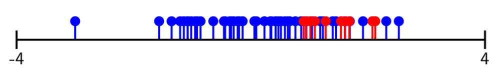
Quantum Entropy Scoring for Fast Robust Mean Estimation and Outlier Detection

Yihe Dong (Microsoft Research), Sam Hopkins (UC Berkeley), Jerry Li (Microsoft Research)

Problem setup

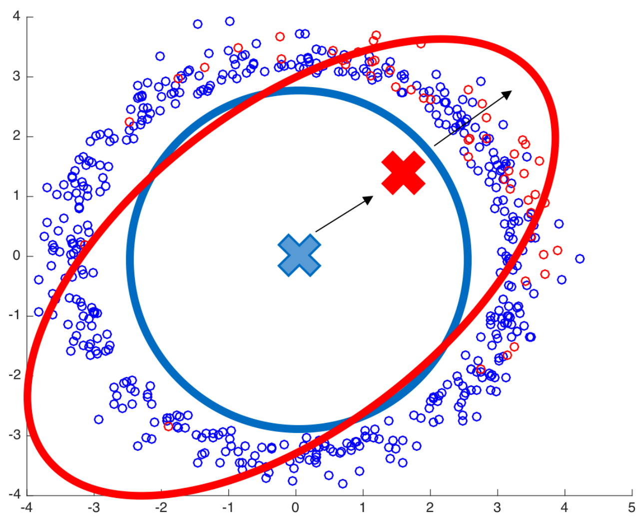
- Given a distribution D over $\mathbb{R}^d$ with mean μ, let $X \in \mathbb{R}^{n\times d}$ be n i.i.d. samples from D, with ϵ-fraction of them corrupted, efficiently detect the corrupted data and estimate the sample mean $\hat{\mu}$, with good error bound.



- This is **nontrivial**: Naïve estimates have errors that scale with dimension.
- **Many applications**: robust regression; detecting fraud, medical anomalies, network traffic irregularities; etc.

Prior Work

- Naïve spectral: filtering method based on projection onto top eigenvector, $\tilde{O}(nd^2)$ complexity.
- Best prior result:
    - $\tilde{O}(\min(nd^2, nd/\epsilon^6))$ time complexity.
    - d-independent error bound: $\|\mu - \hat{\mu}\|_2 \leq O(\epsilon)$.



- Collective inductive bias can be detected by the spectra.

Our contribution

- QUE-scoring: **nearly linear time complexity $\tilde{O}(nd)$.**
- d-independent error bound.
- QUE **interpolates** between scores based on l_2-norm and projection onto the top eigenvalue, controlled by α.
    - Inspecting multiple directions at once.

$$QUE(X_i) = (X_i - \mathbb{E}[X])^\top U (X_i - \mathbb{E}[X]),\ U = \frac{\exp(\alpha\ \mathrm{cov}(X))}{\mathrm{tr}\left(\exp(\alpha\ \mathrm{cov}(X))\right)}$$

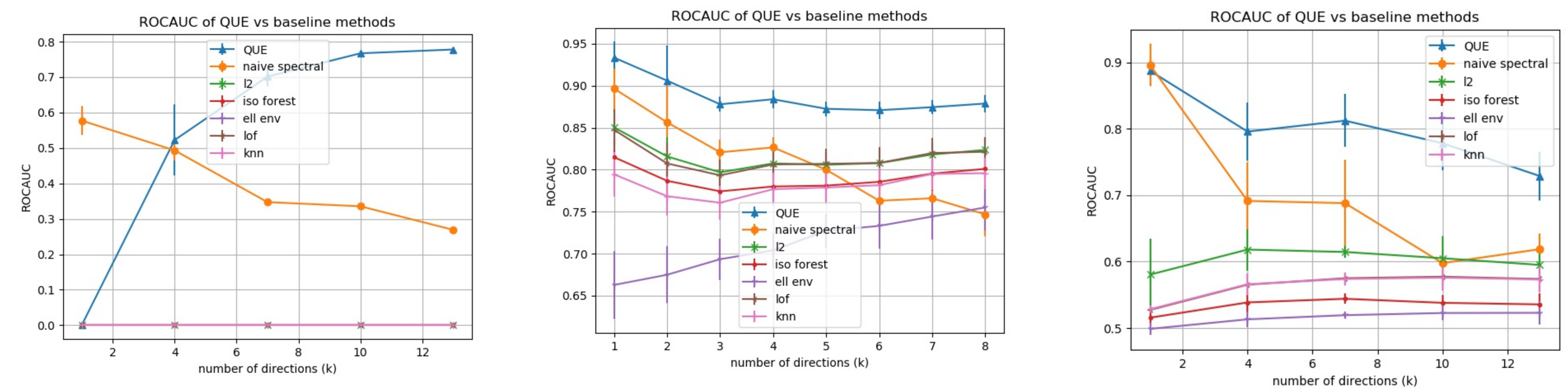
- Lower α -> like l_2, higher α -> like naïve spectral.
- **Fast computation** possible by combining
    - fast Johnson-Lindenstrauss.
    - Chebyshev expansion of $\exp(\mathrm{cov}(X))$.
    - fast Hadamard transform.
- Works well in high dimensions

Experiments: datasets

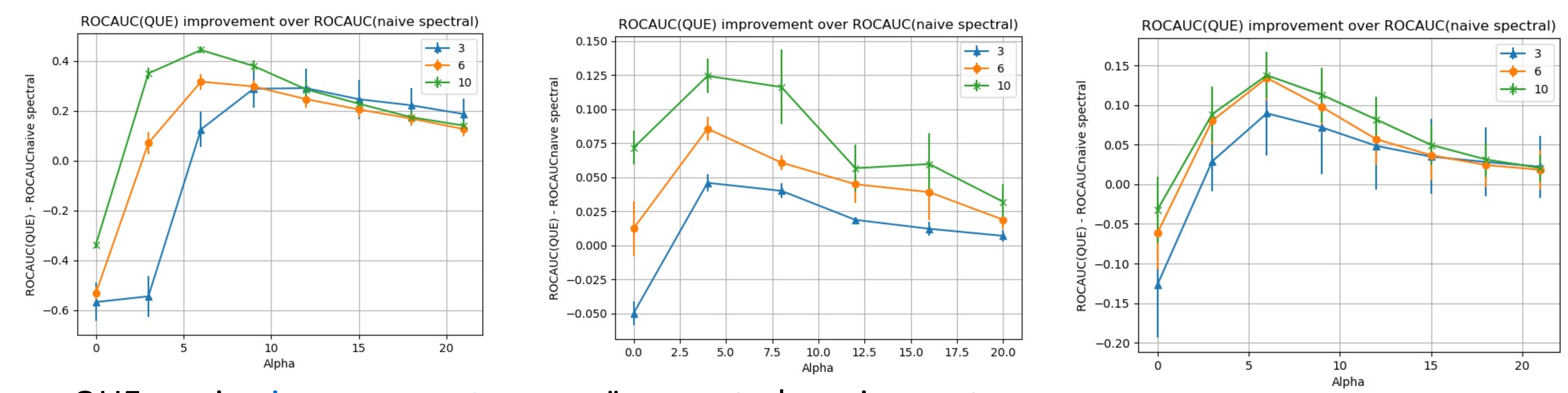
- Synthetic:
    - Inliers: i.i.d. samples from $N(0,\ \mathrm{Id})$.
    - Outliers: i.i.d. samples from mixture of Gaussians $N(\sqrt{\frac{k}{\epsilon}}e_i, \sigma^2\mathrm{Id})$.
- Text:
    - Inliers: word embeddings of sections of *Sherlock Holmes.*
    - Outliers: word embeddings of Wikipedia articles.
- Image:
    - Inliers: CIFAR images
    - Outliers: images with corrupted pixels

Experiments: Results

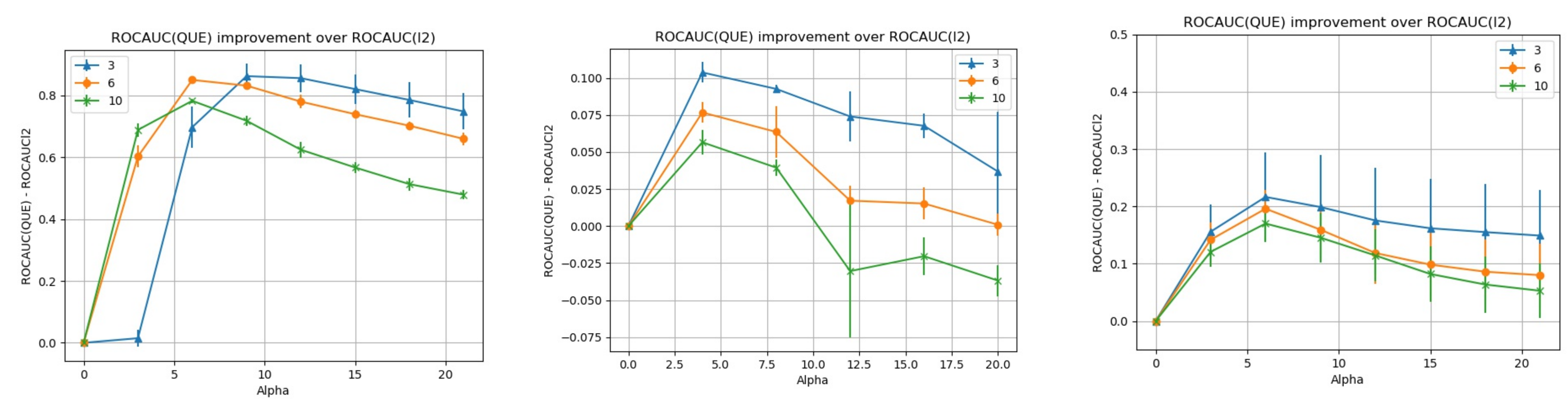
Synthetic | **Text** | **Images**



QUE scoring and baselines v.s. number of outlier directions.



QUE scoring improvement over naïve spectral scoring w.r.t. α:



QUE scoring improvement over l_2 scoring w.r.t. α: