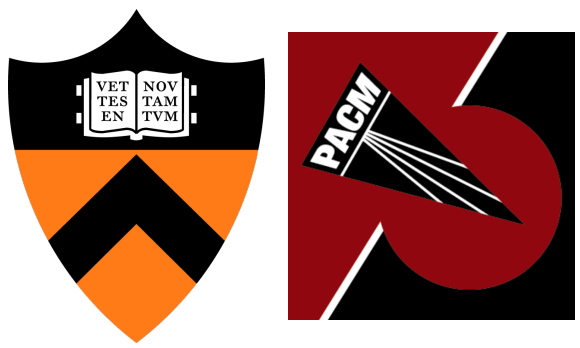




Generalized Method of Moments for Heteroscedastic Low-Rank Gaussian Mixtures

Spectral-Gradient-Reweight GMM (SGR-GMM): robust moment gradient weighting

Liu Zhang¹ and Amit Singer^{1,2}

¹Program in Applied and Computational Mathematics (PACM), Princeton University

²Department of Mathematics, Princeton University



Model: heteroscedastic low-rank Gaussian mixture (GM)

► K -component GM:

$$\mathbf{y}_n \mid (h_n = j) \sim \mathcal{N}(\boldsymbol{\mu}_j, \Sigma_j), \quad \mathbb{P}(h_n = j) = \pi_j,$$

for $n = 1, \dots, N$, $j = 1, \dots, K$,

► Heteroscedastic:

component-specific covariances Σ_j ,

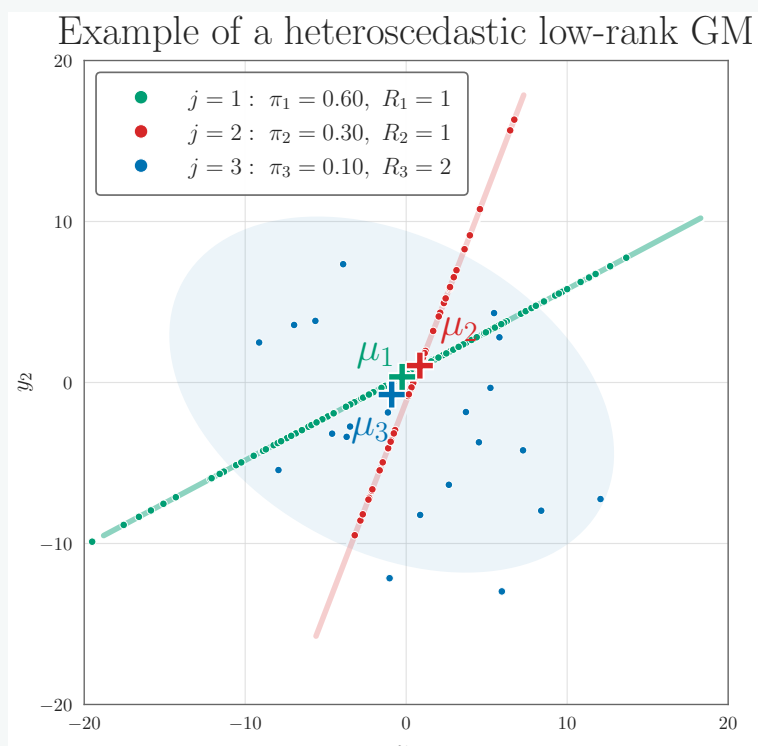
► Low-rank:

$$\Sigma_j = V_j V_j^\top, \quad \text{rank}(\Sigma_j) = R_j \leq R_{\max} \leq d.$$

Goal: to estimate

$$\boldsymbol{\theta}^* = (\pi_1^*, \dots, \pi_K^*; \boldsymbol{\mu}_1^*, \dots, \boldsymbol{\mu}_K^*; V_1^*, \dots, V_K^*)$$

$\in \Theta \subset \mathbb{R}^p$ from N observations, given K and $R_{\max}$.



Additive noise + strong ϵ -contamination

$$\boldsymbol{\xi}_n \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(\mathbf{0}, \Sigma_{\boldsymbol{\xi}}), \quad \tilde{\mathbf{y}}_n = \begin{cases} \mathbf{y}_n + \boldsymbol{\xi}_n, & (\text{if } n \in \mathcal{I}_{\text{in}}) \\ \mathbf{a}_n, & (\text{if } n \in \mathcal{I}_{\text{out}}) \end{cases},$$

with a partition $\{1, \dots, N\} = \mathcal{I}_{\text{in}} \sqcup \mathcal{I}_{\text{out}}, |\mathcal{I}_{\text{out}}| \leq \epsilon N$.

Generalized method of moments (GMM): outlier sensitivity

► Multi-step GMM optimization:

$$\hat{S}^{[t]} = \frac{1}{N} \sum_{n=1}^N g(\hat{\boldsymbol{\theta}}^{[t-1]}; \mathbf{y}_n) g(\hat{\boldsymbol{\theta}}^{[t-1]}; \mathbf{y}_n)^\top,$$

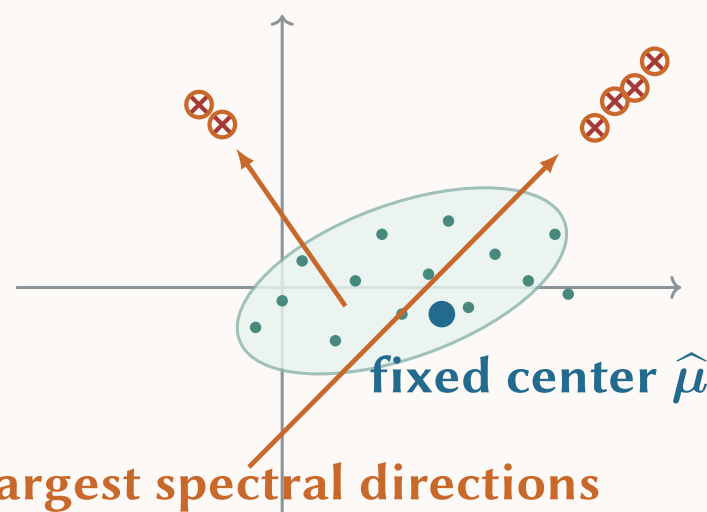
$$\hat{\boldsymbol{\theta}}^{[t]} = \arg \min_{\boldsymbol{\theta} \in \Theta} Q_N^{[t]}(\boldsymbol{\theta}) := \bar{g}_N(\boldsymbol{\theta})^\top \hat{W}^{[t]} \bar{g}_N(\boldsymbol{\theta}), \quad \hat{W}^{[t]} = (\hat{S}^{[t]})^{-1}.$$

Outlier sensitivity:

- Sample moment conditions $\bar{g}_N$: each observation has equal contribution.
- Effects of outliers get amplified.

SGR-GMM: motivations

Spectral principle (e.g., [LRV16; DKKLMS19; DKKLSS19; DM22; DK23]) for moment gradients: if a subset of outliers substantially shifts the gradient mean, it must create a large spectral direction in the gradient covariance.



largest spectral directions

► The feasibility problem is nonconvex in $\mathbf{w}$: to find $\mathbf{w} \in \Delta_{N,\epsilon}$, such that

$$\left\| \sum_n w_n (\bar{\mathbf{g}}_n^{(k)} - \bar{\mathbf{g}}_{\mathbf{w}}^{(k)}) (\bar{\mathbf{g}}_n^{(k)} - \bar{\mathbf{g}}_{\mathbf{w}}^{(k)})^\top \right\|_{\text{op}} \leq C_{\text{stop},k}, \quad \bar{\mathbf{g}}_{\mathbf{w}}^{(k)} = \sum_n w_n \bar{\mathbf{g}}_n^{(k)}.$$

► For a provisional fixed-center $\hat{\boldsymbol{\mu}}$, the feasibility problem is convex in $\mathbf{w}$:

$$\gamma(\mathbf{w}; \hat{\boldsymbol{\mu}}) = \left\| \sum_n w_n (\bar{\mathbf{g}}_n^{(k)} - \hat{\boldsymbol{\mu}}) (\bar{\mathbf{g}}_n^{(k)} - \hat{\boldsymbol{\mu}})^\top \right\|_{\text{op}} \leq C_{\text{stop},k}.$$

► Rewritten as a convex-concave game:

$$\text{OPT}(\hat{\boldsymbol{\mu}}) = \min_{\mathbf{w} \in \Delta_{N,\epsilon}} \gamma(\mathbf{w}; \hat{\boldsymbol{\mu}}) = \min_{\mathbf{w} \in \Delta_{N,\epsilon}} \max_{p \in \mathcal{D}_p} \sum_n w_n (\bar{\mathbf{g}}_n^{(k)} - \hat{\boldsymbol{\mu}})^\top \rho(\bar{\mathbf{g}}_n^{(k)} - \hat{\boldsymbol{\mu}}).$$

► Entropy-regularized coupled multiplicative-weights (MW) - matrix multiplicative-weights (MMW) algorithm (building on [AHK12]):

- **Density-matrix player:** expose the spectral directions with the largest weighted energy.
- **Gradient-weight player:** suppress the gradients most responsible for the exposed directions.

SGR primitive: solve, recenter, certify

Initialize center:
 $\hat{\boldsymbol{\mu}}^{[1]} =$
geometric median

Solve fixed-center game:
run T MW-MMW
rounds to obtain
 $\hat{\mathbf{w}}^{[s]} \approx \arg \min_{\mathbf{w} \in \Delta_{N,\epsilon}} \gamma(\mathbf{w}; \hat{\boldsymbol{\mu}}^{[s]})$

Recenter:
 $\hat{\boldsymbol{\mu}}^{[s+1]} = \sum_n \hat{w}_n^{[s]} \bar{\mathbf{g}}_n^{(k)}$

repeat until the stopping certificate is met
inspired by iteratively reweighted least squares [LM18; LLMZ25]

$$\text{Stopping certificate: } \gamma(\hat{\mathbf{w}}^{[s]}; \hat{\boldsymbol{\mu}}^{[s]}) = \left\| \sum_n \hat{w}_n^{[s]} (\bar{\mathbf{g}}_n^{(k)} - \hat{\boldsymbol{\mu}}^{[s]}) (\bar{\mathbf{g}}_n^{(k)} - \hat{\boldsymbol{\mu}}^{[s]})^\top \right\|_{\text{op}} \leq C_{\text{stop},k}.$$

SGR-GMM: robustify, reweight, certify

GMM optimization
via L-BFGS

Whenever L-BFGS locally
stabilizes, run SGR on
the moment gradients

Update gradient
weights, reset L-BFGS

continue GMM optimization until the stopping certificate is met

$$\text{Stopping certificate: } i = i_{\max} \text{ or } \left\| \hat{\boldsymbol{\theta}}^{[i]} - \hat{\boldsymbol{\theta}}^{[i-1]} \right\| \leq \epsilon_{\text{tol}}.$$

SGR-GMM: theoretical guarantees

1. Fixed-center game regret bound ([ZS26, Thm. 3.14])

$$\gamma(\hat{\mathbf{w}}^{[s]}; \hat{\boldsymbol{\mu}}^{[s]}) - \text{OPT}(\hat{\boldsymbol{\mu}}^{[s]}) \leq 4\nu \left[\sqrt{\frac{\log(1/(1-\epsilon))}{T}} + \sqrt{\frac{\log p}{T}} \right].$$

↓ inlier gradient stability

2. Convergence of the fixed-center updates ([ZS26, Thm. 3.23])

$$e^{[s+1]} \leq \alpha_\epsilon e^{[s]} + R_{\epsilon,T}, \quad e^{[s]} = \left\| \hat{\boldsymbol{\mu}}^{[s]} - \boldsymbol{\mu}_g^{(k)} \right\|_2, \quad \alpha_\epsilon = \sqrt{\frac{\epsilon}{1-2\epsilon}} < 1 \iff \epsilon < \frac{1}{3}.$$

↓ local GMM identification

3. Local finite-sample parameter error ([ZS26, Thm. 3.33])

$$\left\| \hat{\boldsymbol{\theta}}^{(\text{SGR-GMM})} - \boldsymbol{\theta}^* \right\|_2 \leq \frac{2}{\lambda^*} \left[\sum_{k=1}^L a_k (\delta_{\mu,k}(\zeta) + \alpha_\epsilon \sqrt{C_k}) + \delta_{\text{opt}} \right].$$

- λ^* : local GMM identification strength,
- $\delta_{\mu,k}$: inlier gradient stability,
- δ_{opt} : numerical optimizer accuracy.

Note. Theorem 3.33 is a local theorem and does not claim global convergence of the nonconvex GMM objective from arbitrary initialization.

Robust DGMM

DGMM weights: $\hat{w}_1^{[t]}, \dots, \hat{w}_L^{[t]} \in \mathbb{R}$



SGR-GMM weights: $\hat{\mathbf{w}}^{(k)} = (\hat{w}_1^{(k)}, \dots, \hat{w}_N^{(k)}) \in \Delta_{N,\epsilon}$

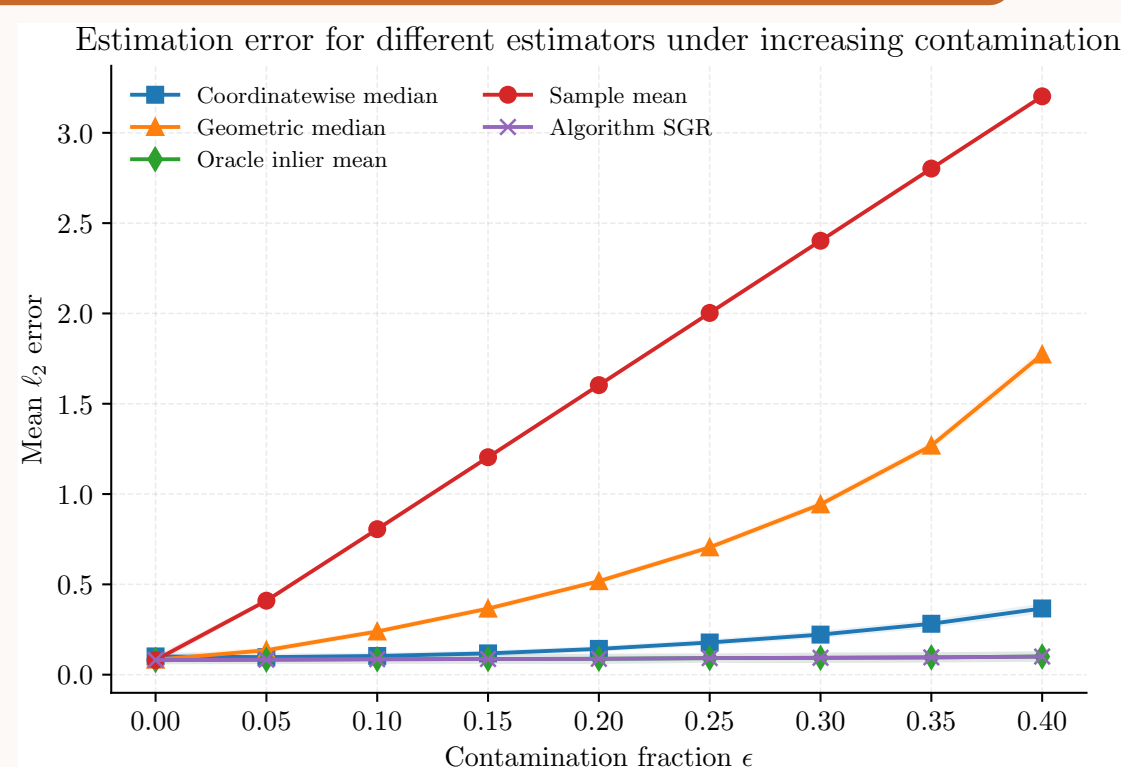


Robust DGMM gradient:

$$\nabla_{\boldsymbol{\theta}} Q_N^{[t]}(\boldsymbol{\theta}) = \sum_{k=1}^L \hat{w}_k^{[t]} \left(\nabla_{\boldsymbol{\theta}} \check{\phi}^{(k)}(\boldsymbol{\theta}; \boldsymbol{\xi}) - 2 \sum_{n=1}^N \hat{w}_n^{(k)} \nabla_{\boldsymbol{\theta}} \check{\psi}^{(k)}(\boldsymbol{\theta}; \boldsymbol{\xi}, \tilde{\mathbf{y}}_n) \right)$$

$\check{\phi}$ and $\check{\psi}$ incorporate $\Sigma_{\boldsymbol{\xi}}$ through the Gaussian cumulants.

SGR primitive numerical results



Isolated SGR experiment

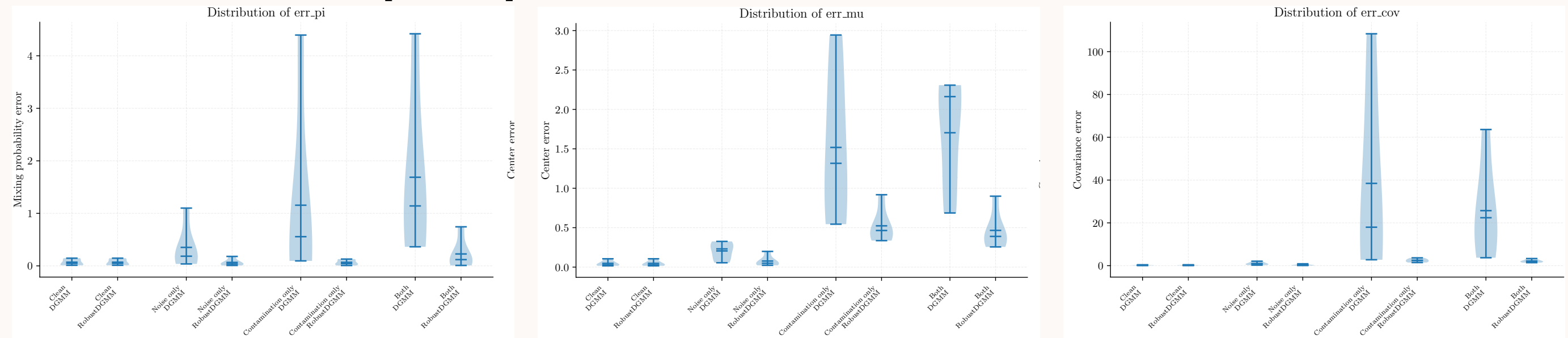
- $p = 10$,
- $N = 600$,
- directional Gaussian replacement outliers.

Near-oracle weighting

Even with increasing ϵ , the SGR error differs from the oracle inlier-mean error by at most 7.24×10^{-5} .

Robust DGMM numerical results

Baseline DGMM [ZMXS25] vs. RobustDGMM: $d = 5, K = 2, N = 1000, L = 4$, five random trials.



Clean GM
Numerically identical.

GM + contamination
baseline DGMM error vs. RobustDGMM error:
 $17.1 \times (\text{Err}_{\pi}), 2.9 \times (\text{Err}_{\mu}), 15.6 \times (\text{Err}_{\Sigma})$.

GM + additive noise
baseline DGMM error vs. RobustDGMM error:
 $5.4 \times (\text{Err}_{\pi}), 2.7 \times (\text{Err}_{\mu}), 2.9 \times (\text{Err}_{\Sigma})$.

GM + additive noise + contamination
baseline DGMM error vs. RobustDGMM error:
 $6.2 \times (\text{Err}_{\pi}), 3.5 \times (\text{Err}_{\mu}), 12.6 \times (\text{Err}_{\Sigma})$.

Acknowledgements and references

Acknowledgements:

The SGR-GMM work is supported in part by AFOSR FA9550-23-1-0249, DARPA HR0011-25-3-E002, NSF DMS 2510039, and the Simons Foundation Math+X Investigator Award.

References:

[AHK12]

[LRV16]

[LM18]

S. Arora, E. Hazan, and S. Kale. "The multiplicative weights update method: a meta-algorithm and applications". In: *Theory Comput.* 8 (2012), pp. 121–164. issn: 1557-2862. doi: 10.4086/toc.2012.v008a006. [DKKLMS19]
K. A. Lai, A. B. Rao, and S. Vempala. "Agnostic estimation of mean and covariance". In: *2016 IEEE 57th Annual Symposium on Foundations of Computer Science (FOCS)*. IEEE, 2016, pp. 665–674. [DKKLSS19]
G. Lerman and T. Maunu. "Fast, robust and non-convex subspace recovery". In: *Inf. Inference* 7.2 (2018), pp. 277–336. issn: 2049-8764,2049-8772. doi: 10.1093/imaiai/iax012. [DM22]

I. Diakonikolas, G. Kamath, D. Kane, J. Li, A. Moitra, and A. Stewart. "Robust estimators in high-dimensions without the computational intractability". In: *SIAM Journal on Computing* 48.2 (2019), pp. 742–864. [DK23]
I. Diakonikolas, G. Kamath, D. Kane, J. Li, J. Steinhardt, and A. Stewart. "SEVER: A robust meta-algorithm for stochastic optimization". In: *International Conference on Machine Learning*. PMLR, 2019, pp. 1596–1606. [LLMZ25]
A. S. Dalalyan and A. Minasyan. "All-in-one robust estimator of the Gaussian mean". In: *Ann. Statist.* 50.2 (2022), pp. 1193–1219. issn: 0090-5364,2168-8966. doi: 10.1214/21-aos2145. [ZMXS25]
[ZS26]

I. Diakonikolas and D. M. Kane. *Algorithmic high-dimensional robust statistics*. Cambridge university press, 2023.
G. Lerman, K. Li, T. Maunu, and T. Zhang. "Global Convergence of Iteratively Reweighted Least Squares for Robust Subspace Recovery". In: *arXiv preprint arXiv:2506.20533* (2025).
L. Zhang, O. Mickelin, S. Xu, and A. Singer. "Diagonally-Weighted Generalized Method of Moments Estimation for Gaussian Mixture Modeling". In: *arXiv preprint arXiv:2507.20439* (2025).
L. Zhang and A. Singer. "Robust Moment-Based Estimation via Spectral Gradient Reweighting". In: *arXiv preprint arXiv:2605.27718* (2026).