Constrained Denoising, Empirical Bayes and Optimal Transport

Bodhisattva Sen[1]
Columbia University, New York

STAMPS@CMU

May 14, 2026

[1]Supported by grants DMS-2311062 and DMS-2515520

Nicolás García Trillos	Adam Jaffe	Nikos Ignatiadis
U of Wisconsin	Columbia U	U Chicago

García Trillos, N. and **Sen, B.** (2024). A New Perspective On Denoising Based On Optimal Transport. *Information and Inference*, **13**.
https://arxiv.org/abs/2312.08135

Jaffe, A., Ignatiadis, N., and **Sen, B.** (2025+). Constrained Denoising, Empirical Bayes and Optimal Transport. *JRSS-B* (under revision)
https://arxiv.org/abs/2506.09986

Denoising: a simple latent variable model

- **Model**: $(Z, \Theta) \sim P_{Z,\Theta}$ where

$$Z \mid \Theta = \theta \sim p(\cdot \mid \theta), \qquad \text{and} \qquad \Theta \sim G^*$$

- **Observation**: $Z \in \mathbb{R}^d$
- **Latent variable**: $\Theta \in \Omega \subset \mathbb{R}^m$ (unobserved)

- $\{p(\cdot \mid \theta)\}_{\theta \in \Omega}$ — pdf/pmf of known parametric family
- Examples: $N(\theta, \Sigma)$ (Σ known), Poisson(θ), $N(0, \theta)$, Uniform($0, \theta$) ...

- G^*: Distribution of the latent variable (unknown)

Denoising: a simple latent variable model

- **Model**: $(Z, \Theta) \sim P_{Z,\Theta}$ where

$$Z \mid \Theta = \theta \sim p(\cdot \mid \theta), \qquad \text{and} \qquad \Theta \sim G^*$$

- **Observation**: $Z \in \mathbb{R}^d$
- **Latent variable**: $\Theta \in \Omega \subset \mathbb{R}^m$ (unobserved)

- $\{p(\cdot \mid \theta)\}_{\theta \in \Omega}$ — pdf/pmf of known parametric family
- Examples: $N(\theta, \Sigma)$ (Σ known), Poisson(θ), $N(0, \theta)$, Uniform$(0, \theta)$ ...

- G^*: Distribution of the latent variable (unknown)

- **Examples**: Measurement error models, astronomy, genomics, ...

Denoising: a simple latent variable model

- **Model**: $(Z, \Theta) \sim P_{Z,\Theta}$ where

$$Z \mid \Theta = \theta \sim p(\cdot \mid \theta), \qquad \text{and} \qquad \Theta \sim G^*$$

- **Observation**: $Z \in \mathbb{R}^d$
- **Latent variable**: $\Theta \in \Omega \subset \mathbb{R}^m$ (unobserved)

- $\{p(\cdot \mid \theta)\}_{\theta \in \Omega}$ — pdf/pmf of known parametric family
- Examples: $N(\theta, \Sigma)$ (Σ known), Poisson(θ), $N(0, \theta)$, Uniform$(0, \theta)$ ...

- G^*: Distribution of the latent variable (unknown)

- **Examples**: Measurement error models, astronomy, genomics, ...

Data: $Z_1, \ldots, Z_n$ from the above model

Goal: Denoise $\{Z_i\}_{i=1}^n$, i.e., predict the latent/unobserved $\{\Theta_i\}_{i=1}^n$

Oracle Bayes estimator under squared error loss

Model: $(Z, \Theta) \sim P_{Z,\Theta}$ where

$$Z \mid \Theta = \theta \sim p(\cdot \mid \theta), \qquad \text{and} \qquad \Theta \sim G^* \quad (\text{known})$$

Goal: Given Z predict Θ

Oracle Bayes estimator

A standard approach is to consider:

$$\inf_{\delta:\mathbb{R}^d \to \mathbb{R}^m} \mathbb{E}_{(Z,\Theta)\sim P_{Z,\Theta}} \left[\|\delta(Z) - \Theta\|^2 \right],$$

Oracle Bayes estimator under squared error loss

Model: $(Z, \Theta) \sim P_{Z,\Theta}$ where

$$Z \mid \Theta = \theta \sim p(\cdot \mid \theta), \qquad \text{and} \qquad \Theta \sim G^* \quad (\text{known})$$

Goal: Given Z predict Θ

Oracle Bayes estimator

A standard approach is to consider:

$$\inf_{\delta:\mathbb{R}^d \to \mathbb{R}^m} \mathbb{E}_{(Z,\Theta)\sim P_{Z,\Theta}} \left[\|\delta(Z) - \Theta\|^2\right],$$

whose solution is the posterior mean:

$$\overline{\theta}(z) := \mathbb{E}_{G^*}[\Theta \mid Z = z] = \frac{\int \theta p(z \mid \theta)\, dG^*(\theta)}{\int p(z \mid \theta)\, dG^*(\theta)}.$$

$\overline{\theta}(Z)$: Optimal in terms of minimizing the (Bayes) risk w.r.t. squared loss

Oracle Bayes estimator under squared error loss

Model: $(Z, \Theta) \sim P_{Z,\Theta}$ where

$$Z \mid \Theta = \theta \sim p(\cdot \mid \theta), \qquad \text{and} \qquad \Theta \sim G^* \quad (\text{known})$$

Goal: Given Z predict Θ

Oracle Bayes estimator

A standard approach is to consider:

$$\inf_{\delta : \mathbb{R}^d \to \mathbb{R}^m} \mathbb{E}_{(Z,\Theta) \sim P_{Z,\Theta}} \left[\|\delta(Z) - \Theta\|^2 \right],$$

whose solution is the posterior mean:

$$\overline{\theta}(z) := \mathbb{E}_{G^*}[\Theta \mid Z = z] = \frac{\int \theta p(z \mid \theta)\, dG^*(\theta)}{\int p(z \mid \theta)\, dG^*(\theta)}.$$

$\overline{\theta}(Z)$: Optimal in terms of minimizing the (Bayes) risk w.r.t. squared loss

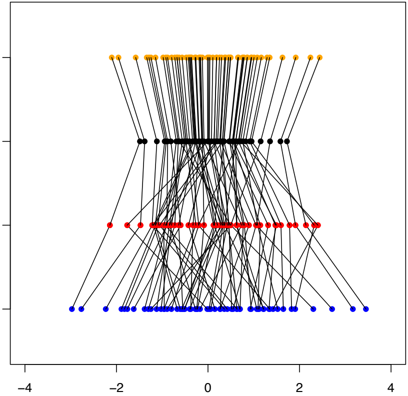
Empirical Bayes: Estimate $\overline{\theta}(\cdot)$ by $\hat{\overline{\theta}}(\cdot)$ (as G^* is usually unknown)

References: Robbins (1951, 1956), Dyson (1926), Jiang & Zhang (2009), Efron (2019)

Drawbacks of $\overline{\theta}(\cdot)$: denoising an ensemble

Example 1: (Normal location mixture) $\{(\Theta_i, Z_i)\}_{i=1}^{n}$ i.i.d. (here $n = 60$):

$$Z \mid \Theta = \theta \sim N(\theta, 1),\ \theta \in \mathbb{R}; \qquad \Theta \sim G^* \equiv N(0, 1)$$

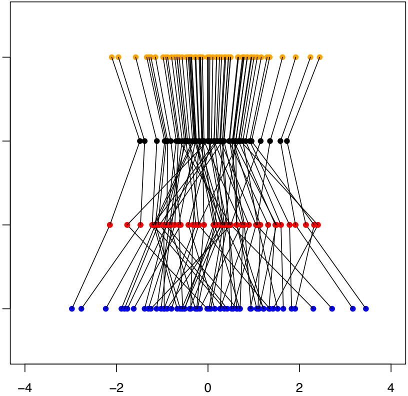


Observations. Latent variables. Bayes estimator. OT-based denoiser.

Example 1: (Normal location mixture) $\{(\Theta_i, Z_i)\}_{i=1}^n$ i.i.d. (here $n = 60$):

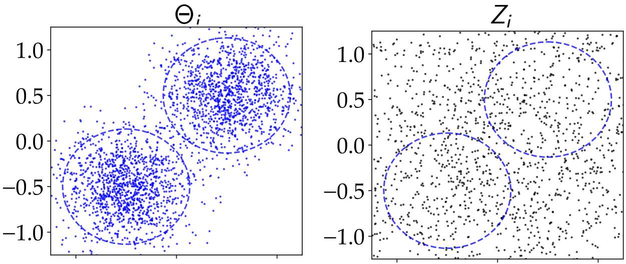
$$Z \mid \Theta = \theta \sim N(\theta, 1),\ \theta \in \mathbb{R}; \qquad \Theta \sim G^* \equiv N(0,1)$$

Overshrunk! Observe that $\overline{\theta}(Z) = \frac{1}{2}Z \sim N(0, \frac{1}{2})$

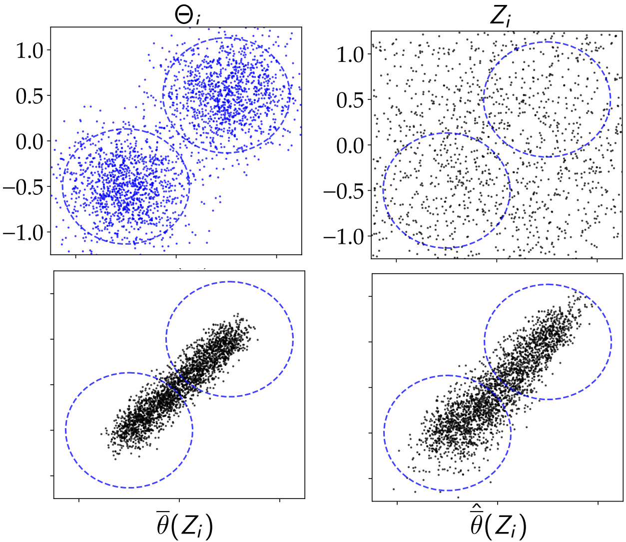


Observations. Latent variables. Bayes estimator. OT-based denoiser.

In general: $\mathrm{Var}(\Theta) = \mathrm{Var}(\overline{\theta}(Z)) + \mathbb{E}[\mathrm{Var}(\Theta \mid Z)] \geq \mathrm{Var}(\overline{\theta}(Z))$



Θ_i
Z_i
1.0
0.5
0.0
−0.5
−1.0



Empirical Bayes: Estimate $\overline{\theta}(\cdot)$ by $\hat{\overline{\theta}}(\cdot)$ (as G^* is usually unknown)

Posterior means shrink the ensemble

Some history

- Astronomers use the empirical distribution of Bayes estimates $\hat{\bar{\theta}}(Z_1), \ldots, \hat{\bar{\theta}}(Z_n)$ and interpret it as an improved estimate of the distribution of $\Theta_1, \ldots, \Theta_n$ (Loredo, 2007)
- The distribution of $\{\hat{\bar{\theta}}(Z_i)\}_{i=1}^n$ is **NOT** a good estimator of G^*
- **Eddington (1940)**: Lamented the use of this "fallacious" practice

Some history

- Astronomers use the empirical distribution of Bayes estimates $\hat{\hat{\theta}}(Z_1), \ldots, \hat{\hat{\theta}}(Z_n)$ and interpret it as an improved estimate of the distribution of $\Theta_1, \ldots, \Theta_n$ (Loredo, 2007)
- The distribution of $\{\hat{\hat{\theta}}(Z_i)\}_{i=1}^n$ is **NOT** a good estimator of G^*
- **Eddington (1940)**: Lamented the use of this "fallacious" practice
- Louis (1984): "... the primary goal is to produce parameter estimates that can be considered as an ensemble. For example, the histogram of estimates should be a good estimate of the histogram of parameters [i.e., Θ_i's] ..."

Some history

- Astronomers use the empirical distribution of Bayes estimates $\hat{\hat{\theta}}(Z_1), \ldots, \hat{\hat{\theta}}(Z_n)$ and interpret it as an improved estimate of the distribution of $\Theta_1, \ldots, \Theta_n$ (Loredo, 2007)

- The distribution of $\{\hat{\hat{\theta}}(Z_i)\}_{i=1}^n$ is **NOT** a good estimator of G^*

- **Eddington (1940)**: Lamented the use of this “fallacious” practice

- Louis (1984): “... the primary goal is to produce parameter estimates that can be considered as an ensemble. For example, the histogram of estimates should be a good estimate of the histogram of parameters [i.e., Θ_i’s] ...”

- **Statistics literature**: Various ways to formalize this idea have been taken up in some existing works, e.g., Louis (1984), Ghosh (1992), Shen and Louis (1998), Ghosh and Maiti (1999), ...

- **Model**: $(Z_i, \Theta_i) \overset{\text{iid}}{\sim} P_{Z,\Theta}$ where

$$Z_i \mid \Theta_i = \theta \sim p(\cdot \mid \theta), \qquad \text{and} \qquad \Theta_i \sim G^*$$

- **Data**: $Z_1, \ldots, Z_n \in \mathbb{R}^d$
- **Goal**: Estimate/Predict $\Theta_1, \ldots, \Theta_n \in \mathbb{R}^m$
- **Original problem** targets the oracle Bayes estimator $\overline{\theta}(\cdot)$:

- **Model**: $(Z_i, \Theta_i) \overset{\text{iid}}{\sim} P_{Z,\Theta}$ where

$$Z_i \mid \Theta_i = \theta \sim p(\cdot \mid \theta), \qquad \text{and} \qquad \Theta_i \sim G^*$$

- **Data**: $Z_1, \ldots, Z_n \in \mathbb{R}^d$

- **Goal**: Estimate/Predict $\Theta_1, \ldots, \Theta_n \in \mathbb{R}^m$

- **Original problem** targets the oracle Bayes estimator $\overline{\theta}(\cdot)$:

$$\begin{cases} \text{minimize} & \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^n \|\delta(Z_i) - \Theta_i\|^2\right] \\ \text{over} & \delta : \mathbb{R}^d \to \mathbb{R}^m \end{cases}$$

- $\overline{\theta}(\cdot)$ overshrinks!

Question: How to rectify this drawback of $\overline{\theta}(\cdot)$?

Our proposal: OT-based Denoiser

Add constraints:

$$\begin{cases} \text{minimize} & \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n}\|\delta(Z_i)-\Theta_i\|^2\right] \\ \text{over} & \delta : \mathbb{R}^d \to \mathbb{R}^m \\ \text{with} & \delta(Z) \stackrel{\mathcal{D}}{=} \Theta \end{cases}$$

i.e., the denoised ensemble has the same distribution as the latent ensemble

Our proposal: Variance constrained denoiser

Add constraints:

$$\begin{cases} \text{minimize} & \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n}\|\delta(Z_i)-\Theta_i\|^2\right] \\ \text{over} & \delta:\mathbb{R}^d\to\mathbb{R}^m \\ \text{with} & \mathbb{E}[\delta(Z)]=\mathbb{E}[\Theta] \\ \text{and} & \mathrm{Var}(\delta(Z))=\mathrm{Var}(\Theta) \end{cases}$$

Our proposal: General constrained denoiser

Fix a linearly independent set of functions $\phi_1, \ldots, \phi_k : \mathbb{R}^m \to \mathbb{R}$

Add constraints:

$$\begin{cases} \text{minimize} & \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^n \|\delta(Z_i) - \Theta_i\|^2\right] \\ \text{over} & \delta : \mathbb{R}^d \to \mathbb{R}^m \\ \text{with} & \mathbb{E}[\phi_\ell(\delta(Z))] = \mathbb{E}[\phi_\ell(\Theta)] \quad \text{for all } 1 \leq \ell \leq k \end{cases}$$

Our proposal

Add constraints:

$$\begin{cases} \text{minimize} & \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n}\|\delta(Z_i)-\Theta_i\|^2\right] \\ \text{over} & \delta:\mathbb{R}^d\to\mathbb{R}^m \\ \text{with} & \cdots \end{cases}$$

Questions

- **Oracle solution?** (i.e., when G^* is known)

Our proposal

Add constraints:

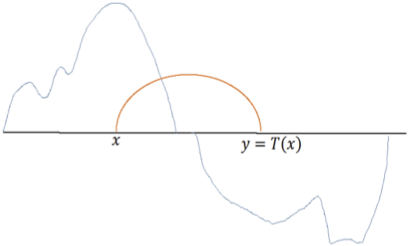
$$\begin{cases} \text{minimize} & \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n}\|\delta(Z_i)-\Theta_i\|^2\right] \\ \text{over} & \delta:\mathbb{R}^d\to\mathbb{R}^m \\ \text{with} & \cdots \end{cases}$$

Questions

- **Oracle solution?** (i.e., when G^* is known)
- **Empirical Bayes approximation?**

Optimal Transport (OT): Monge's problem

What is the cheapest way to transport a pile of sand to cover a sinkhole?



- $\nu, \tilde{\nu}$: two probability measures on Polish spaces $\mathcal{X}$ and $\mathcal{Y}$
- $c(x, y)$: cost of transporting x to y (e.g., $c(x, y) = \|x - y\|^2$)
- **Goal**: $$\inf_{T : T_\sharp \nu = \tilde{\nu}} \mathbb{E}_\nu[c(X, T(X))]$$
- T transports ν to $\tilde{\nu}$: $T_\sharp \nu = \tilde{\nu}$ (i.e., $T(X) \sim \tilde{\nu}$ when $X \sim \nu$)

Let ν and $\tilde{\nu}$ be two distributions over $\mathbb{R}^m$ with finite second moments

Kantorovich relaxation:

$$W_2^2(\nu, \tilde{\nu}) := \min_{\pi \in \Gamma(\nu, \tilde{\nu})} \int \int \|x - y\|^2 \, d\pi(x, y), \tag{1}$$

where $\Gamma(\nu, \tilde{\nu})$ denotes set of all joint distributions on $\mathbb{R}^m \times \mathbb{R}^m$ with marginals ν and $\tilde{\nu}$. (1) is a linear program.

Let ν and $\tilde{\nu}$ be two distributions over $\mathbb{R}^m$ with finite second moments

Kantorovich relaxation:

$$W_2^2(\nu, \tilde{\nu}) := \min_{\pi \in \Gamma(\nu,\tilde{\nu})} \int \int \|x - y\|^2 \, d\pi(x, y), \tag{1}$$

where $\Gamma(\nu, \tilde{\nu})$ denotes set of all joint distributions on $\mathbb{R}^m \times \mathbb{R}^m$ with marginals ν and $\tilde{\nu}$. (1) is a linear program.

Brenier's theorem

Suppose that ν has a Lebesgue density. Then,

- $\exists$ a convex function ψ whose gradient $\nabla\psi$ pushes ν forward to $\tilde{\nu}$.
- $\nabla\psi$ (ν-a.e.) uniquely minimizes Monge's problem, i.e.,

$$\nabla\psi = \operatorname*{argmin}_{T \,:\, T_\sharp \nu = \tilde{\nu}} \int \|x - T(x)\|^2 d\nu(x)$$

- The coupling[a] $(\mathrm{Id} \times \nabla\psi)_\sharp \nu$ uniquely minimizes (1).

[a] The map $(\mathrm{Id} \times T) : \mathbb{R}^m \to \mathbb{R}^m \times \mathbb{R}^m$ is defined as $(\mathrm{Id} \times T)(x) = (x, T(x))$.

Brenier's theorem says that the Kantorovich problem is a tight convex relaxation of the nonconvex Monge problem:

Brenier's theorem says that the Kantorovich problem is a tight convex relaxation of the nonconvex Monge problem:

- Both problems have the same value (under the Brenier assumptions)

Brenier's theorem says that the Kantorovich problem is a tight convex relaxation of the nonconvex Monge problem:

- Both problems have the same value (under the Brenier assumptions)
- Both problems have a unique solution

Brenier's theorem says that the Kantorovich problem is a tight convex relaxation of the nonconvex Monge problem:

- Both problems have the same value (under the Brenier assumptions)
- Both problems have a unique solution
- There is a natural correspondence between the solutions:
  - If T is a solution to the Monge problem, then the distribution of $(X, T(X))$ for $X \sim \nu$ is a solution to the Kantorovich problem.

Brenier's theorem says that the Kantorovich problem is a tight convex relaxation of the nonconvex Monge problem:

- Both problems have the same value (under the Brenier assumptions)
- Both problems have a unique solution
- There is a natural correspondence between the solutions:
  - If T is a solution to the Monge problem, then the distribution of $(X, T(X))$ for $X \sim \nu$ is a solution to the Kantorovich problem.
  - If π is a solution to the Kantorovich problem, then the *barycentric projection*

    $$T(x) := \mathbb{E}_{\pi}[X' \mid X = x]$$

    for $(X, X') \sim \pi$ is a solution to the Monge problem

Brenier's theorem says that the Kantorovich problem is a tight convex relaxation of the nonconvex Monge problem:

- Both problems have the same value (under the Brenier assumptions)
- Both problems have a unique solution
- There is a natural correspondence between the solutions:
  - If T is a solution to the Monge problem, then the distribution of $(X, T(X))$ for $X \sim \nu$ is a solution to the Kantorovich problem.
  - If π is a solution to the Kantorovich problem, then the *barycentric projection*
  $$T(x) := \mathbb{E}_\pi[X' \mid X = x]$$
  for $(X, X') \sim \pi$ is a solution to the Monge problem

Thus, the Monge problem can be solved via the Kantorovich problem

1. Existence & Characterization of OT-based Denoiser

- **Model**: $Z \mid \Theta = \theta \sim p(\cdot \mid \theta)$, and $\Theta \sim G^*$
- **Problem:** $\inf_{\delta : \delta_\sharp \mu = G^*} \mathbb{E}\left[\|\delta(Z) - \Theta\|^2\right] \qquad (Z \sim \mu, \Theta \sim G^*)$
- This is not quite a Monge OT problem (yet!)

- **Model**: $Z \mid \Theta = \theta \sim p(\cdot \mid \theta),$ and $\Theta \sim G^*$

- **Problem:** $\inf_{\delta : \delta_\sharp \mu = G^*} \mathbb{E}\left[\|\delta(Z) - \Theta\|^2\right]$ $(Z \sim \mu, \Theta \sim G^*)$

- This is not quite a Monge OT problem (yet!)

- The objective can be simplified (recall $\bar{\theta}(Z) = \mathbb{E}[\Theta \mid Z]$):

$$\mathbb{E}\left[\|\delta(Z) - \Theta\|^2\right] = \mathbb{E}[\|\delta(Z) - \bar{\theta}(Z)\|^2] + \mathbb{E}[\|\bar{\theta}(Z) - \Theta\|^2]$$

- **Model**: $Z \mid \Theta = \theta \sim p(\cdot \mid \theta)$, and $\Theta \sim G^*$
- **Problem:** $\inf_{\delta : \delta_\sharp \mu = G^*} \mathbb{E}\left[\|\delta(Z) - \Theta\|^2\right]$ $(Z \sim \mu, \Theta \sim G^*)$
- This is not quite a Monge OT problem (yet!)
- The objective can be simplified (recall $\overline{\theta}(Z) = \mathbb{E}[\Theta \mid Z]$):

$$\mathbb{E}\left[\|\delta(Z) - \Theta\|^2\right] = \mathbb{E}[\|\delta(Z) - \overline{\theta}(Z)\|^2] + \mathbb{E}[\|\overline{\theta}(Z) - \Theta\|^2]$$

- **Equivalent problem:** $\inf_{\delta : \delta_\sharp \mu = G^*} \mathbb{E}_{Z \sim \mu}[\|\delta(Z) - \overline{\theta}(Z)\|^2]$ $(\star\star)$

- **Model**: $Z \mid \Theta = \theta \sim p(\cdot \mid \theta)$, and $\Theta \sim G^*$

- **Problem:** $\inf_{\delta : \delta_\sharp \mu = G^*} \mathbb{E}\left[\|\delta(Z) - \Theta\|^2\right] \qquad (Z \sim \mu, \Theta \sim G^*)$

- This is not quite a Monge OT problem (yet!)

- The objective can be simplified (recall $\overline{\theta}(Z) = \mathbb{E}[\Theta \mid Z]$):

$$\mathbb{E}\left[\|\delta(Z) - \Theta\|^2\right] = \mathbb{E}[\|\delta(Z) - \overline{\theta}(Z)\|^2] + \mathbb{E}[\|\overline{\theta}(Z) - \Theta\|^2]$$

- **Equivalent problem:** $\inf_{\delta : \delta_\sharp \mu = G^*} \mathbb{E}_{Z \sim \mu}[\|\delta(Z) - \overline{\theta}(Z)\|^2] \qquad (\star\star)$

OT formulation

- Observe: $(\star\star) \equiv \inf_{\delta : \delta_\sharp \mu = G^*} \mathbb{E}_{Z \sim \mu}[c_{G^*}(Z, \delta(Z))]$ where

$$c_{G^*}(z, \vartheta) := \|\vartheta - \overline{\theta}(z)\|^2$$

- The cost function $c_{G^*}(\cdot, \cdot)$ depends on G^* via $\overline{\theta}(\cdot) = \mathbb{E}_{G^*}[\Theta \mid Z = \cdot]$

Characterization of the OT-based denoiser

Theorem 1 (García Trillos and S., 2024)

Suppose $(Z, \Theta) \sim P_{Z,\Theta}$ with marginals $\Theta \sim G^*$ and $Z \sim \mu$. Under suitable assumptions,

1. there exists μ-a.e. unique solution $\delta^*(\cdot)$:

$$\delta^* := \operatorname*{argmin}_{\delta : \delta_\sharp \mu = G^*} \mathbb{E}\left[\|\delta(Z) - \Theta\|^2\right]$$

Characterization of the OT-based denoiser

Theorem 1 (García Trillos and S., 2024)

Suppose $(Z, \Theta) \sim P_{Z,\Theta}$ with marginals $\Theta \sim G^*$ and $Z \sim \mu$. Under suitable assumptions,

1. there exists μ-a.e. unique solution $\delta^*(\cdot)$:

$$\delta^* := \underset{\delta : \delta_\sharp \mu = G^*}{\operatorname{argmin}} \, \mathbb{E}\left[\|\delta(Z) - \Theta\|^2\right] = \underset{\delta : \delta_\sharp \mu = G^*}{\operatorname{argmin}} \, \mathbb{E}\left[\|\delta(Z) - \overline{\theta}(Z)\|^2\right].$$

Characterization of the OT-based denoiser

Theorem 1 (García Trillos and S., 2024)

Suppose $(Z, \Theta) \sim P_{Z,\Theta}$ with marginals $\Theta \sim G^*$ and $Z \sim \mu$. Under suitable assumptions,

1. there exists μ-a.e. unique solution $\delta^*(\cdot)$:

$$\delta^* := \underset{\delta:\delta_\sharp \mu = G^*}{\operatorname{argmin}} \mathbb{E}\left[\|\delta(Z) - \Theta\|^2\right] = \underset{\delta:\delta_\sharp \mu = G^*}{\operatorname{argmin}} \mathbb{E}\left[\|\delta(Z) - \overline{\theta}(Z)\|^2\right].$$

2. Furthermore, $\delta^*(\cdot)$ can be written as, for some convex function φ,

$$\delta^*(z) = \nabla\varphi^*(\overline{\theta}(z)), \qquad \text{for} \quad z \in \mathbb{R}^d,$$

where $\nabla\varphi^*$ is the OT map from $\overline{\theta}_\sharp\mu$ to G^*.

Characterization of the OT-based denoiser

Theorem 1 (García Trillos and S., 2024)

Suppose $(Z, \Theta) \sim P_{Z,\Theta}$ with marginals $\Theta \sim G^*$ and $Z \sim \mu$. Under suitable assumptions,

1. there exists μ-a.e. unique solution $\delta^*(\cdot)$:

$$\delta^* := \underset{\delta:\delta_\sharp \mu = G^*}{\operatorname{argmin}} \mathbb{E}\left[\|\delta(Z) - \Theta\|^2\right] = \underset{\delta:\delta_\sharp \mu = G^*}{\operatorname{argmin}} \mathbb{E}\left[\|\delta(Z) - \overline{\theta}(Z)\|^2\right].$$

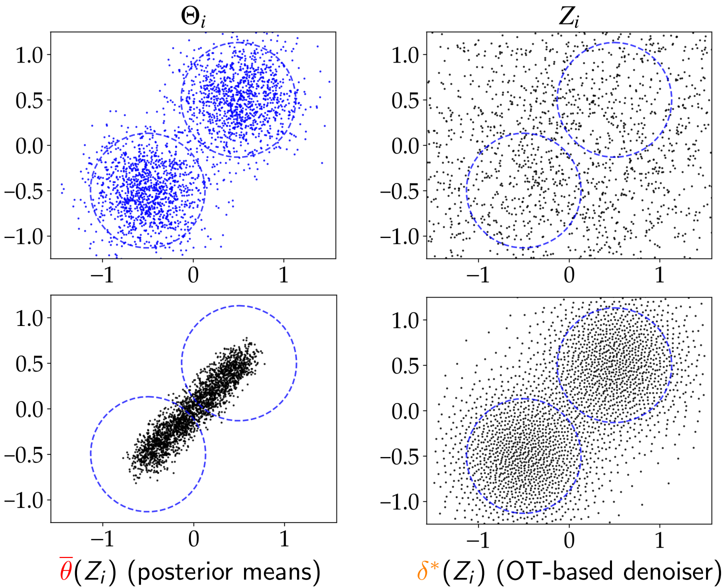
2. Furthermore, $\delta^*(\cdot)$ can be written as, for some convex function φ,

$$\delta^*(z) = \nabla\varphi^*(\overline{\theta}(z)), \qquad \text{for} \quad z \in \mathbb{R}^d,$$

   where $\nabla\varphi^*$ is the OT map from $\overline{\theta}_\sharp \mu$ to G^*.

3. Note: $\delta^*(z) = \mathbb{E}_{\pi^*}[\vartheta \mid Z = z]$ where

$$\pi^* = \underset{\pi \in \Gamma(\mu, G^*)}{\operatorname{argmin}} \mathbb{E}_\pi[\|\vartheta - \overline{\theta}(Z)\|^2]$$



OT restores the latent distribution while staying close to posterior means

2. Finite Sample Denoising via Empirical Bayes

Model: $Z \mid \Theta = \theta \sim p(\cdot \mid \theta)$, and $\Theta \sim G^*$

(Oracle) OT-based denoiser:

$$\delta^* \equiv \operatorname*{argmin}_{\delta:\mathbb{R}^d \to \mathbb{R}^m} \mathbb{E}_{G^*}\left[\|\delta(Z) - \Theta\|^2\right] \quad \text{s.t. } \delta(Z) \sim G^*$$

The proposal so far explicitly uses G^* to define OT-based denoiser

Model: $Z \mid \Theta = \theta \sim p(\cdot \mid \theta)$, and $\Theta \sim G^*$

(Oracle) OT-based denoiser:

$$\delta^* \equiv \operatorname*{argmin}_{\delta:\mathbb{R}^d \to \mathbb{R}^m} \mathbb{E}_{G^*}\left[\|\delta(Z) - \Theta\|^2\right] \quad \text{s.t. } \delta(Z) \sim G^*$$

The proposal so far explicitly uses G^* to define OT-based denoiser

Question: How would one estimate the OT-based denoiser δ^* from i.i.d. data $Z_1, \ldots, Z_n$, if G^* is unknown?

Model: $Z \mid \Theta = \theta \sim p(\cdot \mid \theta)$, and $\Theta \sim G^*$

(Oracle) OT-based denoiser:

$$\delta^* \equiv \underset{\delta:\mathbb{R}^d \to \mathbb{R}^m}{\operatorname{argmin}} \ \mathbb{E}_{G^*}\left[\|\delta(Z) - \Theta\|^2\right] \quad \text{s.t. } \delta(Z) \sim G^*$$

The proposal so far explicitly uses G^* to define OT-based denoiser

Question: How would one estimate the OT-based denoiser δ^* from i.i.d. data $Z_1, \ldots, Z_n$, if G^* is unknown?

Idea: Estimate G^* based on data $Z_1, \ldots, Z_n$, and then plug in the estimate $\widehat{G}$ instead of G^* in all expectation calculations

Model: $Z \mid \Theta = \theta \sim p(\cdot \mid \theta)$, and $\Theta \sim G^*$

(Oracle) OT-based denoiser:

$$\delta^* \equiv \underset{\delta:\mathbb{R}^d \to \mathbb{R}^m}{\operatorname{argmin}} \ \mathbb{E}_{G^*}\left[\|\delta(Z) - \Theta\|^2\right] \quad \text{s.t. } \delta(Z) \sim G^*$$

The proposal so far explicitly uses G^* to define OT-based denoiser

Question: How would one estimate the OT-based denoiser δ^* from i.i.d. data $Z_1, \ldots, Z_n$, if G^* is unknown?

Idea: Estimate G^* based on data $Z_1, \ldots, Z_n$, and then plug in the estimate $\widehat{G}$ instead of G^* in all expectation calculations

Leads us to problems in deconvolution (estimation of G^*) and/or empirical Bayes (estimation of the oracle posterior means) methods

An empirical Bayes approach to estimating G^*

- Observe $Z_1, \ldots, Z_n$ i.i.d. from model: $Z \mid \Theta = \theta \sim p(\cdot \mid \theta), \;\; \Theta \sim G^*$

- Assume here that G^* is unknown and belongs to a family $\mathcal{P}$

- $\mathcal{P}$: parametric/nonparametric (e.g., all probability dist. on $\mathbb{R}^m$)

An empirical Bayes approach to estimating G^*

- Observe $Z_1, \ldots, Z_n$ i.i.d. from model: $Z \mid \Theta = \theta \sim p(\cdot \mid \theta), \;\; \Theta \sim G^*$
- Assume here that G^* is unknown and belongs to a family $\mathcal{P}$
- $\mathcal{P}$: parametric/nonparametric (e.g., all probability dist. on $\mathbb{R}^m$)
- **Observation**: Z_i's are i.i.d. with marginal density f_{G^*} where

$$f_G(z) := \int p(z \mid \theta)\, dG(\theta), \qquad \text{for } z \in \mathbb{R}^d$$

An empirical Bayes approach to estimating G^*

- Observe $Z_1, \ldots, Z_n$ i.i.d. from model: $Z \mid \Theta = \theta \sim p(\cdot \mid \theta), \;\; \Theta \sim G^*$
- Assume here that G^* is unknown and belongs to a family $\mathcal{P}$
- $\mathcal{P}$: parametric/nonparametric (e.g., all probability dist. on $\mathbb{R}^m$)
- **Observation**: Z_i's are i.i.d. with marginal density f_{G^*} where

$$f_G(z) := \int p(z \mid \theta)\, dG(\theta), \qquad \text{for } z \in \mathbb{R}^d$$

- **Maximum likelihood (ML)** estimator: $\widehat{G} \in \mathcal{P}$ maximizes the marginal (log)-likelihood of the observations $(Z_i)_{i=1}^n$, i.e.,

$$\widehat{G} \in \operatorname*{argmax}_{G \in \mathcal{P}} \sum_{i=1}^{n} \log f_G(Z_i)$$

- **References**: Robbins (1950), Kiefer and Wolfowitz (1956), ..., Lindsay (1995), Jiang-Zhang (2009), Koenker-Mizera (2014), ...

Recall: OT-based denoiser $\delta^*(z) = \mathbb{E}_{\pi^*}[\vartheta \mid Z = z]$ where

$$\pi^* = \underset{\pi \in \Gamma(\mu, G^*)}{\operatorname{argmin}} \int \|\overline{\theta}(z) - \vartheta\|^2 \, d\pi(z, \vartheta)$$

Recall: OT-based denoiser $\delta^*(z) = \mathbb{E}_{\pi^*}[\vartheta \mid Z = z]$ where

$$\pi^* = \operatorname*{argmin}_{\pi \in \Gamma(\mu, G^*)} \int \|\overline{\theta}(z) - \vartheta\|^2 \, d\pi(z, \vartheta)$$

1. Empirical Bayes estimate of posterior mean $\overline{\theta}(\cdot)$:

$$\hat{\overline{\theta}}(z) := \mathbb{E}_{\widehat{G}}[\Theta \mid Z = z]$$

Recall: OT-based denoiser $\delta^*(z) = \mathbb{E}_{\pi^*}[\vartheta \mid Z = z]$ where

$$\pi^* = \operatorname*{argmin}_{\pi \in \Gamma(\mu, G^*)} \int \|\overline{\theta}(z) - \vartheta\|^2 \, d\pi(z, \vartheta)$$

1. Empirical Bayes estimate of posterior mean $\overline{\theta}(\cdot)$:

$$\hat{\overline{\theta}}(z) := \mathbb{E}_{\widehat{G}}[\Theta \mid Z = z] = \frac{\int \theta p(z \mid \theta) \, d\widehat{G}(\theta)}{\int p(z \mid \theta) \, d\widehat{G}(\theta)},$$

2. Solve the empirical OT problem:

$$\hat{\pi} \in \operatorname*{argmin}_{\pi \in \Gamma(\mu_n, \widehat{G})} \int \|\hat{\overline{\theta}}(z) - \vartheta\|^2 \, d\pi(z, \vartheta),$$

where $\mu_n := \frac{1}{n}\sum_{i=1}^n \delta_{Z_i}$ is the empirical distribution of the Z_i's

Recall: OT-based denoiser $\delta^*(z) = \mathbb{E}_{\pi^*}[\vartheta \mid Z = z]$ where

$$\pi^* = \operatorname*{argmin}_{\pi \in \Gamma(\mu, G^*)} \int \|\overline{\theta}(z) - \vartheta\|^2 \, d\pi(z, \vartheta)$$

1. Empirical Bayes estimate of posterior mean $\overline{\theta}(\cdot)$:

$$\hat{\overline{\theta}}(z) := \mathbb{E}_{\widehat{G}}[\Theta \mid Z = z] = \frac{\int \theta p(z \mid \theta) \, d\widehat{G}(\theta)}{\int p(z \mid \theta) \, d\widehat{G}(\theta)},$$

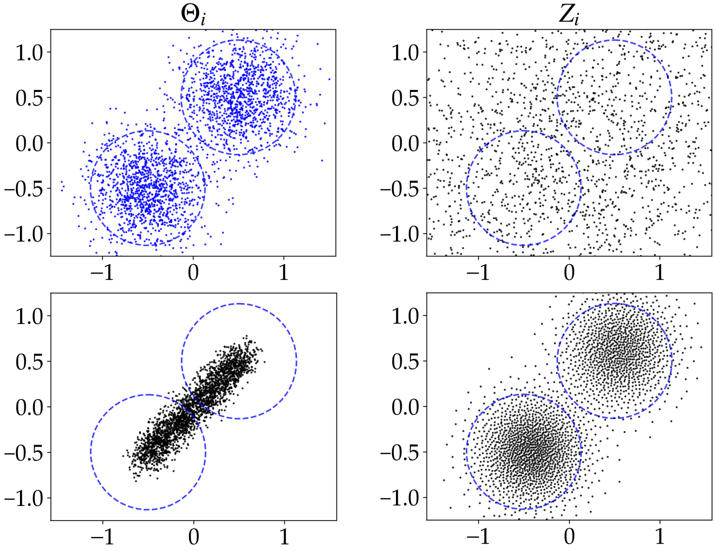
2. Solve the empirical OT problem:

$$\hat{\pi} \in \operatorname*{argmin}_{\pi \in \Gamma(\mu_n, \widehat{G})} \int \|\hat{\overline{\theta}}(z) - \vartheta\|^2 \, d\pi(z, \vartheta),$$

   where $\mu_n := \frac{1}{n} \sum_{i=1}^n \delta_{Z_i}$ is the empirical distribution of the Z_i's

3. Estimate $\delta^*(Z_i)$ by the barycentric projection

$$\hat{\delta}(Z_i) := \mathbb{E}_{\hat{\pi}}[\vartheta \mid Z = Z_i] \qquad \text{where} \qquad (Z, \vartheta) \sim \hat{\pi}$$



$\overline{\theta}(Z_i)$ (posterior means) $\hat{\delta}(Z_i)$ (OT-based denoiser)

OT-based denoiser: $\pi^* \in \mathrm{argmin}_{\pi \in \Gamma(\mu, G^*)} \int \|\overline{\theta}(z) - \vartheta\|^2 \, d\pi(z, \vartheta)$

Empirical problem: $\hat{\pi} \in \mathrm{argmin}_{\pi \in \Gamma(\mu_n, \widehat{G})} \int \|\hat{\overline{\theta}}(z) - \vartheta\|^2 \, d\pi(z, \vartheta)$

Recall: $\delta^*(Z_i) = \mathbb{E}_{\pi^*}[\vartheta \mid Z = Z_i]$; $\hat{\delta}(Z_i) := \mathbb{E}_{\hat{\pi}}[\vartheta \mid Z = Z_i]$

Assumptions

$\frac{1}{n} \sum_{i=1}^n \|\hat{\overline{\theta}}(Z_i) - \overline{\theta}(Z_i)\|^2 = O_{\mathbb{P}}(r_n)$, $W_2^2(\widehat{G}, G^*) = O_{\mathbb{P}}(q_n)$

Theorem [Jaffe, Ignatiadis, S. (2025+)]

Under suitable assumptions, as $n \to \infty$, we have:

$$\frac{1}{n} \sum_{i=1}^n \|\hat{\delta}(Z_i) - \delta^*(Z_i)\|^2 = O_{\mathbb{P}}\left(\sqrt{r_n \vee q_n}\right)$$

OT-based denoiser: $\pi^* \in \mathrm{argmin}_{\pi \in \Gamma(\mu, G^*)} \int \|\overline{\theta}(z) - \vartheta\|^2 \, d\pi(z, \vartheta)$

Empirical problem: $\hat{\pi} \in \mathrm{argmin}_{\pi \in \Gamma(\mu_n, \widehat{G})} \int \|\hat{\overline{\theta}}(z) - \vartheta\|^2 \, d\pi(z, \vartheta)$

Recall: $\delta^*(Z_i) = \mathbb{E}_{\pi^*}[\vartheta \mid Z = Z_i]$; $\hat{\delta}(Z_i) := \mathbb{E}_{\hat{\pi}}[\vartheta \mid Z = Z_i]$

Assumptions

$\frac{1}{n} \sum_{i=1}^{n} \|\hat{\overline{\theta}}(Z_i) - \overline{\theta}(Z_i)\|^2 = O_{\mathbb{P}}(r_n)$, $W_2^2(\widehat{G}, G^*) = O_{\mathbb{P}}(q_n)$

Theorem [Jaffe, Ignatiadis, S. (2025+)]

Under suitable assumptions, as $n \to \infty$, we have:

$$\frac{1}{n} \sum_{i=1}^{n} \|\hat{\delta}(Z_i) - \delta^*(Z_i)\|^2 = O_{\mathbb{P}}\left(\sqrt{r_n \vee q_n}\right)$$

The rate of convergence is determined by whichever of unconstrained Bayes denoising and deconvolution is slower

3. Variance-Constrained Denoising

Variance-constrained denoiser

- **Model**: $Z \mid \Theta = \theta \sim p(\cdot \mid \theta)$, and $\Theta \sim G^*$
- **Data**: $Z \in \mathbb{R}^d$; **latent variable**: $\Theta \in \Omega \subset \mathbb{R}^m$

- Consider the following optimization problem:

$$\begin{cases} \min\limits_{\delta:\mathbb{R}^d \to \mathbb{R}^m} & \mathbb{E}\left[\|\delta(Z) - \Theta\|^2\right] \\ \text{s.t.} & \mathbb{E}[\delta(Z)] = \mathbb{E}[\Theta] \\ \text{and} & \mathrm{Var}(\delta(Z)) = \mathrm{Var}(\Theta) \end{cases}$$

Variance-constrained denoiser

- **Model**: $Z \mid \Theta = \theta \sim p(\cdot \mid \theta)$, and $\Theta \sim G^*$
- **Data**: $Z \in \mathbb{R}^d$; **latent variable**: $\Theta \in \Omega \subset \mathbb{R}^m$

- Consider the following optimization problem:

$$\begin{cases} \min_{\delta:\mathbb{R}^d \to \mathbb{R}^m} & \mathbb{E}\left[\|\delta(Z) - \Theta\|^2\right] \\ \text{s.t.} & \mathbb{E}[\delta(Z)] = \mathbb{E}[\Theta] \\ \text{and} & \mathrm{Var}(\delta(Z)) = \mathrm{Var}(\Theta) \end{cases} \qquad (*)$$

- **Idea**: Minimize the risk while ensuring that the denoiser $\delta(Z)$ has first and second moments matching that of the latent variable Θ

- **Recall**: $\mathrm{Var}(\Theta) = \mathrm{Var}(\bar{\theta}(Z)) + \mathbb{E}[\mathrm{Var}(\Theta \mid Z)] \;\geq\; \mathrm{Var}(\bar{\theta}(Z))$

Variance constrained denoiser: any solution $\delta_{VC}(\cdot)$ to $(*)$

Variance-constrained denoiser

- **Model**: $Z \mid \Theta = \theta \sim p(\cdot \mid \theta)$, and $\Theta \sim G^*$
- **Data**: $Z \in \mathbb{R}^d$; **latent variable**: $\Theta \in \Omega \subset \mathbb{R}^m$

- Consider the following optimization problem:

$$\begin{cases} \min_{\delta:\mathbb{R}^d \to \mathbb{R}^m} & \mathbb{E}\left[\|\delta(Z) - \Theta\|^2\right] \\ \text{s.t.} & \mathbb{E}[\delta(Z)] = \mathbb{E}[\Theta] \\ \text{and} & \mathrm{Var}(\delta(Z)) = \mathrm{Var}(\Theta) \end{cases} \qquad (*)$$

- **Idea**: Minimize the risk while ensuring that the denoiser $\delta(Z)$ has first and second moments matching that of the latent variable Θ

- **Recall**: $\mathrm{Var}(\Theta) = \mathrm{Var}(\overline{\theta}(Z)) + \mathbb{E}[\mathrm{Var}(\Theta \mid Z)] \geq \mathrm{Var}(\overline{\theta}(Z))$

Variance constrained denoiser: any solution $\delta_{VC}(\cdot)$ to $(*)$

Questions: Existence of solutions? Characterization?

Model: $Z \mid \Theta = \theta \sim p(\cdot \mid \theta)$, and $\Theta \sim G^*$

Variance-constrained denoiser $\delta_{VC}(z)$ solves:

$$\begin{cases} \min_{\delta:\mathbb{R}^m \to \mathbb{R}^m} & \mathbb{E}\left[\|\delta(Z) - \Theta\|^2\right] \\ \text{s.t.} & \mathbb{E}[\delta(Z)] = \mathbb{E}[\Theta] \\ \text{and} & \mathrm{Var}(\delta(Z)) = \mathrm{Var}(\Theta) \end{cases}$$

Theorem [Jaffe, Ignatiadis, S. (2025+)]

The variance-constrained Bayes denoiser is given by:

$$\delta_{VC}(z) := \mathrm{t}_{\mathrm{Var}(\overline{\theta}(Z))}^{\mathrm{Var}(\Theta)}(\overline{\theta}(z) - \mathbb{E}[\Theta]) + \mathbb{E}[\Theta],$$

Model: $Z \mid \Theta = \theta \sim p(\cdot \mid \theta)$, and $\Theta \sim G^*$

Variance-constrained denoiser $\delta_{VC}(z)$ solves:

$$\begin{cases} \min_{\delta:\mathbb{R}^m \to \mathbb{R}^m} & \mathbb{E}\left[\|\delta(Z) - \Theta\|^2\right] \\ \text{s.t.} & \mathbb{E}[\delta(Z)] = \mathbb{E}[\Theta] \\ \text{and} & \mathrm{Var}(\delta(Z)) = \mathrm{Var}(\Theta) \end{cases}$$

Theorem [Jaffe, Ignatiadis, S. (2025+)]

The variance-constrained Bayes denoiser is given by:

$$\delta_{VC}(z) := \mathrm{t}_{\mathrm{Var}(\overline{\theta}(Z))}^{\mathrm{Var}(\Theta)}(\overline{\theta}(z) - \mathbb{E}[\Theta]) + \mathbb{E}[\Theta],$$

(connection to Gaussian OT) where for p.d. matrices A, B,

$$\mathrm{t}_A^B := A^{-1/2}\left(A^{1/2}\, B\, A^{1/2}\right)^{\frac{1}{2}} A^{-1/2},$$

which is the unique solution to the equation $\mathrm{t}\, A\, \mathrm{t} = B$

Model: $Z \mid \Theta = \theta \sim p(\cdot \mid \theta)$, and $\Theta \sim G^*$

Variance-constrained denoiser $\delta_{VC}(z)$ solves:

$$\begin{cases} \min_{\delta:\mathbb{R}^m \to \mathbb{R}^m} & \mathbb{E}\left[\|\delta(Z) - \Theta\|^2\right] \\ \text{s.t.} & \mathbb{E}[\delta(Z)] = \mathbb{E}[\Theta] \\ \text{and} & \mathrm{Var}(\delta(Z)) = \mathrm{Var}(\Theta) \end{cases}$$

Theorem [Jaffe, Ignatiadis, S. (2025+)]

The variance-constrained Bayes denoiser is given by:

$$\delta_{VC}(z) := \mathrm{t}_{\mathrm{Var}(\overline{\theta}(Z))}^{\mathrm{Var}(\Theta)}(\overline{\theta}(z) - \mathbb{E}[\Theta]) + \mathbb{E}[\Theta],$$

(connection to Gaussian OT) where for p.d. matrices A, B,

$$\mathrm{t}_A^B := A^{-1/2}\left(A^{1/2}\, B\, A^{1/2}\right)^{\frac{1}{2}} A^{-1/2},$$

which is the unique solution to the equation $\mathrm{t}\, A\, \mathrm{t} = B$

Question: Given data $Z_1, \ldots, Z_n$, how to estimate δ_{VC}?

Algorithm: Gaussian Noise + NP prior

Model: $Z \mid \Theta = \theta \sim N(\theta, \Sigma)$, and $\Theta \sim G^*$ (Σ known)

Recall: $\delta_{VC}(z) := \mathrm{t}_{\mathrm{Var}(\bar{\theta}(Z))}^{\mathrm{Var}(\Theta)}(\bar{\theta}(z) - \mathbb{E}[\Theta]) + \mathbb{E}[\Theta]$

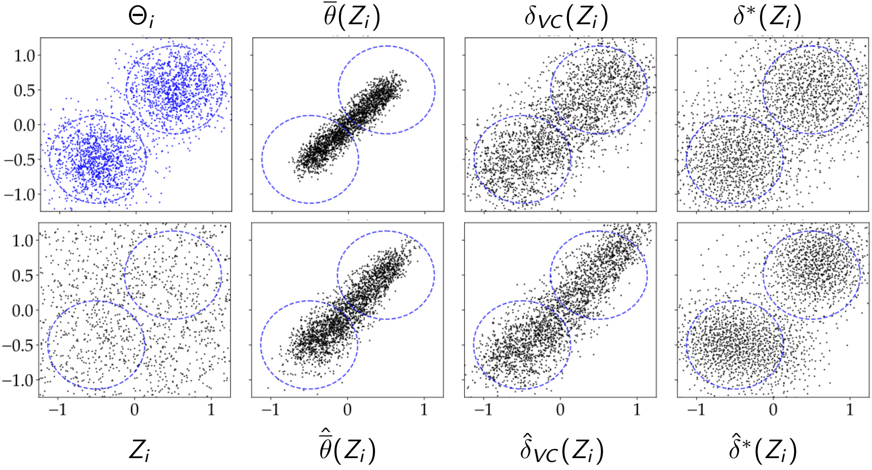
Algorithm: Gaussian Noise + NP prior

Model: $Z \mid \Theta = \theta \sim N(\theta, \Sigma)$, and $\Theta \sim G^*$ (Σ known)

Recall: $\delta_{VC}(z) := \mathrm{t}^{\mathrm{Var}(\Theta)}_{\mathrm{Var}(\overline{\theta}(Z))}(\overline{\theta}(z) - \mathbb{E}[\Theta]) + \mathbb{E}[\Theta]$

procedure VarianceConstrainedEB$(Z_1, \ldots, Z_n; \Sigma)$
    input: samples $Z_1, \ldots, Z_n \in \mathbb{R}^d$
    output: denoising function $\hat{\delta}_{VC} : \mathbb{R}^d \to \mathbb{R}^m$

    $\hat{\overline{\theta}}(\,\cdot\,) \leftarrow$ empirical Bayes approximation of $\overline{\theta}(\,\cdot\,)$
    $\hat{M} \leftarrow$ sample covariance matrix of $\hat{\overline{\theta}}(Z_1), \ldots \hat{\overline{\theta}}(Z_n)$
    $\hat{\mu} \leftarrow$ sample mean of $Z_1, \ldots, Z_n$
    $\hat{S} \leftarrow$ sample covariance matrix of $Z_1, \ldots, Z_n$
    $\hat{A} \leftarrow (\hat{S} - \Sigma)_+$ ## As $\mathrm{Var}(Z) = \mathrm{Var}(\Theta) + \Sigma$
    $\hat{\mathrm{t}} \leftarrow \hat{M}^{-1/2}(\hat{M}^{1/2}\hat{A}\hat{M}^{1/2})^{1/2}\hat{M}^{-1/2}$
    $\hat{\delta}_{VC}(\,\cdot\,) \leftarrow \hat{\mathrm{t}}\,(\hat{\overline{\theta}}(\,\cdot\,) - \hat{\mu}) + \hat{\mu}$
    return $\hat{\delta}_{VC}$
end procedure



$\hat{\delta}^*$ is obtained from smooth NPMLE [Magder-Zeger,1996; Kim-S.,2026)]

$\hat{\delta}_{VC}$ partially restore spread; $\hat{\delta}^*$ restores shape

General constrained denoising

- Fix a linearly independent set of functions $\phi_1, \ldots, \phi_k : \mathbb{R}^m \to \mathbb{R}$
- We consider the problem

$$\begin{cases} \text{minimize} & \mathbb{E}\left[\|\delta(Z) - \Theta\|^2\right] \\ \text{over} & \delta : \mathbb{R}^d \to \mathbb{R}^m \\ \text{with} & \mathbb{E}[\phi_\ell(\delta(Z))] = \mathbb{E}[\phi_\ell(\Theta)] \quad \text{for all } 1 \leq \ell \leq k \end{cases} \tag{2}$$

General constrained denoising

- Fix a linearly independent set of functions $\phi_1, \dots, \phi_k : \mathbb{R}^m \to \mathbb{R}$
- We consider the problem

$$\begin{cases} \text{minimize} & \mathbb{E}\left[\|\delta(Z) - \Theta\|^2\right] \\ \text{over} & \delta : \mathbb{R}^d \to \mathbb{R}^m \\ \text{with} & \mathbb{E}[\phi_\ell(\delta(Z))] = \mathbb{E}[\phi_\ell(\Theta)] \quad \text{for all } 1 \le \ell \le k \end{cases} \tag{2}$$

Theorem [Jaffe, Ignatiadis, S. (2025+)]

Under suitable assumptions, (2) has a unique solution δ_{GC}.

General constrained denoising

- Fix a linearly independent set of functions $\phi_1, \dots, \phi_k : \mathbb{R}^m \to \mathbb{R}$
- We consider the problem

$$\begin{cases} \text{minimize} & \mathbb{E}\left[\|\delta(Z) - \Theta\|^2\right] \\ \text{over} & \delta : \mathbb{R}^d \to \mathbb{R}^m \\ \text{with} & \mathbb{E}[\phi_\ell(\delta(Z))] = \mathbb{E}[\phi_\ell(\Theta)] \quad \text{for all } 1 \le \ell \le k \end{cases} \tag{2}$$

Theorem [Jaffe, Ignatiadis, S. (2025+)]

Under suitable assumptions, (2) has a unique solution δ_{GC}.

$\hat{\delta}_{GC}$: consistent estimator of δ_{GC} from solving a sample version of (2)

Summary

1. The OT-based denoiser δ^* can rectify drawbacks of posterior mean
2. We studied existence, characterization and estimation of δ^*

Summary

1. The OT-based denoiser δ^* can rectify drawbacks of posterior mean
2. We studied existence, characterization and estimation of δ^*
3. Investigated the variance constrained denoiser δ_{VC}
4. Framework can be extended to study general constrained denoiser

Thank you for your attention!

Questions?

Original problem

$$\min \mathbb{E}\left[\|\delta(Z) - \overline{\theta}(Z)\|^2\right]$$
over $\delta : \mathbb{R}^d \to \mathbb{R}^m$

s.t. $\mathbb{E}[\delta(Z)] = \mathbb{E}[\Theta]$
and $\mathrm{Var}(\delta(Z)) = \mathrm{Var}(\Theta)$

Gaussian OT

$$\min_\pi \mathbb{E}_\pi\left[\|U - V\|^2\right]$$
jointly Gaussian $(U, V) \sim \pi$

$U \sim N(\mathbb{E}[\Theta], \mathrm{Var}(\Theta))$
$V \sim N(\mathbb{E}[\Theta], \mathrm{Var}(\overline{\theta}(Z)))$

Original problem

$$\min \mathbb{E}\left[\|\delta(Z) - \overline{\theta}(Z)\|^2\right]$$
over $\delta : \mathbb{R}^d \to \mathbb{R}^m$

s.t. $\mathbb{E}[\delta(Z)] = \mathbb{E}[\Theta]$
and $\mathrm{Var}(\delta(Z)) = \mathrm{Var}(\Theta)$

Gaussian OT

$$\min_\pi \mathbb{E}_\pi\left[\|U - V\|^2\right]$$
jointly Gaussian $(U, V) \sim \pi$

$$U \sim N(\mathbb{E}[\Theta], \mathrm{Var}(\Theta))$$
$$V \sim N(\mathbb{E}[\Theta], \mathrm{Var}(\overline{\theta}(Z)))$$

RHS $\leq$ LHS

For any δ feasible for **LHS**, consider

$$\begin{pmatrix} U \\ V \end{pmatrix} \sim N\left(\begin{pmatrix} \mathbb{E}[\Theta] \\ \mathbb{E}[\Theta] \end{pmatrix}, \begin{pmatrix} \mathrm{Var}(\Theta) & \mathrm{Cov}(\delta(Z), \overline{\theta}(Z)) \\ \mathrm{Cov}(\overline{\theta}(Z), \delta(Z)) & \mathrm{Var}(\overline{\theta}(Z)) \end{pmatrix}\right)$$

Further, $\mathbb{E}\left[\|\delta(Z) - \overline{\theta}(Z)\|^2\right] = \mathbb{E}\left[\|U - V\|^2\right]$

Original problem	**Gaussian OT**
$\min \mathbb{E}\left[\|\delta(Z) - \overline{\theta}(Z)\|^2\right]$	$\min_\pi \mathbb{E}_\pi\left[\|U - V\|^2\right]$
over $\delta : \mathbb{R}^d \to \mathbb{R}^m$	jointly Gaussian $(U, V) \sim \pi$
s.t. $\mathbb{E}[\delta(Z)] = \mathbb{E}[\Theta]$	$U \sim N(\mathbb{E}[\Theta], \mathrm{Var}(\Theta))$
and $\mathrm{Var}(\delta(Z)) = \mathrm{Var}(\Theta)$	$V \sim N(\mathbb{E}[\Theta], \mathrm{Var}(\overline{\theta}(Z)))$

RHS is minimized by $(T(V), V) \sim \pi$ induced by the transport map

$$T(V) = \mathrm{t}_{\mathrm{Var}(\overline{\theta}(Z))}^{\mathrm{Var}(\Theta)}(V - \mathbb{E}[\Theta]) + \mathbb{E}[\Theta]$$

For strictly p.d. matrices A, B, $\mathrm{t}_A^B := A^{-1/2}\left(A^{1/2}\, B\, A^{1/2}\right)^{\frac{1}{2}} A^{-1/2}$,
which is the unique solution to the equation $\mathrm{t}\, A\, \mathrm{t} = B$

Original problem

$$\min \mathbb{E}\left[\|\delta(Z) - \overline{\theta}(Z)\|^2\right]$$
over $\delta : \mathbb{R}^d \to \mathbb{R}^m$

s.t. $\mathbb{E}[\delta(Z)] = \mathbb{E}[\Theta]$
and $\mathrm{Var}(\delta(Z)) = \mathrm{Var}(\Theta)$

Gaussian OT

$$\min_\pi \mathbb{E}_\pi\left[\|U - V\|^2\right]$$
jointly Gaussian $(U, V) \sim \pi$

$U \sim N(\mathbb{E}[\Theta], \mathrm{Var}(\Theta))$
$V \sim N(\mathbb{E}[\Theta], \mathrm{Var}(\overline{\theta}(Z)))$

RHS is minimized by $(T(V), V) \sim \pi$ induced by the transport map

$$T(V) = \mathrm{t}_{\mathrm{Var}(\overline{\theta}(Z))}^{\mathrm{Var}(\Theta)}(V - \mathbb{E}[\Theta]) + \mathbb{E}[\Theta]$$

For strictly p.d. matrices A, B, $\mathrm{t}_A^B := A^{-1/2}\left(A^{1/2} B A^{1/2}\right)^{\frac{1}{2}} A^{-1/2}$,
which is the unique solution to the equation $\mathrm{t}\, A\, \mathrm{t} = B$

Consider $\delta(z) := \mathrm{t}_{\mathrm{Var}(\overline{\theta}(Z))}^{\mathrm{Var}(\Theta)}(\overline{\theta}(z) - \mathbb{E}[\Theta]) + \mathbb{E}[\Theta]$

Then,

$$\textbf{LHS} \le \mathbb{E}\left[\|\delta(Z) - \overline{\theta}(Z)\|^2\right] = \mathbb{E}\left[\|T(V) - V\|^2\right] = \textbf{RHS}$$