DECENTRALIZED COMPUTATION OF TYLER'S M-ESTIMATOR

Anonymous authors
Paper under double-blind review

ABSTRACT

We study decentralized computation of Tyler's M-estimator over a peer-to-peer network. A central difficulty is that the estimator may exist for the pooled data while being undefined at individual nodes, whose local weighted-scatter matrices can be singular. We propose a decentralized majorization–minimization method (D-MM) that combines local weighted-scatter updates with finitely many rounds of neighbor averaging and requires the Tyler existence condition only for the pooled observations. Relative Loewner-order bounds preserve positive definiteness after finite consensus, while Hilbert's projective metric controls both inter-node disagreement and the error caused by heterogeneous local iterates. Under summable consensus errors, all node estimates converge to the pooled Tyler estimator, with explicit finite-step and communication guarantees for polynomial schedules. We further show that, on the determinant-one manifold, the strict occupancy condition induces positive data-dependent geodesic curvature on the invariant region. This yields an instance-dependent linear convergence rate for the unregularized centralized Tyler iteration, sharpening the recent deterministic $O(1/k)$ objective-gap guarantee without regularization. Under geometrically decaying consensus errors, D-MM inherits this contraction and converges linearly. Experiments confirm the predicted effects of communication, network connectivity, and data geometry, and demonstrate agreement with pooled Tyler estimation in a rolling portfolio application.

1 INTRODUCTION

Robust estimation of covariance and shape matrices is a fundamental problem in multivariate statistics and machine learning (Huber & Ronchetti, 2011; Maronna et al., 2019). Tyler's M-estimator is a classical estimator of multivariate shape for elliptical distributions (Cederberg, 2026; Tyler, 1987; Wiesel & Zhang, 2015). It estimates a positive definite shape matrix up to scale and is invariant to arbitrary nonzero rescaling of individual observations, making it particularly suitable for heavy-tailed data. Given nonzero observations $\{\boldsymbol{x}_s\}_{s=1}^N \subset \mathbb{R}^d$, the classical Tyler iteration is

$$\mathcal{T}(\boldsymbol{\Sigma}) = \frac{d}{N}\sum_{s=1}^{N}\frac{\boldsymbol{x}_s\boldsymbol{x}_s^\top}{\boldsymbol{x}_s^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{x}_s}, \qquad \boldsymbol{\Sigma}^{k+1} = \frac{\mathcal{T}(\boldsymbol{\Sigma}^k)}{\det(\mathcal{T}(\boldsymbol{\Sigma}^k))^{1/d}}. \tag{1}$$

Under the standard strict subspace-occupancy condition, the normalized estimator exists and is unique (Franks & Moitra, 2020; Tyler, 1987).

We study Tyler estimation when the observations are distributed over a peer-to-peer network. The pooled data may satisfy the existence condition even when the samples stored at an individual node do not; in particular, a local weighted-scatter matrix may be singular. Consequently, decentralized Tyler estimation cannot in general be reduced to computing local Tyler estimators and subsequently aggregating them.

To compute the pooled estimator directly over the network, we propose a decentralized majorization–minimization method (D-MM). At each outer iteration, every node forms a weighted scatter from its local observations, performs finitely many rounds of neighbor averaging, and determinant-normalizes the resulting aggregate. D-MM requires the strict occupancy condition only for the pooled observations and does not require local Tyler estimators to exist. Analyzing D-MM

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

with finite communication raises two coupled difficulties. First, finite consensus produces only an approximate average of the local weighted scatters. Second, these scatters are evaluated at different node iterates, so their aggregate is not simply an inexact evaluation of the centralized Tyler map. These effects must be controlled while ensuring positive definiteness of every update, even though the individual local contributions may be singular. Standard additive consensus bounds do not directly provide such a guarantee.

We address these difficulties through relative perturbation analysis on the positive definite cone. A relative mixing condition yields Loewner-order control of the finite-consensus outputs and preserves positive definiteness after aggregation. Hilbert's projective metric then controls both inter-node disagreement and the deviation from a centralized Tyler step, allowing the entire decentralized trajectory to be confined to a common compact region.

This framework yields two convergence regimes. Under summable consensus errors, all node estimates converge to the pooled Tyler estimator, with explicit finite-step and communication guarantees for polynomial schedules. We further show that restricting the scale-invariant Tyler problem to the determinant-one manifold removes the flat scaling direction and reveals positive, data-dependent geodesic curvature on the invariant region. This yields an instance-dependent linear convergence rate for the unregularized centralized Tyler iteration and links its rate to the geometry of the observed data. Under geometrically decaying consensus errors, D-MM inherits this contraction and converges linearly as well. Our main contributions are summarized as follows.

(i) Decentralized Tyler estimation without local estimators. We propose D-MM for computing the pooled Tyler estimator using only finite neighbor communication, without requiring local Tyler estimators to exist. Relative consensus accuracy guarantees well-defined positive definite updates despite singular local weighted-scatter matrices.

(ii) Instance-dependent linear convergence of the centralized Tyler iteration. On the determinant-one manifold, strict occupancy yields positive geodesic curvature over the compact invariant region. This gives an instance-dependent linear rate for the unregularized centralized Tyler iteration and makes explicit how the rate depends on the geometry of the observed data.

(iii) Finite-communication convergence of D-MM. We establish stability and convergence under summable consensus errors and derive explicit finite-step guarantees for polynomial schedules. Under geometric consensus accuracy, D-MM preserves the centralized contraction and converges R-linearly.

(iv) Communication complexity and empirical evaluation. We derive explicit communication budgets for polynomial and geometric schedules and evaluate D-MM on synthetic and real-data problems, including the effects of data geometry, network connectivity, and communication accuracy.

2 Related work

Tyler estimation and convergence. Tyler's M-estimator and its existence theory are classical (Tyler, 1987; Wiesel & Zhang, 2015). Its fixed-point iteration admits an MM interpretation (Wiesel & Zhang, 2015), while regularized and structured variants address high-dimensional or constrained settings (Goes et al., 2020; Ollila & Tyler, 2014; Sun et al., 2014). Recent work has studied its computational behavior from several perspectives. Franks & Moitra (2020) connect Tyler estimation to operator scaling and establish linear convergence under a generative model with suitable expansion properties; related algorithmic and statistical analyses include Danon & Garber (2022) and Lau & Ramachandran (2025). Most closely, Cederberg (2026) proves a deterministic $O(1/k)$ objective-gap rate for the unregularized iteration and a linear rate for a regularized variant. In contrast, our linear result is deterministic and applies to every fixed dataset satisfying the standard strict occupancy condition, although its contraction factor is instance dependent and need not be uniform.

Decentralized optimization and estimation. Neighbor averaging is central to distributed subgradient methods, ADMM, EXTRA, DIGing, NEXT, and nested consensus schemes (Berahas et al., 2019; Boyd et al., 2011; Di Lorenzo & Scutari, 2016; Nedić & Ozdaglar, 2009; Nedić et al., 2017; Shi et al., 2015; Xiao & Boyd, 2004). Related statistical settings include distributed sparse regression, penalized Gaussian precision estimation, and decentralized robust covariance estimation (Ma-

teos et al., 2010; Xia et al., 2025), as well as decentralized optimization on matrix manifolds (Chen et al., 2021). These methods do not directly address pooled unregularized Tyler estimation when local weighted scatters may be singular. Our analysis instead uses relative Loewner-order perturbations and Hilbert's projective metric to preserve positive definiteness and control finite-consensus errors without local identifiability assumptions.

3 Problem Formulation and Decentralized MM

3.1 Pooled Tyler estimation

Consider a network of $m \geq 2$ nodes that collectively store N nonzero observations $\{\boldsymbol{x}_s\}_{s=1}^N \subset \mathbb{R}^d$. Node i holds the samples indexed by $\mathcal{I}_i$, where $\{\mathcal{I}_i\}_{i=1}^m$ form a disjoint partition of $\{1, \dots, N\}$. We allow the local sample sizes $N_i = |\mathcal{I}_i|$ to be arbitrarily small, including $N_i < d$, and impose no local identifiability condition. Our goal is to compute the Tyler shape associated with the pooled observations over the network.

Besides the fixed-point characterization in (1), Tyler's estimator admits the equivalent variational characterization

$$\boldsymbol{\Sigma}^\star \in \arg\min_{\boldsymbol{\Sigma} \in \mathcal{M}} G(\boldsymbol{\Sigma}), \qquad G(\boldsymbol{\Sigma}) \coloneqq \log\det \boldsymbol{\Sigma} + \frac{d}{N} \sum_{s=1}^N \log\big(\boldsymbol{x}_s^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{x}_s\big)\,, \tag{2}$$

where $\mathcal{M} \coloneqq \{\boldsymbol{\Sigma} \in \mathbb{S}_{++}^d : \det \boldsymbol{\Sigma} = 1\}$ fixes the scale ambiguity. Indeed, $G(c\boldsymbol{\Sigma}) = G(\boldsymbol{\Sigma})$ for every $c > 0$. Although G is generally not convex in the Euclidean sense, it is geodesically convex over the positive definite cone (Wiesel & Zhang, 2015).

The existence and uniqueness of the normalized solution are guaranteed by the standard strict subspace-occupancy condition on the pooled data.

Assumption 1 (Pooled data)**.** For every linear subspace $V \subset \mathbb{R}^d$ with $0 < \dim V < d$,

$$\frac{\#\{s : \boldsymbol{x}_s \in V\}}{N} < \frac{\dim V}{d}. \tag{3}$$

Under Assumption 1, the minimizer in (2) exists and is unique (Franks & Moitra, 2020; Tyler, 1987); we henceforth use $\boldsymbol{\Sigma}^\star$ to denote this unique determinant-normalized Tyler estimator. Importantly, Assumption 1 is imposed only on the pooled observations; no analogous condition is required at any individual node.

The variational problem (2) is equivalent to the Tyler fixed-point characterization in (1). Moreover, the centralized Tyler iteration is the standard MM iteration associated with G. This interpretation is particularly useful for our analysis, as the resulting surrogate decrease provides the descent property underlying the finite-step and linear convergence guarantees in Section 4. The fixed-point equivalence and the MM surrogate construction are recalled in Appendix A.

To expose the decentralized structure, we decompose the pooled weighted-scatter map according to the data partition:

$$\mathcal{T}(\boldsymbol{\Sigma}) = \sum_{i=1}^m \boldsymbol{H}_i(\boldsymbol{\Sigma}), \qquad \boldsymbol{H}_i(\boldsymbol{\Sigma}) = \frac{d}{N} \sum_{s \in \mathcal{I}_i} \frac{\boldsymbol{x}_s \boldsymbol{x}_s^\top}{\boldsymbol{x}_s^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{x}_s}. \tag{4}$$

Under Assumption 1, the pooled matrix $\mathcal{T}(\boldsymbol{\Sigma})$ is positive definite for every $\boldsymbol{\Sigma} \in \mathbb{S}_{++}^d$. In contrast, an individual $\boldsymbol{H}_i(\boldsymbol{\Sigma})$ may be singular because the local observations are not required to satisfy Assumption 1. The decentralized problem is therefore to recover the pooled MM update from the local contributions $\{\boldsymbol{H}_i\}_{i=1}^m$ using only neighbor-to-neighbor communication.

3.2 Network model

The nodes communicate over a fixed undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ through a symmetric mixing matrix $\boldsymbol{W} \in \mathbb{R}^{m \times m}$.

Assumption 2 (Network)**.** The matrix $\boldsymbol{W}$ is nonnegative, doubly stochastic, and compatible with $\mathcal{G}$. With $\boldsymbol{J} = \frac{1}{m}\mathbf{1}\mathbf{1}^\top$, we assume

$$\rho := \|\boldsymbol{W} - \boldsymbol{J}\|_2 < 1. \tag{5}$$

This condition is satisfied, for example, by lazy Metropolis weights on a connected graph (Xiao & Boyd, 2004). For implementation, it suffices to know a certified upper bound $\rho \le \bar{\rho} < 1$. We assume synchronous and reliable communication throughout.

3.3 Decentralized MM with finite consensus

We construct *decentralized MM* (D-MM) by combining the local MM computations in (4) with finite-round consensus. Starting from a common $\boldsymbol{\Sigma}^0 \in \mathcal{M}$, each node i maintains a local iterate $\boldsymbol{\Sigma}_i^k$. At each outer iteration k, D-MM performs a local weighted-scatter update, q_k rounds of neighbor averaging, and determinant normalization:

$$\textit{Local update}: \quad \boldsymbol{Y}_i^{k,0} = m\boldsymbol{H}_i(\boldsymbol{\Sigma}_i^k) = \frac{md}{N}\sum_{s\in\mathcal{I}_i}\frac{\boldsymbol{x}_s\boldsymbol{x}_s^\top}{\boldsymbol{x}_s^\top(\boldsymbol{\Sigma}_i^k)^{-1}\boldsymbol{x}_s}, \tag{6a}$$

$$\textit{Consensus}: \quad \boldsymbol{Y}_i^{k,r+1} = \sum_{j=1}^m w_{ij}\boldsymbol{Y}_j^{k,r}, \qquad r = 0,\dots,q_k-1, \tag{6b}$$

$$\textit{Normalization}: \quad \boldsymbol{\Sigma}_i^{k+1} = \frac{\boldsymbol{Y}_i^{k,q_k}}{\left(\det \boldsymbol{Y}_i^{k,q_k}\right)^{1/d}}. \tag{6c}$$

After q_k communication rounds, the consensus output at node i can be written as

$$\boldsymbol{Y}_i^{k,q_k} = m\sum_{j=1}^m (\boldsymbol{W}^{q_k})_{ij}\boldsymbol{H}_j(\boldsymbol{\Sigma}_j^k). \tag{7}$$

Equation (7) highlights the two sources of inexactness introduced by decentralization. First, finite consensus only approximately averages the local matrices. Second, the local maps $\boldsymbol{H}_j(\boldsymbol{\Sigma}_j^k)$ are generally evaluated at different node iterates.

We quantify the accuracy of finite consensus through the relative mixing condition

$$\max_{i,j}|m(\boldsymbol{W}^{q_k})_{ij} - 1| \le \delta_k, \qquad 0 < \delta_k \le \bar{\delta} < 1. \tag{8}$$

Since $|(\boldsymbol{W}^q - \boldsymbol{J})_{ij}| \le \|\boldsymbol{W}^q - \boldsymbol{J}\|_2 \le \rho^q$, it is sufficient to choose

$$q_k = \left\lceil \frac{\log(m/\delta_k)}{-\log\bar{\rho}} \right\rceil, \tag{9}$$

where $\rho \le \bar{\rho} < 1$ is the certified network bound introduced above. Reported D-MM experiments use lazy Metropolis weights and set $\bar{\rho} = \rho = \|\boldsymbol{W} - \boldsymbol{J}\|_2$ by an offline spectral calculation for the known graph. Appendix F gives a spectrum-free bound for the same weights and its matched-horizon communication cost. Hence, any prescribed consensus accuracy δ_k can be achieved using finitely many communication rounds.

A key feature of D-MM is that the local samples need not satisfy Assumption 1. Consequently, the matrices $\boldsymbol{H}_i(\boldsymbol{\Sigma}_i^k)$ in (6a) may be singular. D-MM never inverts these local matrices, nor does it require the intermediate consensus states to be positive definite. Positive definiteness is required only for the finite-consensus output entering the normalization step (6c). In Section 4, we show that the relative mixing condition (8) is sufficient to guarantee positive definiteness at every node despite singular local contributions.

4 Convergence Analysis

We now establish convergence guarantees for D-MM with finitely many consensus rounds at each outer iteration. The analysis proceeds in two regimes. We first consider a general class of communication schedules whose accumulated projective consensus error is summable and derive a curvature-free finite-step convergence bound. We then exploit the instance-dependent curvature of the Tyler

objective on the invariant region to obtain a perturbed linear contraction, which becomes an R-linear convergence guarantee when the consensus error decays geometrically. Polynomial and geometric communication schedules are given explicitly in the corresponding subsections.

For convenience, define the normalized centralized Tyler map

$$\mathcal{F}(\mathbf{\Sigma}) \coloneqq \frac{\mathcal{T}(\mathbf{\Sigma})}{\big(\det \mathcal{T}(\mathbf{\Sigma})\big)^{1/d}},$$

so that the centralized iteration in (1) is $\mathbf{\Sigma}^{k+1} = \mathcal{F}(\mathbf{\Sigma}^k)$. We also write

$$\Delta(\mathbf{\Sigma}) \coloneqq G(\mathbf{\Sigma}) - G(\mathbf{\Sigma}^\star).$$

For $\boldsymbol{A}, \boldsymbol{B} \in \mathbb{S}^d_{++}$, define Hilbert's projective distance and the affine-invariant distance by

$$d_{\mathrm{H}}(\boldsymbol{A}, \boldsymbol{B}) \coloneqq \log \frac{\lambda_{\max}(\boldsymbol{A}^{-1/2}\boldsymbol{B}\boldsymbol{A}^{-1/2})}{\lambda_{\min}(\boldsymbol{A}^{-1/2}\boldsymbol{B}\boldsymbol{A}^{-1/2})}, \qquad d_{\mathrm{AI}}(\boldsymbol{A}, \boldsymbol{B}) \coloneqq \|\log(\boldsymbol{A}^{-1/2}\boldsymbol{B}\boldsymbol{A}^{-1/2})\|_F.$$

The projective distance is invariant to positive rescaling, making it particularly convenient for Tyler estimation. On $\mathcal{M}$,

$$d_{\mathrm{AI}}(\boldsymbol{A}, \boldsymbol{B}) \le \sqrt{d}\, d_{\mathrm{H}}(\boldsymbol{A}, \boldsymbol{B}),$$

while $d_{\mathrm{H}}(\boldsymbol{A}, \boldsymbol{B}) \le r$ implies

$$e^{-r}\boldsymbol{A} \preceq \boldsymbol{B} \preceq e^{r}\boldsymbol{A}.$$

Consequently, a uniform projective-distance bound on the node iterates, together with determinant normalization, confines the entire decentralized trajectory to a common compact subset of $\mathcal{M}$.

For later use, define the projective mixing error and the resulting one-step perturbation level by

$$a_k \coloneqq \log \frac{1+\delta_k}{1-\delta_k}, \qquad a_{-1} = 0, \qquad \eta_k \coloneqq 4a_{k-1} + a_k.$$

Throughout the general convergence analysis, we impose the summability condition

$$S \coloneqq \sum_{k \ge 0} a_k < \infty.$$

This condition is directly satisfied by both the polynomial and geometric consensus schedules considered below.

4.1 Sublinear convergence under summable consensus errors

We begin by establishing feasibility and stability of D-MM. These properties do not require the local weighted-scatter matrices to be positive definite.

Theorem 4.1 (Stability under finite consensus)**.** *Under Assumptions 1–2 and* (8)*, every D-MM iterate is well defined and belongs to $\mathcal{M}$. For all $k \ge 0$,*

$$\max_{i,j} d_{\mathrm{H}}(\mathbf{\Sigma}_i^{k+1}, \mathbf{\Sigma}_j^{k+1}) \le 2a_k, \tag{10}$$

$$\max_{i} d_{\mathrm{H}}\big(\mathbf{\Sigma}_i^{k+1}, \mathcal{F}(\mathbf{\Sigma}_1^k)\big) \le \eta_k. \tag{11}$$

Under the summability condition $S < \infty$, all node iterates and the corresponding centralized images remain in

$$\mathcal{K}_R \coloneqq \{\mathbf{\Sigma} \in \mathcal{M} : d_{\mathrm{H}}(\mathbf{\Sigma}, \mathbf{\Sigma}^\star) \le R\}, \qquad R \coloneqq d_{\mathrm{H}}(\mathbf{\Sigma}^0, \mathbf{\Sigma}^\star) + 7S. \tag{12}$$

Moreover, $\mathcal{K}_R$ is compact and

$$e^{-R}\mathbf{\Sigma}^\star \preceq \mathbf{\Sigma}_i^k \preceq e^{R}\mathbf{\Sigma}^\star.$$

Equations (10)–(11) respectively control inter-node disagreement and deviation from a centralized Tyler step. In particular, the relative consensus bound ensures that each finite-consensus output entering the normalization step is positive definite even when individual local weighted scatters are singular; see Appendix C.

Theorem 4.2 (Convergence under summable consensus errors)**.** *Under the conditions of Theorem 4.1 and the summability condition $S < \infty$, let R be given by* (12) *and define*

$$C_R := \frac{d}{2} e^R R^2.$$

Then, for every $K \geq 1$,

$$\max_i \Delta(\boldsymbol{\Sigma}_i^K) \leq \frac{4C_R}{K+1} + \frac{d}{K(K+1)} \sum_{t=0}^{K-1} (t+1)(t+2)\eta_t + 2da_{K-1}.$$

In particular,

$$\max_i \Delta(\boldsymbol{\Sigma}_i^K) \to 0 \qquad \text{as } K \to \infty.$$

Explicit rates follow by specifying the decay of the consensus error.

Corollary 4.3 (Polynomial consensus schedule)**.** *Let*

$$\delta_k = \frac{\bar{\delta}}{(k+2)^p}, \qquad p > 1,$$

and choose q_k according to (9)*. Then $S < \infty$ and*

$$\max_i \Delta(\boldsymbol{\Sigma}_i^K) = O\Big(K^{-\min\{1, p-1\}}\Big).$$

In particular, every $p \geq 2$ yields an $O(1/K)$ objective rate.

For K outer iterations, the total number of communication rounds satisfies

$$Q_K := \sum_{k=0}^{K-1} q_k \leq K + \frac{K \log(m/\bar{\delta}) + pK \log(K+1)}{-\log \bar{\rho}}.$$

Thus, every $p \geq 2$ attains the $O(1/K)$ objective rate with $q_k = O(\log k)$; increasing p further improves consensus accuracy but not the order of the objective bound.

4.2 Linear convergence under geometric consensus accuracy

Cederberg (2026) recently established a deterministic $O(1/k)$ objective-gap bound for the unregularized Tyler fixed-point iteration, while linear convergence was obtained for its regularized counterpart. The unregularized objective is scale invariant, $G(c\boldsymbol{\Sigma}) = G(\boldsymbol{\Sigma})$ for every $c > 0$, and hence cannot be strongly geodesically convex over the full positive definite cone. We instead restrict the dynamics to the determinant-one manifold $\mathcal{M}$, which removes the flat scaling direction. Under the strict occupancy condition, the geodesic Hessian is strictly positive along every nonzero tangent direction; compactness of the invariant region then yields a positive curvature modulus.

For $\boldsymbol{\Sigma} \in \mathcal{M}$ and symmetric $\boldsymbol{U}$ with $\operatorname{tr}(\boldsymbol{U}) = 0$, define

$$\mathcal{V}_{\boldsymbol{\Sigma}}(\boldsymbol{U}) := \frac{d}{N} \sum_{s=1}^{N} \left[\boldsymbol{u}_{\boldsymbol{\Sigma},s}^\top \boldsymbol{U}^2 \boldsymbol{u}_{\boldsymbol{\Sigma},s} - \left(\boldsymbol{u}_{\boldsymbol{\Sigma},s}^\top \boldsymbol{U} \boldsymbol{u}_{\boldsymbol{\Sigma},s} \right)^2 \right], \qquad \boldsymbol{u}_{\boldsymbol{\Sigma},s} := \frac{\boldsymbol{\Sigma}^{-1/2} \boldsymbol{x}_s}{\left\| \boldsymbol{\Sigma}^{-1/2} \boldsymbol{x}_s \right\|_2}.$$

The quantity $\mathcal{V}_{\boldsymbol{\Sigma}}(\boldsymbol{U})$ is the geodesic second derivative of G along $\boldsymbol{\Sigma}^{1/2} \exp(t\boldsymbol{U}) \boldsymbol{\Sigma}^{1/2}$. We define the *curvature modulus* over the compact region $\mathcal{K}_R$ as

$$\mu_R := \min_{\substack{\boldsymbol{\Sigma} \in \mathcal{K}_R,\ \boldsymbol{U} = \boldsymbol{U}^\top \\ \operatorname{tr} \boldsymbol{U} = 0,\ \|\boldsymbol{U}\|_F = 1}} \mathcal{V}_{\boldsymbol{\Sigma}}(\boldsymbol{U}).$$

Under Assumption 1,

$$0 < \mu_R \leq d.$$

The curvature modulus μ_R depends on the realized data and on $\mathcal{K}_R$, and may be arbitrarily small even when the strict occupancy condition holds.

Theorem 4.4 (Linear convergence of the centralized Tyler iteration)**.** *Consider the determinant-normalized centralized Tyler iteration*

$$\mathbf{\Sigma}^{k+1} = \mathcal{F}(\mathbf{\Sigma}^k), \qquad \mathbf{\Sigma}^0 \in \mathcal{M}.$$

Under Assumption 1, define $R_0 \coloneqq d_{\mathrm{H}}(\mathbf{\Sigma}^0, \mathbf{\Sigma}^\star)$, $\mathcal{K}_{R_0} \coloneqq \{\mathbf{\Sigma} \in \mathcal{M} : d_{\mathrm{H}}(\mathbf{\Sigma}, \mathbf{\Sigma}^\star) \le R_0\}$ *and let* $\mu_{R_0} > 0$ *denote the corresponding curvature modulus. Then, for every* $k \ge 0$*,*

$$\Delta(\mathbf{\Sigma}^{k+1}) \le \left(1 - \frac{\mu_{R_0}}{d^2}\right) \Delta(\mathbf{\Sigma}^k).$$

Moreover,

$$d_{\mathrm{AI}}(\mathbf{\Sigma}^k, \mathbf{\Sigma}^\star) \le \sqrt{\frac{2}{\mu_{R_0}}} \left(1 - \frac{\mu_{R_0}}{d^2}\right)^{k/2} \sqrt{\Delta(\mathbf{\Sigma}^0)}.$$

Hence the objective gap converges Q-linearly, while the iterates converge R-linearly to $\mathbf{\Sigma}^\star$*.*

Theorem 4.4 sharpens the deterministic $O(1/k)$ rate of Cederberg (2026) without regularization. Its data-dependent modulus μ_{R_0} links sample geometry to the contraction rate and need not be uniform over admissible datasets.

Theorem 4.5 (Perturbed linear contraction)**.** *Under the conditions of Theorem 4.1 and the summability condition* $S < \infty$*, let* μ_R *denote the curvature modulus over* $\mathcal{K}_R$ *in* (12)*, and define*

$$\theta \coloneqq \sqrt{1 - \mu_R/d^2} \in (0, 1), \qquad c_d \coloneqq d/\sqrt{2}.$$

Then, for every node i *and every* $k \ge 1$*,*

$$\sqrt{\Delta(\mathbf{\Sigma}_i^k)} \le B_k \coloneqq \theta^k \sqrt{\Delta(\mathbf{\Sigma}^0)} + c_d \sum_{t=0}^{k-1} \theta^{k-1-t} \eta_t + 2c_d a_{k-1}. \tag{13}$$

Moreover,

$$\max_i d_{\mathrm{AI}}(\mathbf{\Sigma}_i^k, \mathbf{\Sigma}^\star) \le \sqrt{\frac{2}{\mu_R}}\, B_k.$$

In (13), the three terms represent centralized contraction, accumulated consensus perturbations, and current disagreement.

Corollary 4.6 (Geometric consensus schedule)**.** *Let*

$$\delta_k = \bar{\delta}\beta^k, \qquad 0 < \beta < 1,$$

and choose q_k *according to* (9)*. Then* $S < \infty$ *and, with* $r \coloneqq \max\{\theta, \beta\} < 1$*,*

$$B_k = O(kr^{k-1}).$$

Consequently, for every fixed $\widehat{r} \in (r, 1)$*,*

$$\max_i d_{\mathrm{AI}}(\mathbf{\Sigma}_i^k, \mathbf{\Sigma}^\star) = O(\widehat{r}^k), \qquad \max_i \Delta(\mathbf{\Sigma}_i^k) = O(\widehat{r}^{2k}),$$

so every node converges R-linearly to the pooled Tyler estimator.

For K *outer iterations, the total number of communication rounds satisfies*

$$Q_K \le K + \frac{K \log(m/\bar{\delta}) + \frac{1}{2}K(K-1)\log(1/\beta)}{-\log \bar{\rho}}.$$

Hence an objective accuracy ε *is reached, for a fixed problem instance, with*

$$K = O(\log(1/\varepsilon)), \qquad Q_K = O(\log^2(1/\varepsilon)).$$

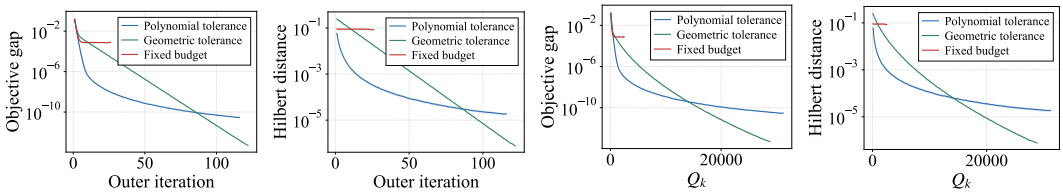
The two regimes are complementary: polynomial schedules require weaker communication accuracy, whereas geometric schedules preserve the instance-dependent linear contraction. The latter need not be faster in communication time because each outer iteration requires increasingly accurate consensus; schedule selection is discussed in Appendix F.

5 NUMERICAL EXPERIMENTS

We first examine the convergence behavior of finite-consensus D-MM and the trade-off between optimization progress and communication. We then investigate the two mechanisms governing its convergence: the geometry of the observations, which controls the intrinsic contraction of the centralized Tyler iteration, and network connectivity, which controls the reduction of consensus errors. Finally, we evaluate D-MM across a broader range of decentralized and statistical configurations and in a rolling portfolio application. Unless stated otherwise, synthetic data have a determinant-one population shape, all iterates are determinant-normalized, and D-MM terminates when $\max_i \left\|\boldsymbol{\Sigma}_i^{k+1} - \boldsymbol{\Sigma}_i^k\right\|_F \leq 10^{-7}$. We compare polynomial, geometric, and fixed-budget consensus schedules, using centralized Tyler MM as the pooled-Tyler reference.

5.1 SYNTHETIC EXPERIMENTS

Convergence under finite consensus. Figure 1 compares polynomial, geometric, and fixed-budget schedules on a path graph. Polynomial and geometric schedules drive the objective gap and disagreement to zero, whereas the fixed-budget schedule reaches an error floor. Polynomial consensus is more communication-efficient at moderate accuracy, while geometric consensus reaches smaller residuals. The slope change in Figure 1a is consistent with a crossover between the θ- and β-governed contraction mechanisms in Theorem 4.5.



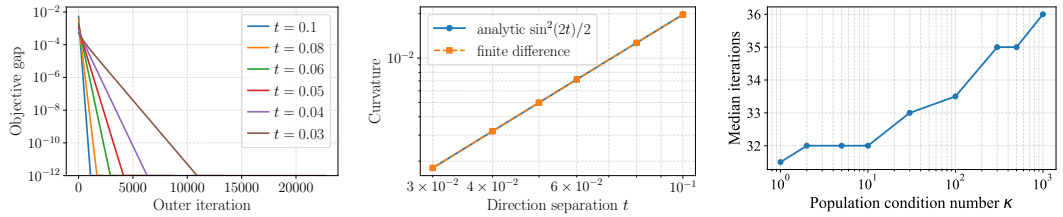
(a) Objective gap vs. iteration. (b) Disagreement vs. iteration. (c) Objective gap vs. Q_K. (d) Disagreement vs. Q_K.

Figure 1: Schedule trajectories for the balanced partition on a path graph. Here Q_K denotes the cumulative number of communication rounds. Curves show the median over repeated runs; shaded bands span the 10%–90% quantiles.

Effect of data geometry. We next isolate the role of data geometry in the centralized contraction. Consider the $d = 2$, $N = 4$ family

$$\boldsymbol{x}_1 = \boldsymbol{e}_1, \quad \boldsymbol{x}_2 = \boldsymbol{e}_2, \quad \boldsymbol{x}_3 = (\cos t, \sin t)^\top, \quad \boldsymbol{x}_4 = (-\sin t, \cos t)^\top.$$

For every $0 < t < \pi/4$, the normalized Tyler estimator is $\boldsymbol{\Sigma}^\star = \boldsymbol{I}$, while the fixed-direction witness $\mathcal{V}_{\boldsymbol{I}}(\boldsymbol{U}) = \sin^2(2t)/2$ upper-bounds μ_R; see Appendix E. Figure 2 therefore illustrates curvature degeneration rather than estimates μ_R: the numerical witness follows the analytic expression and convergence slows as $t \downarrow 0$. The final panel separately varies the population condition number.



(a) Convergence under directional separation t. (b) Analytic and numerical curvature witnesses. (c) Effect of the population condition number.

Figure 2: Effects of data geometry: directional separation provides a curvature-degeneration witness, while a separate experiment isolates population conditioning.

Effect of network connectivity. Figure 3 compares network topologies at a common objective-gap target 10^{-6}. Better-connected graphs require fewer consensus rounds, whereas sparse graphs can reduce the transmitted payload, illustrating the mixing–communication trade-off.

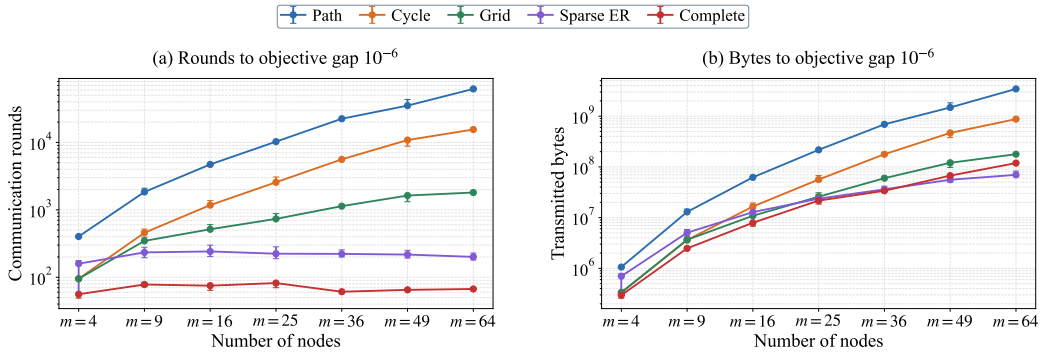


Figure 3: Communication cost of polynomial-schedule D-MM at the common pooled-Tyler objective-gap target 10^{-6}. Points are medians and error bars are interquartile ranges over 100 paired trials for each graph–node-count setting. The left panel reports synchronous communication rounds; the right panel reports transmitted matrix-payload bytes.

Additional experiments across data partitions, network structures, consensus schedules, dimensions, population shapes, and radial laws are reported in Appendix H.

5.2 Real-data experiments

We evaluate D-MM on the first 30 industries with complete observations from a frozen Fama–French 49-industry file, using 120-month rolling windows from 1980-01 to 2026-07. Table 1 shows that D-MM reproduces the pooled-Tyler portfolio statistics. Compared with Shape-SCM, the Tyler portfolios achieve a higher annualized mean return and Sharpe ratio, together with lower annualized volatility. Shape-SCM is included only as a downstream comparator with a different statistical target. Wealth and approximation trajectories are reported in Appendix G.

Table 1: Rolling-GMVP results over 559 held-out months. Mean return and volatility are annualized; the Sharpe ratio uses no risk-free-rate subtraction. D-MM agreement is its median affine-invariant distance to pooled Tyler.

method	annualized mean	annualized volatility	Sharpe	agreement / communication cost
Pooled Tyler	0.1277	0.1349	0.9465	reference
Polynomial D-MM	0.1277	0.1349	0.9465	1.48×10^{-5} / 35,048 rounds
Shape-SCM	0.1270	0.1358	0.9352	downstream comparator

6 Conclusion

We have developed D-MM for decentralized Tyler estimation when the pooled problem is well posed but local Tyler estimators need not exist. Relative Loewner-order bounds and Hilbert's projective metric have yielded well-posedness and convergence despite singular local weighted scatters, with explicit guarantees under polynomial consensus schedules. We have further shown that restricting the scale-invariant Tyler problem to the determinant-one manifold reveals positive data-dependent curvature and yields an instance-dependent linear rate for the unregularized centralized iteration; geometrically decaying consensus errors preserve this contraction in D-MM. The analysis thus connects data geometry, optimization, and communication, consistent with the behavior observed in our synthetic and portfolio experiments.

AI USE STATEMENT

Generative AI (OpenAI ChatGPT) was used only for language polishing, including improvements to English expression, clarity, concision, and presentation. All research ideas, theoretical arguments, proofs, experiments, results, and scientific claims were developed and verified by the authors, who take full responsibility for the final manuscript.

REPRODUCIBILITY STATEMENT

The algorithm, initialization, network assumptions, and communication schedules are specified in Section 3. Complete derivations and proofs are provided in Appendices A–E. Section 5 describes the reported experiments, with additional experimental designs, implementation details, and supplementary results provided in the corresponding appendices.

ETHICS STATEMENT

This work involves no human-subject research. Decentralized storage of observations does not by itself provide a formal privacy guarantee: the communicated weighted-scatter matrices may reveal information about local data. Applications to sensitive data would therefore require additional privacy mechanisms and appropriate governance.

REFERENCES

Albert S. Berahas, Raghu Bollapragada, Nitish Shirish Keskar, and Ermin Wei. Balancing communication and computation in distributed optimization. *IEEE Transactions on Automatic Control*, 64(8):3141–3155, 2019. doi: 10.1109/TAC.2018.2880407.

Stephen Boyd, Neal Parikh, Eric Chu, Borja Peleato, and Jonathan Eckstein. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends in Machine Learning*, 3(1):1–122, 2011. doi: 10.1561/2200000016.

Daniel Cederberg. Tyler's M-estimator through the lens of convex-concave programming. In *Proceedings of the 29th International Conference on Artificial Intelligence and Statistics*, volume 300 of *Proceedings of Machine Learning Research*, pp. 64–72. PMLR, 2026. URL https://proceedings.mlr.press/v300/cederberg26a.html.

Shixiang Chen, Alfredo Garcia, Mingyi Hong, and Shahin Shahrampour. Decentralized Riemannian gradient descent on the Stiefel manifold. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pp. 1594–1605. PMLR, 2021. URL https://proceedings.mlr.press/v139/chen21g.html.

Lior Danon and Dan Garber. Frank–Wolfe-based algorithms for approximating Tyler's M-estimator. In *Advances in Neural Information Processing Systems*, volume 35, pp. 3637–3648, 2022. URL https://arxiv.org/abs/2206.09370.

Paolo Di Lorenzo and Gesualdo Scutari. NEXT: In-network nonconvex optimization. *IEEE Transactions on Signal and Information Processing over Networks*, 2(2):120–136, 2016. doi: 10.1109/TSIPN.2016.2524588.

William Cole Franks and Ankur Moitra. Rigorous guarantees for Tyler's M-estimator via quantum expansion. In *Proceedings of the 33rd Conference on Learning Theory*, volume 125 of *Proceedings of Machine Learning Research*, pp. 1601–1632. PMLR, 2020. URL https://proceedings.mlr.press/v125/franks20a.html.

John Goes, Gilad Lerman, and Boaz Nadler. Robust sparse covariance estimation by thresholding Tyler's M-estimator. *The Annals of Statistics*, 48(1):86–110, 2020. doi: 10.1214/18-AOS1793.

Peter J Huber and Elvezio M Ronchetti. *Robust statistics*. John Wiley & Sons, 2011.

Lap Chi Lau and Akshay Ramachandran. Optimal bounds for Tyler's M-estimator for elliptical distributions. arXiv:2510.13751, 2025. URL https://arxiv.org/abs/2510.13751.

Ricardo A Maronna, R Douglas Martin, Victor J Yohai, and Matías Salibián-Barrera. *Robust statistics: theory and methods (with R)*. John Wiley & Sons, 2019.

Gonzalo Mateos, Juan Andrés Bazerque, and Georgios B. Giannakis. Distributed sparse linear regression. *IEEE Transactions on Signal Processing*, 58(10):5262–5276, 2010. doi: 10.1109/TSP.2010.2055862.

Angelia Nedić and Asuman Ozdaglar. Distributed subgradient methods for multi-agent optimization. *IEEE Transactions on Automatic Control*, 54(1):48–61, 2009. doi: 10.1109/TAC.2008.2009515.

Angelia Nedić, Alex Olshevsky, and Wei Shi. Achieving geometric convergence for distributed optimization over time-varying graphs. *SIAM Journal on Optimization*, 27(4):2597–2633, 2017. doi: 10.1137/16M1084316.

Esa Ollila and David E. Tyler. Regularized M-estimators of scatter matrix. *IEEE Transactions on Signal Processing*, 62(22):6059–6070, 2014. doi: 10.1109/TSP.2014.2360826.

Wei Shi, Qing Ling, Gang Wu, and Wotao Yin. EXTRA: An exact first-order algorithm for decentralized consensus optimization. *SIAM Journal on Optimization*, 25(2):944–966, 2015. doi: 10.1137/14096668X.

Ying Sun, Prabhu Babu, and Daniel P. Palomar. Regularized Tyler's scatter estimator: Existence, uniqueness, and algorithms. *IEEE Transactions on Signal Processing*, 62(19):5143–5156, 2014. doi: 10.1109/TSP.2014.2348944.

David E. Tyler. A distribution-free M-estimator of multivariate scatter. *The Annals of Statistics*, 15 (1):234–251, 1987. doi: 10.1214/aos/1176350263.

Ami Wiesel and Teng Zhang. Structured robust covariance estimation. *Foundations and Trends in Signal Processing*, 8(3):127–216, 2015. doi: 10.1561/2000000053.

Wenfu Xia, Fengpei Li, Ying Sun, and Ziping Zhao. Covariance selection over networks. In *Proceedings of the 28th International Conference on Artificial Intelligence and Statistics*, volume 258 of *Proceedings of Machine Learning Research*, pp. 3385–3393. PMLR, 2025. URL `https://proceedings.mlr.press/v258/xia25a.html`.

Lin Xiao and Stephen Boyd. Fast linear iterations for distributed averaging. *Systems & Control Letters*, 53(1):65–78, 2004. doi: 10.1016/j.sysconle.2004.02.022.