_* = \operatorname{argmin}_\sigma W_{p,\epsilon}^p(\nu_\sigma, \nu) + \epsilon \mathrm{KL}(\nu_\sigma \,|\, \hat{m}) + \epsilon KL(\sigma \,|\, \hat{\sigma})$$

With this regularization, the problem becomes an entropy regularized unbalanced transport problem [Chizat, Peyré, Schmitzer, Vialard '18].

- **Computational benefits:** provably convergent algorithms for approximating minimizer via Bregman projections.
- **Analytical benefits:** Many noise maps are badly unstable, i.e., $\nu_\sigma \approx^{W_2} \nu_{\sigma'} \not\Rightarrow \sigma \approx \sigma'$, and entropic regularization selects a low entropy minimizer.

Numerical Discretization

- Measurement $\nu \in \mathscr{P}(Y)$
- Noise model $\rho : X \to \mathscr{P}(Y)$

Goal: Find data $\sigma \in \mathscr{P}(X)$ so that the noisy data $\nu_\sigma \in \mathscr{P}(Y)$ satisfies

$$\sigma_* = \operatorname{argmin}_\sigma W^p_{p,\epsilon}(\nu_\sigma, \nu) + \epsilon \mathrm{KL}(\nu_\sigma | \hat{m}) + \epsilon KL(\sigma | \hat{\sigma})$$

Numerical Discretization

- Measurement $\nu \in \mathscr{P}(Y) \implies \Sigma_j \delta_{y_j} \nu_j$
- Noise model $\rho : X \to \mathscr{P}(Y)$

Goal: Find data $\sigma \in \mathscr{P}(X)$ so that the noisy data $\nu_\sigma \in \mathscr{P}(Y)$ satisfies

$$\sigma_* = \operatorname{argmin}_\sigma W^p_{p,\epsilon}(\nu_\sigma, \nu) + \epsilon \mathrm{KL}(\nu_\sigma | \hat{m}) + \epsilon KL(\sigma | \hat{\sigma})$$

Numerical Discretization

- Measurement $\nu \in \mathscr{P}(Y) \implies \Sigma_j \delta_{y_j} \boldsymbol{\nu}_j$
- Noise model $\rho : X \to \mathscr{P}(Y) \implies x \mapsto \Sigma_i \delta_{y'_i} (\boldsymbol{\rho}_x)_i$

Goal: Find data $\sigma \in \mathscr{P}(X)$ so that the noisy data $\nu_\sigma \in \mathscr{P}(Y)$ satisfies

$$\sigma_* = \operatorname{argmin}_\sigma W^p_{p,\epsilon}(\nu_\sigma, \nu) + \epsilon \mathrm{KL}(\nu_\sigma \,|\, \hat{m}) + \epsilon KL(\sigma \,|\, \hat{\sigma})$$

Numerical Discretization

- Measurement $\nu \in \mathscr{P}(Y) \implies \Sigma_j \delta_{y_j} \boldsymbol{\nu}_j$
- Noise model $\rho : X \to \mathscr{P}(Y) \implies x \mapsto \Sigma_i \delta_{y'_i} (\boldsymbol{\rho}_x)_i$

Goal: Find data $\sigma \in \mathscr{P}(X)$ so that the noisy data $\nu_\sigma \in \mathscr{P}(Y)$ satisfies

$$\sigma_* = \mathrm{argmin}_\sigma \; W_{p,\epsilon}^p(\nu_\sigma, \nu) + \epsilon \mathrm{KL}(\nu_\sigma \,|\, \hat{m}) + \epsilon KL(\sigma \,|\, \hat{\sigma})$$

$$\hat{\sigma} = \frac{1}{L} \sum_{k=1}^{L} \delta_{x_k} \quad \text{and} \quad \hat{m} = \frac{1}{m} \sum_{i=1}^{m} \delta_{y_i}.$$

Numerical Discretization

- Measurement $\nu \in \mathscr{P}(Y) \implies \Sigma_j \delta_{y_j} \boldsymbol{\nu}_j$
- Noise model $\rho : X \to \mathscr{P}(Y) \implies x \mapsto \Sigma_i \delta_{y'_i} (\boldsymbol{\rho}_x)_i$

Goal: Find data $\sigma \in \mathscr{P}(X)$ so that the noisy data $\nu_\sigma \in \mathscr{P}(Y)$ satisfies

$$\sigma_* = \text{argmin}_\sigma \; W^p_{p,\epsilon}(\nu_\sigma, \nu) + \epsilon \text{KL}(\nu_\sigma | \hat{m}) + \epsilon KL(\sigma | \hat{\sigma})$$

$$\hat{\sigma} = \frac{1}{L} \sum_{k=1}^{L} \delta_{x_k} \quad \text{and} \quad \hat{m} = \frac{1}{m} \sum_{i=1}^{m} \delta_{y_i}.$$

This leads to the following convex optimization problem:

Numerical Discretization

- Measurement $\nu \in \mathscr{P}(Y) \implies \Sigma_j \delta_{y_j} \boldsymbol{\nu}_j$
- Noise model $\rho : X \to \mathscr{P}(Y) \implies x \mapsto \Sigma_i \delta_{y'_i} (\boldsymbol{\rho}_x)_i$

Goal: Find data $\sigma \in \mathscr{P}(X)$ so that the noisy data $\nu_\sigma \in \mathscr{P}(Y)$ satisfies

$$\sigma_* = \mathrm{argmin}_\sigma \; W^p_{p,\epsilon}(\nu_\sigma, \nu) + \epsilon \mathrm{KL}(\nu_\sigma | \hat{m}) + \epsilon KL(\sigma | \hat{\sigma})$$

$$\hat{\sigma} = \frac{1}{L} \sum_{k=1}^{L} \delta_{x_k} \quad \text{and} \quad \hat{m} = \frac{1}{m} \sum_{i=1}^{m} \delta_{y_i}.$$

This leads to the following convex optimization problem:

$$\underset{\boldsymbol{\sigma}, \boldsymbol{\Gamma}}{\mathrm{argmin}} \; \Sigma_{i,j} |y_j - y'_i|^p \, d\boldsymbol{\Gamma}_{ij} + \epsilon \Sigma_k \boldsymbol{\sigma}_k \log(\boldsymbol{\sigma}_k) + \epsilon \Sigma_{ij} \boldsymbol{\Gamma}_{ij} \log(\boldsymbol{\Gamma}_{ij})$$

$$\Sigma_j \boldsymbol{\Gamma}_{ij} = (\boldsymbol{\nu}_{\boldsymbol{\sigma}})_i$$

$$\Sigma_i \boldsymbol{\Gamma}_{ij} = \boldsymbol{\nu}_j$$

Numerical Discretization

- Measurement $\nu \in \mathscr{P}(Y) \implies \Sigma_j \delta_{y_j} \boldsymbol{\nu}_j$
- Noise model $\rho : X \to \mathscr{P}(Y) \implies x \mapsto \Sigma_i \delta_{y'_i} (\boldsymbol{\rho}_x)_i$

Goal: Find data $\sigma \in \mathscr{P}(X)$ so that the noisy data $\nu_\sigma \in \mathscr{P}(Y)$ satisfies

$$\sigma_* = \operatorname{argmin}_\sigma W^p_{p,\epsilon}(\nu_\sigma, \nu) + \epsilon \mathrm{KL}(\nu_\sigma \,|\, \hat{m}) + \epsilon KL(\sigma \,|\, \hat{\sigma})$$

$$\hat{\sigma} = \frac{1}{L} \sum_{k=1}^{L} \delta_{x_k} \quad \text{and} \quad \hat{m} = \frac{1}{m} \sum_{i=1}^{m} \delta_{y_i}.$$

This leads to the following convex optimization problem:

$$\underset{\boldsymbol{\sigma}, \boldsymbol{\Gamma}}{\operatorname{argmin}} \; \Sigma_{i,j} |y_j - y'_i|^p \, d\boldsymbol{\Gamma}_{ij} + \epsilon \Sigma_k \boldsymbol{\sigma}_k \log(\boldsymbol{\sigma}_k) + \epsilon \Sigma_{ij} \boldsymbol{\Gamma}_{ij} \log(\boldsymbol{\Gamma}_{ij})$$

$$\Sigma_j \boldsymbol{\Gamma}_{ij} = (\boldsymbol{\nu}_{\boldsymbol{\sigma}})_i$$

$$\Sigma_i \boldsymbol{\Gamma}_{ij} = \boldsymbol{\nu}_j$$

Note: $(\boldsymbol{\nu}_{\boldsymbol{\sigma}})_i = \sum_{ij} \mathbf{R}_{ij} \boldsymbol{\sigma}_j = (\mathbf{R}\boldsymbol{\sigma})_i$, for $\mathbf{R}_{ij} = (\boldsymbol{\rho}_{x_j})_i$

Numerical Optimization

Numerical Optimization

Given the convex optimization problem,

Numerical Optimization

Given the convex optimization problem,

$$\underset{\boldsymbol{\sigma}, \boldsymbol{\Gamma}}{\operatorname{argmin}} \ \Sigma_{i,j} | y_j - y'_i |^p d\boldsymbol{\Gamma}_{ij} + \epsilon \Sigma_k \boldsymbol{\sigma}_k \log(\boldsymbol{\sigma}_k) + \epsilon \Sigma_{ij} \boldsymbol{\Gamma}_{ij} \log(\boldsymbol{\Gamma}_{ij})$$

$$\Sigma_j \boldsymbol{\Gamma}_{ij} = (\boldsymbol{R}\boldsymbol{\sigma})_i$$

$$\Sigma_i \boldsymbol{\Gamma}_{ij} = \boldsymbol{\nu}_j$$

Numerical Optimization

Given the convex optimization problem,

$$\underset{\boldsymbol{\sigma}, \boldsymbol{\Gamma}}{\operatorname{argmin}} \ \Sigma_{i,j} | y_j - y'_i |^p d\boldsymbol{\Gamma}_{ij} + \epsilon \Sigma_k \boldsymbol{\sigma}_k \log(\boldsymbol{\sigma}_k) + \epsilon \Sigma_{ij} \boldsymbol{\Gamma}_{ij} \log(\boldsymbol{\Gamma}_{ij})$$

$$\Sigma_j \boldsymbol{\Gamma}_{ij} = (\boldsymbol{R}\boldsymbol{\sigma})_i$$

$$\Sigma_i \boldsymbol{\Gamma}_{ij} = \boldsymbol{\nu}_j$$

under the change of variables,

Numerical Optimization

Given the convex optimization problem,

$$\underset{\boldsymbol{\sigma},\boldsymbol{\Gamma}}{\text{argmin}}\ \Sigma_{i,j}|y_j - y'_i|^p d\boldsymbol{\Gamma}_{ij} + \epsilon\Sigma_k \boldsymbol{\sigma}_k \log(\boldsymbol{\sigma}_k) + \epsilon\Sigma_{ij}\boldsymbol{\Gamma}_{ij}\log(\boldsymbol{\Gamma}_{ij})$$

$$\Sigma_j \boldsymbol{\Gamma}_{ij} = (\boldsymbol{R}\boldsymbol{\sigma})_i$$

$$\Sigma_i \boldsymbol{\Gamma}_{ij} = \boldsymbol{\nu}_j$$

under the change of variables,

$$\tilde{\mathbf{C}} = \begin{bmatrix} \mathbf{C} & +\infty_{(m\times 1)} \\ +\infty_{(L\times n)} & \mathbf{0}_{(L\times 1)} \end{bmatrix}, \quad \mathbf{P} = \begin{bmatrix} \boldsymbol{\Gamma} & \mathbf{0}_{(m\times 1)} \\ \mathbf{0}_{(L\times n)} & \boldsymbol{\sigma} \end{bmatrix},$$

$$\tilde{\mathbf{K}} = e^{-\tilde{\mathbf{C}}/\varepsilon}, \quad \mathbf{B} = \left[\mathbf{Id}_m; -\mathbf{R}\right], \quad \mathbf{b} = [\boldsymbol{\nu};\ 1],$$

Numerical Optimization

Given the convex optimization problem,

$$\underset{\boldsymbol{\sigma},\boldsymbol{\Gamma}}{\operatorname{argmin}}\ \Sigma_{i,j}|y_j - y'_i|^p d\boldsymbol{\Gamma}_{ij} + \epsilon\Sigma_k\boldsymbol{\sigma}_k\log(\boldsymbol{\sigma}_k) + \epsilon\Sigma_{ij}\boldsymbol{\Gamma}_{ij}\log(\boldsymbol{\Gamma}_{ij})$$

$$\Sigma_j\boldsymbol{\Gamma}_{ij} = (\boldsymbol{R\sigma})_i$$

$$\Sigma_i\boldsymbol{\Gamma}_{ij} = \boldsymbol{\nu}_j$$

under the change of variables,

$$\tilde{\mathbf{C}} = \begin{bmatrix} \mathbf{C} & +\infty_{(m\times 1)} \\ +\infty_{(L\times n)} & \mathbf{0}_{(L\times 1)} \end{bmatrix}, \quad \mathbf{P} = \begin{bmatrix} \boldsymbol{\Gamma} & \mathbf{0}_{(m\times 1)} \\ \mathbf{0}_{(L\times n)} & \boldsymbol{\sigma} \end{bmatrix},$$

$$\tilde{\mathbf{K}} = e^{-\tilde{\mathbf{C}}/\varepsilon}, \quad \mathbf{B} = \left[\mathbf{Id}_m; -\mathbf{R}\right], \quad \mathbf{b} = [\boldsymbol{\nu};\ 1],$$

the problem becomes:

Numerical Optimization

Given the convex optimization problem,

$$\underset{\boldsymbol{\sigma},\boldsymbol{\Gamma}}{\text{argmin}}\ \Sigma_{i,j}|y_j - y'_i|^p\, d\boldsymbol{\Gamma}_{ij} + \epsilon\Sigma_k \boldsymbol{\sigma}_k \log(\boldsymbol{\sigma}_k) + \epsilon\Sigma_{ij}\boldsymbol{\Gamma}_{ij}\log(\boldsymbol{\Gamma}_{ij})$$

$$\Sigma_j \boldsymbol{\Gamma}_{ij} = (\boldsymbol{R\sigma})_i$$

$$\Sigma_i \boldsymbol{\Gamma}_{ij} = \boldsymbol{\nu}_j$$

under the change of variables,

$$\tilde{\mathbf{C}} = \begin{bmatrix} \mathbf{C} & +\infty_{(m\times 1)} \\ +\infty_{(L\times n)} & \mathbf{0}_{(L\times 1)} \end{bmatrix}, \quad \mathbf{P} = \begin{bmatrix} \boldsymbol{\Gamma} & \mathbf{0}_{(m\times 1)} \\ \mathbf{0}_{(L\times n)} & \boldsymbol{\sigma} \end{bmatrix},$$

$$\tilde{\mathbf{K}} = e^{-\tilde{\mathbf{C}}/\varepsilon}, \quad \mathbf{B} = \left[\mathbf{Id}_m; -\mathbf{R}\right], \quad \mathbf{b} = [\boldsymbol{\nu};\ 1],$$

the problem becomes:

$$\underset{\mathbf{P}\in\mathcal{A}\cap\mathcal{B}}{\arg\min}\ \varepsilon\mathbf{KL}(\mathbf{P}|\tilde{\mathbf{K}})$$

$$\mathcal{A} = \{\mathbf{P} : \mathbf{P}\mathbf{1}_{n+1} \in \ker \mathbf{B}\}, \quad \mathcal{B} = \{\mathbf{P} : \mathbf{P}^T\mathbf{1}_{m+L} = \mathbf{b}\}$$

Bregman Projections

Bregman Projections

Optimizers of problems of the form

Bregman Projections

Optimizers of problems of the form

$$\underset{\mathbf{P}\in\mathcal{A}\cap\mathcal{B}}{\arg\min}\, \varepsilon \mathbf{KL}(\mathbf{P}|\tilde{\mathbf{K}})$$

$$\mathcal{A} = \{\mathbf{P} : \mathbf{P}\mathbf{1}_{n+1} \in \ker \mathbf{B}\}, \quad \mathcal{B} = \{\mathbf{P} : \mathbf{P}^T\mathbf{1}_{m+L} = \mathbf{b}\}$$

Bregman Projections

Optimizers of problems of the form

$$\arg\min_{\mathbf{P}\in\mathcal{A}\cap\mathcal{B}} \varepsilon \mathbf{KL}(\mathbf{P}|\tilde{\mathbf{K}})$$

$$\mathcal{A} = \{\mathbf{P} : \mathbf{P}\mathbf{1}_{n+1} \in \ker \mathbf{B}\}, \quad \mathcal{B} = \{\mathbf{P} : \mathbf{P}^T \mathbf{1}_{m+L} = \mathbf{b}\}$$

can be approximated via Bregman iterations,

Bregman Projections

Optimizers of problems of the form

$$\underset{\mathbf{P}\in\mathcal{A}\cap\mathcal{B}}{\arg\min}\, \varepsilon\mathbf{KL}(\mathbf{P}|\tilde{\mathbf{K}})$$

$$\mathcal{A} = \{\mathbf{P} : \mathbf{P}\mathbf{1}_{n+1} \in \ker \mathbf{B}\}, \quad \mathcal{B} = \{\mathbf{P} : \mathbf{P}^T\mathbf{1}_{m+L} = \mathbf{b}\}$$

can be approximated via Bregman iterations,

$$\mathbf{P}^{2l+1} = \mathrm{Proj}^{\mathbf{KL}}_{\mathcal{A}}(\mathbf{P}^{2l}), \quad \mathbf{P}^{2l+2} = \mathrm{Proj}^{\mathbf{KL}}_{\mathcal{B}}(\mathbf{P}^{2l+1})$$

Bregman Projections

Optimizers of problems of the form

$$\underset{\mathbf{P}\in\mathcal{A}\cap\mathcal{B}}{\arg\min}\, \varepsilon\mathbf{KL}(\mathbf{P}|\tilde{\mathbf{K}})$$

$$\mathcal{A} = \{\mathbf{P} : \mathbf{P}\mathbf{1}_{n+1} \in \ker \mathbf{B}\}, \quad \mathcal{B} = \{\mathbf{P} : \mathbf{P}^T\mathbf{1}_{m+L} = \mathbf{b}\}$$

can be approximated via Bregman iterations,

$$\mathbf{P}^{2l+1} = \mathrm{Proj}_{\mathcal{A}}^{\mathbf{KL}}(\mathbf{P}^{2l}), \quad \mathbf{P}^{2l+2} = \mathrm{Proj}_{\mathcal{B}}^{\mathbf{KL}}(\mathbf{P}^{2l+1})$$

where $\mathrm{Proj}_{\mathcal{C}}^{\mathbf{KL}}(\mathbf{P}) = \arg\min_{\mathbf{P}'\in\mathcal{C}} \mathbf{KL}(\mathbf{P}'|\mathbf{P})$.

Bregman Projections

Optimizers of problems of the form

$$\arg\min_{\mathbf{P}\in\mathcal{A}\cap\mathcal{B}} \varepsilon\mathbf{KL}(\mathbf{P}|\tilde{\mathbf{K}})$$

$$\mathcal{A} = \{\mathbf{P} : \mathbf{P}\mathbf{1}_{n+1} \in \ker\mathbf{B}\}, \quad \mathcal{B} = \{\mathbf{P} : \mathbf{P}^T\mathbf{1}_{m+L} = \mathbf{b}\}$$

can be approximated via Bregman iterations,

$$\mathbf{P}^{2l+1} = \mathrm{Proj}_{\mathcal{A}}^{\mathbf{KL}}(\mathbf{P}^{2l}), \quad \mathbf{P}^{2l+2} = \mathrm{Proj}_{\mathcal{B}}^{\mathbf{KL}}(\mathbf{P}^{2l+1})$$

where $\mathrm{Proj}_{\mathcal{C}}^{\mathbf{KL}}(\mathbf{P}) = \arg\min_{\mathbf{P}'\in\mathcal{C}} \mathbf{KL}(\mathbf{P}'|\mathbf{P})$.

Remark: For all $\epsilon > 0$, the Bregman iterations converge to optimum with rate $\mathcal{O}(1/l)$ in dual variables, c.f. [Peyré, 2026].

Bregman Projections

Optimizers of problems of the form

$$\underset{\mathbf{P}\in\mathcal{A}\cap\mathcal{B}}{\arg\min}\, \varepsilon \mathbf{KL}(\mathbf{P}|\tilde{\mathbf{K}})$$

$$\mathcal{A} = \{\mathbf{P} : \mathbf{P}\mathbf{1}_{n+1} \in \ker \mathbf{B}\}, \quad \mathcal{B} = \{\mathbf{P} : \mathbf{P}^T\mathbf{1}_{m+L} = \mathbf{b}\}$$

can be approximated via Bregman iterations,

$$\mathbf{P}^{2l+1} = \mathrm{Proj}_{\mathcal{A}}^{\mathbf{KL}}(\mathbf{P}^{2l}), \quad \mathbf{P}^{2l+2} = \mathrm{Proj}_{\mathcal{B}}^{\mathbf{KL}}(\mathbf{P}^{2l+1})$$

where $\mathrm{Proj}_{\mathcal{C}}^{\mathbf{KL}}(\mathbf{P}) = \arg\min_{\mathbf{P}'\in\mathcal{C}} \mathbf{KL}(\mathbf{P}'|\mathbf{P})$.

Remark: For all $\epsilon > 0$, the Bregman iterations converge to optimum with rate $\mathcal{O}(1/l)$ in dual variables, c.f. [Peyré, 2026].

Lacking a closed form expression for $\mathrm{Proj}_{\mathcal{A}}^{\mathbf{KL}}$, we approximate via Douglas-Rachford splitting algorithm.

Bregman Projections

Optimizers of problems of the form

$$\underset{\mathbf{P}\in\mathcal{A}\cap\mathcal{B}}{\arg\min}\, \varepsilon \mathbf{KL}(\mathbf{P}|\tilde{\mathbf{K}})$$

$$\mathcal{A} = \{\mathbf{P} : \mathbf{P}\mathbf{1}_{n+1} \in \ker \mathbf{B}\}, \quad \mathcal{B} =$$

Computational complexity scales with the size of the matrices, not the ambient dimension.

can be approximated via Bregman iterations

$$\mathbf{P}^{2l+1} = \mathrm{Proj}_{\mathcal{A}}^{\mathbf{KL}}(\mathbf{P}^{2l}), \quad \mathbf{P}^{2l+2} = \mathrm{Proj}_{\mathcal{B}}^{\mathbf{KL}}(\mathbf{P}^{2l+1})$$

where $\mathrm{Proj}_{\mathcal{C}}^{\mathbf{KL}}(\mathbf{P}) = \arg\min_{\mathbf{P}'\in\mathcal{C}} \mathbf{KL}(\mathbf{P}'|\mathbf{P})$.

Remark: For all $\epsilon > 0$, the Bregman iterations converge to optimum with rate $\mathcal{O}(1/l)$ in dual variables, c.f. [Peyré, 2026].

Lacking a closed form expression for $\mathrm{Proj}_{\mathcal{A}}^{\mathbf{KL}}$, we approximate via Douglas-Rachford splitting algorithm.

Plan

Plan

1. Motivation

2. Previous Work

3. Well-posedness vs Ill-posedness

4. Numerical Method

5. Numerical Results

Numerics: 1-D model problem

Numerics: 1-D model problem

Continuum:

Noise model: $\tilde{\rho}_x = \mathcal{N}(t(x), \beta) d\mathcal{L}, \; t = \mathrm{id} + \frac{1}{2}\mathrm{sgn}$

True data: $\tilde{\sigma}_{\mathrm{true}} = \sum_{i=1}^{3} w_i \mathcal{N}(c_i, v_i)(x) d\mathcal{L}(x) \text{ for } c_i \in \mathbb{R}, v_i > 0.$

Measurement: $\tilde{\nu} = (\mathcal{N}(0, \beta) d\mathcal{L}) * (t\#\sigma_{\mathrm{true}})$

Numerics: 1-D model problem

Continuum:

Noise model: $\tilde{\rho}_x = \mathcal{N}(t(x), \beta) d\mathcal{L}, \; t = \mathrm{id} + \frac{1}{2}\mathrm{sgn}$

True data: $\tilde{\sigma}_{\mathrm{true}} = \sum_{i=1}^{3} w_i \mathcal{N}(c_i, v_i)(x) d\mathcal{L}(x) \text{ for } c_i \in \mathbb{R}, v_i > 0.$

Measurement: $\tilde{\nu} = (\mathcal{N}(0, \beta) d\mathcal{L}) * (t \# \sigma_{\mathrm{true}})$

Discrete:

Choose L' samples of $\tilde{\boldsymbol{\sigma}}_{\text{true}}$ to construct $\boldsymbol{\sigma}_{\text{true}}$

Combine with M' samples of $\tilde{\rho}_{x_k}$ to construct and $L'M'$ samples of $\tilde{\nu}$

Given $\hat{\sigma} = \frac{1}{L} \sum_{k=1}^{L} \delta_{x_k}$, choose M samples of $\tilde{\rho}_{x_k}$ to construct $\mathbf{R}$

Numerics: 1-D model problem

Continuum:

Noise model: $\tilde{\rho}_x = \mathcal{N}(t(x), \beta)d\mathcal{L}, \; t = \mathrm{id} + \frac{1}{2}\mathrm{sgn}$

True data: $\tilde{\sigma}_{\mathrm{true}} = \sum_{i=1}^{3} w_i \mathcal{N}(c_i, v_i)(x)d\mathcal{L}(x)$ for $c_i \in \mathbb{R}, v_i > 0.$

Measurement: $\tilde{\nu} = (\mathcal{N}(0, \beta)d\mathcal{L}) * (t\#\sigma_{\mathrm{true}})$

Discrete:

Choose L' samples of $\tilde{\boldsymbol{\sigma}}_{\mathrm{true}}$ to construct $\boldsymbol{\sigma}_{\mathrm{true}}$

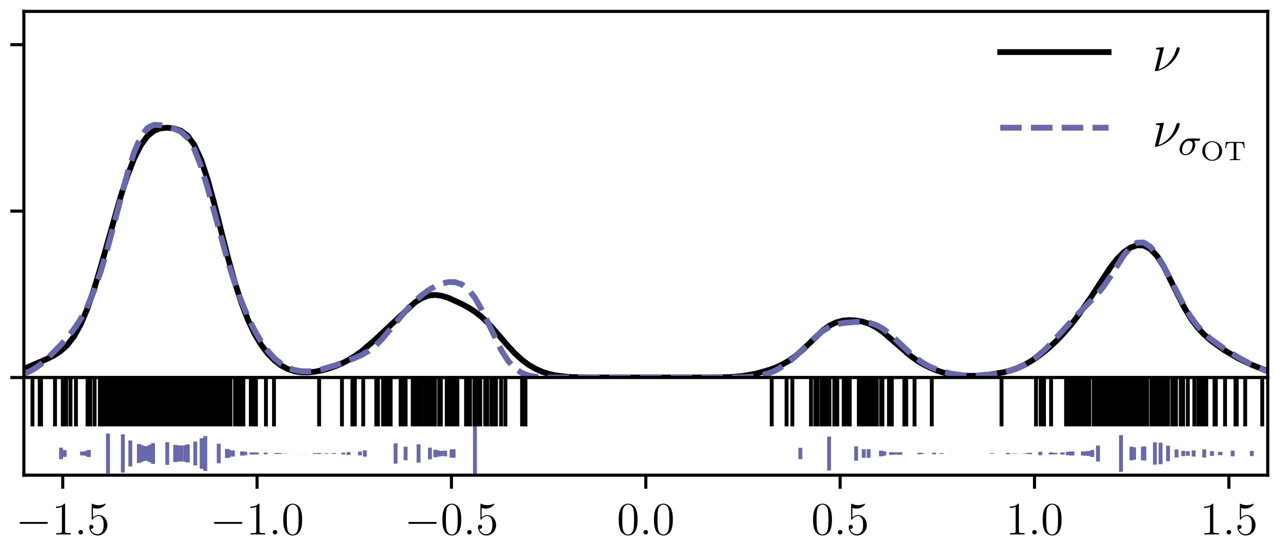
Combine with M' samples of $\tilde{\rho}_{x_k}$ to construct and $L'M'$ samples of $\tilde{\nu}$

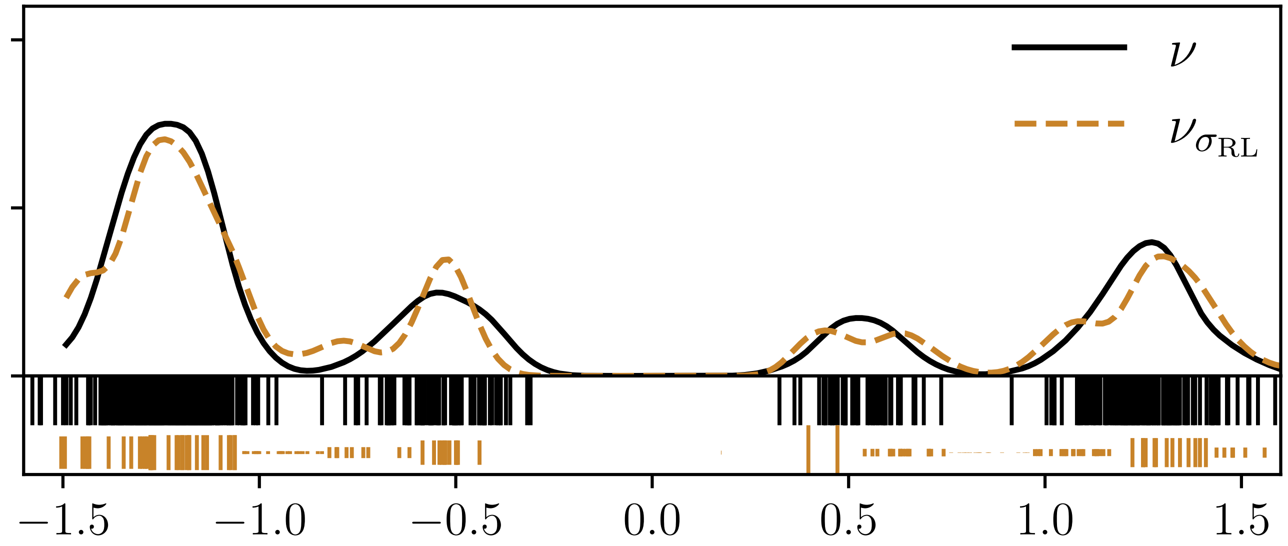
Given $\hat{\sigma} = \frac{1}{L}\sum_{k=1}^{L} \delta_{x_k}$, choose M samples of $\tilde{\rho}_{x_k}$ to construct $\mathbf{R}$

Binned approximations (with noise model perturbed to satisfy absolute continuity hypotheses) are used to compare with Richardson-Lucy.

Numerics: 1-D model problem

Measured data ν vs. unfolding approximations ν_{OT} and ν_{RL}

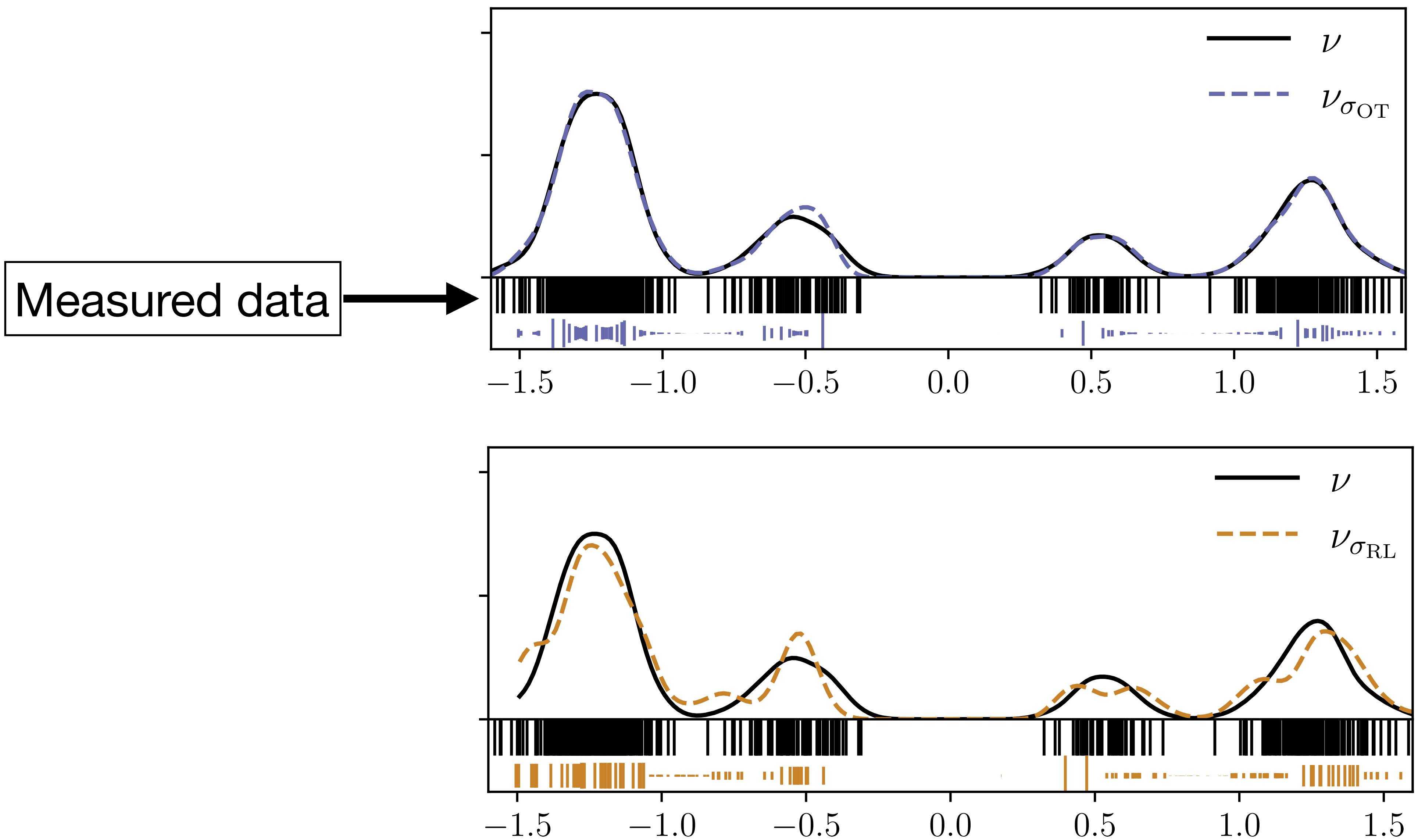




$$L = L' = 150,\ M = 1,\ M' = 3$$

Numerics: 1-D model problem

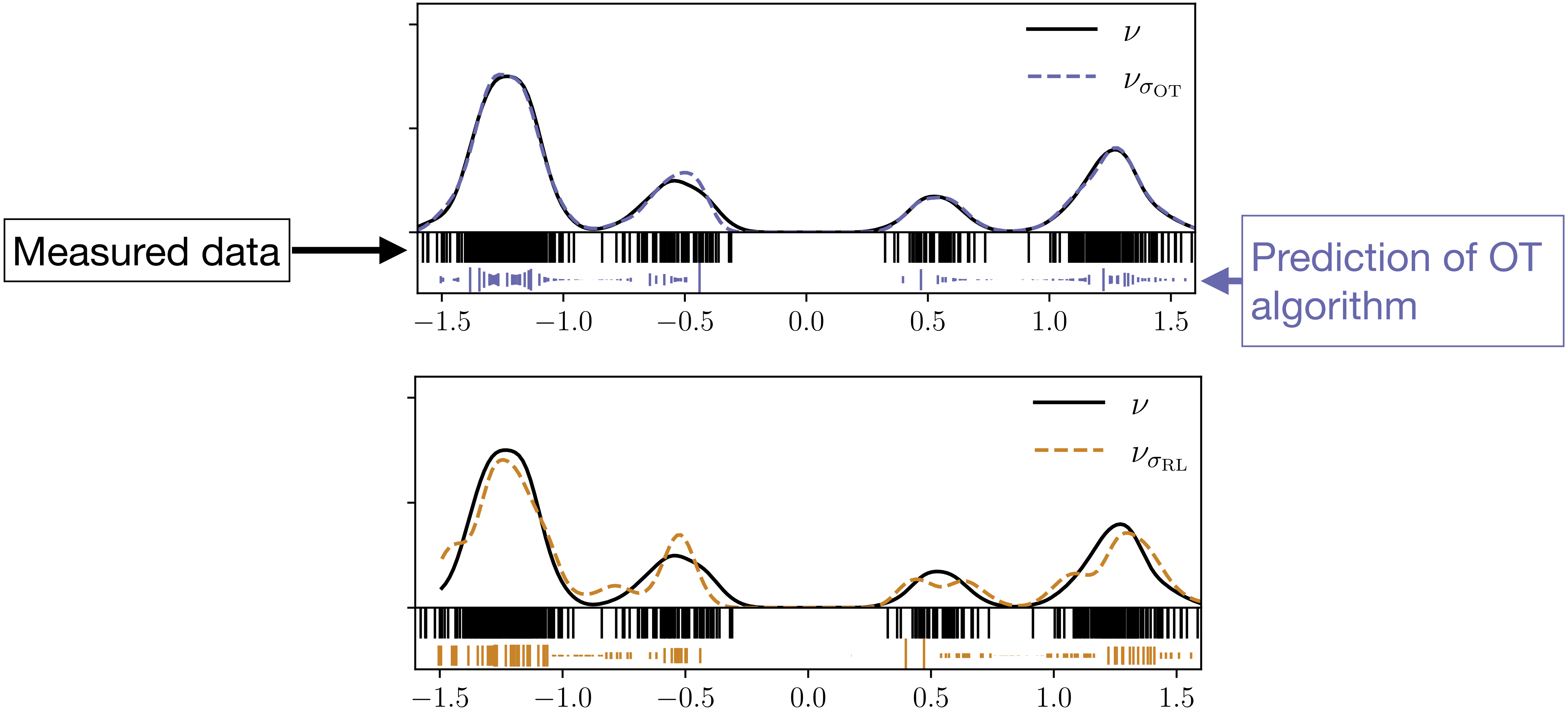
Measured data ν vs. unfolding approximations ν_{OT} and ν_{RL}



$$L = L' = 150,\ M = 1,\ M' = 3$$

Numerics: 1-D model problem

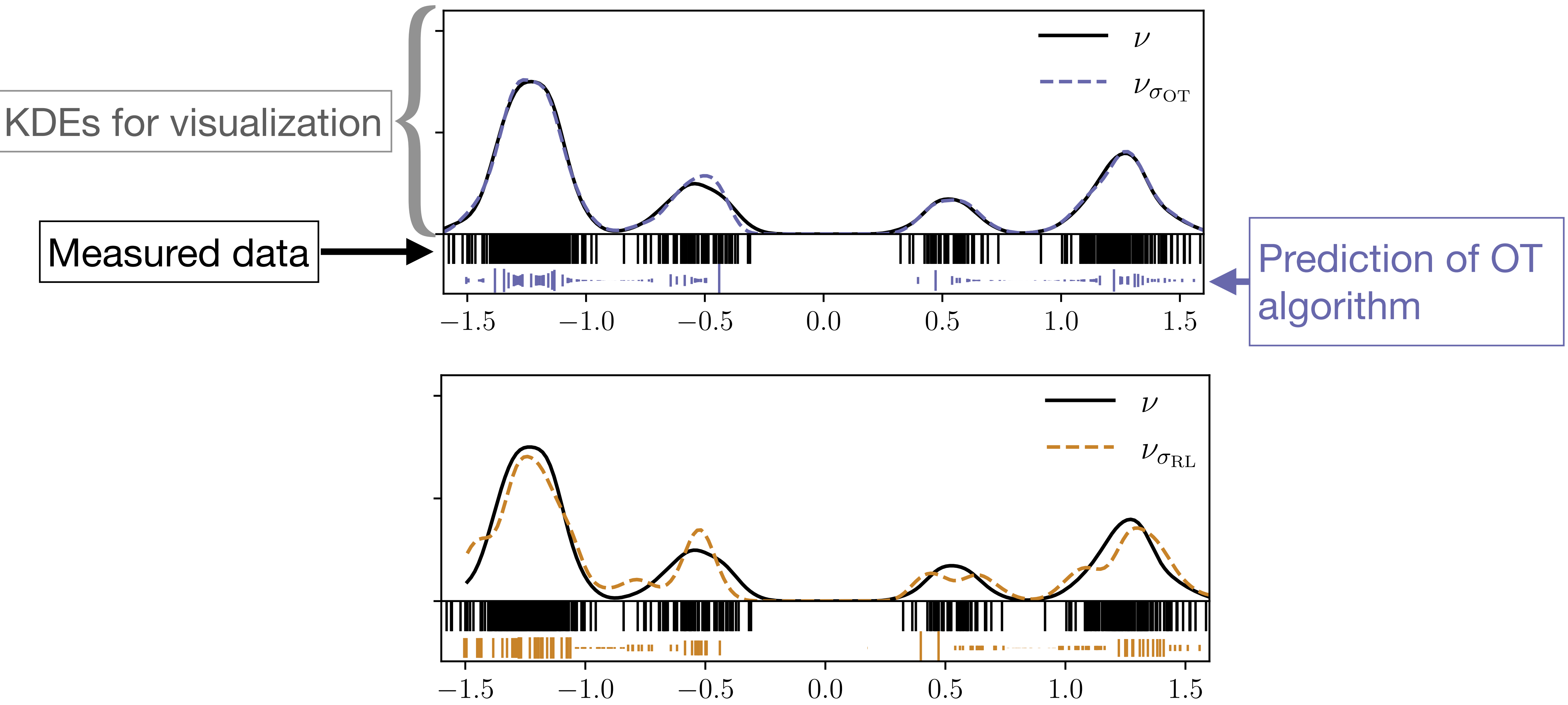
Measured data ν vs. unfolding approximations ν_{OT} and ν_{RL}



$$L = L' = 150,\ M = 1,\ M' = 3$$

Numerics: 1-D model problem

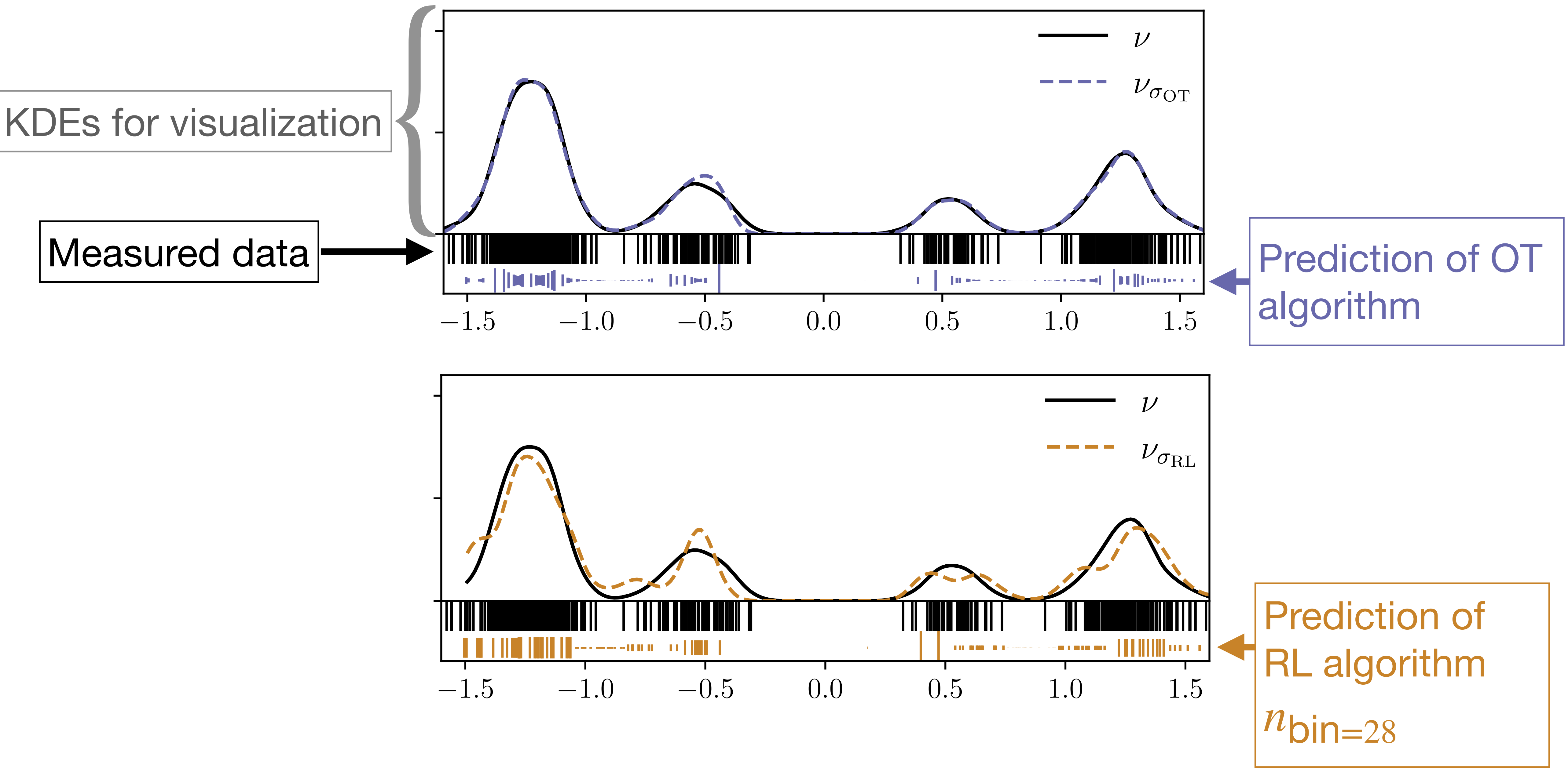
Measured data ν vs. unfolding approximations ν_{OT} and ν_{RL}



$$L = L' = 150,\ M = 1,\ M' = 3$$

Numerics: 1-D model problem

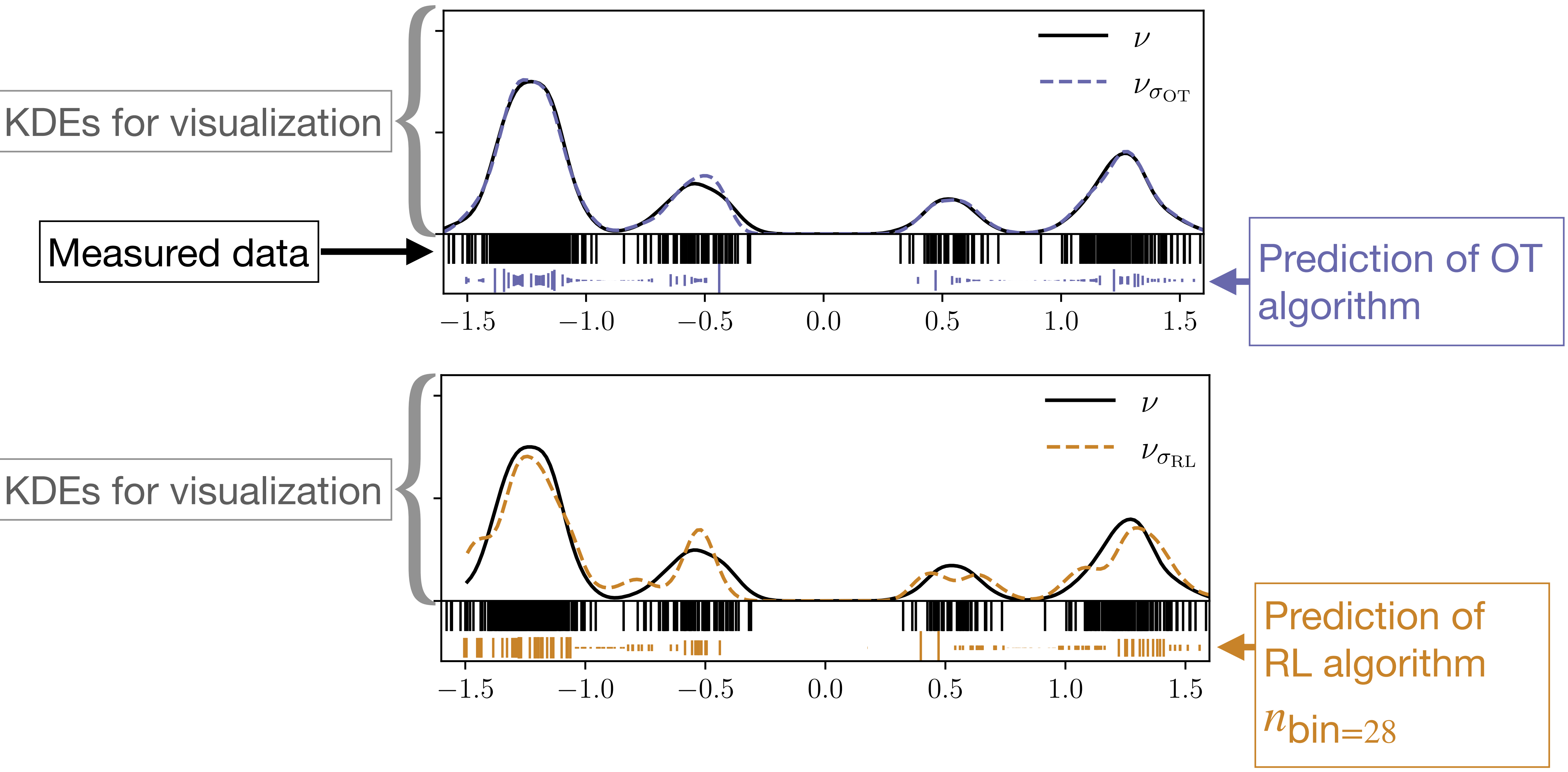
Measured data ν vs. unfolding approximations ν_{OT} and ν_{RL}



$$L = L' = 150,\ M = 1,\ M' = 3$$

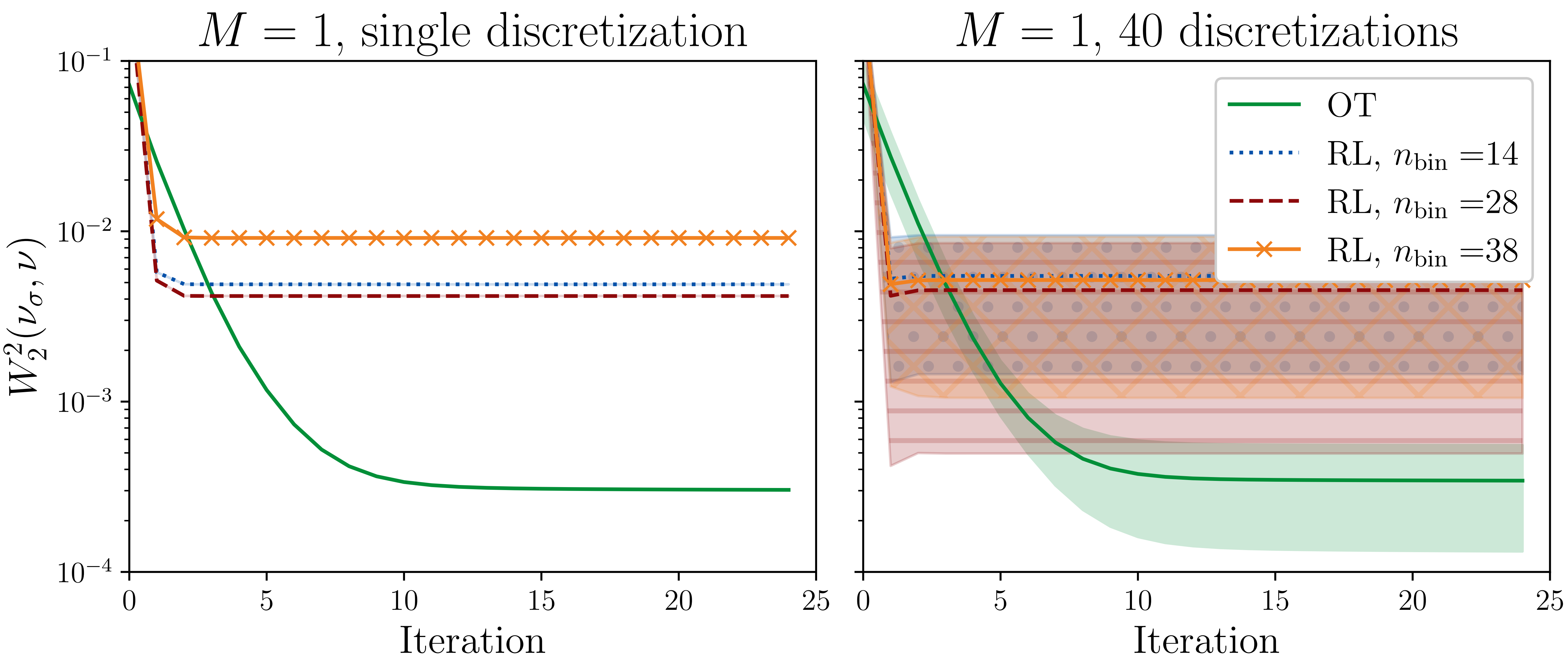
Numerics: 1-D model problem

Measured data ν vs. unfolding approximations ν_{OT} and ν_{RL}

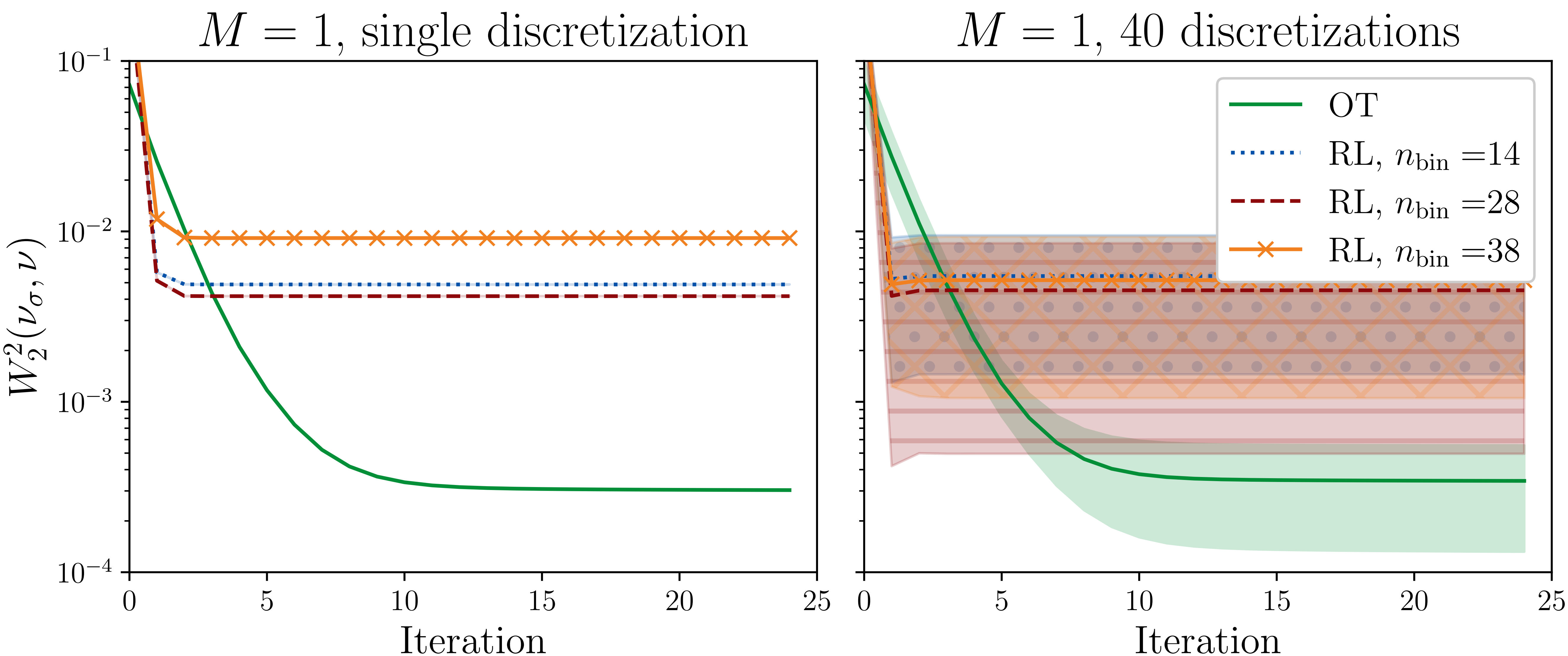


$$L = L' = 150,\ M = 1,\ M' = 3$$

Numerics: 1-D model problem

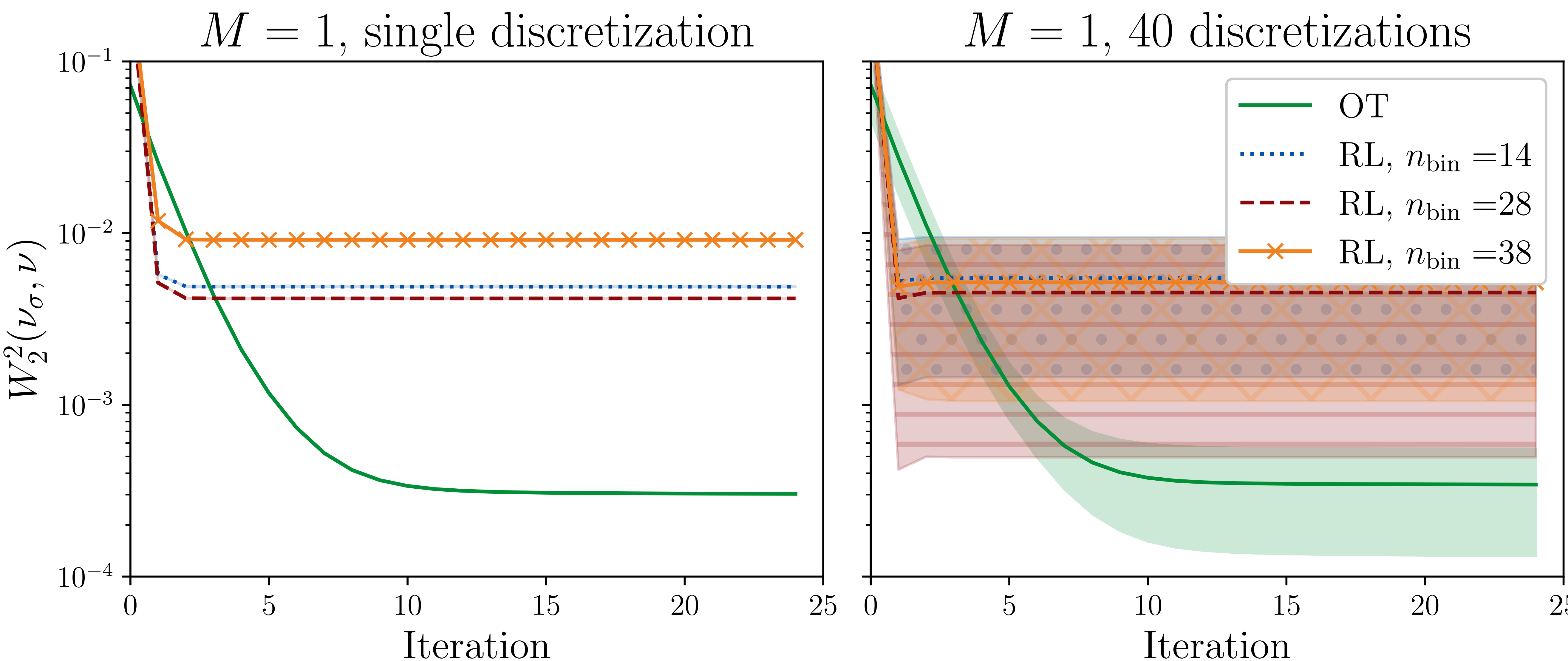


Numerics: 1-D model problem



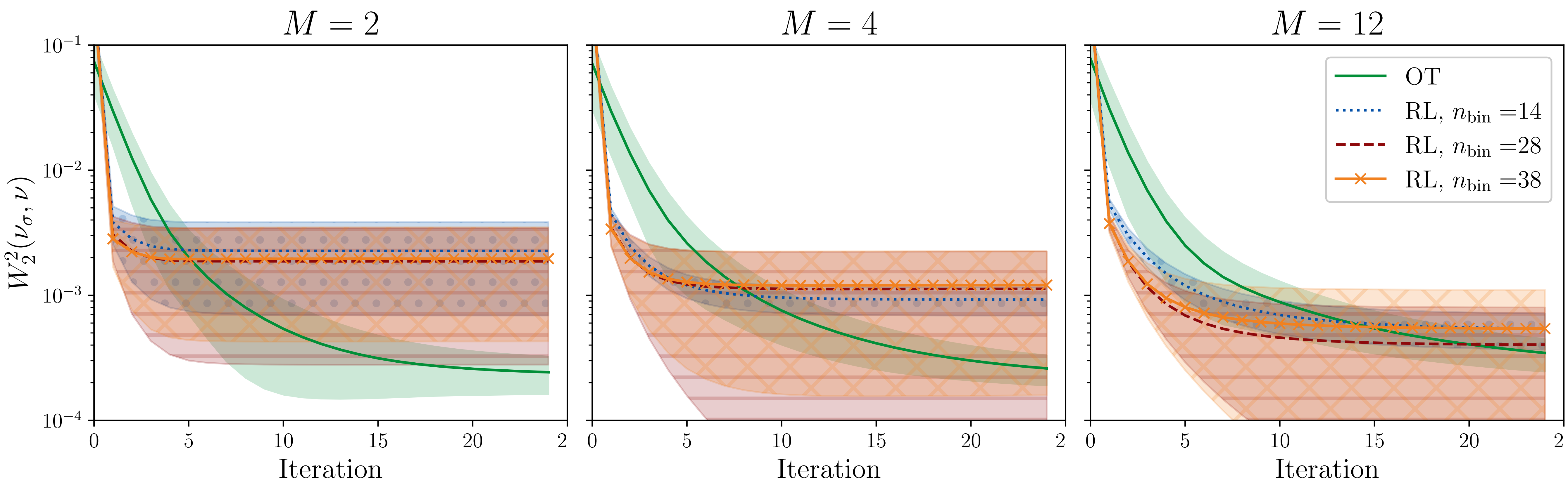
- RL performance can deteriorate with too many bins

Numerics: 1-D model problem



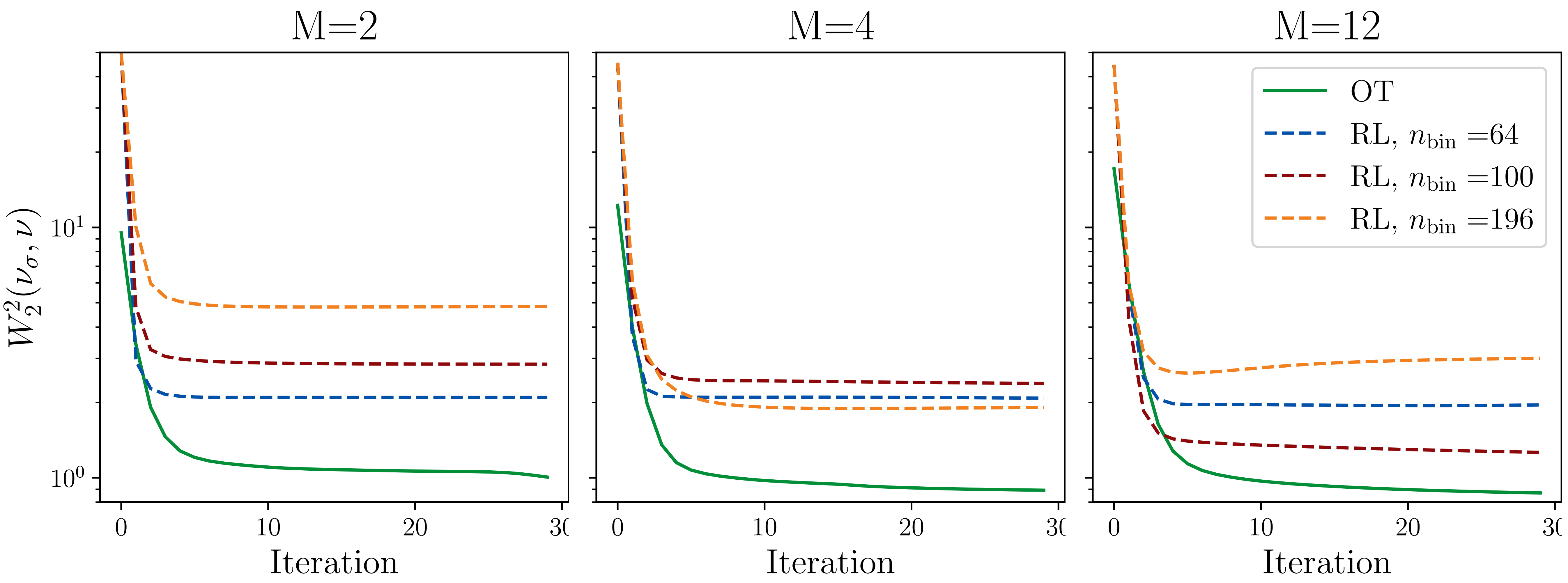
- RL performance can deteriorate with too many bins
- OT algorithm takes longer to converge than RL; achieves higher/comparable accuracy

Numerics: 1-D model problem



Relative benefit of OT appears highest for course discretization of noise model used in optimization (OT better for smaller M).

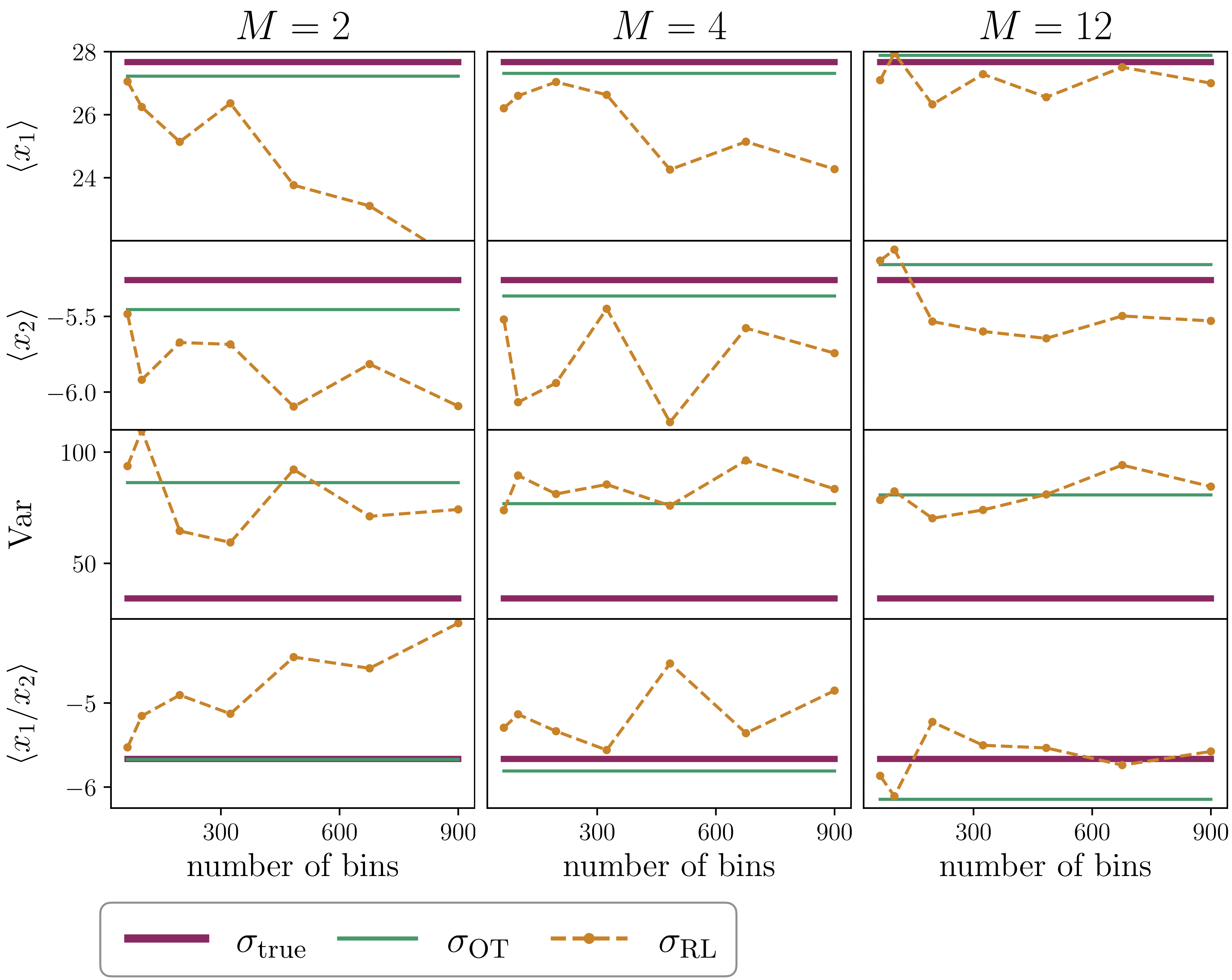
Numerics: 2-D jet mass problem



Performance on jet mass and groomed jet mass observables; synthetic data generated using Pythia, detector effects Delphes.

$$L = L' = 100 \text{ and } M' = 3$$

Numerics: 2-D jet mass problem



$$L = L' = 100 \text{ and } M' = 3$$

Future work

Future work

Denoising via Optimal Transport

Future work

Denoising via Optimal Transport

- Application to collider data

Future work

Denoising via Optimal Transport

- Application to collider data
- Improved algorithms (computation of $\mathrm{Proj}_{\mathcal{A}}^{\mathbf{KL}}$?)

Future work

Denoising via Optimal Transport

- Application to collider data
- Improved algorithms (computation of $\mathrm{Proj}_{\mathcal{A}}^{\mathbf{KL}}$?)
- Convergence and stability analysis

Future work

Denoising via Optimal Transport

- Application to collider data
- Improved algorithms (computation of $\mathrm{Proj}_{\mathcal{A}}^{\mathbf{KL}}$?)
- Convergence and stability analysis
  - Dependence on ϵ in rate of convergence of Bregman iterations?

Future work

Denoising via Optimal Transport

- Application to collider data
- Improved algorithms (computation of $\mathrm{Proj}_{\mathcal{A}}^{\mathbf{KL}}$?)
- Convergence and stability analysis
  - Dependence on ϵ in rate of convergence of Bregman iterations?
  - Stability of numerical optimizers with respect to perturbations/ samples of ν and $x \mapsto \rho_x$?

Future work

Denoising via Optimal Transport

- Application to collider data
- Improved algorithms (computation of $\mathrm{Proj}^{\mathbf{KL}}_{\mathcal{A}}$?)
- Convergence and stability analysis
  - Dependence on ϵ in rate of convergence of Bregman iterations?
  - Stability of numerical optimizers with respect to perturbations/ samples of ν and $x \mapsto \rho_x$?
  - Dynamic choice of $\hat{\sigma}$?

Future work

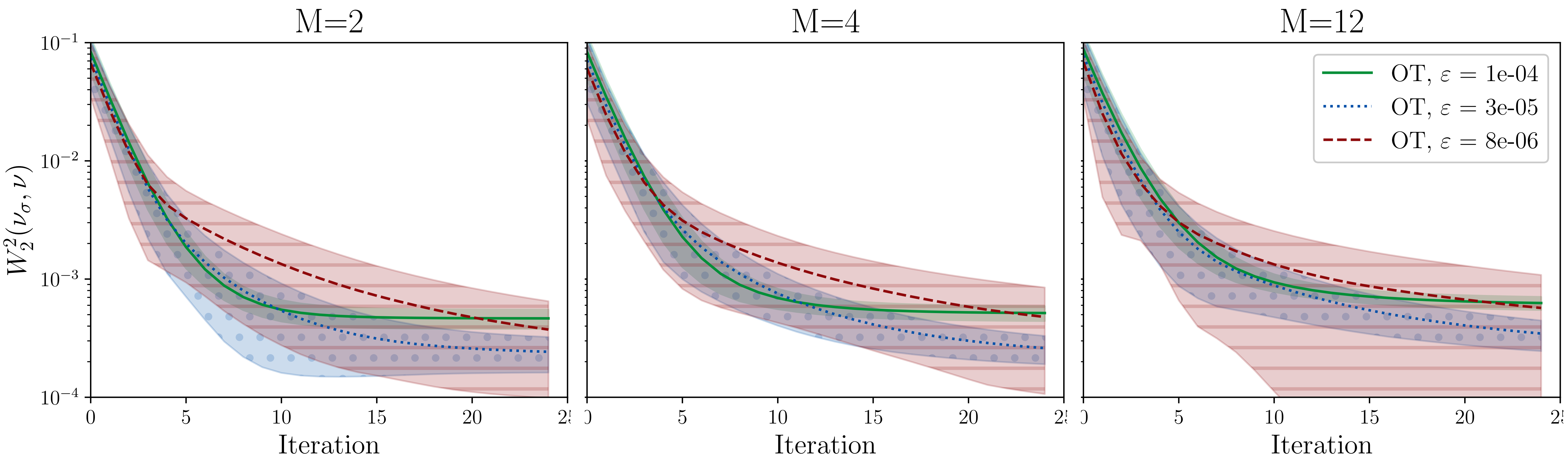
Denoising via Optimal Transport

- Application to collider data
- Improved algorithms (computation of $\mathrm{Proj}_{\mathcal{A}}^{\mathbf{KL}}$?)
- Convergence and stability analysis
  - Dependence on ϵ in rate of convergence of Bregman iterations?
  - Stability of numerical optimizers with respect to perturbations/ samples of ν and $x \mapsto \rho_x$?
  - Dynamic choice of $\hat{\sigma}$?

Thank you!

Backup

Backup



$$L = L' = 150 \text{ and } M' = 3$$