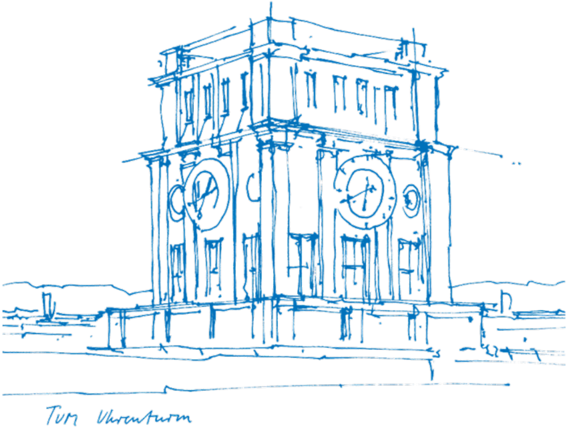
Robust Score Matching

Using the geometric median

Richard Schwank

Mathematical Statistics
Department of Mathematics
Technical University of Munich

September 1st, 2025

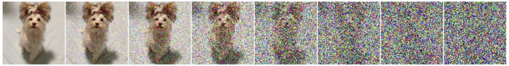


Outline

1. Score Matching using the Geometric Median (AISTATS 2025)
2. Some phenomena in high dimensions (arXiv:2508.12926)

Setup

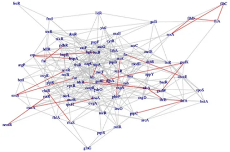
Score Matching



$$\nabla \log(c \cdot p(x)) = \nabla \log(p(x))$$

Robustness

Exponential family/
Graphical model



Exponential family score matching

Density model: $p(\mathbf{x} \mid \boldsymbol{\theta}) = \exp\left(\boldsymbol{\theta}^\top \mathbf{t}(\mathbf{x}) - a(\boldsymbol{\theta}) + b(\mathbf{x})\right)$ with $\mathbf{x} \in \mathcal{X}$ and parameter $\boldsymbol{\theta} \in \mathbb{R}^r$.

$$\boxed{\boldsymbol{\theta}_0} = \operatorname{argmin}_{\boldsymbol{\theta}} \ \frac{1}{2} \mathrm{E}_{\mathbf{x}\sim p(\cdot|\boldsymbol{\theta}_0)}\left[\, \|\nabla_{\mathbf{x}} \log(p(\mathbf{x} \mid \boldsymbol{\theta})) - \nabla_{\mathbf{x}} \log(p(\mathbf{x} \mid \boldsymbol{\theta}_0))\|^2 \,\right]$$

$$= \operatorname{argmin}_{\boldsymbol{\theta}} \ \frac{1}{2} \mathrm{E}_{\boldsymbol{\theta}_0}\left[\, \|\nabla_{\mathbf{x}} \log(p(\mathbf{x} \mid \boldsymbol{\theta}))\|^2 \,\right] + \mathrm{E}_{\boldsymbol{\theta}_0}\left[\, \Delta_{\mathbf{x}} \log(p(\mathbf{x} \mid \boldsymbol{\theta})) \,\right]$$

$$\boxed{=: \operatorname{argmin}_{\boldsymbol{\theta}} \ \tfrac{1}{2}\boldsymbol{\theta}^T \ \mathrm{E}_{\boldsymbol{\theta}_0}[\,\boldsymbol{\Gamma}(\mathbf{x})\,]\,\boldsymbol{\theta} + \boldsymbol{\theta}^T \ \mathrm{E}_{\boldsymbol{\theta}_0}[\,\mathbf{g}(\mathbf{x})\,]}$$

for some $\boldsymbol{\Gamma}(\mathbf{x}) \in \mathbb{R}^{r\times r}$ and $\mathbf{g}(\mathbf{x}) \in \mathbb{R}^r$.

Example (Square root graphical model)

Let $p(\mathbf{x} \mid \boldsymbol{\theta}) \propto \exp\left(-\sqrt{\mathbf{x}}^T \boldsymbol{\Theta} \sqrt{\mathbf{x}}\right)$, $\mathbf{x} \in \mathbb{R}^m_+$. Then,

$$\boldsymbol{\Gamma}(\mathbf{x}) = \operatorname{diag}(1/\mathbf{x}) \otimes \sqrt{\mathbf{x}}\sqrt{\mathbf{x}}^T \in \mathbb{R}^{m^2\times m^2}$$

Robust score matching - considerations

Standard score matching estimator $\hat{\boldsymbol{\theta}} := \mathrm{argmin}_{\boldsymbol{\theta}} \; \frac{1}{2}\boldsymbol{\theta}^T \bar{\boldsymbol{\Gamma}} \boldsymbol{\theta} + \boldsymbol{\theta}^T \bar{\mathbf{g}}$ with $\bar{\boldsymbol{\Gamma}} := \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{\Gamma}(\mathbf{x}_i)$.
Estimator not robust, e.g. against $\|\boldsymbol{\Gamma}(\mathbf{x}_1)\| \to \infty$.

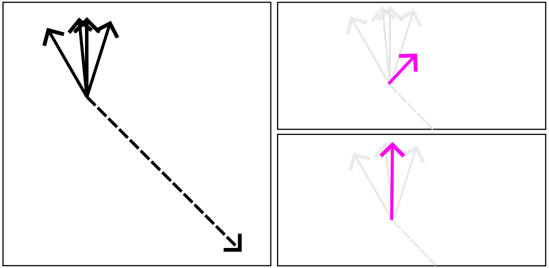
Desiderata for alternatives $\hat{\boldsymbol{\Gamma}}, \hat{\boldsymbol{g}}$

- Robust against arbitrary contamination of a few observations $\mathbf{x}_i$ (i.e. rowwise contamination)
- Few assumptions on $\boldsymbol{\Gamma}, \boldsymbol{g}$ required
- Computable in high dimensions
- $\hat{\boldsymbol{\Gamma}}$ should be positive (semi)definite

The geometric median

Definition: $\mathrm{Med}(\,\mathbf{x}_1, \dots, \mathbf{x}_n\,) := \mathrm{argmin}_{\mathbf{m}} \; \frac{1}{n}\sum_{i=1}^{n} \|\mathbf{x}_i - \mathbf{m}\|_2$

Illustration:



Mean

Geometric median

- Robustness and computation well studied in the literature
- Geometric median exists & unique under mild assumptions
- $\mathrm{Med}(\,\mathbf{x}_1, \dots, \mathbf{x}_n\,) \in$ ConvHull$(\mathbf{x}_1, \dots, \mathbf{x}_n)$

  $\Rightarrow$ When $\mathbf{\Gamma}(\mathbf{x}_i)$ is pos. sem. def., then so is $\mathrm{Med}(\mathbf{\Gamma}(\mathbf{x}_1), \dots, \mathbf{\Gamma}(\mathbf{x}_n))$

Robust Score Matching

We consider $\hat{\boldsymbol{\theta}} := \mathrm{argmin}_{\boldsymbol{\theta}} \ \frac{1}{2}\boldsymbol{\theta}^T \hat{\boldsymbol{\Gamma}} \boldsymbol{\theta} + \boldsymbol{\theta}^T \hat{\mathbf{g}}$ with $\hat{\boldsymbol{\Gamma}} := \mathrm{Med}(\boldsymbol{\Gamma}(\mathbf{x}_1), \ldots, \boldsymbol{\Gamma}(\mathbf{x}_n))$.

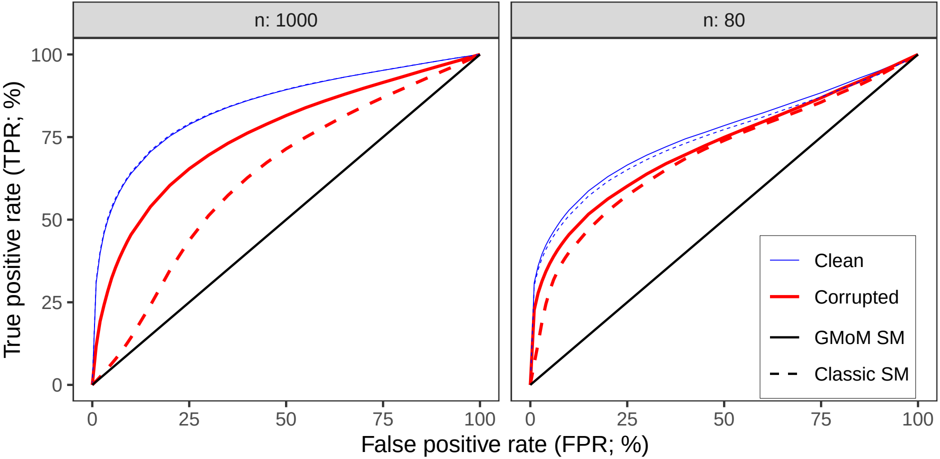
Additionally, we allow for

- ℓ_1-regularization
- support constraints (e.g. $\mathbf{x} > 0$; *generalized* score matching)
- median-of-means to balance bias with robustness.

For pairwise interaction models, we prove edge recovery under standard Lasso-assumptions, even if a fraction of the n observations is arbitrarily contaminated.

Simulation results for graphical models

Task: find edges in square root graphical model. Below: ROC curves (ℓ_1-penalty varies)



Outline

1. Score Matching using the Geometric Median (AISTATS 2025)
2. Some phenomena in high dimensions (arXiv:2508.12926)

Bias in high dimensions

Observation from robust score matching:

On uncorrupted data, geometric median-based $\hat{\mathbf{\Gamma}}$ surprisingly close to sample mean $\bar{\mathbf{\Gamma}}$ when

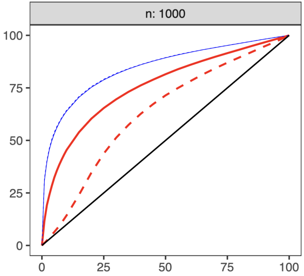
- sample size n is large *and*
- number of observed variables is high.

Large-n-limit: Given random variable $\mathbf{x} \in \mathbb{R}^p$, define

- $\mathbf{m} := \mathrm{Med}(\mathbf{x}) := \mathrm{argmin}_{\tilde{\mathbf{m}}} \mathrm{E}[\, \|\mathbf{x} - \tilde{\mathbf{m}}\| \,]$
- $\boldsymbol{\mu} := \mathrm{E}[\, \mathbf{x} \,]$.

Research Question: Quantify $\|\mathbf{m} - \boldsymbol{\mu}\|$ as $p \to \infty$.

From last slide:



A simple example

Consider $\mathbf{x} \in \mathbb{R}^p$ with x_i iid. Exp(1). Geometric median $\mathbf{m}$ exists & unique.

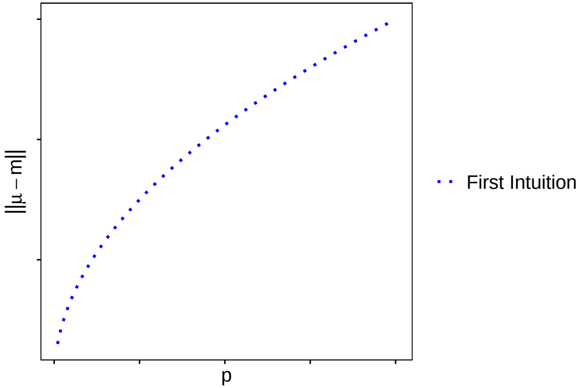
How does $\|\mathbf{m} - \boldsymbol{\mu}\|$ scale with p?

A simple example

Consider $\mathbf{x} \in \mathbb{R}^p$ with x_i iid. Exp(1). Geometric median $\mathbf{m}$ exists & unique.

How does $\|\mathbf{m} - \boldsymbol{\mu}\|$ scale with p?

First intuition: Componentwise median $\tilde{\mathbf{m}}$ has $\|\tilde{\mathbf{m}} - \boldsymbol{\mu}\| = |\log(2) - 1| \cdot \sqrt{p}$

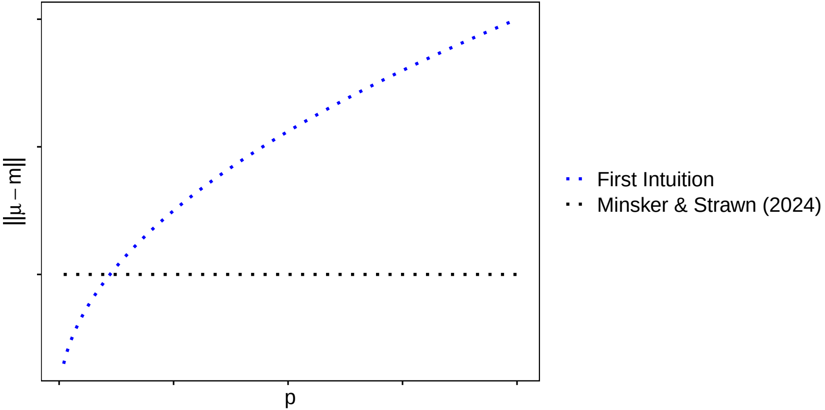


A simple example

Consider $\mathbf{x} \in \mathbb{R}^p$ with x_i iid. Exp(1). Geometric median $\mathbf{m}$ exists & unique.

How does $\|\mathbf{m} - \boldsymbol{\mu}\|$ scale with p?

First intuition: Componentwise median $\tilde{\mathbf{m}}$ has $\|\tilde{\mathbf{m}} - \boldsymbol{\mu}\| = |\log(2) - 1| \cdot \sqrt{p}$

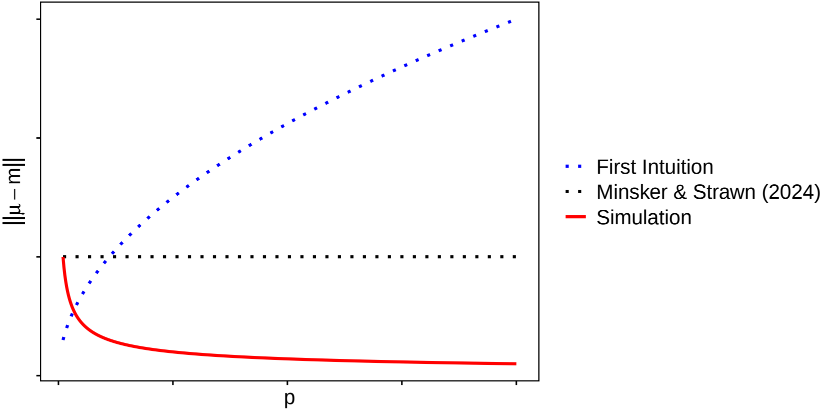


A simple example

Consider $\mathbf{x} \in \mathbb{R}^p$ with x_i iid. Exp(1). Geometric median $\mathbf{m}$ exists & unique.

How does $\|\mathbf{m} - \boldsymbol{\mu}\|$ scale with p?

First intuition: Componentwise median $\tilde{\mathbf{m}}$ has $\|\tilde{\mathbf{m}} - \boldsymbol{\mu}\| = |\log(2) - 1| \cdot \sqrt{p}$



Theoretical insights

arXiv > math > arXiv:2508.12926

Search... | All fields | Search

Help | Advanced Search

Mathematics > Statistics Theory

[Submitted on 18 Aug 2025]

On the distance between mean and geometric median in high dimensions

Richard Schwank, Mathias Drton

The geometric median, a notion of center for multivariate distributions, has gained recent attention in robust statistics and machine learning. Although conceptually distinct from the mean (i.e., expectation), we demonstrate that both are very close in high dimensions when the dependence between the distribution components is suitably controlled. Concretely, we find an upper bound on the distance that vanishes with the dimension asymptotically, and derive a rate-matching first order expansion of the geometric median components. Simulations illustrate and confirm our results.

Subjects: **Statistics Theory (math.ST)**

Cite as: arXiv:2508.12926 **[math.ST]**

(or arXiv:2508.12926v1 **[math.ST]** for this version)

https://doi.org/10.48550/arXiv.2508.12926

Submission history

From: Richard Schwank [view email]

[v1] Mon, 18 Aug 2025 13:49:00 UTC (65 KB)

Access Paper:

View PDF

HTML (experimental)

TeX Source

Other Formats

view license

Current browse context:

math.ST

< prev | next >

new | recent | 2025-08

Change to browse by:

math

stat

stat.TH

References & Citations

NASA ADS

Google Scholar

Semantic Scholar

Export BibTeX Citation

Export citation

We prove $\|\boldsymbol{\mu} - \mathbf{m}\| = \mathcal{O}\left(\frac{1}{\sqrt{p}}\right)$ when components are M-dependent.

Discussion & Implications

- Similarities with (Brown 1983): sample geometric median becomes efficient for isotropic Gaussian as $p \to \infty$
- If M-dependence $\approx$ edge sparsity, can explain Robust SM's small bias on uncorrupted data
- Consistency check: rowwise contaminated observations generally *not* M-dependent
- Note: certain *cellwise* contaminated observations are M-dependent. Indeed, geometric median *not* robust (Raymaekers and Rousseeuw 2024)
- “Robustness" depends on contamination assumption!

Thank you!

- Schwank, R., McCormack, A., Drton, M. (2025). Robust score matching, AISTATS 2025.
- Schwank, R., Drton, M. (2025). On the distance between mean and geometric median in high dimensions. 10.48550/arXiv.2508.12926.

Contact: `richard.schwank@tum.de`

With funding from the:

Federal Ministry of Research, Technology and Space

This work is supported by the DAAD programme Konrad Zuse Schools of Excellence in Artificial Intelligence, sponsored by the Federal Ministry of Research, Technology and Space.

References

Brown, B. M. (Sept. 1983). "Statistical Uses of the Spatial Median". In: *Journal of the Royal Statistical Society: Series B (Methodological)* 45.1, pp. 25–30. ISSN: 2517-6161. DOI: 10.1111/j.2517-6161.1983.tb01226.x.

Minsker, Stanislav and Nate Strawn (July 2023). "The Geometric Median and Applications to Robust Mean Estimation". In: DOI: 10.48550/ARXIV.2307.03111. arXiv: 2307.03111 [math.ST].

Raymaekers, Jakob and Peter J. Rousseeuw (Feb. 2024). "Challenges of cellwise outliers". In: *Econometrics and Statistics*. ISSN: 2452-3062. DOI: 10.1016/j.ecosta.2024.02.002.