WHEN DOES RECONSTRUCTION IDENTIFY BM TENSOR REPRESENTATIONS?

Anonymous authors
Paper under double-blind review

ABSTRACT

The Bhattacharya–Mesner (BM) decomposition represents a third-order tensor through factors indexed by complementary pairs of data indices. We study identifiability of these factors from a fully observed tensor. On the entrywise-nonzero stratum, we characterize all term-preserving transformations and construct a canonical normalization. A full-column-rank Jacobian in the normalized coordinates certifies local identifiability, determines reconstruction-loss curvature, and yields local perturbation bounds. For cubic tensors, parameter counting gives a necessary term-count ceiling of $N/3 + O(1)$ for generic local identifiability. Below this ceiling, separable template ratios generate continuous families of inequivalent decompositions with identical reconstructions. Under explicit full-column-rank assumptions, the slice-wise mixing families have dimension $pr(r-1)$ in one mode, with corresponding formulas in the other modes; these ambiguities also occur for positive factors. At rank one, we establish global identifiability, closed-form recovery, and a membership test expressed entirely in tensor entries. Exact finite-field witnesses and synthetic experiments examine gauge freedom, structural ambiguity, numerical conditioning, and optimization behavior.

1 INTRODUCTION

Reconstruction-based learning estimates latent factors by fitting observed data. Interpreting these factors, comparing them across training runs, and using them to separate latent sources also depend on identifiability: the observations must determine the representation up to specified symmetries. This requirement is central to work on disentangled and identifiable representations (Locatello et al., 2019; Khemakhem et al., 2020; Roeder et al., 2021). In tensor models, the contraction structure determines the relevant symmetries and uniqueness conditions (Kruskal, 1977; Kolda & Bader, 2009).

The Bhattacharya–Mesner (BM) decomposition represents a third-order tensor as

$$X_{ijk} = \sum_{t=1}^{r} A_{itk} B_{ijt} C_{tjk}. \tag{1}$$

Each factor in a BM term retains two data indices, whereas a canonical polyadic (CP) term uses one index per factor. This pairwise structure supports compressed spatiotemporal representations and video separation (Tian et al., 2025). It also admits index-dependent rescalings that cancel within each component tensor. Diagonal scaling ambiguity is discussed in Tian et al. (2025, Section 7). Identifiability of BM factors therefore concerns uniqueness modulo the full family of term-preserving transformations.

Three properties govern factor recovery in this model. Gauge transformations reparameterize each component while preserving its tensor entries. Structural ambiguities change the component tensors while preserving their sum. Conditioning describes the sensitivity of locally identifiable factors to reconstruction errors. We analyze these properties in a common system of normalized coordinates.

We first characterize the complete term-preserving gauge on nonzero factors and construct a unique normalized representative for each labeled scaling orbit. The reduced Jacobian then yields a certificate for local identifiability and quantitative bounds on loss curvature and factor error. We also construct positive-dimensional families of inequivalent representations from separable template ratios, including a positive rank-two example below the dimension ceiling. At rank one, we obtain global

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.

uniqueness and explicit recovery formulas. Exact rank witnesses and synthetic experiments illustrate the identifiable and ambiguous regimes.

The analysis concerns fully observed tensors and exact decompositions at a fixed term count, with local perturbation bounds near regular solutions. The experiments examine identifiability certificates, factor variation along exact fibers, noise sensitivity, and alternating least-squares behavior.

2 BM REPRESENTATIONS AND THEIR SYMMETRIES

Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$, with factors $\mathcal{A} \in \mathbb{F}^{m\times r\times n}$, $\mathcal{B} \in \mathbb{F}^{m\times p\times r}$, and $\mathcal{C} \in \mathbb{F}^{r\times p\times n}$. Write $\theta = (\mathcal{A}, \mathcal{B}, \mathcal{C})$ and $\Phi_r(\theta) = X$ for (1). All dimensions are positive integers, and all factor entries are nonzero unless a fixed support is specified. The parameter r denotes the prescribed term count, which may exceed the minimum count needed to represent X; minimality is proved for the explicit rank-two example. We use $\odot$ and $\oslash$ for entrywise multiplication and division.

For fixed j, define $A_t = A_{\cdot t\cdot}$, $b_t^{(j)} = B_{\cdot jt}$, and $c_t^{(j)} = C_{tj\cdot}$. Then

$$X^{(j)} = \sum_{t=1}^{r} A_t \odot \left(b_t^{(j)} c_t^{(j)\top}\right). \tag{2}$$

The matrices A_t serve as templates shared across slices, and the outer products provide slice-dependent rank-one modulations. This form exposes the rescaling symmetries and the template relations that generate nonuniqueness.

2.1 GAUGE EQUIVALENCE AND CANONICAL COORDINATES

For nonzero scalars s_{it}, q_{jt}, u_{kt}, the transformation

$$\widehat{A}_{itk} = s_{it}u_{kt}A_{itk}, \qquad \widehat{B}_{ijt} = s_{it}^{-1}q_{jt}B_{ijt}, \qquad \widehat{C}_{tjk} = q_{jt}^{-1}u_{kt}^{-1}C_{tjk} \tag{3}$$

preserves each component tensor $T_t(i, j, k) = A_{itk}B_{ijt}C_{tjk}$. We call this a *gauge transformation*: it changes only the representation of each term. A common permutation of t changes only the component labels.

Definition 1. *Let Θ_r be the entrywise-nonzero factor space and let $\mathscr{G}_r$ contain the transformations* (3) *and common component permutations. A representation θ is* locally essentially unique *if some neighborhood of its orbit $\mathscr{G}_r \cdot \theta$ contains no other orbit mapping to $\Phi_r(\theta)$. It is* globally essentially unique *if no other orbit in Θ_r maps to that tensor. We use local and global identifiability in these respective senses.*

Theorem 2. *Two nonzero factor triples represent the same ordered component tensors if and only if they are related by* (3)*. The effective gauge group, including permutations, is*

$$\mathscr{G}_r \cong (\mathbb{F}^\times)^{r(m+p+n-1)} \rtimes S_r. \tag{4}$$

Every labeled scaling orbit has exactly one representative satisfying

$$A_{it1} = 1, \qquad A_{1tk} = 1 \ (k \geq 2), \qquad B_{1jt} = 1. \tag{5}$$

The gauge dimension g and the number d of remaining coordinates are

$$g = r(m + p + n - 1), \qquad d = r(mn + mp + pn - m - p - n + 1). \tag{6}$$

Proof sketch. Ratios of two factorizations of one ordered term satisfy $U_{ik}V_{ij}W_{jk} = 1$. Comparing this identity across indices gives $U_{ik} = s_i u_k$, $V_{ij} = s_i^{-1}q_j$, and $W_{jk} = q_j^{-1}u_k^{-1}$. The scaling parameters themselves have a one-dimensional kernel per term. Explicit normalization is

$$A^0_{itk} = \frac{A_{1t1}A_{itk}}{A_{it1}A_{1tk}}, \quad B^0_{ijt} = \frac{A_{it1}B_{ijt}}{A_{1t1}B_{1jt}}, \quad C^0_{tjk} = A_{1tk}B_{1jt}C_{tjk}. \tag{7}$$

Any gauge preserving the anchor entries acts trivially on every factor. Appendix A gives the complete argument. □

The normalization removes the full continuous term-preserving gauge. Let $z \in (\mathbb{F}^\times)^d$ collect the free entries and let $F(z)$ be the BM map restricted to this section. The remaining component permutations form a finite group. Distances between normalized factors are gauge-invariant coordinate distances associated with the chosen anchors. Their values, and the conditioning measured in these coordinates, depend on that choice.

3 LOCAL IDENTIFIABILITY AND RECONSTRUCTION-LOSS GEOMETRY

The kernel of the full Jacobian $J_\theta = D\Phi_r(\theta)$ contains the g gauge directions. Restricting to the normalized section gives the reduced Jacobian $J = DF(z) \in \mathbb{F}^{mpn \times d}$, with the same rank as J_θ. For a variation $h = (\Delta A, \Delta B, \Delta C)$,

$$D\Phi_r(\theta)[h]_{ijk} = \sum_{t=1}^{r} \big(\Delta A_{itk} B_{ijt} C_{tjk} + A_{itk} \Delta B_{ijt} C_{tjk} + A_{itk} B_{ijt} \Delta C_{tjk}\big). \tag{8}$$

The reduced derivative sets variations of the anchor entries to zero.

Theorem 3. *If* $\operatorname{rank} DF(z) = d$*, the BM representation is locally essentially unique. Equivalently,* $\ker J_\theta$ *consists exactly of gauge tangents. Conversely, local essential uniqueness implies* $\operatorname{rank} DF(z) = d$ *whenever* F *has constant rank in a neighborhood of* z*.*

Proof sketch. A nonsingular $d \times d$ output minor gives a locally invertible projection of F by the inverse function theorem. If instead the local constant rank is $q < d$, the constant-rank theorem gives a $(d - q)$-dimensional local fiber, which cannot consist of a finite permutation orbit. See Appendix B. □

Under the constant-rank hypothesis, the Jacobian nullity equals the dimension of the local fiber. At singular points, these dimensions can differ: the map $z \mapsto z^2$ has an isolated zero fiber and a zero derivative. Numerical rank estimates near singular points must therefore be interpreted together with the local fiber structure.

3.1 PARAMETER COUNTS AND LOCAL IDENTIFIABILITY

Proposition 4. *Generic local identifiability requires*

$$r \le r^\star(m, p, n) := \left\lfloor \frac{mpn}{mn + mp + pn - m - p - n + 1} \right\rfloor. \tag{9}$$

The same obstruction holds at every locally constant-rank point. For $N \times N \times N$ *tensors,* $r^\star(N) = \lfloor N^3/(3N^2 - 3N + 1) \rfloor = N/3 + O(1)$.

The bound follows from $d \le mpn$ on the open dense maximal-rank locus. Over $\mathbb{C}$, the dimension theorem for fibers gives the same inequality at every locally identifiable point. Over $\mathbb{R}$, the stated obstruction applies at locally constant-rank points. Within the admissible range, local identifiability depends on the fiber structure, as illustrated by the rank-two construction in Section 4.

Scalar-normalized CP parameter counting gives $\lfloor N^3/(3N - 2) \rfloor = N^2/3 + O(N)$. These different term-count scalings reflect the $O(N^2)$ degrees of freedom in a BM component and the $O(N)$ degrees of freedom in a CP component. For BM tensors with $N = 5, r = 2$, the parameter count is $d = 122 < 125$; the example in Section 4 nevertheless admits a continuous ambiguity.

3.2 RECONSTRUCTION ACCURACY AND FACTOR STABILITY

Consider $\ell_X(w) = \frac{1}{2}\|F(w) - X\|_F^2$ at an exact normalized representation $X = F(z)$. For real factors,

$$\nabla^2 \ell_X(z) = J^\top J, \qquad \lambda_{\min}(\nabla^2 \ell_X(z)) = \sigma_d(J)^2. \tag{10}$$

A full-column-rank reduced Jacobian therefore gives a positive-definite Hessian at an exact fit, with smallest eigenvalue $\sigma_d(J)^2$. For complex factors, the analogous assertion holds after realification.

Proposition 5. *Suppose* $\sigma_d(DF(z)) = \sigma > 0$*. For every sufficiently small perturbation* E*, the objective* $\frac{1}{2}\|F(w) - F(z) - E\|_F^2$ *has a unique stationary point* $\widehat{z}(E)$ *in a fixed sufficiently small neighborhood of* z*. It is a strict local minimizer, with*

$$\widehat{z}(E) - z = DF(z)^\dagger E + O(\|E\|_F^2), \qquad \|\widehat{z}(E) - z\|_2 \le \frac{\|E\|_F}{\sigma} + O(\|E\|_F^2). \tag{11}$$

The implicit function theorem applied to the normal equations gives the result; Appendix F contains the proof. The following bound also controls nearby approximate fits.

Corollary 6. *Suppose DF is L-Lipschitz on a ball around z, and w lies in this ball with $\|w - z\|_2 \leq \sigma/L$. If $\|F(w) - (F(z) + E)\|_F \leq \varepsilon$, then*

$$\|w - z\|_2 \leq \frac{2}{\sigma}\big(\varepsilon + \|E\|_F\big). \tag{12}$$

For $L = 0$ the radius restriction is unnecessary within the ball.

Taylor's theorem gives $\|F(w) - F(z)\|_F \geq \sigma\|w - z\|_2 - (L/2)\|w - z\|_2^2$, and the triangle inequality yields (12). After local label alignment, the same estimate controls the minimum normalized distance over permutations. Within the stated neighborhood, the stability margin σ converts reconstruction accuracy into a quantitative factor-error bound.

4 Exact ambiguities beyond scaling

Separable template ratios generate exact nonuniqueness in normalized coordinates, including for positive factors. Suppose the templates in mode A satisfy

$$A_t = A_1 \odot \big(\rho^{(t)}\sigma^{(t)\top}\big), \qquad \rho^{(1)} = \mathbf{1}, \quad \sigma^{(1)} = \mathbf{1}. \tag{13}$$

Set $P^{(j)}_{\cdot t} = \rho^{(t)} \odot b^{(j)}_t$ and $Q^{(j)}_{\cdot t} = \sigma^{(t)} \odot c^{(j)}_t$. Equation (2) becomes

$$X^{(j)} = A_1 \odot (P^{(j)}Q^{(j)\top}), \qquad (P^{(j)}, Q^{(j)}) \mapsto (P^{(j)}S_j, Q^{(j)}S_j^{-\top}), \tag{14}$$

where each $S_j \in GL_r(\mathbb{F})$ can be chosen independently near the identity.

Theorem 7. *Let $r \geq 2$ and all factor entries be nonzero. If* (13) *holds in any one of the three modes, the representation is locally essentially nonunique. In mode A, if all $P^{(j)}$ and $Q^{(j)}$ have column rank r, the fixed-template mixing family has dimension $pr(r-1)$ modulo gauge and permutations. The corresponding dimensions for modes B and C are $nr(r-1)$ and $mr(r-1)$. For positive real factors, these ambiguities persist within the positive stratum.*

Proof sketch. The transformation (14) preserves every slice. A shear $S_j = I + \epsilon e_1 e_2^\top$ changes the first two ordered component matrices by $(-\epsilon D, +\epsilon D)$ with $D \neq 0$, so the curve is not gauge freedom. Under the full-column-rank assumptions, the GL_r action is free. Its overlap with the residual gauge is the r-dimensional group of diagonal column scalings, leaving $r^2 - r$ dimensions per slice. Nonzero entries, and positivity, persist for sufficiently small shears. Appendix C supplies the complete all-rank argument. □

For rank two, define $R_A = A_2 \oslash A_1$, $R_B = B_{\cdot\cdot 2} \oslash B_{\cdot\cdot 1}$, and $R_C = C_{2\cdot\cdot} \oslash C_{1\cdot\cdot}$. Nonseparability of all three ratios is a necessary condition for essential uniqueness. Separability is gauge-invariant and is preserved by component swaps, which take entrywise reciprocals of the ratios.

4.1 A positive ambiguity below the dimension ceiling

Appendix C.1 specifies positive 5×2 matrices U_j and a 2×5 matrix V such that $A_{itk} = 1$ and $X^{(j)} = U_j V$ define a BM-rank-two tensor. For $S_\epsilon = \left[\begin{smallmatrix} 1 & \epsilon \\ 0 & 1 \end{smallmatrix}\right]$, replace U_j by $U_j S_\epsilon$ and V by $S_\epsilon^{-1}V$ independently within each slice. All entries stay positive for $0 \leq \epsilon < 1/2$. At $(1, 1, 1)$, the unordered component values change from $\{1, 1\}$ to $\{1 - \epsilon, 1 + \epsilon\}$, while X is unchanged. The two sides of a rank-one membership identity from Section 5 evaluate to $297 \neq 350$, proving that the minimum term count is two.

Proposition 8. *Along this specified family, the full Jacobian has rank 92. Its 58-dimensional kernel contains 28 gauge directions and 30 excess infinitesimal directions. The fixed-template family accounts for 10 excess directions; the other 20 involve template variations. The same counts hold for the nonconstant-template extension in Appendix C.1.*

Appendix C.4 proves the rank by an exact image-dimension count. The construction gives a 10-dimensional family of inequivalent decompositions, while the remaining 20 null directions describe additional infinitesimal degeneracy. All members of the mixing family have zero reconstruction loss after normalization. Selecting among them by regularization introduces a criterion beyond reconstruction.

Table 1: Exact Jacobian witnesses generated by the supplied verifier. Every rank equals the algebraic upper bound $\min(d, mpn)$. The integers, seeds, and elimination code are included in the supplement.

Shape	r	d	mpn	$\operatorname{rank}_q DF$	Certified generic conclusion
4^3	1	37	64	37	Local identifiability
4^3	2	74	64	64	Fiber dimension 10
5^3	2	122	125	122	Local identifiability
6^3	2	182	216	182	Local identifiability
8^3	3	507	512	507	Local identifiability
$6 \times 3 \times 8$	3	222	144	144	Fiber dimension 78

5 Rank-one representations are globally identifiable

At rank one, normalized factors and a membership test follow directly from the observed tensor entries.

Theorem 9. *For an entrywise-nonzero tensor X, any one-term BM factorization $X_{ijk} = a_{ik}b_{ij}c_{jk}$ is globally essentially unique. In the section $a_{i1} = a_{1k} = b_{1j} = 1$, its factors are*

$$a_{ik} = \frac{X_{i1k}X_{111}}{X_{i11}X_{11k}}, \qquad b_{ij} = \frac{X_{ij1}}{X_{1j1}}, \qquad c_{jk} = X_{1jk}. \tag{15}$$

Such a factorization exists if and only if

$$X_{ijk}X_{i11}X_{11k}X_{1j1} = X_{ij1}X_{i1k}X_{1jk}X_{111} \tag{16}$$

for every (i, j, k).

The termwise ratio argument proves uniqueness; evaluating the anchor slices gives (15), and substitution proves the membership equivalence. The complete proof is in Appendix D. Recovering the factors uses $O(mn + mp + pn)$ operations; testing all identities uses $O(mpn)$ operations.

For noisy data, let $\delta = \|E\|_\infty / \min_{ijk} |X_{ijk}| < 1$. The direct estimator satisfies

$$\max_{ik} \left| \frac{\widehat{a}_{ik} - a_{ik}}{a_{ik}} \right| \le \frac{4\delta + 2\delta^2}{(1-\delta)^2}, \quad \max_{ij} \left| \frac{\widehat{b}_{ij} - b_{ij}}{b_{ij}} \right| \le \frac{2\delta}{1-\delta}, \quad \max_{jk} \left| \frac{\widehat{c}_{jk} - c_{jk}}{c_{jk}} \right| \le \delta. \tag{17}$$

The estimates give worst-case relative factor errors in terms of the entrywise noise level δ. For noisy observations, the direct recovery formulas yield an approximate factorization. The residual in (16) measures rank-one compatibility; its use as a statistical test requires a specified noise model and a calibrated threshold.

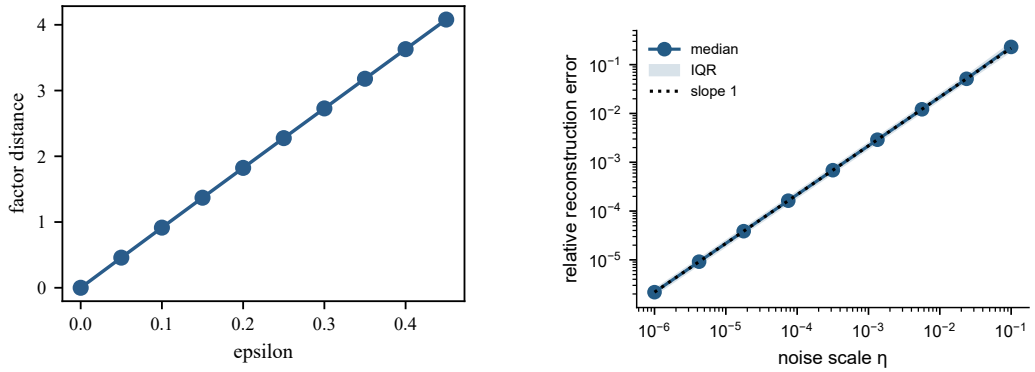
6 Experiments

We use synthetic tensors to examine exact local-identifiability certificates, factor variation along constant-reconstruction families, noise sensitivity in direct recovery, and alternating least-squares (ALS) behavior. Appendix H reports the Monte Carlo settings and aggregate results.

6.1 Exact rank certificates and numerical conditioning

We generate six integer witnesses in the normalized section, using entries in $\{1, \dots, 19\}$ and seeds 270001–270006. Gaussian elimination modulo the prime $q = 1{,}000{,}003$ gives the ranks in Table 1. A nonzero modular minor is nonzero in characteristic zero and establishes that the corresponding minor polynomial is nontrivial. The four full-column-rank cases therefore certify *generic local* identifiability in the stated dimensions over both $\mathbb{R}$ and $\mathbb{C}$. The two full-row-rank cases certify positive-dimensional generic fibers.

Numerical rank diagnostics at relative singular-value tolerances 10^{-8}, 10^{-10}, and 10^{-12} agree on the displayed regular and dimension-obstructed cases (Appendix H.1). Singular values describe the conditioning of the chosen normalized representative, complementing the exact rank certificates obtained by modular elimination.



(a) Factor variation at fixed reconstruction.

(b) Rank-one recovery under observation noise.

Figure 1: Left: inequivalent normalized rank-two factors along an exact constant-reconstruction family. Right: reconstruction error in the 50-seed rank-one noise study, shown as medians and interquartile ranges, with approximately linear dependence on small noise levels. Theoretical bounds on relative factor error are given in (17).

Table 2: ALS fitting results for three rank-two settings, with 50 runs per initialization type. Broad-initialization median fit and truth-local acceptance counts are reported; the acceptance threshold is 10^{-6}.

Setting ($r = 2$)	Broad median fit	Local accepted	Interpretation
Regular 6^3	4.08×10^{-3}	$4/50$	Low local fit acceptance
Dimension-obstructed 4^3	1.71×10^{-4}	$50/50$	Excess parameter dimension
Separable-template 5^3	1.92×10^{-3}	$50/50$	Exact mixing family

6.2 Identical reconstructions, different representations

The positive rank-two family in Section 4 has constant reconstruction and varying normalized factors (Figure 1a). Numerical checks at $\epsilon \in \{0, 0.1, 0.2, 0.3, 0.45\}$ give Jacobian rank 92 and reconstruction errors below 10^{-13} for both the constant- and nonconstant-template constructions. At $\epsilon = 0$, modular elimination together with the analytic upper bound in Proposition 8 also establishes the exact rank.

The template-transition experiment perturbs a separable ratio. When $\sigma_2(R_A)/\sigma_1(R_A) \geq 1.45 \times 10^{-5}$, all 100 reported witnesses have numerical rank $d = 122$ at tolerance 10^{-10}. Near separability, the estimated rank varies with the tolerance, reflecting the small singular values associated with the degenerate template relation.

6.3 Noise sensitivity and optimizer behavior

The rank-one study reports a median log–log reconstruction-error slope of 0.9999 for noise levels $\eta \leq 1.33 \times 10^{-3}$ (Figure 1b). The dynamic-range and rank-one detection studies in Appendix H.2 examine the dependence of the membership residual on the noise scale. The verification script checks exact recovery, canonical normalization, and the three deterministic perturbation bounds on a separately generated rank-one instance.

The ALS stress test uses 50 broad and 50 truth-local initializations in each of three rank-two settings, a 3,000-sweep budget, and a 10^{-8} relative-fit stopping criterion. Table 2 reports the achieved fit and the number of runs satisfying the 10^{-6} acceptance threshold for factor analysis. In the regular case, $4/50$ truth-local runs meet this threshold. In both the dimension-obstructed and separable-template settings, all $50/50$ truth-local runs meet the acceptance threshold. These outcomes distinguish ALS fitting behavior from the identifiability properties of the representation map. Factor-distance summaries are interpreted qualitatively in view of the unequal acceptance counts and the availability of aggregate data.

7 Related work

Identifiable representations. Impossibility results for unsupervised disentanglement (Locatello et al., 2019), identifiable latent-variable models with auxiliary information (Khemakhem et al., 2020), and linear identifiability of learned representation functions (Roeder et al., 2021) formulate recovery relative to specified equivalence classes. For a fully observed tensor parameterized by the BM contraction, the relevant equivalence consists of term-preserving transformations and component permutations. Our analysis characterizes this equivalence and the fibers that remain after normalization.

Tensor and BM decompositions. CP uniqueness has classical, deterministic, and generic formulations (Kruskal, 1977; De Lathauwer, 2006; Domanov & De Lathauwer, 2013; 2015; Chiantini & Ottaviani, 2012; Chiantini et al., 2014). Block-term uniqueness theory addresses the contraction structures of the corresponding models (De Lathauwer, 2008; Yang, 2014; Yang et al., 2016; 2020; Domanov & De Lathauwer, 2020). General conditioning theory for join decompositions relates sensitivity to differential singular values (Breiding & Vannieuwenhoven, 2018). Our local bounds apply inverse- and implicit-function arguments in explicit normalized BM coordinates. The BM product originates in ternary algebra (Mesner & Bhattacharya, 1990; 1994); subsequent work studies its rank (Gnang & Filmus, 2020) and applications to video compression and separation (Tian et al., 2025). The latter study discusses diagonal rescaling and regularized ALS. We characterize the full term-preserving gauge, construct a canonical section, and derive local certificates, global rank-one recovery formulas, and dimensions of explicit higher-rank ambiguity families.

8 Discussion and conclusion

Canonical normalization reduces BM factor recovery to identifiability of the normalized representation map. A full-column-rank reduced Jacobian certifies local uniqueness, and its smallest singular value controls loss curvature and local sensitivity. Separable template ratios generate positive-dimensional exact fibers, including for positive factors below the parameter-count ceiling. At rank one, the factors are globally identifiable and admit closed-form recovery together with a tensor-entry membership test.

Appendix E extends the scaling analysis to prescribed supports, with transformations restricted to each fixed support. The exact witnesses and synthetic experiments illustrate local identifiability, constant-reconstruction families, and noise sensitivity in the chosen normalized coordinates. Together, these results relate recovery of BM factors to the symmetries, fibers, and local conditioning of the representation map.

Reproducibility statement

Appendices A–F contain the complete proofs. Appendix G specifies the integer witnesses, coordinate ordering, random-number generator, and executable checks for the Jacobian ranks, ambiguity construction, and rank-one recovery. Appendix H records the Monte Carlo settings and aggregate summaries.

Ethics statement

The study analyzes BM representations using theoretical arguments and synthetic tensors, with no human-subject study or real-world deployment evaluation. Identifiability is defined relative to gauge transformations and component permutations.

References

Paul Breiding and Nick Vannieuwenhoven. The condition number of join decompositions. *SIAM Journal on Matrix Analysis and Applications*, 39(1):287–309, 2018. doi: 10.1137/17M1142880.

Luca Chiantini and Giorgio Ottaviani. On generic identifiability of 3-tensors of small rank. *SIAM Journal on Matrix Analysis and Applications*, 33(3):1018–1037, 2012. doi: 10.1137/110829180.

Luca Chiantini, Giorgio Ottaviani, and Nick Vannieuwenhoven. An algorithm for generic and low-rank specific identifiability of complex tensors. *SIAM Journal on Matrix Analysis and Applications*, 35(4):1265–1287, 2014. doi: 10.1137/140961389.

Lieven De Lathauwer. A Link between the Canonical Decomposition in Multilinear Algebra and Simultaneous Matrix Diagonalization. *SIAM Journal on Matrix Analysis and Applications*, 28(3): 642–666, 2006. doi: 10.1137/040608830.

Lieven De Lathauwer. Decompositions of a Higher-Order Tensor in Block Terms—Part II: Definitions and Uniqueness. *SIAM Journal on Matrix Analysis and Applications*, 30(3):1033–1066, 2008. doi: 10.1137/070690729.

Ignat Domanov and Lieven De Lathauwer. On the Uniqueness of the Canonical Polyadic Decomposition of Third-Order Tensors—Part II: Uniqueness of the Overall Decomposition. *SIAM Journal on Matrix Analysis and Applications*, 34(3):876–903, 2013. doi: 10.1137/120877258.

Ignat Domanov and Lieven De Lathauwer. Generic Uniqueness Conditions for the Canonical Polyadic Decomposition and INDSCAL. *SIAM Journal on Matrix Analysis and Applications*, 36 (4):1567–1589, 2015. doi: 10.1137/140970276.

Ignat Domanov and Lieven De Lathauwer. On Uniqueness and Computation of the Decomposition of a Tensor into Multilinear Rank-$(1, L_r, L_r)$ Terms. *SIAM Journal on Matrix Analysis and Applications*, 41(2):747–803, 2020. doi: 10.1137/18m1206849.

Edinah K. Gnang and Yuval Filmus. On the Bhattacharya-Mesner rank of third order hypermatrices. *Linear Algebra and its Applications*, 588:391–418, 2020. doi: 10.1016/j.laa.2019.11.023.

Ilyes Khemakhem, Diederik Kingma, Ricardo Monti, and Aapo Hyvarinen. Variational autoencoders and nonlinear ICA: A unifying framework. In *Proceedings of the Twenty Third International Conference on Artificial Intelligence and Statistics*, volume 108 of *Proceedings of Machine Learning Research*, pp. 2207–2217. PMLR, 2020. URL https://proceedings.mlr.press/v108/khemakhem20a.html.

Tamara G. Kolda and Brett W. Bader. Tensor decompositions and applications. *SIAM Review*, 51(3): 455–500, 2009. doi: 10.1137/07070111X.

Joseph B. Kruskal. Three-way arrays: Rank and uniqueness of trilinear decompositions, with application to arithmetic complexity and statistics. *Linear Algebra and its Applications*, 18:95–138, 1977. doi: 10.1016/0024-3795(77)90069-6.

Francesco Locatello, Stefan Bauer, Mario Lucic, Gunnar Raetsch, Sylvain Gelly, Bernhard Schölkopf, and Olivier Bachem. Challenging common assumptions in the unsupervised learning of disentangled representations. In *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pp. 4114–4124. PMLR, 2019. URL https://proceedings.mlr.press/v97/locatello19a.html.

Dale M Mesner and Prabir Bhattacharya. Association schemes on triples and a ternary algebra. *Journal of Combinatorial Theory, Series A*, 55(2):204–234, 1990. doi: 10.1016/0097-3165(90)90068-8.

D.M. Mesner and P. Bhattacharya. A Ternary Algebra Arising from Association Schemes on Triples. *Journal of Algebra*, 164(3):595–613, 1994. doi: 10.1006/jabr.1994.1081.

Geoffrey Roeder, Luke Metz, and Diederik P. Kingma. On linear identifiability of learned representations. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pp. 9030–9039. PMLR, 2021. URL https://proceedings.mlr.press/v139/roeder21a.html.

Fan Tian, Misha E. Kilmer, Eric Miller, and Abani Patra. Tensor BM-Decomposition for Compression and Analysis of Video Data. *Numerical Linear Algebra with Applications*, 32(6):e70043, 2025. doi: 10.1002/nla.70043.

Ming Yang. On partial and generic uniqueness of block term tensor decompositions. *Annali dell'Università di Ferrara*, 60(2):465–493, 2014. doi: 10.1007/s11565-013-0190-z.

Ming Yang, Dunren Che, Wen Liu, Zhao Kang, Chong Peng, Mingqing Xiao, and Qiang Cheng. On identifiability of 3-tensors of multilinear rank $(1,\ L_r,\ L_r)$. *Big Data and Information Analytics*, 1 (4):391–401, 2016. doi: 10.3934/bdia.2016017.

Ming Yang, Wen Li, and Mingqing Xiao. On identifiability of higher order block term tensor decompositions of rank $L_r\otimes$ rank-1. *Linear and Multilinear Algebra*, 68(2):223–245, 2020. doi: 10.1080/03081087.2018.1502251.

A GAUGE CHARACTERIZATION AND CANONICAL NORMALIZATION

Lemma 10. *Let nonzero arrays U_{ik}, V_{ij}, W_{jk} satisfy $U_{ik}V_{ij}W_{jk} = 1$ for all indices. There exist nonzero vectors s_i, q_j, u_k such that*

$$U_{ik} = s_i u_k, \qquad V_{ij} = s_i^{-1} q_j, \qquad W_{jk} = q_j^{-1} u_k^{-1}. \tag{18}$$

Proof. Divide the identity at (i, j, k) by the identity at $(i, j, 1)$ to obtain $U_{ik}/U_{i1} = W_{j1}/W_{jk}$. The first expression is independent of j, the second of i, so their common value depends only on k; denote it by u_k, with $u_1 = 1$. Let $s_i = U_{i1}$ and $q_j = W_{j1}^{-1}$. Then $U_{ik} = s_i u_k$ and $W_{jk} = q_j^{-1} u_k^{-1}$; the original identity yields $V_{ij} = s_i^{-1} q_j$. The construction also applies when a mode has size one. □

Proof of Theorem 2. The transformation (3) preserves each ordered term by cancellation. Conversely, the entrywise ratios between two nonzero representations of the same term satisfy Lemma 10. Applying it independently for every t gives exactly (3).

The scaling-parameter group has $r(m + p + n)$ coordinates. A transformation acts trivially on all nonzero factor entries precisely when

$$s_{it} u_{kt} = 1, \qquad s_{it}^{-1} q_{jt} = 1, \qquad q_{jt}^{-1} u_{kt}^{-1} = 1.$$

The first identity forces $s_{it} = \lambda_t$, $u_{kt} = \lambda_t^{-1}$, and the second forces $q_{jt} = \lambda_t$. The third then follows. Thus its kernel is

$$K = \{s_{it} = q_{jt} = \lambda_t,\ u_{kt} = \lambda_t^{-1}\}_{t=1}^{r} \cong (\mathbb{F}^\times)^r.$$

Quotienting by K gives an effective scaling group isomorphic to $(\mathbb{F}^\times)^{r(m+p+n-1)}$. Component permutations act on these blocks, yielding the semidirect product in (4). Over $\mathbb{R}$, the multiplicative group has the same dimension and multiple connected components.

For normalization, choose $s_{it} = A_{it1}^{-1}$, $u_{kt} = A_{1t1}/A_{1tk}$, and $q_{jt} = (A_{1t1}B_{1jt})^{-1}$. Substitution gives (7) and all anchors in (5). If a gauge transformation preserves those anchors, then

$$s_{it} u_{1t} = 1, \qquad s_{1t} u_{kt} = 1, \qquad s_{1t}^{-1} q_{jt} = 1.$$

These relations put the scaling parameters in K, so every entry is unchanged. There are $(m+n-1)+p$ independent anchors per term. Subtracting their total from the original $r(mn + mp + pn)$ parameters gives (6). This proves uniqueness of the normalized representative for labeled components. □

B LOCAL CERTIFICATION AND DIMENSION OBSTRUCTIONS

Proof of Theorem 3. Normalize θ and parameterize the section by $z \in (\mathbb{F}^\times)^d$. The explicit normalization gives a local product of scaling-orbit coordinates and section coordinates. The derivative vanishes on gauge tangents, so the restricted derivative $DF(z)$ has the same rank as J_θ. If it has full column rank, select d output coordinates whose derivative minor is invertible. The inverse function theorem applied to those coordinates makes F locally injective, proving local essential uniqueness after accounting for the finite permutation orbit. Saturating the corresponding neighborhoods under the gauge action gives a neighborhood of the entire orbit as in Definition 1.

If F has constant rank $q < d$ near z, the constant-rank theorem gives a local fiber of dimension $d - q > 0$. Only finitely many points in the section can belong to one gauge-and-permutation orbit. Therefore that fiber contradicts local essential uniqueness. Over $\mathbb{C}$, use complex coordinates; equivalently the real dimensions and ranks are doubled. □

Proof of Proposition 4. At a locally constant-rank identifiable point, Theorem 3 requires $d = \operatorname{rank} DF(z) \le mpn$. The normalized map is polynomial in its free entries. Its maximal-rank locus is open and dense: some maximal minor is a nonzero polynomial, and the union of the nonvanishing loci of such minors is exactly that locus. The same dimension inequality is therefore necessary for generic local identifiability. Substituting (6) and solving for r gives (9). For cubic shapes, division by $3N^2 - 3N + 1$ gives the stated asymptotic expression.

Over $\mathbb{C}$, a fiber is locally described by mpn polynomial equations in d normalized variables. Its local dimension is at least $d - mpn$, so $d > mpn$ precludes an isolated fiber at every point. Over $\mathbb{R}$, the

pointwise conclusion holds under the stated constant-rank assumption; singular real fibers can be isolated. □

B.1 Generic conclusions from exact witnesses

Suppose an integer normalized point yields a nonzero $d \times d$ Jacobian minor modulo a prime q. The determinant of this minor is therefore a nonzero integer. As a polynomial in the normalized free entries, it defines a nonempty Zariski-open nonvanishing set in the complex torus $(\mathbb{C}^\times)^d$. The torus is irreducible, so this set is dense. For real parameters, the zero set of the polynomial has Lebesgue measure zero and empty interior. Theorem 3 therefore yields generic local identifiability in the dimensions of the constructed witness.

If an $mpn \times mpn$ minor is nonzero with $d > mpn$, the normalized map has full row rank on an open dense set. The constant-rank theorem gives local fibers of dimension $d - mpn$ there. Applied to the two full-row-rank witnesses, this argument gives the generic fiber dimensions in Table 1.

A deficient modular reduction supplies a lower bound for the characteristic-zero rank. Floating-point rank estimates describe the singular-value spectrum at a specified tolerance; exact rank certification uses nonvanishing minors and matching upper bounds.

B.2 A one-slice boundary case

Proposition 11. *For $p = 1$, the normalized dimension is $d = rmn$ and $r^\star = 1$. Every entrywise-nonzero r-term decomposition with $r \geq 2$ is locally, and therefore globally, essentially nonunique.*

Proof. Define $T_t(i,k) = A_{itk}B_{i1t}C_{t1k}$. For any sufficiently small nonzero matrix H, replace

$$A_{i1k} \mapsto A_{i1k} + \frac{H_{ik}}{B_{i11}C_{11k}}, \qquad A_{i2k} \mapsto A_{i2k} - \frac{H_{ik}}{B_{i12}C_{21k}}.$$

All other factors remain fixed, so the sum is unchanged. Nonvanishing persists near $H = 0$, whereas the ordered component tensors change from (T_1, T_2) to $(T_1 + H, T_2 - H)$. A finite permutation group cannot identify this continuous family. The formulas for d and $r^\star$ follow from (6)–(9). □

C Separable templates and exact ambiguity families

Proof of Theorem 7. Assume separability in mode A. Define $P^{(j)}, Q^{(j)}$ as in the main text. Their entries are nonzero, and their product satisfies (14). For independently chosen $S_j \in GL_r(\mathbb{F})$ near the identity, put

$$\widetilde{P}^{(j)} = P^{(j)}S_j, \qquad \widetilde{Q}^{(j)} = Q^{(j)}S_j^{-\top}.$$

Keep all A_t fixed and set

$$\widetilde{b}_t^{(j)} = \widetilde{P}_{\cdot t}^{(j)} \oslash \rho^{(t)}, \qquad \widetilde{c}_t^{(j)} = \widetilde{Q}_{\cdot t}^{(j)} \oslash \sigma^{(t)}.$$

Then $\widetilde{P}^{(j)}\widetilde{Q}^{(j)\top} = P^{(j)}Q^{(j)\top}$, so X is unchanged. Continuity preserves nonzero entries for a sufficiently small neighborhood of the identity; for positive real factors it preserves strict positivity as well.

To construct an inequivalent representation, vary one slice using $S_j = I + \epsilon e_1 e_2^\top$. Its inverse is $I - \epsilon e_1 e_2^\top$. Consequently,

$$\widetilde{P}_{\cdot 1} = P_{\cdot 1}, \quad \widetilde{P}_{\cdot 2} = P_{\cdot 2} + \epsilon P_{\cdot 1}, \quad \widetilde{Q}_{\cdot 1} = Q_{\cdot 1} - \epsilon Q_{\cdot 2}, \quad \widetilde{Q}_{\cdot 2} = Q_{\cdot 2}.$$

The first two component matrices change by $(-\epsilon D, +\epsilon D)$, where $D = A_1 \odot (P_{\cdot 1}Q_{\cdot 2}^\top)$ is entrywise nonzero. Every labeled gauge transformation preserves the ordered component tensors. A finite permutation group identifies at most finitely many points of this nonconstant curve, and the intersection with the fixed finite permutation orbit of the starting tuple is isolated. Hence arbitrarily close, inequivalent representations exist.

For the dimension claim, suppose both $P^{(j)}$ and $Q^{(j)}$ have column rank r. Their GL_r action is free, since $P^{(j)}S_j = P^{(j)}$ implies $S_j = I$. The orbit has dimension r^2 per slice. A labeled gauge

preserving every A_t must satisfy $s_{it} = \lambda_t$ and $u_{kt} = \lambda_t^{-1}$. Its residual action is an arbitrary column scaling of $P^{(j)}$, paired with inverse scaling of $Q^{(j)}$, for each j. Conversely, if two members of the mixing family are related by a labeled gauge, full column rank forces their mixing matrices to differ by precisely such a diagonal right factor. The residual gauge dimension is r per slice. Quotienting leaves $r^2 - r$ dimensions per slice and $pr(r-1)$ overall; finite permutations do not alter dimension. Complementary-index symmetry gives $nr(r-1)$ in mode B and $mr(r-1)$ in mode C. □

C.1 A POSITIVE RANK-TWO CONSTRUCTION

Consider the $m = p = n = 5$, $r = 2$ specialization $A_{itk} = 1$ and define

$$U_1 = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 1 \\ 1 & 1 \\ 1 & 1 \end{bmatrix}, \quad U_2 = \begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 1 & 1 \\ 1 & 1 \\ 1 & 1 \end{bmatrix}, \quad V = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 2 & 1 & 1 & 1 \end{bmatrix},$$

with $U_3 = U_4 = U_5 = U_1$, $B_{ijt} = U_j(i,t)$, and $C_{tjk} = V(t,k)$ for every j. For $0 \le \varepsilon < 1/2$, let

$$S_\varepsilon = \begin{bmatrix} 1 & \varepsilon \\ 0 & 1 \end{bmatrix}, \qquad \widetilde{U}_j = U_j S_\varepsilon, \qquad \widetilde{V}_j = S_\varepsilon^{-1} V.$$

All factors remain positive and $\widetilde{U}_j \widetilde{V}_j = U_j V$. At $(1,1,1)$, the unordered component values change from $\{1,1\}$ to $\{1-\varepsilon, 1+\varepsilon\}$. The tensor has BM rank exactly two: the two sides of (16) at $(2,2,2)$ are 297 and 350. This gives a nonunique decomposition with $d = 122 < 125 = mpn$. The family-distance diagnostic uses this specialization.

C.2 NONCONSTANT TEMPLATES

Let $H \in \mathbb{R}^{5\times 5}$ be any entrywise-positive matrix and let ρ, σ be positive vectors. Set

$$\begin{aligned} A_{i1k} &= H_{ik}, & A_{i2k} &= H_{ik}\rho_i\sigma_k, \\ B_{ij1} &= U_j(i,1), & B_{ij2} &= U_j(i,2)/\rho_i, \\ C_{1jk} &= V(1,k), & C_{2jk} &= V(2,k)/\sigma_k. \end{aligned} \tag{19}$$

Then $X^{(j)} = H \odot (U_j V)$ and the same mixing construction applies. For a deterministic nonconstant choice, take $H_{ik} = 1 + ik/25$, $\rho_i = 1 + i/5$, and $\sigma_k = 1 + k/5$. Multiplication by H multiplies both sides of (16) by the same nonzero factor $H_{ik}H_{i1}H_{1k}H_{11}$, so BM rank remains exactly two. Separable template ratios therefore produce ambiguity for nonconstant templates as well.

C.3 A MODE-B OBSTRUCTION WITH A NONSEPARABLE MODE-A RATIO

For an explicit 5^3, $r = 2$ construction, take $A_{i1k} = 1$, $A_{i2k} = 1 + \delta_{ik}$, $B_{ij1} = B_{ij2} = 1$, $C_{1jk} = 1$, and $C_{2jk} = 1 + j/5$. Here δ_{ik} is the Kronecker delta, so $R = \mathbf{1}\mathbf{1}^\top + I_5$ has rank five. For each fixed k,

$$X_{\cdot\cdot k} = M^{(k)} N^{(k)\top}, \qquad M^{(k)} = [A_{\cdot 1k}, A_{\cdot 2k}], \quad N^{(k)} = [C_{1\cdot k}^\top, C_{2\cdot k}^\top].$$

Both matrices have column rank two. Independent transformations $M^{(k)} \mapsto M^{(k)} S_k$, $N^{(k)} \mapsto N^{(k)} S_k^{-\top}$ preserve B and X. A shear changes the unordered component values, giving inequivalent nearby positive factors. Here a full-rank mode-A ratio coexists with separable mode-B and mode-C ratios. The ambiguity follows from the mode-B obstruction in the theorem.

C.4 EXACT JACOBIAN RANK OF THE EXPLICIT FAMILY

Proof of Proposition 8. First set $\varepsilon = 0$. With A fixed, variations of B and C give, in slice j, the matrix tangent space

$$\mathcal{T}_j = \{D_j V + U_j E_j : D_j \in \mathbb{F}^{5\times 2},\ E_j \in \mathbb{F}^{2\times 5}\}.$$

Since U_j and V have rank two, $\dim \mathcal{T}_j = 10 + 10 - 4 = 16$. The five slices are disjoint output blocks, so their direct sum $\mathcal{T}$ has dimension 80.

Let $Z = [e_3 - e_1, e_4 - e_1, e_5 - e_1] \in \mathbb{F}^{5\times 3}$, where e_i are standard basis vectors. Then $VZ = 0$. Choose full-row-rank left annihilators for U_1 and U_2 as

$$N_1 = \begin{bmatrix} -1 & 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 1 & 0 \\ -1 & 0 & 0 & 0 & 1 \end{bmatrix}, \qquad N_2 = \begin{bmatrix} -1 & 1 & -2 & 0 & 0 \\ 0 & 0 & -1 & 1 & 0 \\ 0 & 0 & -1 & 0 & 1 \end{bmatrix}.$$

For the other slices use $N_j = N_1$. The map $Y \mapsto N_j Y Z$ is onto $\mathbb{F}^{3\times 3}$ and has kernel $\mathcal{T}_j$. It therefore represents the quotient by the slice tangent space.

For template variations $H_t = \delta A_t$, the additional derivative in slice j is

$$\delta X^{(j)} = \sum_{t=1}^{2} \operatorname{diag}(U_j(:,t)) H_t \operatorname{diag}(V(t,:)).$$

Both rows of V equal one on columns $1, 3, 4, 5$, and row 2 of Z is zero. Hence $\operatorname{diag}(V(t,:))Z = Z$. Write $W_t = H_t Z$, which can be any 5×3 matrix. Only two distinct slice types remain. For each of the three columns of (W_1, W_2), their joint quotient map has matrix

$$T = \begin{bmatrix} N_1 \operatorname{diag}(U_1(:,1)) & N_1 \operatorname{diag}(U_1(:,2)) \\ N_2 \operatorname{diag}(U_2(:,1)) & N_2 \operatorname{diag}(U_2(:,2)) \end{bmatrix} \in \mathbb{F}^{6\times 10}.$$

Its first three rows are independent. Row four is independent of them because its second coordinate is 3, whereas theirs are zero. Rows five and six are respectively row two minus row one and row three minus row one. Thus $\operatorname{rank} T = 4$. The three columns vary independently, so template variations add $3 \operatorname{rank} T = 12$ dimensions modulo $\mathcal{T}$. The full Jacobian rank is $80 + 12 = 92$.

For $0 \le \varepsilon < 1/2$, $U_j S_\varepsilon$ has the same column space as U_j and $S_\varepsilon^{-1} V$ has the same row space as V. Thus $\mathcal{T}_j$, N_j, and Z remain valid. On columns $1, 3, 4, 5$, the rows of the transformed V are $1 - \varepsilon$ and 1. The template quotient map reduces to the previous one under

$$(W_1, W_2) \longmapsto ((1-\varepsilon)W_1 + \varepsilon W_2, W_2),$$

an invertible change of variables. Therefore its rank is still 12 throughout the family.

The parameter dimension is 150 and the gauge dimension is 28, leaving $150 - 92 - 28 = 30$ excess kernel dimensions. With A fixed, there are 100 parameters and derivative rank 80. The gauge transformations preserving A have effective dimension $2p = 10$, so this restricted kernel contributes $100 - 80 - 10 = 10$ non-gauge directions. The other 20 excess dimensions involve template variations. The 30-dimensional count describes the excess tangent kernel, while the fixed-template construction gives a 10-dimensional family of exact ambiguities.

For fixed positive H, ρ, σ, the passage to (19) is an invertible diagonal change on the full factor space. It multiplies each output slice entrywise by H, also an invertible linear output transformation. Jacobian rank is therefore preserved, establishing the same counts for nonconstant templates. □

D GLOBAL RANK-ONE RECOVERY AND DETERMINISTIC NOISE BOUNDS

Proof of Theorem 9. Given two factorizations, put $U_{ik} = \widetilde{a}_{ik}/a_{ik}$, $V_{ij} = \widetilde{b}_{ij}/b_{ij}$, and $W_{jk} = \widetilde{c}_{jk}/c_{jk}$. Equality of the tensors gives $U_{ik}V_{ij}W_{jk} = 1$, so Lemma 10 yields exactly the gauge transformation. If X is entrywise nonzero, every factor in any one-term representation is nonzero, so this proves global uniqueness among all one-term representations.

In the normalized section, evaluating $i = 1$ gives $c_{jk} = X_{1jk}$. Evaluating $k = 1$ then gives $b_{ij} = X_{ij1}/X_{1j1}$, and evaluating $j = 1$ yields the formula for a_{ik} in (15). Substituting any factorization into (16) proves necessity. Conversely, multiplying the recovered factors gives

$$a_{ik} b_{ij} c_{jk} = \frac{X_{i1k} X_{111} X_{ij1} X_{1jk}}{X_{i11} X_{11k} X_{1j1}},$$

which equals X_{ijk} precisely when (16) holds for every entry. □

D.1 Perturbation bounds for direct recovery

Write each perturbed tensor entry as $\widehat{X}_{ijk} = X_{ijk}(1 + e_{ijk})$ with $|e_{ijk}| \le \delta < 1$. The ratio $\widehat{a}_{ik}/a_{ik}$ is a quotient of two products, each with two factors of the form $1 + e$. Each product differs from one by at most $2\delta + \delta^2$, while the denominator modulus is at least $(1 - \delta)^2$. This gives the first bound in (17). The corresponding one-factor quotient for b gives $2\delta/(1 - \delta)$, and the direct formula for c gives δ.

The hypothesis $\delta < 1$ keeps every denominator in the recovery map nonzero. Writing $\kappa(X) = \max|X_{ijk}|/\min|X_{ijk}|$ gives $\delta = \kappa(X)\|E\|_\infty/\|X\|_\infty$. The relative-error bound therefore depends on the dynamic range of X. Global rank-one uniqueness holds throughout the entrywise-nonzero stratum. For noisy data, the anchor-slice formulas return factors of a one-term approximation.

E Scaling coordinates on a fixed support

Proposition 12. *Fix the supports of the three factors. For component t, form a graph G_t on all $m + p + n$ vertices $\{a_i\} \cup \{b_j\} \cup \{c_k\}$, including isolated vertices. A nonzero A_{itk} gives edge $a_i c_k$, a nonzero B_{ijt} gives edge $a_i b_j$, and a nonzero C_{tjk} gives edge $b_j c_k$. Let e_t be its edge count and $c(G_t)$ its number of connected components. The effective scaling orbit has dimension*

$$g_t = \operatorname{rank} L_t = m + p + n - c(G_t), \qquad d_{\mathrm{supp}} = \sum_{t=1}^{r} (e_t - g_t). \tag{20}$$

Here L_t maps vertex scaling potentials to relative variations of nonzero factor coordinates. In particular, G_t is connected if and only if $g_t = m + p + n - 1$. Normalizing all factor entries on a spanning forest to one gives a unique representative of each labeled scaling orbit on the fixed support, including disconnected supports.

Proof. Write infinitesimal scaling parameters as $(\alpha_i, \beta_j, \gamma_k)$ and set $z_k = -\gamma_k$. Relative variations are $\alpha_i - z_k$ on A, $\beta_j - \alpha_i$ on B, and $z_k - \beta_j$ on C. Thus L_t is an oriented edge-by-vertex incidence matrix. Its kernel consists of potentials constant on each connected component, proving (20). Multiplicatively, choose a root scaling in each tree and solve successively along its edges to set the corresponding factor entries to one. Any remaining transformation preserving these entries is constant on each component and therefore acts trivially on every edge. This proves existence and uniqueness of the representative. □

The graph and forest can be constructed in $O(m + p + n + e_t)$ time. Forest normalization uses nonzero entries as anchors and applies to disconnected supports as well. Connectivity determines the effective scaling dimension, while identifiability is governed by the restricted BM map. Theorem 3 applies to the fixed-support stratum with d_{supp} coordinates. Perturbations that activate zero entries leave this stratum, and zeros can introduce additional term-preserving freedoms.

F Local least-squares stability and loss curvature

Proof of Proposition 5. Let $J = DF(z)$ and define the normal-equation map

$$G(w, E) = DF(w)^*\big(F(w) - F(z) - E\big).$$

Here $*$ is transpose in the real case and conjugate transpose in the complex case. Treat complex variables as real pairs when applying the implicit function theorem. At $(z, 0)$, $G(z, 0) = 0$ and $D_w G(z, 0) = J^*J$, which is positive definite because J has full column rank. The implicit function theorem gives neighborhoods and a unique smooth solution $\widehat{z}(E)$ to $G = 0$ within them. Positive definiteness of the least-squares Hessian persists after shrinking these neighborhoods, making that stationary point a strict local minimizer.

Differentiation at $E = 0$ yields

$$\widehat{z}(E) - z = (J^*J)^{-1}J^*E + O(\|E\|_F^2) = J^\dagger E + O(\|E\|_F^2).$$

The BM map is polynomial, so the smoothness needed for the quadratic remainder holds. Since $\|J^\dagger\|_2 = 1/\sigma_d$, this proves (11). The neighborhood size and remainder constant depend on the reference point, including its derivatives and conditioning. □

F.1 Loss Hessian and residual-to-factor bounds

For real coordinates, differentiating $\ell_X(w) = \frac{1}{2}\|F(w) - X\|_F^2$ twice gives

$$\nabla^2 \ell_X(w) = DF(w)^\top DF(w) + \sum_\alpha (F_\alpha(w) - X_\alpha)\nabla^2 F_\alpha(w).$$

At $X = F(z)$ the residual term vanishes, proving (10). Continuity makes the Hessian positive definite on a sufficiently small ball whenever $DF(z)$ has full column rank. In complex coordinates one applies this calculation to the real and imaginary parts; singular values of the complex derivative are repeated in its realification.

Let $h = w - z$. The integral Taylor remainder satisfies

$$F(z+h) - F(z) = DF(z)h + R(h), \qquad \|R(h)\|_F \le \int_0^1 Lt\|h\|_2^2 \, dt = \frac{L}{2}\|h\|_2^2.$$

Thus $\|F(w) - F(z)\|_F \ge (\sigma - L\|h\|_2/2)\|h\|_2$. If $\|h\|_2 \le \sigma/L$, this is at least $\sigma\|h\|_2/2$. Combining it with

$$\|F(w) - F(z)\|_F \le \|F(w) - (F(z) + E)\|_F + \|E\|_F$$

proves Corollary 6. The estimate holds when the segment $[z, w]$ lies in the stated ball and in the nonzero coordinate domain. Component alignment gives an upper bound for the quotient distance because the minimum over permutations is no larger than the distance for the selected local labeling.

G Exact certificates and verification procedures

G.1 Implementation and execution

The supplement contains `code/verify_bm.py`, stored integer witnesses in `verification/`, and a machine-readable `verification_results.json`. These files provide deterministic checks of the algebraic results. The script evaluates the BM map, canonical normalization, analytic full and reduced Jacobians, exact modular elimination, the specified ambiguity family, and direct rank-one recovery and noise bounds. It checks both derivatives using an exact centered identity for cubic polynomials, verifies integer gauge tangents, and checks reconstruction and normalization invariance under positive gauge transformations (seed 269999).

Run `python code/verify_bm.py --out verification` from the project root. The recorded run used Python 3.13.5, NumPy 2.3.5, and SciPy 1.17.0, with CPU computation and synthetic inputs. The floating-point checks can be repeated with `OPENBLAS_NUM_THREADS=1`. Modular elimination uses exact arithmetic.

G.2 Integer witnesses and exact elimination

The six shapes are ordered as in Table 1. For case $k = 1, \dots, 6$, the script uses a NumPy `Generator(PCG64(270000+k))`. It draws A, then B, then C independently as integer arrays in $\{1, \dots, 19\}$ with the model's array shapes, and overwrites the anchor entries in (5) with ones. The resulting integer arrays lie in the positive normalized section. Array flattening uses C-order, so the output index k varies fastest.

The reduced Jacobian columns are ordered as follows: A_{itk} with $i \ge 2, k \ge 2$ in (i, t, k) order, then B_{ijt} with $i \ge 2$ in (i, j, t) order, then all C_{tjk} in (t, j, k) order. The derivative is (8). Gaussian elimination over $\mathbb{F}_q$ with $q = 1{,}000{,}003$ records pivot columns and the exact rank. The stored factor arrays and their checksums are recorded in the JSON report.

A full-column-rank reduction proves that some square minor is nonzero modulo q, hence nonzero as an integer. The full-row-rank cases use an $mpn \times mpn$ minor in the same way. The verifier stores the reduced Jacobian densely, requiring $O(mpn\, d)$ storage, and targets the small instances in Table 1. The full derivative has at most $3rmpn$ structural nonzeros.

G.3 VERIFICATION OF THE EXPLICIT FAMILY

At $\epsilon = 0$, the integer full Jacobian of the positive 5^3 example has modular rank 92. The analytic calculation in Appendix C.4 supplies the matching characteristic-zero upper bound and the result for the whole family. The script separately evaluates $\epsilon = 0, 0.1, 0.2, 0.3, 0.45$ for the constant- and nonconstant-template examples, verifies positivity, checks the reconstruction to relative error below 10^{-13}, and observes rank 92 at relative singular-value tolerance 10^{-10}. At $(2, 2, 2)$, the two integer sides of the rank-one identity are checked to be 297 and 350.

G.4 VERIFICATION OF RANK-ONE RECOVERY

With seed 271000, the script generates factors of sizes $5 \times 1 \times 6$, $5 \times 4 \times 1$, and $1 \times 4 \times 6$ by exponentiating independent uniform draws on $[-1/2, 1/2]$. It checks that direct recovery agrees with canonical normalization and reconstructs the tensor to floating-point precision. It then uses entrywise relative perturbations in $[-10^{-4}, 10^{-4}]$, computes $\delta = \|E\|_\infty / \min |X_{ijk}|$ for that realization, and verifies all three inequalities in (17). These checks test the recovery formulas and deterministic perturbation bounds on a single rank-one instance.

G.5 INTERPRETING DIAGNOSTIC OUTPUTS

A local identifiability diagnostic returns `certified` when an exact full-column-rank minor is established. It returns `nonidentifiable` when an explicit inequivalent curve is established, or when exact rank deficiency is accompanied by a valid local constant-rank argument. Otherwise it returns `inconclusive`. These statuses use exact algebraic criteria; tolerance-based numerical ranks serve as conditioning diagnostics.

For rank-one membership, form the two sides L_{ijk} and R_{ijk} of (16). Exact equality at all entries establishes membership. In floating-point arithmetic, the normalized residual

$$s(X) = \max_{ijk} \frac{|L_{ijk} - R_{ijk}|}{|L_{ijk}| + |R_{ijk}|}$$

measures rank-one compatibility. For nonzero data in exact arithmetic, its denominators are nonzero. Floating-point implementations require explicit handling of numerical underflow. Statistical use of $s(X)$ requires a specified noise model and calibrated thresholds.

H SUPPLEMENTAL NUMERICAL DIAGNOSTICS

This section presents the aggregate figures and numerical summaries from the empirical study. The available archive contains these summaries, while seed-level Monte Carlo arrays, complete generating distributions, the full ALS initialization specification, original plotting scripts, and run-level audit and filtering logs are unavailable. The settings and values below follow the archive. Appendix G provides the separate executable checks for the deterministic examples and exact algebraic certificates.

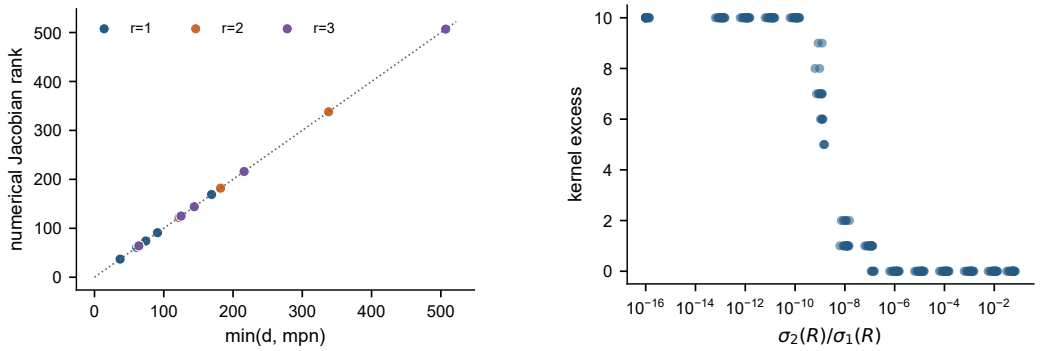
H.1 NUMERICAL JACOBIAN RANKS AND TEMPLATE TRANSITIONS

The numerical ranks in Table 3 use thresholds $10^{-8}\sigma_1$, $10^{-10}\sigma_1$, and $10^{-12}\sigma_1$. The regular instances have full column rank at all three tolerances; the dimension-obstructed instances attain the ambient row rank. Smallest singular values depend on the sampled factors and their normalization. The values below refer to these numerical instances; integer-witness values are recorded separately in the verification JSON.

The transition experiment varies the template ratio away from rank one. In this numerical family, all 100 witnesses attain $d = 122$ at tolerance 10^{-10} once $\sigma_2(R_A)/\sigma_1(R_A) \geq 1.45 \times 10^{-5}$. A rank-one ratio remains separable under the diagonal left and right scalings induced by the gauge.

Table 3: Numerical Jacobian ranks at three relative singular-value tolerances. The smallest singular values belong to the sampled numerical instances; Table 1 gives the separate integer-witness certificates.

(m, p, n)	r	d	mpn	$\min(d, mpn)$	rank_8	rank_{10}	rank_{12}	σ_d
4^3	1	37	64	37	37	37	37	2.68e-1
4^3	2	74	64	64	64	64	64	–
5^3	2	122	125	122	122	122	122	1.66e-2
6^3	2	182	216	182	182	182	182	2.17e-3
8^3	3	507	512	507	507	507	507	9.85e-3
$6 \times 3 \times 8$	3	222	144	144	144	144	144	–



(a) Numerical rank against $\min(d, mpn)$.

(b) A transition away from a separable ratio.

Figure 2: Empirical Jacobian diagnostics. Left: the diagonal marks the parameter-count upper bound. Right: the numerical excess-kernel dimension near the separable-ratio limit depends on the chosen tolerance. The right panel concerns the template-transition experiment; Proposition 8 treats the explicitly constructed rank-92 family.

H.2 Rank-one noise, dynamic range, and detection

The noise-scaling study uses 50 seeds per noise level. The median reconstruction-error slope is 0.9999 for $\eta \le 1.33 \times 10^{-3}$; the seed-wise slope quartiles are $(0.9999, 1.0000, 1.0002)$. The plots summarize reconstruction error, while (17) bounds errors in the individual normalized factors.

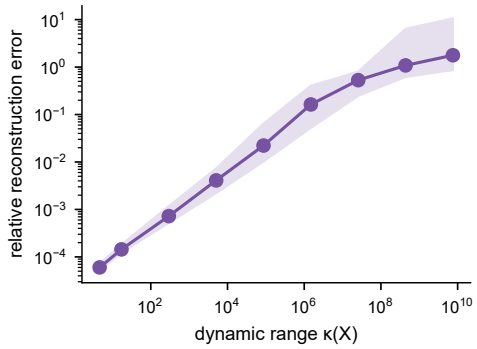
The dynamic-range experiment uses the perturbation convention $10^{-4}\,\mathrm{sd}(X)$. The archived summary records no zero or subnormal entries in the tested range. The first-order fit concerns the small-noise portion of the curve; the final portion lies outside this regime. The deterministic bounds describe worst-case sensitivity across noise realizations.

The detector score is $s(X)$ from Appendix G, with larger values indicating a larger departure from the rank-one identity. The archive summarizes two stress tests. In the noiseless test, X_0 is compared with $X_1 = X_0 + \delta X_2$ while δ varies. AUC exceeds 0.9 between $\delta = 10^{-16}$ and 10^{-15}, probing the finite-precision detection floor in the noiseless setting. At fixed $\delta = 10^{-3}$, AUC is 0.9998 for observation-noise scale $\tau = 3.16 \times 10^{-5}$, decreases to 0.8931 at $\tau = 10^{-4}$, and approaches chance when $\tau \ge 3.16 \times 10^{-4}$. The archive leaves the normalization of the observation-noise scale τ unspecified.

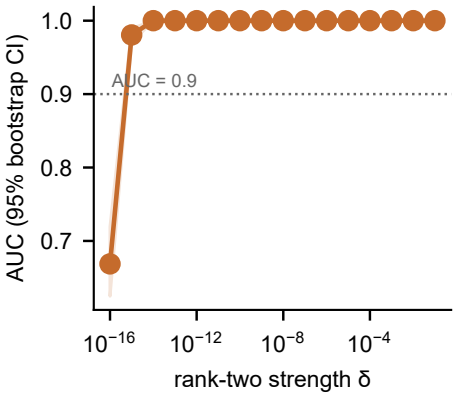
Calibrating the score threshold at the 95th percentile of 300 noisy rank-one null scores gives an empirical null rejection rate of 2% at $\tau = 10^{-3}$. The reported roundoff guard $64\varepsilon_{\text{mach}}\kappa(X)$ is distinct from this statistical threshold.

H.3 ALS stress tests and acceptance filtering

The stress test compares three $r = 2$ settings: regular 6^3, dimension-obstructed 4^3, and separable-template 5^3. It uses 50 broad random initializations and 50 truth-local initializations per setting. The



(a) Sensitivity to tensor dynamic range.



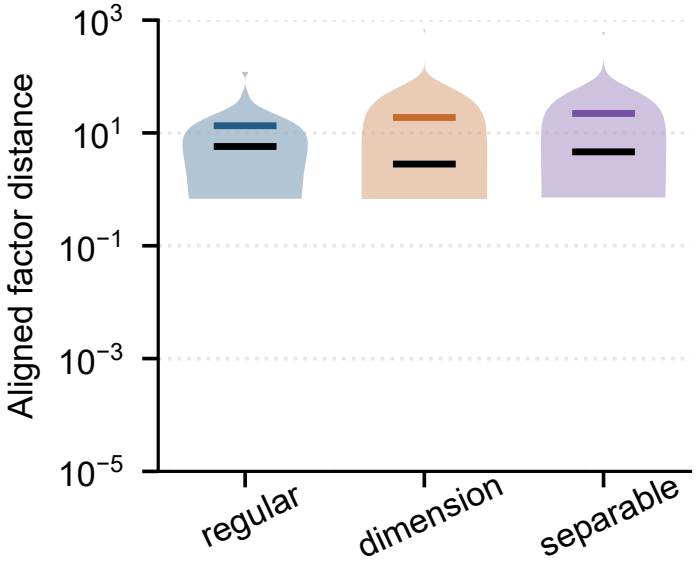
(b) Noiseless near-rank-one detection.

Figure 3: Additional rank-one diagnostics. Left: medians and interquartile ranges over 50 seeds under the stated perturbation convention. Right: AUC for distinguishing X_0 from $X_0 + \delta X_2$ at zero observation noise, with percentile-bootstrap intervals over 200 paired tensors. The horizontal line marks an AUC reference level of 0.9.

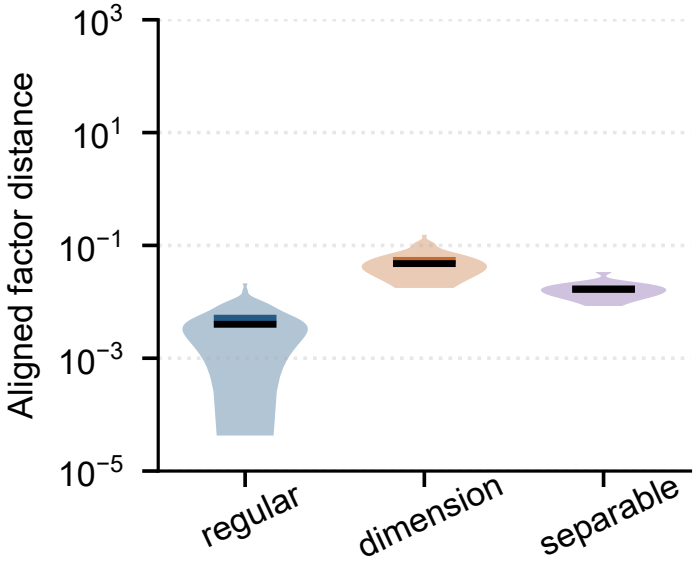
method uses unregularized least-squares block updates, normalization after each recorded sweep, a 3,000-sweep budget, a relative-fit stopping tolerance of 10^{-8}, and a 10^{-6} fit-acceptance threshold for factor analysis. Canonical normalization preserves the reconstruction objective.

For broad initialization, the median final fit errors are 4.08×10^{-3}, 1.71×10^{-4}, and 1.92×10^{-3} in the three settings, respectively. All three medians exceed the stopping tolerance; these runs describe factor variability at the achieved fit levels. Truth-local runs meet the acceptance criterion in $4/50$, $50/50$, and $50/50$ cases. Factor-distance comparisons are qualitative, with four accepted runs in the regular group.

For normalized free-coordinate vectors $z, \widehat{z}$, define the gauge-aligned factor distance by $\min_{\pi \in S_r} \|\widehat{z} - \pi \cdot z\|_2$. Canonical normalization removes scaling freedom, and minimization over common component permutations aligns the labels. The distance scale depends on the anchors defining the section. The archive specifies no additional distance rescaling.



(a) Broad initializations.



(b) Truth-local initializations.

Figure 4: Gauge-aligned factor-distance summaries for broad and truth-local initializations. The distances are interpreted together with the fit errors and acceptance counts in Table 2.