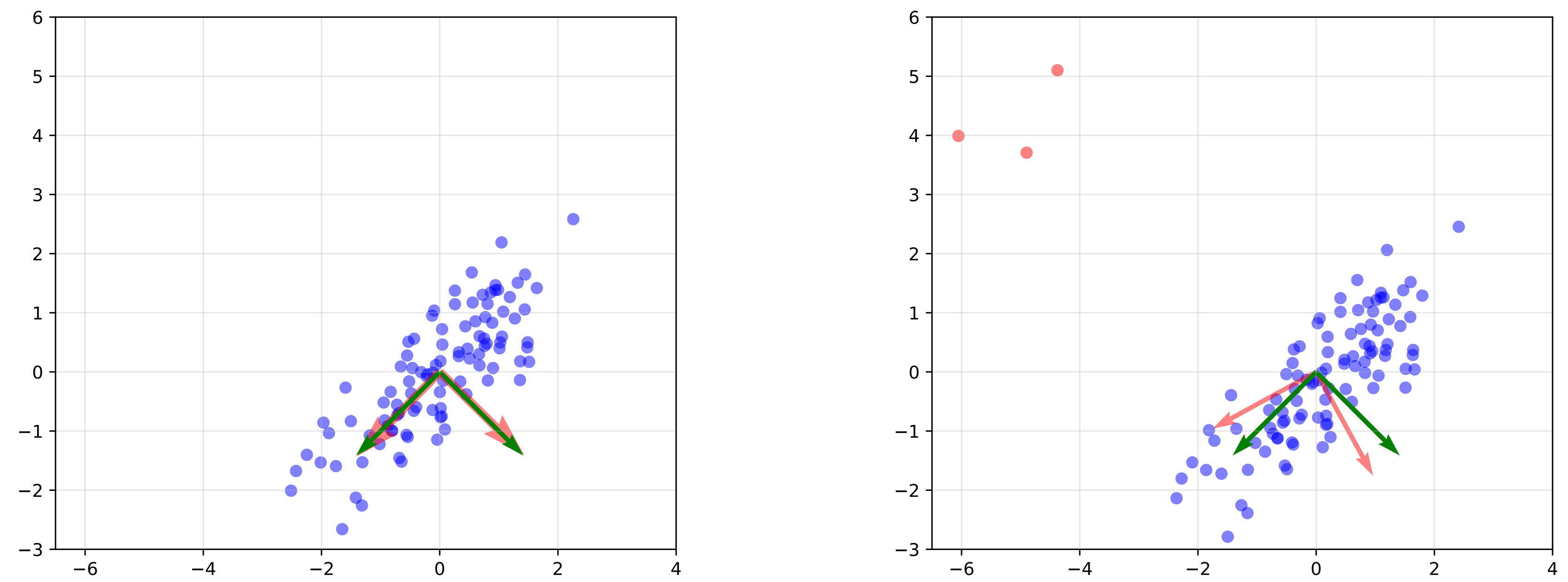
Tyler's M-estimator Through the Lens of Convex-Concave Programming

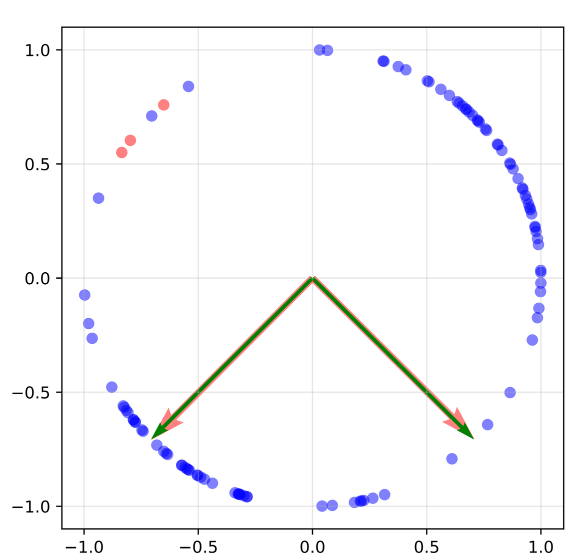
Daniel Cederberg • *Stanford University* • AISTATS 2026

Robust covariance estimation



- normal data points (blue) are corrupted by outliers (red)
- classical covariance estimators are sensitive to outliers
- green arrows = true principal components, red = sample principal components

Angular central Gaussian distribution



- normalize samples to the unit sphere: $X := Z/\|Z\|_2$ where $Z \sim \mathcal{N}(0, \Sigma)$
- outliers can only influence direction, not magnitude
- density of X is *angular central Gaussian* (ACG):

$$p(x; \Sigma) = \frac{\Gamma(n/2)}{2\pi^{n/2}\sqrt{\det \Sigma}}(x^T\Sigma^{-1}x)^{-n/2}$$

Tyler's M-estimator (Tyler, 1987)

$$\text{minimize } G(\Sigma) := \log\det\Sigma + \frac{n}{m}\sum_{i=1}^{m}\log(x_i^T\Sigma^{-1}x_i)$$

- variable $\Sigma \in \mathbf{S}^n$, data $x_1, \ldots, x_m \in \mathbf{R}^n$
- MLE for the angular central Gaussian distribution
- computed via fixed-point iteration:

$$\Sigma_{k+1} = \frac{n}{m}\sum_{i=1}^{m}\frac{x_i x_i^T}{x_i^T\Sigma_k^{-1}x_i}$$

- non-convex problem, but global *asymptotic* convergence known [1]
- we derive the **first deterministic global *non-asymptotic* convergence rate**

Key idea: Tyler's fixed-point iteration is CCP

$$\text{minimize } \underbrace{-\log\det Y}_{f(Y)} - \underbrace{\left(-\tfrac{n}{m}\textstyle\sum_i \log(x_i^T Y x_i)\right)}_{g(Y)}$$

- a difference of convex functions after change of variables $Y = \Sigma^{-1}$
- minimize $f(Y) - g(Y)$ via the **convex-concave procedure**:

$$Y_{k+1} = \operatorname*{argmin}_Y\{f(Y) - (g(Y_k) + \langle\nabla g(Y_k), Y - Y_k\rangle)\}.$$

- equivalent to Tyler's fixed-point iteration
- this perspective unlocks an optimization-based analysis

KL geometry: relative smoothness

- for progress-per-iteration analysis we need a smoothness property
- g is not Euclidean smooth, but it is relatively smooth in KL geometry

$$g(Y) \le g(Z) + \langle\nabla g(Z), Y - Z\rangle + LD_{\mathrm{KL}}(Y\|Z)$$

- the KL divergence is

$$D_{\mathrm{KL}}(Y\|Z) = \tfrac{1}{2}\left(\mathbf{Tr}(Z^{-1}Y) + \log\det(ZY^{-1}) - n\right)$$

Riemannian geometry: geodesic convexity

- for global convergence we need a convexity property
- G is not Euclidean convex, but geodesically convex

$$G(\gamma(t; X, Y)) \le (1-t)G(X) + tG(Y)$$

- the geodesic between X and Y is

$$\gamma(t; X, Y) = X^{1/2}(X^{-1/2}YX^{-1/2})^t X^{1/2}, \qquad t \in [0, 1]$$

- geodesic distance measure

$$D_{\mathrm{geo}}(X, Y) = \|\log(X^{-1/2}YX^{-1/2})\|_F$$

Bridging lemma

Lemma. Along the geodesic from iterate Y_k to the optimum $Y^\star$,

$$D_{\mathrm{KL}}(\gamma(t; Y_k, Y^\star)\,\|\,Y_k) \;\le\; C\,D_{\mathrm{geo}}(\gamma(t; Y_k, Y^\star), Y_k)^2.$$

- the two geometries are not compatible in general
- but along geodesics to the optimum, D_{KL} is controlled by D^2_{geo}
- C depends on the conditioning of the optimal solution $\Sigma^\star$

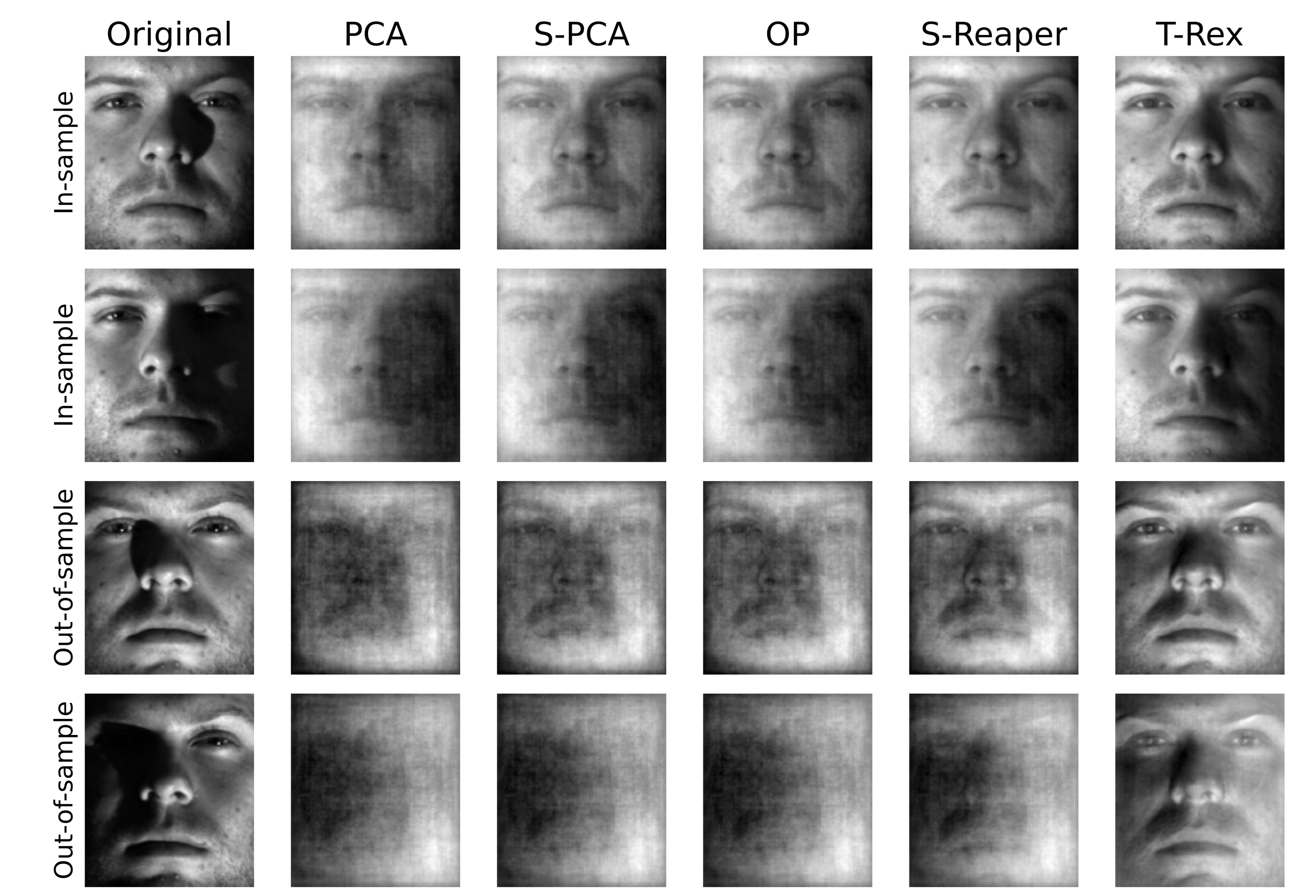
Main result

Theorem 1. For Tyler's fixed-point iteration,

$$G(\Sigma_k) - G(\Sigma^\star) \;\le\; O(1/k), \qquad k \ge 1.$$

- **first deterministic** $O(1/k)$ **convergence rate**
- we also establish **linear convergence** for a regularized variant

Extension: structured estimation



- the CCP framework extends to **structural constraints** on Σ or its inverse
- for example, factor model $\Sigma = FF^T + D$
- each yields a convex CCP subproblem (solvable with CVXPY)
- **rightmost column**: faces recovered by Tyler (factor-model variant from [2])

Takeaways

- first deterministic $O(1/k)$ rate for Tyler's fixed-point iteration (open for nearly 40 years)
- a Euclidean algorithm analyzed by combining relative smoothness with Riemannian optimization — a template likely to generalize.

References

[1] Kent & Tyler (1988). Maximum likelihood estimation for the wrapped Cauchy distribution. *Journal of Applied Statistics*.

[2] Cederberg (2026). T-Rex: Fitting a Robust Factor Model via Expectation-Maximization. *IEEE Transactions on Signal Processing*.