IDENTIFIABILITY OF NONLINEAR MULTI-VIEW CCA UNDER PARTIAL LATENT OBSERVABILITY: LIMITS AND GUARANTEES OF RECOVERY

Anonymous authors
Paper under double-blind review

ABSTRACT

We investigate the identifiability of nonlinear multi-view canonical correlation analysis (CCA) under partial latent observability. We formally show that existing pairwise identifiability does not guarantee simultaneous source recovery using a single encoder per view. When the representation dimension matches the source dimension, sum-of-correlations (SUMCOR), a generalized CCA objective variant, can favor higher-order modes over first-order source directions, causing *representation drift*. Under isotropic Gaussian prior and full cross-view latent coverage, we establish three recovery regimes at population global optima. First, global first-order dominance guarantees recovery of each view-specific source up to an affine transformation. Second, known overlap ranks enable Top-r and masked SUMCOR to achieve the same recovery under pairwise first-order dominance. Third, even without either dominance condition, SUMCOR admits exact affine source readouts at finite, spectrum-dependent representation dimensions under a joint spectral compatibility condition. The enlarged representations retain source directions alongside stronger nonlinear modes. Our results demonstrate how spectral structure, rank information, and representation capacity govern source recovery. Synthetic experiments confirm the predicted recovery regimes and representation drift.

1 INTRODUCTION

Paired and multi-view data arise across multimodal sensing, synchronized instrumentation, and repeated measurements. Canonical correlation analysis (CCA) and its multi-view extensions learn representations by maximizing normalized cross-view dependence (Hotelling, 1936; Kettenring, 1971). Kernel and neural extensions enlarge the representation class (Fukumizu et al., 2007; Andrew et al., 2013; Benton et al., 2019), while related self-supervised objectives control second-order statistics to prevent collapse (Ermolov et al., 2021; Zbontar et al., 2021; Bardes et al., 2022). However, high cross-view correlation alone does not specify which underlying source directions a representation retains. Establishing identifiability requires assumptions on the generative process or latent distributions (Hyvärinen & Pajunen, 1999; Locatello et al., 2020).

Existing guarantees cover post-nonlinear mixtures and nonlinear shared–private models (Lyu & Fu, 2020; Lyu et al., 2022; Karakasis & Sidiropoulos, 2023). Under latent distributional priors, Han et al. (2026a) establish two-view affine identifiability, while Han et al. (2026b) identify correlated source subspaces under view-specific loadings. Partial latent observability introduces a further issue: different view pairs may reveal different source directions. This setting also appears in multi-view representation learning, where one encoder per view with content selectors identifies overlapping latent blocks up to smooth bijections (Yao et al., 2024). For a multi-view correlation objective, one encoder must allocate its representation dimensions across these pairwise interactions. The question is therefore: *When does pairwise recoverability lead to simultaneous source recovery by one encoder per view?*

In this work, we study a coordinate-subset specialization of the loading model (Han et al., 2026b), which performs coordinate selection to govern partial latent observability for affine recovery of noisy view-specific sources. Focusing on population global optima under an isotropic Gaussian model with

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.

full cross-view latent coverage, we analyze pairwise nuclear-norm sum-of-correlations (SUMCOR), a basis-invariant objective summing pairwise canonical correlations.

We identify *representation drift*, a failure mode of affine recovery when the representation and source dimensions match, provide a sufficient failure criterion, and present a three-view counterexample. We then establish three distinct recovery regimes:

- **Global First-order Dominance.** Every SUMCOR maximizer recovers each source up to an invertible affine transformation at matching representation dimensions.
- **Known Overlap Ranks.** Under pairwise first-order dominance, Top-r and masked SUMCOR achieve affine recovery without global dominance or known overlap supports.
- **Finite Overcompleteness.** Under joint spectral compatibility, SUMCOR enables exact affine source readouts at finite, spectrum-dependent representation dimensions, even when both dominance conditions fail. These enlarged representations additionally preserve nonlinear modes.

Finally, we provide conditional non-Gaussian extensions and synthetic experiments that confirm the predicted recovery regimes and representation drift, with near-perfect SUMCOR recovery also observed in the tested non-Gaussian constructions.

2 Related Work

Correlation objectives and spectral analysis. CCA and its multiset, kernel, and deep variants maximize normalized cross-view dependence (Hotelling, 1936; Kettenring, 1971; Fukumizu et al., 2007; Andrew et al., 2013; Benton et al., 2019). Canonical expansions and nonparametric CCA characterize canonical functions spectrally (Lancaster, 1958; Eagleson, 1964; Michaeli et al., 2016). For jointly Gaussian scalars, network maximal correlation gives conditions admitting linear optimizers of a signed correlation sum (Feizi et al., 2017), while MCPCA admits linear optimizers for a joint-covariance eigenvalue objective (Feizi & Tse, 2017). Our pairwise nuclear-norm SUMCOR uses one vector-valued encoder per view and edge-specific canonical bases, targeting source retention at every population optimum.

Identifiability of nonlinear CCA. Prior guarantees cover post-nonlinear mixtures (Lyu & Fu, 2020) and nonlinear shared–private models (Lyu et al., 2022; Karakasis & Sidiropoulos, 2023). Under distributional priors and spectral separation, Han et al. (2026a) establish two-view affine identifiability. Han et al. (2026b) identify correlated source subspaces under loadings that permit unobserved directions. Their infinite-dimensional multi-view theorem targets the intersection of pairwise correlated subspaces. We specialize this loading model to coordinate subsets and study whether finite-dimensional aggregate optima retain the collective pairwise span, which equals each noisy source's first-order space under our covariance and coverage assumptions.

Partial observability and recovery targets. Content–style and multimodal models identify shared information up to invertible transformations (von Kügelgen et al., 2021; Daunhawer et al., 2023). Yao et al. (2024) identify shared blocks of potentially dependent latents up to smooth bijections using one encoder per view and content selectors, including unknown shared-block sizes with known view-specific dimensions. Our rank-constrained CCA regime assumes known overlap ranks but unknown supports. Under isotropic Gaussianity, pairwise first-order dominance, and full cross-view coverage, it recovers noisy view-specific sources up to invertible affine transformations at matched source and representation dimensions, without identifying generative selectors or unique mask supports.

Whitening, spectral selection, and capacity. Whitening and covariance regularization are used to prevent collapse (Ermolov et al., 2021; Zbontar et al., 2021; Bardes et al., 2022). Spectral analyses link positive-pair operators to linear prediction and approximation of view-invariant functions (HaoChen et al., 2021; Johnson et al., 2023). Other work studies feature suppression (Robinson et al., 2021) and dimensional collapse (Jing et al., 2022). Our representation drift loses source directions to higher-order modes despite exact population whitening and full-rank representation covariance. Our finite-overcomplete guarantee requires joint spectral compatibility of the finite cutoffs and yields affine source readouts at prescribed, spectrum-dependent widths while allowing nonlinear features.

3 Problem Setup and Preliminaries

We introduce the multi-view model, learning objective, and recovery targets, then recall pairwise identifiability (Han et al., 2026b) for the simultaneous source recovery analysis.

Notation. Plain letters denote scalars, bold lowercase letters vectors and vector-valued functions, bold uppercase letters matrices, and calligraphic letters sets or spaces.

3.1 Structured Multi-view Model and Learning Objective

Let $N \geq 2$ observed views $\mathbf{x}_i \in \mathcal{X}_i \subseteq \mathbb{R}^{d_{\mathcal{X}_i}}$ be generated from a shared latent vector $\mathbf{c} \in \mathbb{R}^{d_c}$. Each view $i \in [N] := \{1, \ldots, N\}$ observes a subset of latent coordinates indexed by $\mathcal{V}_i \subseteq [d_c]$, with $m_i := |\mathcal{V}_i| > 0$ and a fixed enumeration $(v_{i,1}, \ldots, v_{i,m_i})$. We consider a structured specialization of the general rectangular-loading model (Han et al., 2026b) and factorize its loading matrix as $\mathbf{A}_i = \mathbf{B}_i \mathbf{P}_i$. Here, the selection matrix $\mathbf{P}_i \in \{0,1\}^{m_i \times d_c}$, whose k-th row is $\mathbf{e}_{v_{i,k}}^\top$, encodes partial observability and satisfies $\mathbf{P}_i \mathbf{P}_i^\top = \mathbf{I}_{m_i}$, while $\mathbf{B}_i \in \mathrm{GL}(m_i)$ denotes an invertible view-specific mixing matrix. The view-specific source and its corresponding observation are generated according to:

$$\mathbf{s}_i = \mathbf{A}_i \mathbf{c} + \boldsymbol{\epsilon}_i = \mathbf{B}_i \mathbf{P}_i \mathbf{c} + \boldsymbol{\epsilon}_i, \qquad \mathbf{x}_i = \mathbf{g}_i(\mathbf{s}_i), \tag{1}$$

where $\boldsymbol{\epsilon}_i \in \mathbb{R}^{m_i}$ is a view-specific perturbation. The observation map $\mathbf{g}_i : \mathcal{S}_i \to \mathcal{X}_i$ is a smooth diffeomorphism onto its image, mapping the view-specific source space $\mathcal{S}_i \subseteq \mathbb{R}^{m_i}$ to the observation space $\mathcal{X}_i$, where $d_{(\cdot)}$ denotes ambient dimension.

Remark 1 (Structural implications of the loading factorization). The factorization in Equation 1 separates *latent visibility* from *invertible within-view mixing*. It accommodates the sharing patterns of fully shared, shared–private, and subset-shared latent models within a common coordinate-based description. Under the latent distributional prior stated later in this section, it also allows all pairwise first-order targets at a view to be spanned by coordinate blocks of a single source representation, supporting the subsequent masked-recovery analysis. Connections to existing generative models are discussed in Appendix A.

Definition 1 (Active view-pair set)**.** For distinct views $i \neq j$, let $r_{ij} := |\mathcal{V}_i \cap \mathcal{V}_j|$ denote their overlap cardinality. A view pair $\{i, j\}$ is *active* if $r_{ij} > 0$. The active view-pair set is

$$\mathcal{E} := \big\{\{i, j\} : 1 \leq i < j \leq N,\ r_{ij} > 0\big\}.$$

The learner observes paired samples $\{\mathbf{x}_i\}_{i=1}^N$ over an active view-pair set $\mathcal{E}$, which defines the objective edges. The underlying latent variables, mixing matrices $\{\mathbf{B}_i\}_{i=1}^N$, observation maps $\{\mathbf{g}_i\}_{i=1}^N$, and selection matrices $\{\mathbf{P}_i\}_{i=1}^N$ are entirely unknown. Crucially, $\mathcal{E}$ conveys neither the identities of shared coordinates nor their overlap ranks r_{ij}. Estimating $\mathcal{E}$ from data is outside the scope of this analysis. Both rank-constrained objectives assume known overlap ranks, while masked SUMCOR also learns coordinate masks.

Identifiability target. We seek conditions under which nonlinear CCA in a multi-view setting recovers overlap-induced source subspaces up to a restricted class of invertible transformations, such as affine maps. Our analysis concerns identifiability at global optima of the population objective, rather than the optimization dynamics or convergence of specific estimators, such as neural-network-based encoders.

3.2 Pairwise Identifiability Revisited

Let $\mathbf{f}_i \in L^2(\mathcal{X}_i; \mathbb{R}^{d_{\mathcal{Z}_i}})$ denote the learnable square-integrable encoder for view $i \in \{1, \ldots, N\}$, with representation dimension $d_{\mathcal{Z}_i}$ and output $\mathbf{z}_i := \mathbf{f}_i(\mathbf{x}_i) \in \mathcal{Z}_i \subseteq \mathbb{R}^{d_{\mathcal{Z}_i}}$. Assuming a nonsingular covariance $\boldsymbol{\Sigma}_{ii} := \mathrm{Cov}(\mathbf{z}_i, \mathbf{z}_i) \succ 0$, we formulate the population CCA (Hotelling, 1936) and pairwise nuclear-norm SUMCOR objectives over an active view-pair set $\mathcal{E}$ as:

$$J_{\mathrm{CCA}}(\mathbf{z}_i, \mathbf{z}_j) := \left\| \boldsymbol{\Sigma}_{ii}^{-1/2} \boldsymbol{\Sigma}_{ij} \boldsymbol{\Sigma}_{jj}^{-1/2} \right\|_*, \quad J_{\mathrm{SUMCOR}}^{\mathcal{E}}\big(\{\mathbf{z}_i\}_{i=1}^N\big) := \sum_{\{i,j\} \in \mathcal{E}} J_{\mathrm{CCA}}(\mathbf{z}_i, \mathbf{z}_j), \tag{2}$$

where $\|\cdot\|_*$ is the nuclear norm. Throughout, SUMCOR denotes this basis-invariant objective, with one encoder per view and edge-specific canonical bases. We do not claim general equivalence to classical shared-basis trace SUMCOR, which uses one shared basis per view.

Assuming each observation map $\mathbf{g}_i$ is invertible on its image $\mathcal{X}_i$, optimizing over square-integrable encoders $\mathbf{f}_i$ is equivalent to optimizing over the composite source-space maps $\mathbf{h}_i := \mathbf{f}_i \circ \mathbf{g}_i$ (Han et al., 2026a). The corresponding whitened source representation is given by:

$$\tilde{\mathbf{h}}_i(\mathbf{s}_i) := \mathbf{\Sigma}_{ii}^{-1/2}\big(\mathbf{h}_i(\mathbf{s}_i) - \mathbb{E}[\mathbf{h}_i(\mathbf{s}_i)]\big), \quad \text{where } \operatorname{Cov}\big(\tilde{\mathbf{h}}_i(\mathbf{s}_i)\big) = \mathbf{I}_{d_{\mathcal{Z}_i}}.$$

Population whitening makes the representation coordinates orthonormal in function space.

Assumption 1 (Isotropic Gaussian model)**.** The shared latent vector $\mathbf{c}$ and the perturbations $\{\boldsymbol{\epsilon}_i\}_{i=1}^N$ are mutually independent and satisfy

$$\mathbf{c} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_{d_c}), \qquad \boldsymbol{\epsilon}_i \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_{m_i}), \quad i \in [N].$$

Under Assumption 1, let $\mathbf{W}_i := \operatorname{Cov}(\mathbf{s}_i)^{-1/2} = (\mathbf{A}_i\mathbf{A}_i^\top + \mathbf{I}_{m_i})^{-1/2}$ with loading matrices $\mathbf{A}_i = \mathbf{B}_i\mathbf{P}_i$. The whitened source vector $\tilde{\mathbf{s}}_i := \mathbf{W}_i\mathbf{s}_i$ has zero mean and identity covariance. For each active pair $\{i,j\} \in \mathcal{E}$, the normalized source cross-covariance has rank r_{ij} and singular value decomposition (SVD):

$$\mathbf{R}_{ij} := \operatorname{Cov}(\tilde{\mathbf{s}}_i, \tilde{\mathbf{s}}_j) = \mathbf{W}_i\mathbf{A}_i\mathbf{A}_j^\top\mathbf{W}_j = \mathbf{W}_i\mathbf{B}_i\mathbf{P}_i\mathbf{P}_j^\top\mathbf{B}_j^\top\mathbf{W}_j = \mathbf{U}_{ij}\mathbf{T}_{ij}\mathbf{V}_{ij}^\top,$$

with nonzero singular values $1 > t_{ij,1} \geq \cdots \geq t_{ij,r_{ij}} > 0$.

Assumption 2 (Pairwise First-order Dominance)**.** For every $\{i,j\} \in \mathcal{E}$, $t_{ij,r_{ij}} > t_{ij,1}^2$.

Proposition 1 (Pairwise subspace recovery (Han et al., 2026b))**.** *Fix $\{i,j\} \in \mathcal{E}$ and let $r := r_{ij}$. Under Assumption 1 and Assumption 2, any centered, whitened population maximizer $(\tilde{\mathbf{h}}_i^\star, \tilde{\mathbf{h}}_j^\star)$ of J_{CCA} in Equation 2 with $d_{\mathcal{Z}_i} = d_{\mathcal{Z}_j} = r$ satisfies, almost surely for some $\mathbf{O}_{ij}, \mathbf{O}_{ji} \in O(r)$:*

$$\tilde{\mathbf{h}}_i^\star(\mathbf{s}_i) = \mathbf{O}_{ij}\mathbf{U}_{ij}(:, 1:r)^\top\mathbf{W}_i\mathbf{s}_i, \qquad \tilde{\mathbf{h}}_j^\star(\mathbf{s}_j) = \mathbf{O}_{ji}\mathbf{V}_{ij}(:, 1:r)^\top\mathbf{W}_j\mathbf{s}_j.$$

Thus, pairwise CCA identifies correlated subspaces up to orthogonal transformations.

The proof combines Hermite–Mehler expansion with L^2-whitening orthonormality (Han et al., 2026b). Under Assumption 2, the top r singular modes correspond to first-order canonical functions, and Ky Fan's variational principle restricts any optimal r-dimensional subspace to these modes up to orthogonal rotation.

4 Theory

Pairwise identifiability alone does not guarantee that a single encoder with minimum-sufficient capacity can simultaneously preserve all directions recoverable across multiple view pairs. We analyze this limitation by varying the objective structure and representation dimension to achieve full recovery of view-specific sources under full cross-view latent coverage.

Let $\mathcal{N}_i := \{j : \{i,j\} \in \mathcal{E}\}$ denote the set of views that share at least one latent coordinate with view i.

Assumption 3 (Full cross-view latent coverage)**.** For every view $i \in [N]$, every latent coordinate in $\mathcal{V}_i$ is shared with at least one neighboring view $j \in \mathcal{N}_i$, i.e., $\bigcup_{j\in\mathcal{N}_i}(\mathcal{V}_i \cap \mathcal{V}_j) = \mathcal{V}_i$.

Definition 2 (Minimum-sufficient encoder capacity)**.** An encoder $\mathbf{f}_i : \mathcal{X}_i \to \mathbb{R}^{d_{\mathcal{Z}_i}}$ has *minimum-sufficient capacity* for full affine source recovery if $d_{\mathcal{Z}_i} = m_i$, where m_i is the minimal representation dimension enabling exact recovery of the full source $\mathbf{s}_i$ via an affine readout.

The Gaussian analysis has conditional non-Gaussian extensions in Appendix E.

4.1 Finite-Dimensional SUMCOR: Recovery and Representation Drift

At minimum-sufficient capacity, higher-order modes can displace first-order source directions under SUMCOR, making full affine source recovery strictly suboptimal even with full cross-view latent coverage and Assumption 2. We call this failure mode *representation drift*. The following proposition gives a sufficient condition for it.

Proposition 2 (Representation drift at minimum-sufficient capacity)**.** *Under Assumption 1, Assumption 2, and Assumption 3, let $d_{\mathcal{Z}_i} = m_i$ for all i. For each view, choose $k_i \in \{0, \dots, m_i\}$, $\mathbf{Q}_i \in \mathbb{R}^{m_i \times (m_i - k_i)}$ with orthonormal columns, and a centered, whitened function $\boldsymbol{\psi}_i(\mathbf{s}_i) \in \mathbb{R}^{k_i}$ satisfying $\mathbb{E}[\boldsymbol{\psi}_i \tilde{\mathbf{s}}_i^\top] = \mathbf{0}$. If*

$$\underbrace{\sum_{\{i,j\}\in\mathcal{E}} \left\|\mathrm{Cov}(\boldsymbol{\psi}_i, \boldsymbol{\psi}_j)\right\|_*}_{\textit{higher-order spectral gain}} > \underbrace{\sum_{\{i,j\}\in\mathcal{E}} \left(\|\mathbf{R}_{ij}\|_* - \|\mathbf{Q}_i^\top \mathbf{R}_{ij} \mathbf{Q}_j\|_*\right)}_{\textit{first-order spectral loss}}, \tag{3}$$

then every encoder collection permitting full affine source recovery in all views is strictly suboptimal for SUMCOR.

Proof sketch. Concatenating $\mathbf{Q}_i^\top \tilde{\mathbf{s}}_i$ and $\boldsymbol{\psi}_i$ gives a feasible m_i-dimensional encoder. Gaussian conditional expectations make its edgewise cross-covariances block diagonal between first- and higher-order components. Thus, Equation 3 makes its SUMCOR value strictly exceed $\sum_{\{i,j\}\in\mathcal{E}} \|\mathbf{R}_{ij}\|_*$, the value attained by every fully affine-recovering collection at $d_{\mathcal{Z}_i} = m_i$. The full proof is given in Appendix B.1. A three-view instance in Appendix B.2 realizes this strict gain–loss inequality despite pairwise first-order dominance and full cross-view latent coverage. □

To characterize this allocation, we compare aggregate first-order directional support against higher-order modal contributions.

Assumption 4 (Global First-order Dominance)**.** For each $i \in [N]$, define

$$\mathbf{G}_i := \sum_{j\in\mathcal{N}_i} (\mathbf{R}_{ij}\mathbf{R}_{ij}^\top)^{1/2}, \qquad \alpha_i := \lambda_{\min}(\mathbf{G}_i), \qquad \beta_i := \sum_{j\in\mathcal{N}_i} t_{ij,1}^2,$$

where $t_{ij,1} = \sigma_1(\mathbf{R}_{ij})$ and $(\cdot)^{1/2}$ denotes the principal positive semidefinite matrix square root. We assume $\alpha_i > \beta_i$ for all $i \in [N]$.

The quantity α_i quantifies the weakest aggregate support for a first-order direction, accounting for correlation strength and pairwise subspace geometry. Conversely, β_i bounds higher-order contributions: under joint Gaussianity, $t_{ij,1}^2$ is the norm of the coupling operator restricted to Hermite modes of degree at least two on edge $\{i, j\}$. Assumption 4 thus requires view-wise first-order signal strength to strictly dominate higher-order modal interference.

Theorem 4.1 (Full affine identifiability at minimum-sufficient capacity)**.** *Under Assumption 1, Assumption 3, and Assumption 4, let $d_{\mathcal{Z}_i} = m_i$ for all $i \in [N]$. Any population maximizer $\{\tilde{\mathbf{h}}_i^\star\}_{i=1}^N$ of $J_{\mathrm{SUMCOR}}^{\mathcal{E}}$ over centered, whitened, square-integrable encoders satisfies, almost surely,*

$$\tilde{\mathbf{h}}_i^\star(\mathbf{s}_i) = \mathbf{O}_i \mathbf{W}_i \mathbf{s}_i, \quad i \in [N], \tag{4}$$

for orthogonal matrices $\mathbf{O}_i \in O(m_i)$. Equivalently, the unwhitened representations recover each view-specific source $\mathbf{s}_i$ up to an invertible affine transformation. The optimal objective value is $\sum_{\{i,j\}\in\mathcal{E}} \sum_{k=1}^{r_{ij}} t_{ij,k}$.

Proof sketch. Decompose each whitened encoder into a linear function of its source and an orthogonal higher-order remainder. The Gaussian spectral structure underlying pairwise identifiability (Han et al., 2026b) bounds the cross-view contribution of this remainder. At minimum-sufficient capacity, retaining higher-order features necessarily reduces first-order energy. Global first-order dominance makes the resulting loss strictly exceed any possible gain, so every maximizer contains only linear source functions. Whitening then makes each encoder an orthogonal transformation of its whitened source, yielding full affine recovery and the stated optimum. The full proof is given in Appendix B.3. □

4.2 Rank-constrained SUMCOR: Recovery under Pairwise First-order Dominance

Known overlap ranks provide an alternative sufficient recovery route. Controlling the number of canonical directions rewarded on each edge guarantees full affine identifiability under *pairwise*

first-order dominance at $d_{\mathcal{Z}_i} = m_i$, without known overlap supports. Pairwise and global dominance do not imply one another in general.

Assume that the overlap ranks $\{r_{ij}\}_{\{i,j\}\in\mathcal{E}}$ are known. We analyze two rank-constrained objectives over edges $\{i,j\} \in \mathcal{E}$: (i) **Top-r SUMCOR** truncates the rewarded spectrum of normalized cross-covariance $\mathbf{C}_{ij} := \mathbf{\Sigma}_{ii}^{-1/2}\mathbf{\Sigma}_{ij}\mathbf{\Sigma}_{jj}^{-1/2}$ to its r_{ij} leading modes without constraining the rank of $\mathbf{C}_{ij}$. (ii) **masked SUMCOR** jointly optimizes encoders and sample-independent binary coordinate selectors $\mathbf{M}_{i|j} \in \{0,1\}^{r_{ij}\times m_i}$. Each selector consists of r_{ij} distinct rows of $\mathbf{I}_{m_i}$ and therefore satisfies $\mathbf{M}_{i|j}\mathbf{M}_{i|j}^\top = \mathbf{I}_{r_{ij}}$. It selects exactly r_{ij} raw encoder coordinates before whitening. The two objectives are

$$J_{\mathrm{top}}^{\mathcal{E}} := \sum_{\{i,j\}\in\mathcal{E}} \sum_{k=1}^{r_{ij}} \sigma_k(\mathbf{C}_{ij}), \qquad J_{\mathrm{mask}}^{\mathcal{E}} := \sum_{\{i,j\}\in\mathcal{E}} J_{\mathrm{CCA}}\big(\mathbf{M}_{i|j}\mathbf{z}_i, \mathbf{M}_{j|i}\mathbf{z}_j\big). \tag{5}$$

Each selected block is whitened separately, and masks may overlap across edges. Under Assumption 1, all incident first-order targets are simultaneously realizable as coordinate blocks of one invertible source representation (Lemma 1). Maximizing out the finite mask variables preserves the population maximizing encoders (Lemma 2).

Theorem 4.2 (Full affine identifiability by rank-constrained SUMCOR)**.** *Under Assumption 1, Assumption 2, and Assumption 3, let $d_{\mathcal{Z}_i} = m_i$ for all i and assume known overlap ranks r_{ij}. Over square-integrable encoders with nonsingular covariances, both Top-r SUMCOR and masked SUMCOR, the latter jointly optimized over masks, attain the population maximum $J_{\mathrm{aff}} := \sum_{\{i,j\}\in\mathcal{E}} \sum_{k=1}^{r_{ij}} t_{ij,k}$. The centered, whitened forms of every maximizing encoder collection satisfy, almost surely,*

$$\tilde{\mathbf{h}}_i^\star(\mathbf{s}_i) = \mathbf{O}_i\mathbf{W}_i\mathbf{s}_i, \quad i \in [N], \tag{6}$$

for orthogonal matrices $\mathbf{O}_i \in O(m_i)$. Equivalently, the unwhitened representations recover each view-specific source $\mathbf{s}_i$ up to an invertible affine transformation.

Proof sketch. Pairwise first-order dominance yields $J_{\mathrm{mask}}^{\mathcal{E}} \le J_{\mathrm{top}}^{\mathcal{E}} \le J_{\mathrm{aff}}$. Full source representations achieve the Top-r bound, while simultaneous coordinate realizability enables masks to attain it as well. Thus, any maximizer saturates the Top-r bound on every edge. Pairwise identifiability places all incident first-order targets in the centered representation span. Coverage and $d_{\mathcal{Z}_i} = m_i$ force this span to equal the complete first-order source space, yielding affine recovery (see Appendix C). □

Remark 2. Theorem 4.2 guarantees source recovery rather than identification of generative selectors or unique mask supports. If pairwise first-order dominance fails, rank constraints alone may omit required first-order modes, motivating the larger representation capacities ($d_{\mathcal{Z}_i} > m_i$) explored next.

4.3 Finite-Overcomplete Recovery beyond First-order Dominance

Even when both dominance conditions fail, SUMCOR guarantees affine source readouts at specified finite, spectrum-dependent dimensions under joint spectral compatibility. Gaussianity and coverage ensure finite cutoff spaces containing all source directions, but do not ensure this compatibility.

To quantify capacity, extend the first-order aggregate $\mathbf{G}_i$ to all centered L^2 source functions. With $\mathcal{T}_{ij}f := \mathbb{E}[f(\mathbf{s}_j) \mid \mathbf{s}_i]$, define

$$\Gamma_i := \sum_{j\in\mathcal{N}_i} (\mathcal{T}_{ij}\mathcal{T}_{ij}^*)^{1/2}, \qquad \mathsf{Q}_i^\gamma := \mathbf{1}_{[\gamma,\infty)}(\Gamma_i). \tag{7}$$

Its first-order restriction is $\mathbf{G}_i$, with $\alpha_i > 0$ by coverage. For $0 < \gamma < \min_i \alpha_i$, set $\mathcal{Q}_i^\gamma := \mathrm{ran}(\mathsf{Q}_i^\gamma)$ and $D_i(\gamma) := \mathrm{rank}(\mathsf{Q}_i^\gamma)$. Under Gaussianity, Γ_i is compact, so this space is finite-dimensional. It contains all first-order directions, with $D_i(\gamma) - m_i$ surplus dimensions accommodating nonlinear modes (Appendix D.1).

The following joint spectral compatibility condition requires these cutoff spaces to attain the SUMCOR trace bound.

Assumption 5 (Joint spectral compatibility of finite cutoffs)**.** For some $0 < \gamma < \min_i \alpha_i$,

$$\sum_{\{i,j\}\in\mathcal{E}} \|\mathsf{Q}_i^\gamma \mathcal{T}_{ij} \mathsf{Q}_j^\gamma\|_* = \frac{1}{2}\sum_{i=1}^{N} \mathrm{tr}(\mathsf{Q}_i^\gamma \Gamma_i). \tag{8}$$

This holds if, in suitable edgewise singular systems, the cutoff projections retain both functions of each nonzero singular pair or annihilate both (Appendix D.2).

Theorem 4.3 (Finite-overcomplete affine source recovery)**.** *Under Assumption 1, Assumption 3, and Assumption 5, let $d_{\mathcal{Z}_i} = D_i(\gamma)$ for all i. SUMCOR attains its population maximum over centered, whitened, square-integrable encoders. Every maximizing collection admits matrices $\mathbf{L}_i \in \mathbb{R}^{m_i \times D_i(\gamma)}$ satisfying, almost surely,*

$$\mathbf{L}_i \tilde{\mathbf{h}}_i^\star(\mathbf{s}_i) = \mathbf{W}_i \mathbf{s}_i, \qquad i \in [N]. \tag{9}$$

Equivalently, the unwhitened representations recover each view-specific source $\mathbf{s}_i$ via an affine readout.

Proof sketch. The cutoff spaces maximize the view-wise trace bounds at the prescribed dimensions. Joint spectral compatibility makes the resulting SUMCOR upper bound exact. Spectral gaps below the cutoffs force every maximizer to span these spaces, yielding finite source readouts (Appendix D.3). □

Remark 3 (Capacity replaces spectral priority). In contrast to the preceding regimes, recovery does not require first-order modes to precede their nonlinear competitors. Intuitively, the objective need not discard stronger higher-order modes. Instead, additional dimensions accommodate them alongside weaker first-order directions. The theorem specifies a finite capacity allocation satisfying this compatibility condition, not a universal dimension bound or a guarantee for every larger allocation. Overlap ranks do not enter the objective, although the sufficient capacities depend on the population spectrum. The result does not single out the source coordinates within the learned basis and makes no claim about neural optimization dynamics.

In Setup C, six coordinates per view suffice for three-dimensional source recovery despite failure of both dominance conditions (Appendix D.4).

5 Numerical Experiments

Our synthetic experiments examine how spectral dominance, overlap-rank information, and representation capacity affect recovery of view-specific sources under partial observability. We extend the disentanglement setup of Zimmermann et al. (2021); Matthes et al. (2023) to the multi-view model in Equation 1, complementing the affine recovery studies of Han et al. (2026b;a). Across source distributions, we compare correlation objectives to characterize spectral competition and test the two recovery mechanisms established by our theory.

5.1 Experimental Setup

Data generation. Following the online-sampling protocol of Zimmermann et al. (2021); Matthes et al. (2023), we generate three views via fixed, invertible feedforward neural networks $\mathbf{x}_i = \mathbf{g}_i(\mathbf{s}_i)$ with non-singular weight matrices and elementwise invertible activations. Each source vector $\mathbf{s}_i$ has dimension $m_i = 3$. Our recovery target is the full vector $\mathbf{s}_i$, including view-specific perturbations, rather than only the shared latent variable. To examine recovery across spectral regimes, we evaluate three experimental configurations:

- **Setup A** uses Gaussian, Gamma, Poisson, negative binomial, and hypergeometric sources at $d_{\mathcal{Z}} = 3$. Its joint constructions satisfy the corresponding pairwise and global dominance conditions, verified for the non-Gaussian cases in Proposition 6.
- **Setup B** retains pairwise dominance but violates global dominance.
- **Setup C** violates both pairwise and global dominance.

Table 1: View-averaged R^2 (%) under Gaussian priors across Setups A–C ($m_i = 3$).

	Setup A	**Setup B**		**Setup C**		
Method	$d_{\mathcal{Z}} = 3$	$d_{\mathcal{Z}} = 3$	$d_{\mathcal{Z}} = 4$	$d_{\mathcal{Z}} = 3$	$d_{\mathcal{Z}} = 4$	$d_{\mathcal{Z}} = 6$
BT	$37.90 \pm .32$	$39.01 \pm .28$	$42.35 \pm .28$	$44.25 \pm .29$	$51.97 \pm .25$	$47.05 \pm .39$
VICReg	$40.21 \pm .97$	$48.71 \pm .28$	$53.54 \pm .31$	$50.57 \pm .32$	$49.02 \pm .32$	$51.24 \pm .37$
W-MSE	$\mathbf{99.51 \pm .07}$	$68.51 \pm .20$	$92.83 \pm .11$	$57.67 \pm .27$	$80.81 \pm .44$	$93.63 \pm .09$
SimCLR	$78.51 \pm .28$	$87.05 \pm .27$	$84.90 \pm .10$	$81.56 \pm .11$	$83.95 \pm .13$	$80.04 \pm .23$
DGCCA	$\mathbf{99.72 \pm .05}$	$59.82 \pm .23$	$78.90 \pm .19$	$43.22 \pm .30$	$43.23 \pm .31$	$\mathbf{99.30 \pm .10}$
rDGCCA	$\mathbf{99.17 \pm .06}$	$70.42 \pm .23$	$87.31 \pm .16$	$43.61 \pm .27$	$63.23 \pm .21$	$\mathbf{99.89 \pm .01}$
SUMCOR	$\mathbf{99.69 \pm .04}$	$68.47 \pm .21$	$\mathbf{99.91 \pm .00}$	$43.21 \pm .30$	$43.23 \pm .31$	$\mathbf{99.74 \pm .01}$
tSUMCOR	—	$\mathbf{99.93 \pm .00}$	$\mathbf{99.96 \pm .00}$	$43.28 \pm .26$	$49.75 \pm .22$	46.13 ± 2.21
mSUMCOR	—	$\mathbf{99.94 \pm .00}$	$\mathbf{99.96 \pm .00}$	$43.67 \pm .29$	$52.31 \pm .19$	$62.77 \pm .37$

Table 2: View-averaged R^2 (%) under non-Gaussian priors in Setup A ($m_i = d_{\mathcal{Z}} = 3$).

Method	Gamma	Poisson	Neg. binomial	Hypergeometric
BT	$38.42 \pm .94$	$38.20 \pm .37$	$41.41 \pm .77$	$36.07 \pm .46$
VICReg	39.68 ± 1.05	$37.28 \pm .88$	$40.37 \pm .88$	$34.83 \pm .72$
W-MSE	$\mathbf{99.55 \pm .28}$	$\mathbf{99.33 \pm .05}$	$\mathbf{99.49 \pm .09}$	$97.94 \pm .12$
SimCLR	$73.27 \pm .28$	$75.75 \pm .28$	$73.73 \pm .40$	59.99 ± 1.95
DGCCA	$84.19 \pm .27$	$88.22 \pm .26$	$79.23 \pm .14$	$77.72 \pm .11$
rDGCCA	$96.87 \pm .33$	$97.58 \pm .15$	$91.99 \pm .30$	$90.96 \pm .20$
SUMCOR	$\mathbf{99.79 \pm .03}$	$\mathbf{99.44 \pm .05}$	$\mathbf{99.70 \pm .04}$	$\mathbf{99.33 \pm .06}$

Table 3: Mean correlation diagnostics (%) under Gaussian priors in Setups B/C ($m_i = 3$). $\bar{t} = J/(3k)$ averages rewarded correlations, with $k = d_{\mathcal{Z}}$ for SUMCOR and $k = 2$ otherwise. mSUMCOR J uses supports selected on each evaluation batch. $\gamma_{\min}$ and its SD use the view with the lowest batch-mean minimum source–representation correlation. Training details appear in Appendix F.2.

	B, $d_{\mathcal{Z}} = 3$		C, $d_{\mathcal{Z}} = 3$		C, $d_{\mathcal{Z}} = 6$	
Method	$\bar{t}$ (%)	$\gamma_{\min}$ (%)	$\bar{t}$ (%)	$\gamma_{\min}$ (%)	$\bar{t}$ (%)	$\gamma_{\min}$ (%)
SUMCOR	$39.54 \pm .55$	$1.06 \pm .70$	37.74 ± 1.33	$0.56 \pm .46$	25.14 ± 1.32	$\mathbf{99.72 \pm .02}$
tSUMCOR	$52.13 \pm .61$	$\mathbf{99.95 \pm .01}$	47.23 ± 1.11	$0.49 \pm .49$	54.82 ± 8.41	3.85 ± 1.18
mSUMCOR	$52.09 \pm .62$	$\mathbf{99.94 \pm .01}$	46.96 ± 1.01	$0.54 \pm .50$	47.79 ± 1.62	13.93 ± 1.27

All setups maintain full cross-view latent coverage. Setups B/C use isotropic Gaussian noise and $\mathbf{B}_i = \mathbf{Q}_i \operatorname{diag}(a, b, b)$ with orthogonal $\mathbf{Q}_i$. These rotations preserve the aligned-coordinate spectra (Appendix D.4).

Baselines. We compare pairwise nuclear-norm SUMCOR (Equation 2), DGCCA (Benton et al., 2019), and rDGCCA (Karakasis & Sidiropoulos, 2023). Additionally, we evaluate our Top-2 SUMCOR (tSUMCOR) and masked SUMCOR (mSUMCOR) variants: tSUMCOR rewards the two leading singular values of each normalized cross-covariance, whereas mSUMCOR selects and separately whitens rank-two coordinate blocks on each training and evaluation batch (Appendix F). We evaluate Barlow Twins (BT in tables) (Zbontar et al., 2021), VICReg (Bardes et al., 2022), SimCLR (Chen et al., 2020; Zimmermann et al., 2021), and W-MSE (Ermolov et al., 2021).

Training and evaluation. All models use residual encoders trained with Adam for $50{,}000$ updates (learning rate 10^{-4}, batch size 8192). Following Matthes et al. (2023), we fit affine source readouts and average coordinate-wise R^2 across views and ten independent evaluation batches of 4096 samples.

Each readout is fitted and evaluated on the same batch, giving an in-sample span recovery diagnostic rather than a held-out generalization test.

5.2 Validation of Theoretical Findings

Source recovery under partial observability. In Gaussian Setup A, evaluated CCA-family objectives achieve near-perfect affine recovery of view-specific sources under partial observability (Table 1). For the four non-Gaussian constructions in Table 2, SUMCOR also achieves near-perfect recovery, consistent with Theorem 4.1 and the verification in Proposition 6. DGCCA and rDGCCA, evaluated as empirical baselines, recover less accurately but outperform Barlow Twins, VICReg, and SimCLR across all distributions. W-MSE achieves near-perfect scores except on hypergeometric sources. Separate two-dimensional experiments illustrate similar recovery for Gaussian, Gamma, Poisson, and negative binomial sources (Figure 1): source-coordinate colorings follow fixed directions, with gradients and ordered discrete patterns consistent with an affine relationship up to a basis change.

Representation drift and rank-constrained recovery. In Setup B, pairwise first-order dominance holds but global dominance fails. At minimum-sufficient capacity, SUMCOR exhibits representation drift and incomplete source recovery, whereas tSUMCOR and masked SUMCOR achieve near-perfect affine recovery (Table 1). Correlation diagnostics show that SUMCOR nearly loses a source direction while both rank-constrained objectives retain all directions with high accuracy (Table 3). This is consistent with Theorem 4.2: known overlap ranks restrict rewarded edgewise modes to enable full affine recovery under pairwise dominance and cross-view coverage.

Finite-overcomplete recovery beyond dominance. In Setup C, both pairwise and global first-order dominance fail, and higher-order modes can outrank source directions. Neither SUMCOR nor its rank-constrained variants achieve complete source recovery at minimum-sufficient capacity. SUMCOR remains ineffective at dimension four but achieves near-perfect source recovery at dimension six (Table 1), as do DGCCA and rDGCCA. The rank-two methods and SUMCOR achieve near-perfect recovery in Setup B at dimension four. The rank-two methods yield incomplete recovery in Setup C at dimension six. These SUMCOR results illustrate affine readouts in the compatible family of Appendix D.4, with the Setup C result consistent with Theorem 4.3. They do not establish spectral compatibility for arbitrary mixing.

6 Discussion

Pairwise identifiability can coexist with representation drift because view pairs compete for each encoder's capacity. Without global first-order dominance, known overlap ranks restrict rewarded modes under pairwise dominance, while compatible capacity for full SUMCOR can retain source directions alongside stronger nonlinear modes.

The main guarantees concern population global optima for noisy view-specific sources under the isotropic Gaussian coordinate-subset model and full cross-view latent coverage. Non-Gaussian extensions require the stated operator, spectral, and geometric conditions. The overcomplete result gives affine source readouts at spectrum-dependent dimensions satisfying joint spectral compatibility, while allowing nonlinear representations.

Future Work. Priorities include estimating overlap ranks and suitable representation dimensions, establishing finite-sample guarantees, and analyzing neural optimization. Our synthetic experiments confirm these recovery patterns using in-sample affine readouts. Repeated training and held-out evaluation would provide a stronger assessment of practical recovery.

AI use statement

We used generative AI tools only for language editing and technical discussion. We did not use these tools to draft the manuscript or derive equations. The authors take full responsibility for the final manuscript, including its theoretical claims, proofs, experimental results, and references.

REFERENCES

Galen Andrew, Raman Arora, Jeff Bilmes, and Karen Livescu. Deep canonical correlation analysis. In *Proceedings of the 30th International Conference on Machine Learning*, volume 28 of *Proceedings of Machine Learning Research*, pp. 1247–1255. PMLR, 2013.

Adrien Bardes, Jean Ponce, and Yann LeCun. VICReg: Variance-invariance-covariance regularization for self-supervised learning. In *International Conference on Learning Representations*, 2022.

Adrian Benton, Huda Khayrallah, Biman Gujral, Dee Ann Reisinger, Sheng Zhang, and Raman Arora. Deep generalized canonical correlation analysis. In *Proceedings of the 4th Workshop on Representation Learning for NLP (RepL4NLP-2019)*, pp. 1–6. Association for Computational Linguistics, 2019. doi: 10.18653/v1/W19-4301.

Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A simple framework for contrastive learning of visual representations. In *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pp. 1597–1607. PMLR, 2020.

Imant Daunhawer, Alice Bizeul, Emanuele Palumbo, Alexander Marx, and Julia E. Vogt. Identifiability results for multimodal contrastive learning. In *International Conference on Learning Representations*, 2023.

G. K. Eagleson. Polynomial expansions of bivariate distributions. *The Annals of Mathematical Statistics*, 35(3):1208–1215, 1964. doi: 10.1214/aoms/1177703278.

Aleksandr Ermolov, Aliaksandr Siarohin, Enver Sangineto, and Nicu Sebe. Whitening for self-supervised representation learning. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pp. 3015–3024. PMLR, 2021.

Soheil Feizi and David Tse. Maximally correlated principal component analysis. arXiv preprint arXiv:1702.05471, 2017. Version 2, revised 21 February 2017.

Soheil Feizi, Ali Makhdoumi, Ken Duffy, Manolis Kellis, and Muriel Médard. Network maximal correlation. *IEEE Transactions on Network Science and Engineering*, 4(4):229–247, 2017. doi: 10.1109/TNSE.2017.2716966.

Kenji Fukumizu, Francis R. Bach, and Arthur Gretton. Statistical consistency of kernel canonical correlation analysis. *Journal of Machine Learning Research*, 8(14):361–383, 2007.

Zhiwei Han, Stefan Matthes, and Hao Shen. Provable affine identifiability of nonlinear CCA under latent distributional priors. In *Proceedings of The 29th International Conference on Artificial Intelligence and Statistics*, volume 300 of *Proceedings of Machine Learning Research*, pp. 3835–3843. PMLR, 2026a.

Zhiwei Han, Stefan Matthes, and Hao Shen. Provable subspace identification of nonlinear multi-view CCA. In *Proceedings of the 42nd Conference on Uncertainty in Artificial Intelligence*, volume 337 of *Proceedings of Machine Learning Research*, pp. 1947–1982. PMLR, 2026b.

Jeff Z. HaoChen, Colin Wei, Adrien Gaidon, and Tengyu Ma. Provable guarantees for self-supervised deep learning with spectral contrastive loss. In *Advances in Neural Information Processing Systems*, volume 34, pp. 5000–5011. Curran Associates, Inc., 2021.

Harold Hotelling. Relations between two sets of variates. *Biometrika*, 28(3–4):321–377, 1936. doi: 10.1093/biomet/28.3-4.321.

Aapo Hyvärinen and Petteri Pajunen. Nonlinear independent component analysis: Existence and uniqueness results. *Neural Networks*, 12(3):429–439, 1999. doi: 10.1016/S0893-6080(98)00140-3.

Li Jing, Pascal Vincent, Yann LeCun, and Yuandong Tian. Understanding dimensional collapse in contrastive self-supervised learning. In *International Conference on Learning Representations*, 2022.

Daniel D. Johnson, Ayoub El Hanchi, and Chris J. Maddison. Contrastive learning can find an optimal basis for approximately view-invariant functions. In *International Conference on Learning Representations*, 2023.

Paris A. Karakasis and Nicholas D. Sidiropoulos. Revisiting deep generalized canonical correlation analysis. *IEEE Transactions on Signal Processing*, 71:4392–4406, 2023. doi: 10.1109/TSP.2023.3333212.

J. R. Kettenring. Canonical analysis of several sets of variables. *Biometrika*, 58(3):433–451, 1971. doi: 10.1093/biomet/58.3.433.

H. O. Lancaster. The structure of bivariate distributions. *The Annals of Mathematical Statistics*, 29 (3):719–736, 1958. doi: 10.1214/aoms/1177706532.

Francesco Locatello, Stefan Bauer, Mario Lucic, Gunnar Rätsch, Sylvain Gelly, Bernhard Schölkopf, and Olivier Bachem. A sober look at the unsupervised learning of disentangled representations and their evaluation. *Journal of Machine Learning Research*, 21(209):1–62, 2020.

Qi Lyu and Xiao Fu. Nonlinear multiview analysis: Identifiability and neural network-assisted implementation. *IEEE Transactions on Signal Processing*, 68:2697–2712, 2020. doi: 10.1109/TSP.2020.2986351.

Qi Lyu, Xiao Fu, Weiran Wang, and Songtao Lu. Understanding latent correlation-based multiview learning and self-supervision: An identifiability perspective. In *International Conference on Learning Representations*, 2022.

Stefan Matthes, Zhiwei Han, and Hao Shen. Towards a unified framework of contrastive learning for disentangled representations. In *Advances in Neural Information Processing Systems*, volume 36, pp. 67459–67470. Curran Associates, Inc., 2023. doi: 10.52202/075280-2948.

Tomer Michaeli, Weiran Wang, and Karen Livescu. Nonparametric canonical correlation analysis. In *Proceedings of the 33rd International Conference on Machine Learning*, volume 48 of *Proceedings of Machine Learning Research*, pp. 1967–1976. PMLR, 2016.

Joshua W. Robinson, Li Sun, Ke Yu, Kayhan Batmanghelich, Stefanie Jegelka, and Suvrit Sra. Can contrastive learning avoid shortcut solutions? In *Advances in Neural Information Processing Systems*, volume 34, pp. 4974–4986. Curran Associates, Inc., 2021.

Seppo Virtanen, Arto Klami, Suleiman Khan, and Samuel Kaski. Bayesian group factor analysis. In *Proceedings of the Fifteenth International Conference on Artificial Intelligence and Statistics*, volume 22 of *Proceedings of Machine Learning Research*, pp. 1269–1277. PMLR, 2012.

Julius von Kügelgen, Yash Sharma, Luigi Gresele, Wieland Brendel, Bernhard Schölkopf, Michel Besserve, and Francesco Locatello. Self-supervised learning with data augmentations provably isolates content from style. In *Advances in Neural Information Processing Systems*, volume 34, pp. 16451–16467. Curran Associates, Inc., 2021.

Dingling Yao, Danru Xu, Sébastien Lachapelle, Sara Magliacane, Perouz Taslakian, Georg Martius, Julius von Kügelgen, and Francesco Locatello. Multi-view causal representation learning with partial observability. In *International Conference on Learning Representations*, pp. 34817–34848, 2024.

Jure Zbontar, Li Jing, Ishan Misra, Yann LeCun, and Stéphane Deny. Barlow Twins: Self-supervised learning via redundancy reduction. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pp. 12310–12320. PMLR, 2021.

Roland S. Zimmermann, Yash Sharma, Steffen Schneider, Matthias Bethge, and Wieland Brendel. Contrastive learning inverts the data generating process. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pp. 12979–12990. PMLR, 2021.

CONTENTS OF THE APPENDIX

A Structural Implications of the Loading Factorization **12**
A.1 Visibility and within-view mixing 12
A.2 Fully shared, shared–private, and subset-shared patterns 13
A.3 Connection to coordinate-masked recovery 13
A.4 Connections to existing generative models 14

B Finite-Dimensional SUMCOR: Proofs and Spectral Competition **15**
B.1 Proof of representation drift 15
B.2 A three-view instance of spectral competition 17
B.3 Proof of affine recovery under global dominance 19

C Rank-constrained SUMCOR: Proof of Affine Recovery **21**
C.1 Mask profiling and empirical support selection 22

D Finite-Overcomplete SUMCOR: Proof and Capacity Example **23**
D.1 Finite spectral depth and finite aggregate capacity 23
D.2 A view-wise trace bound for SUMCOR 24
D.3 Proof of the finite-capacity recovery theorem 24
D.4 A finite-capacity recovery instance without dominance 25

E Beyond Gaussian Sources: First-Order Invariant Couplings **27**
E.1 Operator structure and overlap geometry 27
E.2 Finite-dimensional recovery 28
E.3 Full-spectrum recovery under a trace-class condition 30
E.4 Lancaster specialization and interpretation 31
E.5 Spectral verification for non-Gaussian Setup A 32

F Experimental Details and Additional Results **35**
F.1 Data-generating models 35
F.2 Encoders, objectives, and training protocol 36
F.3 Recovery metrics and aggregation 37
F.4 Complete per-view quantitative results 37
F.5 Complete two-dimensional visualizations 41

A STRUCTURAL IMPLICATIONS OF THE LOADING FACTORIZATION

The factorization $\mathbf{A}_i = \mathbf{B}_i\mathbf{P}_i$ in Equation 1 separates the coordinates visible to a view from their invertible mixing in source space. This section describes the sharing patterns represented by this construction, explains the geometric compatibility it provides for coordinate-masked recovery, and relates it to existing generative models. Correspondence between sharing patterns is distinguished from correspondence between complete generative models or recovery guarantees.

A.1 VISIBILITY AND WITHIN-VIEW MIXING

For a fixed visibility set $\mathcal{V}_i$, the selector $\mathbf{P}_i$ specifies which coordinates of $\mathbf{c}$ enter view i, whereas $\mathbf{B}_i \in \mathrm{GL}(m_i)$ determines how those coordinates are mixed. In particular,

$$\mathrm{row}(\mathbf{A}_i) = \mathrm{row}(\mathbf{P}_i) = \mathrm{span}\{\mathbf{e}_k^\top : k \in \mathcal{V}_i\}.$$

Thus $\mathbf{B}_i$ may be dense and need not be orthogonal, but it cannot change the visible latent subspace. Equivalently, $\mathbf{A}_i$ vanishes outside the columns indexed by $\mathcal{V}_i$, and its active $m_i \times m_i$ block is invertible. The requirement that these row spaces be coordinate subspaces in one common latent basis is a substantive restriction on general rectangular loadings.

The explicit mixing matrix also fixes the signal–noise parameterization. Defining

$$\widetilde{\mathbf{g}}_i(\mathbf{u}) := \mathbf{g}_i(\mathbf{B}_i\mathbf{u})$$

gives the exact identity

$$\mathbf{x}_i = \widetilde{\mathbf{g}}_i\big(\mathbf{P}_i\mathbf{c} + \mathbf{B}_i^{-1}\boldsymbol{\epsilon}_i\big)\,.$$

Consequently, $\mathbf{B}_i$ can be absorbed into the observation map in the noiseless case. Under Assumption 1, however, the transformed perturbation has covariance $\mathbf{B}_i^{-1}\mathbf{B}_i^{-\top}$, which is the identity only when $\mathbf{B}_i$ is orthogonal. Keeping $\mathbf{B}_i$ explicit therefore separates visibility from within-view geometry while preserving the stipulated source-noise model. It does not by itself enlarge the noiseless observation family.

A.2 Fully shared, shared–private, and subset-shared patterns

For each nonempty subset $\mathcal{H} \subseteq [N]$, introduce a latent block $\mathbf{c}^{(\mathcal{H})} \in \mathbb{R}^{d_{\mathcal{H}}}$ that is visible exactly in the views indexed by $\mathcal{H}$. Allow $d_{\mathcal{H}} = 0$ for absent blocks, concatenate the remaining blocks into $\mathbf{c}$, and let $\mathcal{J}_{\mathcal{H}} \subseteq [d_c]$ denote the coordinate indices of each block. Taking

$$\mathcal{V}_i = \bigcup_{\mathcal{H} \ni i} \mathcal{J}_{\mathcal{H}}, \qquad m_i = \sum_{\mathcal{H} \ni i} d_{\mathcal{H}} \tag{10}$$

realizes any fixed coordinate-sharing pattern with nonempty visibility in each view. The corresponding $\mathbf{P}_i$ selects these coordinates, and any invertible $\mathbf{B}_i$ preserves their view incidence.

Fully shared models retain only the block indexed by $[N]$, giving $\mathbf{P}_i = \mathbf{I}_{d_c}$ after a common ordering. *Shared–private* models retain a fully shared block together with singleton blocks $\mathbf{c}^{(\{i\})}$. *Subset-shared* models additionally allow blocks with $2 \le |\mathcal{H}| < N$. Pairwise sharing is the case $|\mathcal{H}| = 2$. For example, three views may share

$$\mathbf{c} = (a, b_{12}, b_{13}, b_{23})^\top,$$

with

$$\mathbf{P}_1\mathbf{c} = (a, b_{12}, b_{13})^\top,$$
$$\mathbf{P}_2\mathbf{c} = (a, b_{12}, b_{23})^\top,$$
$$\mathbf{P}_3\mathbf{c} = (a, b_{13}, b_{23})^\top.$$

Here a is globally shared, while each b_{ij} is visible to only two views. Every coordinate visible to a view is shared with some other view, even though only a belongs to the intersection of all three visibility sets.

A singleton block is an explicit view-private latent factor, not the additive perturbation $\boldsymbol{\epsilon}_i$. Such blocks are permitted by the generative model but appear in no pairwise latent overlap. Their presence therefore violates Assumption 3. The full-source recovery guarantees requiring that assumption do not apply. This distinction separates the patterns admitted by the model from the directions covered by cross-view recovery.

A.3 Connection to coordinate-masked recovery

For ordered active pairs, set $\mathbf{R}_{ji} := \mathbf{R}_{ij}^\top$. Define the full first-order source space and its pairwise target by

$$\mathcal{L}_i := \{\mathbf{a}^\top \tilde{\mathbf{s}}_i : \mathbf{a} \in \mathbb{R}^{m_i}\}, \qquad \mathcal{L}_{i|j} := \{\mathbf{a}^\top \tilde{\mathbf{s}}_i : \mathbf{a} \in \mathrm{col}(\mathbf{R}_{ij})\}.$$

The span of a vector-valued function means the linear span of its scalar coordinates in $L^2(P_{\mathbf{s}_i})$. Write $\mathcal{M}(r, m) := \{\mathbf{M} \in \{0,1\}^{r \times m} : \mathbf{M}\mathbf{M}^\top = \mathbf{I}_r\}$ for the admissible coordinate selectors.

Lemma 1 (Simultaneous coordinate realizability)**.** *Under Assumption 1, define*

$$\mathbf{q}_i(\mathbf{s}_i) := \mathbf{B}_i^\top \mathbf{W}_i^2 \mathbf{s}_i, \qquad \mathcal{I}_{i|j} := \{k \in [m_i] : v_{i,k} \in \mathcal{V}_j\}. \tag{11}$$

The map $\mathbf{q}_i$ is invertible and linear, and every ordered active pair satisfies

$$\mathrm{span}\{q_{i,k} : k \in \mathcal{I}_{i|j}\} = \mathcal{L}_{i|j}.$$

If Assumption 3 also holds, then $\mathrm{span}_{j \in \mathcal{N}_i} \mathcal{L}_{i|j} = \mathcal{L}_i$ *for every view i.*

Proof. Under the model, $\mathbf{W}_i = (\mathbf{B}_i\mathbf{B}_i^\top + \mathbf{I}_{m_i})^{-1/2}$ is invertible. Hence $\mathbf{B}_i^\top \mathbf{W}_i^2$ is invertible, and the coordinates of $\mathbf{q}_i$ form a basis of $\mathcal{L}_i$.

Let $\mathbf{E}_{i|j} \in \mathcal{M}(r_{ij}, m_i)$ select the indices in $\mathcal{I}_{i|j}$. The columns of $\mathbf{P}_i\mathbf{P}_j^\top$ span the same coordinate space as those of $\mathbf{E}_{i|j}^\top$. Since $\mathbf{B}_j^\top\mathbf{W}_j$ is invertible,

$$\operatorname{col}(\mathbf{R}_{ij}) = \operatorname{col}\big(\mathbf{W}_i\mathbf{B}_i\mathbf{P}_i\mathbf{P}_j^\top\mathbf{B}_j^\top\mathbf{W}_j\big) = \operatorname{col}\Big(\mathbf{W}_i\mathbf{B}_i\mathbf{E}_{i|j}^\top\Big).$$

Using $\tilde{\mathbf{s}}_i = \mathbf{W}_i\mathbf{s}_i$ therefore gives

$$\begin{aligned}\mathcal{L}_{i|j} &= \operatorname{span}\big(\mathbf{E}_{i|j}\mathbf{B}_i^\top\mathbf{W}_i\tilde{\mathbf{s}}_i\big)\\ &= \operatorname{span}\big(\mathbf{E}_{i|j}\mathbf{q}_i(\mathbf{s}_i)\big).\end{aligned}$$

Finally, coverage implies $\bigcup_{j\in\mathcal{N}_i}\mathcal{I}_{i|j} = [m_i]$. The incident target spaces thus collectively contain every basis coordinate of $\mathbf{q}_i$, proving the final claim. □

The construction is an existence witness, not information supplied to the learner, and imposes no diagonal restriction on $\mathbf{B}_i$. It establishes simultaneous feasibility of the coordinate blocks, not identification of the generative selectors or unique mask supports. In general, $\mathbf{q}_i$ is not globally whitened: Equation 5 whitens each selected block separately. The covariance argument also extends, after centering, to the non-Gaussian setting of Lemma 5. Nonlinear CCA recovery additionally requires the operator invariance and spectral conditions in Appendix E.

A.4 Connections to existing generative models

General additive-loading models. Han et al. (2026b) consider $\mathbf{s}_i = \overline{\mathbf{A}}_i\mathbf{c} + \boldsymbol{\epsilon}_i$ and $\mathbf{x}_i = \mathbf{g}_i(\mathbf{s}_i)$ with general view-specific loadings $\overline{\mathbf{A}}_i$. Setting $\overline{\mathbf{A}}_i = \mathbf{B}_i\mathbf{P}_i$ recovers our structural model and, with the same independent standard Gaussian latents and perturbations, its Gaussian instance. Our model is therefore a structured specialization of that loading family. The additional structure makes the pairwise targets compatible with coordinate selection from a common view-wise representation, rather than extending the class of arbitrary loadings.

Subset-based partial observability. Yao et al. (2024) consider $\mathbf{x}_i = \boldsymbol{\varphi}_i(\mathbf{z}_{S_i})$, where S_i indexes a fixed subset of potentially dependent latents and $\boldsymbol{\varphi}_i$ is a diffeomorphism. The parameter correspondence

$$\mathbf{z} = \mathbf{c}, \qquad S_i = \mathcal{V}_i, \qquad \boldsymbol{\varphi}_i(\mathbf{u}) = \mathbf{g}_i(\mathbf{B}_i\mathbf{u})$$

identifies their structural equation with the noiseless version of Equation 1, obtained by setting $\boldsymbol{\epsilon}_i = \mathbf{0}$. This correspondence does not identify the full model classes: their latent law can contain dependencies, whereas our Gaussian instance stipulates independent coordinates and nondegenerate additive source perturbations. They also use one encoder per view together with content selectors to identify shared latent blocks up to smooth bijections. Our recovery analysis instead concerns first-order subspaces of the noisy sources under CCA objectives.

Shared–private nonlinear models. The two-view model of Lyu et al. (2022) has the form

$$\mathbf{x}_i = \boldsymbol{\varphi}_i(\mathbf{u}, \mathbf{v}_i), \qquad i \in \{1, 2\},$$

with mutually independent shared and private blocks $\mathbf{u}, \mathbf{v}_1, \mathbf{v}_2$ and smooth invertible view maps. Concatenate these blocks into $\mathbf{c} = (\mathbf{u}^\top, \mathbf{v}_1^\top, \mathbf{v}_2^\top)^\top$, and let $\mathbf{P}_i$ select $(\mathbf{u}, \mathbf{v}_i)$. Taking $\mathbf{B}_i = \mathbf{I}_{m_i}$, $\boldsymbol{\epsilon}_i = \mathbf{0}$, and $\mathbf{g}_i = \boldsymbol{\varphi}_i$ reproduces their generative equations. This is again a noiseless structural correspondence. Their block distributions need not satisfy our Gaussian assumption. In particular, their private block $\mathbf{v}_i$ corresponds to singleton visibility, not to our additive perturbation.

Group factor models. Bayesian Group Factor Analysis (Virtanen et al., 2012) uses group-wise shrinkage to model factors active in different subsets of views. For fixed loadings with exact inactive columns, these activity patterns admit the incidence construction in Equation 10. Its general active loading blocks, however, can be rectangular and need not satisfy our square-invertible restriction. A restricted correspondence can be made explicit. Consider a view equation

$$\mathbf{y}_i = \mathbf{F}_i\mathbf{c} + \boldsymbol{\xi}_i, \qquad \boldsymbol{\xi}_i \sim \mathcal{N}(\mathbf{0}, \sigma_i^2\mathbf{I}_{m_i}),$$

with independent standard Gaussian factors and independent residuals. Suppose that $\mathbf{F}_i$ has exactly m_i active columns, indexed by $\mathcal{V}_i$, and that $\mathbf{F}_i(:, \mathcal{V}_i)$ is invertible. For fixed loadings and $\sigma_i > 0$, take

$$\mathbf{B}_i = \sigma_i^{-1}\mathbf{F}_i(:, \mathcal{V}_i), \qquad \boldsymbol{\epsilon}_i = \sigma_i^{-1}\boldsymbol{\xi}_i, \qquad \mathbf{g}_i(\mathbf{s}) = \sigma_i \mathbf{s}.$$

Then Equation 1 reproduces the conditional observation law with $\mathbf{x}_i = \mathbf{y}_i$. This embeds a restricted square-active-loading Gaussian subfamily, not the general loading family or its full Bayesian hierarchy.

Noiseless correspondences and the spectral boundary. The noiseless correspondences above concern generative structure, not direct applicability of our nondegenerate Gaussian recovery results. If $\boldsymbol{\epsilon}_i = \mathbf{0}$ while $\mathbf{c} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_{d_c})$, then

$$\mathbf{R}_{ij} = \mathbf{O}_i \mathbf{P}_i \mathbf{P}_j^\top \mathbf{O}_j^\top, \qquad \mathbf{O}_i := (\mathbf{B}_i \mathbf{B}_i^\top)^{-1/2} \mathbf{B}_i$$

with orthogonal $\mathbf{O}_i$. Every nonzero canonical correlation is therefore one. In particular, the strict pairwise first-order dominance condition $t_{ij,r_{ij}} > t_{ij,1}^2$ fails on each active edge. Noiseless latent-block identifiability and our noisy-source spectral guarantees must therefore be distinguished, even when their visibility patterns agree.

B Finite-Dimensional SUMCOR: Proofs and Spectral Competition

The proofs follow subsection 4.1: representation drift and its three-view instance precede affine recovery under global first-order dominance.

B.1 Proof of representation drift

Proof of Proposition 2. All expectations and covariance matrices are population quantities, and equalities between random variables hold almost surely. Since each observation map $\mathbf{g}_i$ is invertible on its image, it suffices to work with source-space encoders. Moreover, invertible affine reparameterizations of the encoder outputs preserve both the CCA objective and affine source recoverability. We may therefore restrict attention to centered, whitened encoders $\tilde{\mathbf{h}}_i(\mathbf{s}_i) \in \mathbb{R}^{m_i}$ satisfying

$$\mathbb{E}[\tilde{\mathbf{h}}_i] = \mathbf{0}, \qquad \mathbb{E}[\tilde{\mathbf{h}}_i \tilde{\mathbf{h}}_i^\top] = \mathbf{I}_{m_i}.$$

For such encoders, Equation 2 reduces to

$$J^{\mathcal{E}}_{\mathrm{SUMCOR}}(\{\tilde{\mathbf{h}}_i\}) = \sum_{\{i,j\} \in \mathcal{E}} \left\| \mathbb{E}[\tilde{\mathbf{h}}_i \tilde{\mathbf{h}}_j^\top] \right\|_* .$$

We abbreviate $\boldsymbol{\psi}_i = \boldsymbol{\psi}_i(\mathbf{s}_i)$. When $k_i = 0$ or $k_i = m_i$, the corresponding empty blocks are omitted, and their nuclear norms are understood to be zero.

Step 1: The objective value of every fully affine-recovering collection. Because $\mathbf{W}_i$ is invertible, affine recovery of $\mathbf{s}_i$ is equivalent to affine recovery of $\tilde{\mathbf{s}}_i = \mathbf{W}_i \mathbf{s}_i$. Suppose that $\tilde{\mathbf{h}}_i$ permits such recovery. Then there exist $\mathbf{D}_i \in \mathbb{R}^{m_i \times m_i}$ and $\mathbf{b}_i \in \mathbb{R}^{m_i}$ such that

$$\tilde{\mathbf{s}}_i = \mathbf{D}_i \tilde{\mathbf{h}}_i + \mathbf{b}_i.$$

Centering gives $\mathbf{b}_i = \mathbf{0}$, while whitening gives

$$\mathbf{I}_{m_i} = \mathbb{E}[\tilde{\mathbf{s}}_i \tilde{\mathbf{s}}_i^\top] = \mathbf{D}_i \mathbf{D}_i^\top.$$

Since $\mathbf{D}_i$ is square, it is orthogonal. Consequently,

$$\tilde{\mathbf{h}}_i = \mathbf{O}_i \tilde{\mathbf{s}}_i, \qquad \mathbf{O}_i := \mathbf{D}_i^\top \in O(m_i).$$

Conversely, every encoder of this form permits full affine source recovery. Thus, at $d_{\mathcal{Z}_i} = m_i$, the centered, whitened fully affine-recovering encoders are exactly the orthogonal transformations of the whitened sources.

For a collection that permits full affine recovery in every view, its cross-covariance on an active edge is therefore

$$\mathbb{E}[\tilde{\mathbf{h}}_i \tilde{\mathbf{h}}_j^\top] = \mathbf{O}_i \mathbf{R}_{ij} \mathbf{O}_j^\top.$$

Orthogonal invariance of the nuclear norm gives

$$J^{\mathcal{E}}_{\mathrm{SUMCOR}}(\{\tilde{\mathbf{h}}_i\}) = J_{\mathrm{aff}} := \sum_{\{i,j\}\in\mathcal{E}} \|\mathbf{R}_{ij}\|_*. \tag{12}$$

In particular, this value is independent of the chosen affine-recovering collection.

Step 2: A feasible spectral substitution within the same dimension budget. For the matrices $\mathbf{Q}_i$ and functions $\boldsymbol{\psi}_i$ in the proposition, define

$$\tilde{\mathbf{h}}^{\mathrm{alt}}_i(\mathbf{s}_i) := \begin{pmatrix} \mathbf{Q}_i^\top \tilde{\mathbf{s}}_i \\ \boldsymbol{\psi}_i(\mathbf{s}_i) \end{pmatrix} \in \mathbb{R}^{m_i}.$$

This encoder allocates $m_i - k_i$ coordinates to a first-order subspace and k_i coordinates to the orthogonal complement of the full first-order source space.

By the hypotheses,

$$\begin{aligned} \mathbf{Q}_i^\top \mathbf{Q}_i &= \mathbf{I}_{m_i-k_i}, \\ \mathbb{E}[\boldsymbol{\psi}_i] &= \mathbf{0}, \\ \mathbb{E}[\boldsymbol{\psi}_i \boldsymbol{\psi}_i^\top] &= \mathbf{I}_{k_i}, \\ \mathbb{E}[\boldsymbol{\psi}_i \tilde{\mathbf{s}}_i^\top] &= \mathbf{0}. \end{aligned}$$

Hence $\tilde{\mathbf{h}}^{\mathrm{alt}}_i$ is centered and

$$\mathbb{E}\Big[\tilde{\mathbf{h}}^{\mathrm{alt}}_i (\tilde{\mathbf{h}}^{\mathrm{alt}}_i)^\top\Big] = \begin{pmatrix} \mathbf{I}_{m_i-k_i} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{k_i} \end{pmatrix} = \mathbf{I}_{m_i}.$$

It is therefore feasible without increasing the representation dimension. Its observation-space realization is $\tilde{\mathbf{h}}^{\mathrm{alt}}_i \circ \mathbf{g}_i^{-1}$.

Step 3: Cross-view first-order–higher-order orthogonality. Under Assumption 1, $(\tilde{\mathbf{s}}_i, \tilde{\mathbf{s}}_j)$ is jointly Gaussian, with identity marginal covariances and cross-covariance $\mathbf{R}_{ij}$. Therefore,

$$\begin{aligned} \mathbb{E}[\tilde{\mathbf{s}}_i \mid \mathbf{s}_j] &= \mathbf{R}_{ij}\tilde{\mathbf{s}}_j, \\ \mathbb{E}[\tilde{\mathbf{s}}_j \mid \mathbf{s}_i] &= \mathbf{R}_{ij}^\top \tilde{\mathbf{s}}_i. \end{aligned} \tag{13}$$

For completeness, the first identity follows by considering the Gaussian residual $\tilde{\mathbf{s}}_i - \mathbf{R}_{ij}\tilde{\mathbf{s}}_j$. Its covariance with $\tilde{\mathbf{s}}_j$ is zero, so it is independent of $\tilde{\mathbf{s}}_j$ and has conditional mean zero. Conditioning on $\mathbf{s}_j$ is equivalent to conditioning on $\tilde{\mathbf{s}}_j$ because $\mathbf{W}_j$ is invertible. The second identity follows symmetrically.

Using Equation 13 and the within-view orthogonality of $\boldsymbol{\psi}_j$, we obtain

$$\begin{aligned} \mathbb{E}[\tilde{\mathbf{s}}_i \boldsymbol{\psi}_j^\top] &= \mathbb{E}\big[\mathbb{E}[\tilde{\mathbf{s}}_i \mid \mathbf{s}_j]\boldsymbol{\psi}_j^\top\big] \\ &= \mathbf{R}_{ij}\mathbb{E}[\tilde{\mathbf{s}}_j \boldsymbol{\psi}_j^\top] = \mathbf{0}. \end{aligned}$$

Similarly,

$$\begin{aligned} \mathbb{E}[\boldsymbol{\psi}_i \tilde{\mathbf{s}}_j^\top] &= \mathbb{E}\big[\boldsymbol{\psi}_i \mathbb{E}[\tilde{\mathbf{s}}_j \mid \mathbf{s}_i]^\top\big] \\ &= \mathbb{E}[\boldsymbol{\psi}_i \tilde{\mathbf{s}}_i^\top]\mathbf{R}_{ij} = \mathbf{0}. \end{aligned}$$

These identities hold for arbitrary square-integrable $\boldsymbol{\psi}_i$ satisfying the stated orthogonality conditions. No finite polynomial expansion is required.

Define

$$\mathbf{K}_{ij} := \mathbb{E}[\boldsymbol{\psi}_i \boldsymbol{\psi}_j^\top] = \mathrm{Cov}(\boldsymbol{\psi}_i, \boldsymbol{\psi}_j).$$

The cross-covariance of the alternative encoders is then exactly block diagonal:

$$\mathbb{E}\Big[\tilde{\mathbf{h}}^{\mathrm{alt}}_i (\tilde{\mathbf{h}}^{\mathrm{alt}}_j)^\top\Big] = \begin{pmatrix} \mathbf{Q}_i^\top \mathbf{R}_{ij} \mathbf{Q}_j & \mathbf{0} \\ \mathbf{0} & \mathbf{K}_{ij} \end{pmatrix}. \tag{14}$$

Step 4: Exact spectral gain–loss comparison. The nuclear norm is additive across diagonal blocks, including rectangular blocks. Indeed, for any such blocks $\mathbf{A}$ and $\mathbf{B}$,

$$\left\| \begin{pmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{0} & \mathbf{B} \end{pmatrix} \right\|_* = \mathrm{tr}\big((\mathbf{A}\mathbf{A}^\top)^{1/2}\big) + \mathrm{tr}\big((\mathbf{B}\mathbf{B}^\top)^{1/2}\big) = \|\mathbf{A}\|_* + \|\mathbf{B}\|_*.$$

Applying this identity to Equation 14 yields

$$J_{\mathrm{CCA}}(\tilde{\mathbf{h}}_i^{\mathrm{alt}}, \tilde{\mathbf{h}}_j^{\mathrm{alt}}) = \|\mathbf{Q}_i^\top \mathbf{R}_{ij} \mathbf{Q}_j\|_* + \|\mathbf{K}_{ij}\|_*.$$

Write the aggregate higher-order gain and first-order loss as

$$G_{\geq 2} := \sum_{\{i,j\} \in \mathcal{E}} \|\mathbf{K}_{ij}\|_*,$$
$$L_1 := \sum_{\{i,j\} \in \mathcal{E}} \left(\|\mathbf{R}_{ij}\|_* - \|\mathbf{Q}_i^\top \mathbf{R}_{ij} \mathbf{Q}_j\|_* \right).$$

Each summand in L_1 is nonnegative because compression by $\mathbf{Q}_i^\top$ and $\mathbf{Q}_j$ cannot increase the nuclear norm.

Summing the edgewise identities gives the exact comparison

$$J^{\mathcal{E}}_{\mathrm{SUMCOR}}(\{\tilde{\mathbf{h}}_i^{\mathrm{alt}}\}) = \sum_{\{i,j\} \in \mathcal{E}} \left(\|\mathbf{Q}_i^\top \mathbf{R}_{ij} \mathbf{Q}_j\|_* + \|\mathbf{K}_{ij}\|_* \right)$$
$$= J_{\mathrm{aff}} - L_1 + G_{\geq 2}.$$

The hypothesis Equation 3 is precisely $G_{\geq 2} > L_1$. Therefore,

$$J^{\mathcal{E}}_{\mathrm{SUMCOR}}(\{\tilde{\mathbf{h}}_i^{\mathrm{alt}}\}) > J_{\mathrm{aff}}.$$

By Equation 12, every collection permitting full affine source recovery in all views attains exactly J_{aff}, and is consequently strictly suboptimal.

The argument does not require the alternative collection to be globally optimal, nor does it require existence of a global maximizer. If a global maximizer exists, it cannot permit full affine source recovery in every view. □

B.2 A THREE-VIEW INSTANCE OF SPECTRAL COMPETITION

We instantiate Proposition 2 with three-dimensional sources and encoders. The construction shows how higher-order modes can collectively outweigh the first-order directions they displace, despite pairwise first-order dominance on every individual edge.

Model and pairwise spectra. Consider Equation 1 under Assumption 1, with $\mathbf{c} \in \mathbb{R}^4$, $m_i = d_{\mathcal{Z}_i} = 3$, and visibility sets

$$\mathcal{V}_1 = \{1,2,3\}, \qquad \mathcal{V}_2 = \{1,2,4\}, \qquad \mathcal{V}_3 = \{1,3,4\},$$

where each selection matrix lists its coordinates in increasing order. Thus, c_1 is visible in all three views, whereas each remaining latent coordinate is visible in exactly two views. Set

$$\mathbf{B}_i = \mathrm{diag}(a,b,b), \qquad a = \sqrt{\frac{\rho}{1-\rho}}, \qquad b = \sqrt{\frac{\tau}{1-\tau}},$$

with

$$0 < \rho < 1, \qquad \rho^2 < \tau < \min\{\rho, \rho^2 + \rho^3\}. \tag{15}$$

Every pair is active and has nonzero first-order canonical correlations

$$(t_{ij,1}, t_{ij,2}) = (\rho, \tau).$$

Hence pairwise first-order dominance holds because $\tau > \rho^2$, and every view satisfies full cross-view latent coverage.

Spectral substitution. Let $\zeta_i = \mathbf{e}_1^\top \tilde{\mathbf{s}}_i$. These variables are standard normal, with $\mathrm{Cov}(\zeta_i, \zeta_j) = \rho$ for every $i \neq j$. In the notation of Proposition 2, choose

$$k_i = 2, \qquad \mathbf{Q}_i = \mathbf{e}_1, \qquad \boldsymbol{\psi}_i(\mathbf{s}_i) = \begin{pmatrix} \psi_2(\zeta_i) \\ \psi_3(\zeta_i) \end{pmatrix},$$

where

$$\psi_2(z) = \frac{z^2 - 1}{\sqrt{2}}, \qquad \psi_3(z) = \frac{z^3 - 3z}{\sqrt{6}}$$

are normalized Hermite polynomials. Gaussian orthogonality and independence of the source coordinates within each view give

$$\mathbb{E}[\boldsymbol{\psi}_i] = \mathbf{0}, \qquad \mathrm{Cov}(\boldsymbol{\psi}_i) = \mathbf{I}_2, \qquad \mathbb{E}[\boldsymbol{\psi}_i \tilde{\mathbf{s}}_i^\top] = \mathbf{0}.$$

Consequently, the alternative encoder

$$\tilde{\mathbf{h}}_i^{\mathrm{drift}} := \begin{pmatrix} \mathbf{Q}_i^\top \tilde{\mathbf{s}}_i \\ \boldsymbol{\psi}_i(\mathbf{s}_i) \end{pmatrix} = \begin{pmatrix} \zeta_i \\ (\zeta_i^2 - 1)/\sqrt{2} \\ (\zeta_i^3 - 3\zeta_i)/\sqrt{6} \end{pmatrix}$$

is centered and whitened, and uses the same three-dimensional representation budget.

For every active edge, the retained first-order block is $\mathbf{Q}_i^\top \mathbf{R}_{ij} \mathbf{Q}_j = \rho$. The Hermite–Mehler identity further gives

$$\mathrm{Cov}(\boldsymbol{\psi}_i, \boldsymbol{\psi}_j) = \mathrm{diag}(\rho^2, \rho^3).$$

Thus,

$$\mathrm{Cov}(\tilde{\mathbf{h}}_i^{\mathrm{drift}}, \tilde{\mathbf{h}}_j^{\mathrm{drift}}) = \mathrm{diag}(\rho, \rho^2, \rho^3).$$

First-order loss versus higher-order gain. The aggregate first-order spectral loss is

$$L_1 := \sum_{\{i,j\} \in \mathcal{E}} \left(\|\mathbf{R}_{ij}\|_* - \|\mathbf{Q}_i^\top \mathbf{R}_{ij} \mathbf{Q}_j\|_* \right) = 3\tau,$$

whereas the aggregate higher-order spectral gain is

$$G_{\geq 2} := \sum_{\{i,j\} \in \mathcal{E}} \|\mathrm{Cov}(\boldsymbol{\psi}_i, \boldsymbol{\psi}_j)\|_* = 3(\rho^2 + \rho^3).$$

Equation 15 therefore yields $G_{\geq 2} > L_1$, realizing the strict spectral competition condition in Equation 3.

Every collection permitting full affine source recovery at $d_{\mathcal{Z}_i} = m_i = 3$ has objective value $J_{\mathrm{aff}} = 3(\rho + \tau)$. The alternative collection instead satisfies

$$\begin{aligned} J_{\mathrm{drift}} &= 3(\rho + \rho^2 + \rho^3), \\ J_{\mathrm{drift}} - J_{\mathrm{aff}} &= G_{\geq 2} - L_1 \\ &= 3(\rho^2 + \rho^3 - \tau) > 0. \end{aligned}$$

For $\rho = 0.6$ and $\tau = 0.45$, this gives

$$G_{\geq 2} = 1.728, \qquad L_1 = 1.35, \qquad J_{\mathrm{drift}} = 3.528 > 3.15 = J_{\mathrm{aff}}.$$

Interpretation. Each individual higher-order mode remains weaker than the pair-specific first-order mode: $\rho^3 < \rho^2 < \tau$. Nevertheless, each view replaces its two pair-specific source directions with two higher-order modes of ζ_i. The displaced directions each contribute on only one incident edge, whereas both replacement modes contribute on every incident edge. Their combined spectral gain therefore exceeds the first-order loss within the same dimension budget. By Proposition 2, every fully affine-recovering collection is strictly suboptimal. This conclusion does not require $\{\tilde{\mathbf{h}}_i^{\mathrm{drift}}\}$ itself to be a global maximizer.

B.3 PROOF OF AFFINE RECOVERY UNDER GLOBAL DOMINANCE

Proof of Theorem 4.1. Since each observation map $\mathbf{g}_i$ is invertible on its image, the source-space reduction of Han et al. (2026a) allows us to work with arbitrary centered, whitened encoders $\tilde{\mathbf{h}}_i(\mathbf{s}_i) \in \mathbb{R}^{m_i}$. All expectations below are population quantities, and we suppress function arguments when no ambiguity arises. For ordered pairs, set $\mathbf{R}_{ji} := \mathbf{R}_{ij}^\top$. By Equation 2, the objective is

$$J_{\mathrm{SUMCOR}}^{\mathcal{E}}(\{\tilde{\mathbf{h}}_i\}) = \sum_{\{i,j\}\in\mathcal{E}} \left\| \mathbb{E}[\tilde{\mathbf{h}}_i \tilde{\mathbf{h}}_j^\top] \right\|_* .$$

The feasible collection $\tilde{\mathbf{h}}_i = \tilde{\mathbf{s}}_i := \mathbf{W}_i \mathbf{s}_i$ attains

$$J_{\mathrm{aff}} := \sum_{\{i,j\}\in\mathcal{E}} \|\mathbf{R}_{ij}\|_* = \sum_{\{i,j\}\in\mathcal{E}} \sum_{k=1}^{r_{ij}} t_{ij,k}. \tag{16}$$

We show that every feasible collection with a nonzero higher-order component has a strictly smaller value.

Step 1: First-order projection and residual energy. For each view, decompose an arbitrary feasible encoder as

$$\tilde{\mathbf{h}}_i = \mathbf{F}_i \tilde{\mathbf{s}}_i + \boldsymbol{\xi}_i, \qquad \mathbf{F}_i := \mathbb{E}[\tilde{\mathbf{h}}_i \tilde{\mathbf{s}}_i^\top].$$

Since $\tilde{\mathbf{s}}_i$ is centered and whitened, $\mathbf{F}_i \tilde{\mathbf{s}}_i$ is the coordinatewise L^2 projection onto the linear source functions. Thus

$$\mathbb{E}[\boldsymbol{\xi}_i] = \mathbf{0}, \qquad \mathbb{E}[\boldsymbol{\xi}_i \tilde{\mathbf{s}}_i^\top] = \mathbf{0}.$$

Whitening of $\tilde{\mathbf{h}}_i$ gives

$$\mathbf{F}_i \mathbf{F}_i^\top + \mathbb{E}[\boldsymbol{\xi}_i \boldsymbol{\xi}_i^\top] = \mathbf{I}_{m_i}.$$

In particular, all singular values of $\mathbf{F}_i$ are at most one. Because $d_{\mathcal{Z}_i} = m_i$, the matrix $\mathbf{F}_i$ is square, and hence

$$\mathbf{D}_i := \mathbf{I}_{m_i} - \mathbf{F}_i^\top \mathbf{F}_i \succeq \mathbf{0}, \qquad \ell_i := \mathbb{E}\|\boldsymbol{\xi}_i\|_2^2 = m_i - \|\mathbf{F}_i\|_F^2 = \mathrm{tr}(\mathbf{D}_i).$$

Thus the residual energy ℓ_i equals the total first-order energy deficit.

Step 2: Higher-order contributions. Under Assumption 1, each source pair is jointly Gaussian. Its conditional means satisfy

$$\mathbb{E}[\tilde{\mathbf{s}}_i \mid \mathbf{s}_j] = \mathbf{R}_{ij} \tilde{\mathbf{s}}_j, \qquad \mathbb{E}[\tilde{\mathbf{s}}_j \mid \mathbf{s}_i] = \mathbf{R}_{ij}^\top \tilde{\mathbf{s}}_i.$$

Together with the projection orthogonality in Step 1, these identities imply

$$\mathbb{E}[\tilde{\mathbf{s}}_i \boldsymbol{\xi}_j^\top] = \mathbf{0}, \qquad \mathbb{E}[\boldsymbol{\xi}_i \tilde{\mathbf{s}}_j^\top] = \mathbf{0}.$$

Consequently,

$$\mathbb{E}[\tilde{\mathbf{h}}_i \tilde{\mathbf{h}}_j^\top] = \mathbf{F}_i \mathbf{R}_{ij} \mathbf{F}_j^\top + \mathbf{K}_{ij}, \qquad \mathbf{K}_{ij} := \mathbb{E}[\boldsymbol{\xi}_i \boldsymbol{\xi}_j^\top]. \tag{17}$$

We use the Gaussian Hermite–Mehler spectral characterization underlying Proposition 1 (Han et al., 2026b). In canonical coordinates, the nonzero singular values of the conditional-expectation operator on Hermite modes of total degree at least two are

$$\prod_{k=1}^{r_{ij}} t_{ij,k}^{n_k}, \qquad \mathbf{n} \in \mathbb{N}_0^{r_{ij}}, \quad |\mathbf{n}| \geq 2.$$

Each is at most $t_{ij,1}^2$. Modes involving an uncoupled canonical coordinate have zero coupling. Therefore, for any centered square-integrable scalar functions $p(\mathbf{s}_i)$ and $q(\mathbf{s}_j)$ orthogonal to all linear source functions in their respective views,

$$|\mathbb{E}[p(\mathbf{s}_i) q(\mathbf{s}_j)]| \leq t_{ij,1}^2 \|p\|_{L^2(P_{\mathbf{s}_i})} \|q\|_{L^2(P_{\mathbf{s}_j})}. \tag{18}$$

This bound does not require pairwise first-order dominance.

To obtain a nuclear-norm bound, let $\{(\mathbf{a}_k, \mathbf{b}_k)\}_{k=1}^{q}$ be orthonormal left and right singular-vector pairs of $\mathbf{K}_{ij}$, where $q = \operatorname{rank}(\mathbf{K}_{ij})$. Applying Equation 18 and Cauchy–Schwarz yields

$$\begin{aligned}\|\mathbf{K}_{ij}\|_* &= \sum_{k=1}^{q} \mathbb{E}[(\mathbf{a}_k^\top \boldsymbol{\xi}_i)(\mathbf{b}_k^\top \boldsymbol{\xi}_j)] \\ &\le t_{ij,1}^2 \left(\sum_{k=1}^{q} \mathbb{E}[(\mathbf{a}_k^\top \boldsymbol{\xi}_i)^2]\right)^{1/2} \left(\sum_{k=1}^{q} \mathbb{E}[(\mathbf{b}_k^\top \boldsymbol{\xi}_j)^2]\right)^{1/2} \\ &\le t_{ij,1}^2 \sqrt{\ell_i \ell_j} \le \frac{t_{ij,1}^2}{2}(\ell_i + \ell_j).\end{aligned}$$

The zero-rank case is immediate.

Step 3: First-order loss. For each ordered active pair, define

$$\mathbf{S}_{ij} := (\mathbf{R}_{ij}\mathbf{R}_{ij}^\top)^{1/2}, \qquad \mathbf{S}_{ji} := (\mathbf{R}_{ij}^\top \mathbf{R}_{ij})^{1/2}.$$

Take a compact SVD $\mathbf{R}_{ij} = \mathbf{U}\mathbf{T}\mathbf{V}^\top$, where $\mathbf{T}$ contains its positive singular values. Using $\|\mathbf{X}\mathbf{Y}^\top\|_* \le \|\mathbf{X}\|_F \|\mathbf{Y}\|_F \le (\|\mathbf{X}\|_F^2 + \|\mathbf{Y}\|_F^2)/2$, we obtain

$$\begin{aligned}\|\mathbf{F}_i \mathbf{R}_{ij} \mathbf{F}_j^\top\|_* &\le \frac{1}{2}\left(\|\mathbf{F}_i \mathbf{U}\mathbf{T}^{1/2}\|_F^2 + \|\mathbf{F}_j \mathbf{V}\mathbf{T}^{1/2}\|_F^2\right) \\ &= \frac{1}{2}\left\{\operatorname{tr}(\mathbf{S}_{ij}\mathbf{F}_i^\top \mathbf{F}_i) + \operatorname{tr}(\mathbf{S}_{ji}\mathbf{F}_j^\top \mathbf{F}_j)\right\} \\ &= \|\mathbf{R}_{ij}\|_* - \frac{1}{2}\left\{\operatorname{tr}(\mathbf{S}_{ij}\mathbf{D}_i) + \operatorname{tr}(\mathbf{S}_{ji}\mathbf{D}_j)\right\},\end{aligned}$$

where the last equality uses $\operatorname{tr}(\mathbf{S}_{ij}) = \operatorname{tr}(\mathbf{S}_{ji}) = \|\mathbf{R}_{ij}\|_*$.

Step 4: Global first-order dominance and the equality case. Combining Equation 17, the triangle inequality, and the bounds from Steps 2–3, then summing over unordered active edges, gives

$$\begin{aligned}J_{\mathrm{SUMCOR}}^{\mathcal{E}} &\le J_{\mathrm{aff}} - \frac{1}{2}\sum_{i=1}^{N} \operatorname{tr}(\mathbf{G}_i \mathbf{D}_i) + \frac{1}{2}\sum_{i=1}^{N} \beta_i \ell_i \\ &\le J_{\mathrm{aff}} - \frac{1}{2}\sum_{i=1}^{N} (\alpha_i - \beta_i)\ell_i. \end{aligned} \tag{19}$$

Indeed, $\mathbf{G}_i = \sum_{j \in \mathcal{N}_i} \mathbf{S}_{ij}$, $\beta_i = \sum_{j \in \mathcal{N}_i} t_{ij,1}^2$, and

$$\operatorname{tr}(\mathbf{G}_i \mathbf{D}_i) \ge \alpha_i \operatorname{tr}(\mathbf{D}_i) = \alpha_i \ell_i,$$

since $\mathbf{G}_i \succeq \alpha_i \mathbf{I}_{m_i}$ and $\mathbf{D}_i \succeq \mathbf{0}$.

The collection $\tilde{\mathbf{h}}_i = \tilde{\mathbf{s}}_i$ attains J_{aff} by Equation 16. Hence this is the population maximum. Since $\alpha_i > \beta_i$ for every view, Equation 19 forces $\ell_i = 0$ for all i at every maximizer. Thus $\boldsymbol{\xi}_i = \mathbf{0}$ almost surely, and whitening gives

$$\mathbf{F}_i \mathbf{F}_i^\top = \mathbf{I}_{m_i}.$$

Because $\mathbf{F}_i$ is square, it is orthogonal. Writing $\mathbf{O}_i := \mathbf{F}_i$ proves

$$\tilde{\mathbf{h}}_i^\star(\mathbf{s}_i) = \mathbf{O}_i \tilde{\mathbf{s}}_i = \mathbf{O}_i \mathbf{W}_i \mathbf{s}_i, \qquad \mathbf{O}_i \in O(m_i),$$

almost surely. Conversely, every such orthogonal collection attains J_{aff} by orthogonal invariance of the nuclear norm.

Finally, undoing representation whitening yields

$$\mathbf{h}_i^\star(\mathbf{s}_i) = \boldsymbol{\Sigma}_{ii}^{1/2} \mathbf{O}_i \mathbf{W}_i \mathbf{s}_i + \mathbb{E}[\mathbf{h}_i^\star(\mathbf{s}_i)],$$

where $\boldsymbol{\Sigma}_{ii} := \operatorname{Cov}(\mathbf{h}_i^\star(\mathbf{s}_i)) \succ 0$. The linear factor is invertible, proving full affine identifiability of each view-specific source. □

C RANK-CONSTRAINED SUMCOR: PROOF OF AFFINE RECOVERY

Proof of Theorem 4.2. All expectations and covariances are population quantities. Invertibility of $\mathbf{g}_i$ on its image allows us to work with source-space encoders $\mathbf{z}_i = \mathbf{h}_i(\mathbf{s}_i) \in \mathbb{R}^{m_i}$. Set $\mathbf{\Sigma}_{ii} := \mathrm{Cov}(\mathbf{z}_i) \succ 0$ and $\tilde{\mathbf{z}}_i := \mathbf{\Sigma}_{ii}^{-1/2}(\mathbf{z}_i - \mathbb{E}\mathbf{z}_i)$, so that $\mathbf{C}_{ij} = \mathbb{E}[\tilde{\mathbf{z}}_i\tilde{\mathbf{z}}_j^\top]$.

Step 1: Linear readouts and objective comparison. Fix an active edge $\{i, j\}$ and write $r = r_{ij}$. For $\ell \in \{i, j\}$, let $\mathbf{D}_\ell \in \mathbb{R}^{r\times m_\ell}$ have full row rank. The centered, whitened version of $\mathbf{D}_\ell\mathbf{z}_\ell$ is $\mathbf{F}_\ell\tilde{\mathbf{z}}_\ell$, where

$$\mathbf{F}_\ell := (\mathbf{D}_\ell\mathbf{\Sigma}_{\ell\ell}\mathbf{D}_\ell^\top)^{-1/2}\mathbf{D}_\ell\mathbf{\Sigma}_{\ell\ell}^{1/2}, \qquad \mathbf{F}_\ell\mathbf{F}_\ell^\top = \mathbf{I}_r.$$

The singular-value variational principle yields

$$J_{\mathrm{CCA}}(\mathbf{D}_i\mathbf{z}_i, \mathbf{D}_j\mathbf{z}_j) = \|\mathbf{F}_i\mathbf{C}_{ij}\mathbf{F}_j^\top\|_* \le \sum_{k=1}^{r}\sigma_k(\mathbf{C}_{ij}).$$

Equality is attained by $\mathbf{D}_i = \widehat{\mathbf{U}}_{ij}^\top\mathbf{\Sigma}_{ii}^{-1/2}$ and $\mathbf{D}_j = \widehat{\mathbf{V}}_{ij}^\top\mathbf{\Sigma}_{jj}^{-1/2}$, where $\widehat{\mathbf{U}}_{ij}$ and $\widehat{\mathbf{V}}_{ij}$ contain corresponding leading r left and right singular vectors of $\mathbf{C}_{ij}$. Consequently,

$$\sum_{k=1}^{r}\sigma_k(\mathbf{C}_{ij}) = \max_{\substack{\mathbf{D}_i\in\mathbb{R}^{r\times m_i},\,\mathbf{D}_j\in\mathbb{R}^{r\times m_j}\\ \mathrm{rank}(\mathbf{D}_i)=\mathrm{rank}(\mathbf{D}_j)=r}} J_{\mathrm{CCA}}(\mathbf{D}_i\mathbf{z}_i, \mathbf{D}_j\mathbf{z}_j). \tag{20}$$

Every admissible coordinate mask is a full-row-rank linear readout. Summing over edges gives, for any feasible encoders and masks,

$$J_{\mathrm{mask}}^{\mathcal{E}}\big(\{\mathbf{z}_i\}, \{\mathbf{M}_{i|j}\}\big) \le J_{\mathrm{top}}^{\mathcal{E}}\big(\{\mathbf{z}_i\}\big). \tag{21}$$

This comparison does not require either dominance condition.

Step 2: Simultaneous attainment of the pairwise bounds. Under Assumption 1, the Hermite–Mehler decomposition underlying Proposition 1 (Han et al., 2026b) gives first-order singular values $t_{ij,1}, \ldots, t_{ij,r_{ij}}$ for the centered pairwise coupling operator. All remaining nonconstant singular values are at most $t_{ij,1}^2$. Pairwise first-order dominance therefore makes these r_{ij} first-order modes its leading modes. In particular, every r_{ij}-dimensional pair of encoders has CCA score at most $b_{ij} := \sum_{k=1}^{r_{ij}} t_{ij,k}$, with equality characterized by Proposition 1. Applying this bound to Equation 20 gives

$$J_{\mathrm{mask}}^{\mathcal{E}} \le J_{\mathrm{top}}^{\mathcal{E}} \le \sum_{\{i,j\}\in\mathcal{E}} b_{ij} = J_{\mathrm{aff}}. \tag{22}$$

For Top-r SUMCOR, choosing $\mathbf{h}_i(\mathbf{s}_i) = \mathbf{s}_i$ makes $\mathbf{C}_{ij} = \mathbf{R}_{ij}$ and attains every b_{ij}. For masked SUMCOR, use $\mathbf{h}_i = \mathbf{q}_i$ and $\mathbf{M}_{i|j} = \mathbf{E}_{i|j}$ from Lemma 1. Each selected block has nonsingular covariance and spans $\mathcal{L}_{i|j}$. After separate block whitening, the two blocks on an edge are orthonormal bases of its first-order canonical spaces, so their cross-covariance has singular values $t_{ij,1}, \ldots, t_{ij,r_{ij}}$. Thus all edge bounds are attained simultaneously. Both constructions are feasible observation-space encoders through composition with $\mathbf{g}_i^{-1}$. Hence both population maxima exist and equal J_{aff}.

Step 3: Every maximizing encoder is affine. Let $\{\mathbf{h}_i^\star\}$ maximize Top-r SUMCOR or be the encoder component of a jointly maximizing masked SUMCOR solution. In the latter case, Equation 22 implies $J_{\mathrm{top}}^{\mathcal{E}}(\{\mathbf{z}_i^\star\}) = J_{\mathrm{aff}}$ as well. In either case,

$$0 = \sum_{\{i,j\}\in\mathcal{E}}\left(b_{ij} - \sum_{k=1}^{r_{ij}}\sigma_k(\mathbf{C}_{ij}^\star)\right).$$

Each summand is nonnegative, so every edge attains its individual upper bound. Using leading singular-vector matrices of $\mathbf{C}_{ij}^\star$, define

$$\mathbf{p}_{i|j} := \widehat{\mathbf{U}}_{ij}^\top\tilde{\mathbf{h}}_i^\star(\mathbf{s}_i), \qquad \mathbf{p}_{j|i} := \widehat{\mathbf{V}}_{ij}^\top\tilde{\mathbf{h}}_j^\star(\mathbf{s}_j).$$

These are centered, whitened r_{ij}-dimensional readouts whose CCA score is b_{ij}. By Proposition 1, $\mathrm{span}(\mathbf{p}_{i|j}) = \mathcal{L}_{i|j}$ and $\mathrm{span}(\mathbf{p}_{j|i}) = \mathcal{L}_{j|i}$.

Let $\mathcal{H}_i^\star$ be the span of the centered coordinates of $\mathbf{h}_i^\star(\mathbf{s}_i)$. Every $\mathbf{p}_{i|j}$ is a linear readout of these coordinates, so $\mathcal{L}_{i|j} \subseteq \mathcal{H}_i^\star$. Coverage and Lemma 1 therefore imply

$$\mathcal{L}_i = \mathrm{span}_{j \in \mathcal{N}_i} \mathcal{L}_{i|j} \subseteq \mathcal{H}_i^\star.$$

Nonsingularity of the encoder covariance and $d_{\mathcal{Z}_i} = m_i$ give $\dim \mathcal{H}_i^\star = m_i = \dim \mathcal{L}_i$. Thus $\mathcal{H}_i^\star = \mathcal{L}_i$, and

$$\mathbf{h}_i^\star(\mathbf{s}_i) = \mathbf{T}_i \mathbf{s}_i + \mathbf{b}_i, \qquad \mathbf{b}_i = \mathbb{E}[\mathbf{h}_i^\star(\mathbf{s}_i)].$$

The nonsingular encoder covariance forces $\mathbf{T}_i \in \mathrm{GL}(m_i)$.

Finally, write $\mathbf{\Sigma}_{ii}^\star = \mathbf{T}_i \mathbf{W}_i^{-2} \mathbf{T}_i^\top$ and set $\mathbf{O}_i := (\mathbf{\Sigma}_{ii}^\star)^{-1/2} \mathbf{T}_i \mathbf{W}_i^{-1}$. Then $\mathbf{O}_i \mathbf{O}_i^\top = \mathbf{I}_{m_i}$ and $\tilde{\mathbf{h}}_i^\star(\mathbf{s}_i) = \mathbf{O}_i \mathbf{W}_i \mathbf{s}_i$, proving Equation 4. This final whitening describes the recovery ambiguity. It is not a global whitening constraint on the coordinate-masked encoders. □

C.1 Mask profiling and empirical support selection

Fix encoder output dimensions $d_{\mathcal{Z}_i}$ and known ranks $1 \le r_{ij} \le \min(d_{\mathcal{Z}_i}, d_{\mathcal{Z}_j})$ on $\mathcal{E}$. Let $\mathfrak{M}$ be the finite, nonempty set of admissible coordinate-mask collections. Selectors have r_{ij} distinct rows of $\mathbf{I}_{d_{\mathcal{Z}_i}}$ at endpoint i. Row order may be fixed because CCA scores are invariant to coordinate permutations. For square-integrable encoder collections $\mathbf{f} = (\mathbf{f}_i)_{i=1}^N$ with nonsingular output covariances, write $J(\mathbf{f}, \mathbf{M})$ for the population masked objective in Equation 5, evaluated at $\mathbf{z}_i = \mathbf{f}_i(\mathbf{x}_i)$. Every selected block has positive-definite covariance. Define the profiled objective

$$F_{\mathrm{prof}}(\mathbf{f}) := \max_{\mathbf{M} \in \mathfrak{M}} J(\mathbf{f}, \mathbf{M}). \tag{23}$$

Masks on different edges are uncoupled and may overlap, so this maximum can be computed separately on each edge.

Lemma 2 (Population mask profiling)**.** *Over the encoder domain above,*

$$\sup_{\mathbf{f}} F_{\mathrm{prof}}(\mathbf{f}) = \sup_{\mathbf{f},\, \mathbf{M} \in \mathfrak{M}} J(\mathbf{f}, \mathbf{M}).$$

An encoder collection maximizes F_{prof} if and only if it is the encoder component of a jointly maximizing encoder–mask pair.

Proof. For each $\mathbf{f}$, finiteness and nonemptiness of $\mathfrak{M}$ ensure that some $\mathbf{M}_{\mathbf{f}}$ attains the inner maximum. Thus $J(\mathbf{f}, \mathbf{M}) \le F_{\mathrm{prof}}(\mathbf{f}) = J(\mathbf{f}, \mathbf{M}_{\mathbf{f}})$. Taking suprema proves the equality. If $\mathbf{f}$ maximizes F_{prof}, then $(\mathbf{f}, \mathbf{M}_{\mathbf{f}})$ attains this same supremum. Conversely, a jointly maximizing pair has an encoder component whose profiled value cannot be smaller or exceed that supremum. □

Under the assumptions of Theorem 4.2, including $d_{\mathcal{Z}_i} = m_i$, the population maximum is attained. Consequently, every maximizing encoder of F_{prof} has the same affine recovery guarantee. Profiling does not identify unique supports or estimate the supplied ranks. Sample independence means that a population mask is constant across input samples. It does not prohibit updating masks while optimizing the encoders.

Lemma 3 (Empirical mask selection at fixed encoders)**.** *Fix an admissible encoder collection $\mathbf{f}$. Let $\widehat{J}_n(\mathbf{f}, \mathbf{M})$ be empirical scores defined for all $\mathbf{M} \in \mathfrak{M}$, and choose any empirical maximizer $\widehat{\mathbf{M}}_n$. Set*

$$\varepsilon_n := \max_{\mathbf{M} \in \mathfrak{M}} \left| \widehat{J}_n(\mathbf{f}, \mathbf{M}) - J(\mathbf{f}, \mathbf{M}) \right|.$$

Then

$$0 \le F_{\mathrm{prof}}(\mathbf{f}) - J(\mathbf{f}, \widehat{\mathbf{M}}_n) \le 2\varepsilon_n. \tag{24}$$

Proof. Let $\mathbf{M}^\star$ maximize the population score for $\mathbf{f}$. The lower bound follows from its optimality. For the upper bound,

$$\begin{aligned} J(\mathbf{f}, \mathbf{M}^\star) &\le \widehat{J}_n(\mathbf{f}, \mathbf{M}^\star) + \varepsilon_n \\ &\le \widehat{J}_n(\mathbf{f}, \widehat{\mathbf{M}}_n) + \varepsilon_n \\ &\le J(\mathbf{f}, \widehat{\mathbf{M}}_n) + 2\varepsilon_n. \end{aligned}$$

□

This is a deterministic comparison, not a bound on ε_n or a statistical convergence rate. Its fixed-encoder interpretation holds all encoder state, including batch-normalization statistics, fixed. For the regularized empirical CCA score used in our implementation, ε_n includes both covariance estimation error and the discrepancy from regularized whitening. No vanishing-error claim is made at fixed ridge, and no uniform control over encoders visited during training is established. The bound controls mask-selection loss at a fixed representation. It does not show that the representation attains a population optimum or satisfies the source-recovery conclusion of the theorem.

For fixed encoders, the expected minibatch score $\mathbb{E}_B[\max_{\mathbf{M}\in\mathfrak{M}} \widehat{J}_B(\mathbf{f},\mathbf{M})]$ need not equal $F_{\mathrm{prof}}(\mathbf{f})$. In particular,

$$\mathbb{E}_B\left[\max_{\mathbf{M}\in\mathfrak{M}} \widehat{J}_B(\mathbf{f},\mathbf{M})\right] \geq \max_{\mathbf{M}\in\mathfrak{M}} \mathbb{E}_B[\widehat{J}_B(\mathbf{f},\mathbf{M})].$$

The right-hand side also need not equal the population profile, since empirical CCA scores are not assumed unbiased. Thus exact enumeration solves the empirical support subproblem, while population profiling alone does not justify exchanging expectation and maximization or establish convergence of neural training.

D Finite-Overcomplete SUMCOR: Proof and Capacity Example

Throughout this section, all function spaces are real and centered. Orthogonal complements are taken inside $\mathcal{F}_i = L_0^2(P_{\mathbf{s}_i})$, and P_i denotes the orthogonal projection onto a finite-dimensional feature space $\mathcal{H}_i$. Write $\mathcal{A}_{ij} := (\mathcal{T}_{ij}\mathcal{T}_{ij}^*)^{1/2}$ for the directed positive operator, so that $\Gamma_i = \sum_{j\in\mathcal{N}_i} \mathcal{A}_{ij}$. By source-space reduction and population whitening, choosing an encoder of dimension $d_{\mathcal{Z}_i}$ is equivalent to choosing an orthonormal family of that size in $\mathcal{F}_i$. In these coordinates, the compressed operator $\mathsf{P}_i\mathcal{T}_{ij}\mathsf{P}_j$ and the representation cross-covariance have the same nonzero singular values. Hence

$$J^{\mathcal{E}}_{\mathrm{SUMCOR}} = \sum_{\{i,j\}\in\mathcal{E}} \|\mathsf{P}_i\mathcal{T}_{ij}\mathsf{P}_j\|_*. \tag{25}$$

D.1 Finite spectral depth and finite aggregate capacity

Trace-classness. For each active edge, the Gaussian Hermite–Mehler expansion yields

$$\begin{aligned}\|\mathcal{T}_{ij}\|_* &= \sum_{\mathbf{n}\in\mathbb{N}_0^{r_{ij}}\setminus\{\mathbf{0}\}} \prod_{k=1}^{r_{ij}} t_{ij,k}^{n_k} \\ &= \prod_{k=1}^{r_{ij}} (1-t_{ij,k})^{-1} - 1 = M_{ij} < \infty.\end{aligned} \tag{26}$$

Thus $\mathcal{A}_{ij}$ and Γ_i are positive trace-class operators, and

$$\mathrm{tr}(\Gamma_i) = \sum_{j\in\mathcal{N}_i} M_{ij}.$$

First-order containment. Gaussian conditional expectations and their adjoints preserve first-order functions and their orthogonal complements. Accordingly, $\mathcal{L}_i$ reduces Γ_i. In the orthonormal coordinates $\tilde{\mathbf{s}}_i$, the restriction of $\mathcal{A}_{ij}$ to $\mathcal{L}_i$ is $(\mathbf{R}_{ij}\mathbf{R}_{ij}^\top)^{1/2}$, and the restriction of Γ_i is $\mathbf{G}_i$. Coverage and Lemma 1 give

$$\mathrm{span}_{j\in\mathcal{N}_i} \mathrm{col}(\mathbf{R}_{ij}) = \mathbb{R}^{m_i}.$$

The positive semidefinite summands of $\mathbf{G}_i$ therefore have trivial common kernel, so $\alpha_i > 0$. For $\gamma < \alpha_i$, every eigenvector of this first-order block lies in $\mathcal{Q}_i^\gamma$.

Since Γ_i is compact, its spectral projection above a positive threshold has finite rank. Moreover,

$$\gamma D_i(\gamma) \leq \mathrm{tr}(\mathsf{Q}_i^\gamma \Gamma_i) \leq \mathrm{tr}(\Gamma_i).$$

The reducing decomposition also gives the exact identity

$$D_i(\gamma) = m_i + \mathrm{rank}\big(\mathbf{1}_{[\gamma,\infty)}(\Gamma_i|_{\mathcal{L}_i^\perp})\big).$$

Thus the surplus capacity is precisely the number of retained higher-order aggregate directions, counted with multiplicity.

D.2 A VIEW-WISE TRACE BOUND FOR SUMCOR

Lemma 4 (View-wise trace bound)**.** *For arbitrary finite-rank orthogonal projections* $\{\mathsf{P}_i\}_{i=1}^N$,

$$J^{\mathcal{E}}_{\mathrm{SUMCOR}} \le \frac{1}{2}\sum_{i=1}^{N} \mathrm{tr}(\mathsf{P}_i \Gamma_i). \tag{27}$$

Proof. Take a singular expansion on one edge,

$$\mathcal{T}_{ij} = \sum_{k\ge 1} \sigma_k \, u_k \langle v_k, \cdot \rangle, \qquad \sigma_k > 0,$$

which converges in trace norm. The trace norm of a rank-one operator $u\langle v, \cdot\rangle$ equals $\|u\|\|v\|$. The triangle inequality and $2ab \le a^2 + b^2$ yield

$$\begin{aligned} \|\mathsf{P}_i \mathcal{T}_{ij} \mathsf{P}_j\|_* &\le \sum_k \sigma_k \|\mathsf{P}_i u_k\| \|\mathsf{P}_j v_k\| \\ &\le \frac{1}{2}\sum_k \sigma_k \big(\|\mathsf{P}_i u_k\|^2 + \|\mathsf{P}_j v_k\|^2\big) \\ &= \frac{1}{2}\big\{\mathrm{tr}(\mathsf{P}_i \mathcal{A}_{ij}) + \mathrm{tr}(\mathsf{P}_j \mathcal{A}_{ji})\big\}. \end{aligned} \tag{28}$$

Summing over unordered edges proves the claim. □

A sufficient compatibility criterion. The joint spectral compatibility condition is checkable from the coupling operators and their prescribed finite spectral projections. It does not refer to unknown optimal encoders. In particular, suppose that for each edge one can choose a singular system and an index set I_{ij} such that

$$\mathsf{Q}_i^{\gamma} u_k = \begin{cases} u_k, & k \in I_{ij}, \\ 0, & k \notin I_{ij}, \end{cases} \quad \mathsf{Q}_j^{\gamma} v_k = \begin{cases} v_k, & k \in I_{ij}, \\ 0, & k \notin I_{ij}. \end{cases}$$

The compressed coupling then retains exactly the matched singular pairs indexed by I_{ij}. Its trace norm and both directed traces in Equation 28 equal $\sum_{k\in I_{ij}} \sigma_k$. Thus Assumption 5 holds. Directions in an edge's kernel may be retained independently, and different edges may use different singular bases.

D.3 PROOF OF THE FINITE-CAPACITY RECOVERY THEOREM

Proof of Theorem 4.3. Write $D_i = D_i(\gamma)$ and $\mathsf{Q}_i = \mathsf{Q}_i^{\gamma}$. Let $\lambda_{i,1} \ge \lambda_{i,2} \ge \cdots \ge 0$ be the eigenvalues of Γ_i, counted with multiplicity and extended by zeros if necessary. Define

$$U_\gamma := \frac{1}{2}\sum_i \sum_{k=1}^{D_i} \lambda_{i,k} = \frac{1}{2}\sum_i \mathrm{tr}(\mathsf{Q}_i \Gamma_i).$$

For any feasible encoder collection, Lemma 4 and the spectral variational principle give

$$J^{\mathcal{E}}_{\mathrm{SUMCOR}} \le \frac{1}{2}\sum_i \mathrm{tr}(\mathsf{P}_i \Gamma_i) \le U_\gamma. \tag{29}$$

Choose an orthonormal basis of each $\mathcal{Q}_i^{\gamma}$ as its encoder coordinates. These finite-dimensional encoders are feasible and, by Assumption 5, attain equality in both inequalities. Hence the maximum exists and equals U_γ.

It remains to characterize every maximizer. For a maximizing collection, all view-wise trace deficits in Equation 29 are nonnegative and sum to zero. Therefore

$$\mathrm{tr}(\mathsf{P}_i \Gamma_i) = \mathrm{tr}(\mathsf{Q}_i \Gamma_i) \quad \text{for every } i.$$

Because the cutoff retains the entire spectral subspace for $[\gamma, \infty)$, compactness implies

$$\delta_i := \lambda_{i,D_i} - \lambda_{i,D_i+1} > 0, \qquad \lambda_{i,D_i} \geq \gamma > \lambda_{i,D_i+1}.$$

For any rank-D_i projection P_i, diagonalizing Γ_i and using $\operatorname{tr}(\mathsf{Q}_i(\mathsf{I} - \mathsf{P}_i)) = \operatorname{tr}((\mathsf{I} - \mathsf{Q}_i)\mathsf{P}_i)$ gives

$$\begin{aligned}\operatorname{tr}(\mathsf{Q}_i\Gamma_i) - \operatorname{tr}(\mathsf{P}_i\Gamma_i) &\geq \delta_i \operatorname{tr}(\mathsf{Q}_i(\mathsf{I} - \mathsf{P}_i)) \\ &= \frac{\delta_i}{2}\|\mathsf{Q}_i - \mathsf{P}_i\|_{\mathrm{HS}}^2.\end{aligned}$$

Equality of the traces thus forces $\mathsf{P}_i = \mathsf{Q}_i$. Consequently, $\mathcal{H}_i^\star = \mathcal{Q}_i^\gamma \supseteq \mathcal{L}_i$.

Finally, expand the whitened source coordinates in the finite orthonormal feature family. Explicitly, set

$$\mathbf{L}_i := \mathbb{E}\big[\tilde{\mathbf{s}}_i \tilde{\mathbf{h}}_i^{\star\top}\big].$$

Source-space containment implies $\mathbf{L}_i\tilde{\mathbf{h}}_i^\star = \tilde{\mathbf{s}}_i$ in L^2, hence almost surely. Whitening gives $\mathbf{L}_i\mathbf{L}_i^\top = \operatorname{Cov}(\tilde{\mathbf{s}}_i) = \mathbf{I}_{m_i}$. Undoing centering and whitening yields an affine readout of $\mathbf{s}_i$ from the original representation. □

Scope. The argument concerns the specified capacities $D_i(\gamma)$. It does not assert recovery for every componentwise larger allocation, nor a spectrum-independent finite dimension bound. Gaussianity is used to obtain reducing first-order spaces and trace-class couplings with explicit spectra. The same trace-bound argument applies to non-Gaussian couplings satisfying these operator properties, positive first-order coverage, and joint spectral compatibility. A Lancaster expansion alone does not supply the last condition.

D.4 A FINITE-CAPACITY RECOVERY INSTANCE WITHOUT DOMINANCE

Consider the visibility pattern

$$\mathcal{V}_1 = \{1, 2, 3\}, \qquad \mathcal{V}_2 = \{1, 2, 4\}, \qquad \mathcal{V}_3 = \{1, 3, 4\},$$

and the Gaussian source model under Assumption 1, $\mathbf{s}_i = \mathbf{B}_i\mathbf{P}_i\mathbf{c} + \boldsymbol{\epsilon}_i$, where

$$\mathbf{B}_i = \mathbf{Q}_i \operatorname{diag}(a, b, b), \qquad a = \sqrt{\frac{\rho}{1-\rho}}, \qquad b = \sqrt{\frac{\tau}{1-\tau}}, \qquad 0 < \tau < \rho < 1.$$

Here $\mathbf{Q}_i$ is any orthogonal matrix, including the dense rotations used in Setups B and C. Independent isotropic noise makes these rotations immaterial to the following spectral calculation. Write

$$(\zeta_i, \eta_i, \xi_i)^\top := \mathbf{Q}_i^\top \tilde{\mathbf{s}}_i.$$

These are independent standard normal coordinates within each view. The ζ_i coordinates have correlation ρ on every edge, while each of η_i, ξ_i has correlation τ with its counterpart in exactly one neighboring view. The nonzero first-order edge spectrum is (ρ, τ) and $\mathbf{G}_i$ is orthogonally similar to $\operatorname{diag}(2\rho, \tau, \tau)$. Thus $\alpha_i = \tau$ and $\beta_i = 2\rho^2$.

Let H_a denote the normalized probabilists' Hermite polynomial of degree a, with $H_0 = 1$ and $H_1(z) = z$. The tensor Hermite basis diagonalizes Γ_i. For $a, b, c \in \mathbb{N}_0$ with $a + b + c > 0$, its eigenvalue on $H_a(\zeta_i)H_b(\eta_i)H_c(\xi_i)$ is

$$\mu_{a,b,c} = \begin{cases} 2\rho^a, & b = c = 0,\ a \geq 1, \\ \rho^a\tau^b, & b \geq 1,\ c = 0, \\ \rho^a\tau^c, & c \geq 1,\ b = 0, \\ 0, & b \geq 1,\ c \geq 1. \end{cases} \tag{30}$$

The factor two in the first line records the two incident edges. A mode involving both pair-specific coordinates couples to neither neighbor and therefore has zero aggregate eigenvalue.

Proposition 3 (An explicit finite recovery capacity)**.** *For the above family, define*

$$p_\star := \max\{p \in \mathbb{N} : 2\rho^p \geq \tau\} = \left\lfloor \frac{\log(\tau/2)}{\log \rho} \right\rfloor.$$

SUMCOR with $d_{\mathcal{Z}_i} = p_\star + 2$ for every view has maximum

$$J_\star = 3\left(\sum_{a=1}^{p_\star} \rho^a + \tau\right).$$

Every global maximizer permits full affine source recovery, without assuming either first-order dominance condition.

Proof. Among the eigenvalues in Equation 30, the only ones at least τ are $2\rho^a$ for $1 \le a \le p_\star$ and the two eigenvalues τ associated with η_i and ξ_i. All remaining eigenvalues are at most

$$b_\star := \max\{2\rho^{p_\star+1}, \rho\tau, \tau^2\} < \tau.$$

Choose $b_\star < \gamma < \tau = \alpha_i$. Then

$$\begin{aligned} \mathcal{Q}_i^\gamma &= \operatorname{span}\{H_1(\zeta_i), \dots, H_{p_\star}(\zeta_i), \eta_i, \xi_i\}, \\ D_i(\gamma) &= p_\star + 2. \end{aligned}$$

Every retained global mode is present at all three views. Each retained pair-specific mode is present at both endpoints of its active edge. The other pair-specific coordinate lies in that edge's kernel. Hence the matched-singular-mode criterion holds on every edge. The compressed edge singular values are $\rho, \dots, \rho^{p_\star}, \tau$. Their sum is $\sum_{a=1}^{p_\star} \rho^a + \tau$. Summing over the three edges and applying Theorem 4.3 proves the claim. □

Setup C: six coordinates are sufficient and minimal. For $(\rho, \tau) = (0.6, 0.2)$, both dominance gaps are negative:

$$\tau - \rho^2 = -0.16, \qquad \alpha_i - \beta_i = \tau - 2\rho^2 = -0.52.$$

Nevertheless,

$$2\rho^4 = 0.2592 > \tau = 0.2 > 2\rho^5 = 0.15552,$$

so $p_\star = 4$ and, for example, $\gamma = 0.18$ gives $D_i(\gamma) = 6$. An optimal orthonormal feature family is

$$\tilde{\mathbf{h}}_i = \big(H_1(\zeta_i), H_2(\zeta_i), H_3(\zeta_i), H_4(\zeta_i), \eta_i, \xi_i\big)^\top,$$

with objective $J_\star = 4.5168$. The source is recovered from coordinates $1, 5, 6$, followed by an invertible linear transformation.

For completeness, six is also the smallest common output dimension at which a fully recovering collection can be globally optimal in this particular instance. Dimensions below three cannot support a full affine readout. For $d \in \{3, 4, 5\}$, any encoder collection that fully recovers every view-specific source has $\mathcal{H}_i = \mathcal{L}_i \oplus \mathcal{R}_i$ with $\dim \mathcal{R}_i = d - 3$ and $\mathcal{R}_i \subseteq \mathcal{L}_i^\perp$. The two largest higher-order aggregate eigenvalues are $2\rho^2$ and $2\rho^3$. The trace upper bound therefore gives

$$J_{\mathrm{rec},d} \le 3\left(\rho + \tau + \sum_{a=2}^{d-2} \rho^a\right),$$

where the sum is empty for $d = 3$. The global-only encoder with coordinates $H_a(\zeta_i)$, $1 \le a \le d$, is feasible and instead attains

$$J_{\mathrm{global},d} = 3\sum_{a=1}^{d} \rho^a.$$

Their difference is bounded below by

$$J_{\mathrm{global},d} - J_{\mathrm{rec},d} \ge 3(\rho^{d-1} + \rho^d - \tau) > 0, \qquad d = 3, 4, 5.$$

The final inequality holds because its smallest right-hand side, at $d = 5$, is $3(0.1296 + 0.07776 - 0.2) = 0.02208 > 0$. Thus every fully recovering collection is strictly suboptimal at these smaller dimensions, whereas all six-dimensional global maximizers recover the full sources. This minimality claim is specific to the stated parameters and common-dimension constraint.

E BEYOND GAUSSIAN SOURCES: FIRST-ORDER INVARIANT COUPLINGS

This appendix separates the operator structure used by the recovery proofs from its Gaussian realization. We first give sufficient conditions directly on the joint source law, and then specialize them to degree-compatible Lancaster couplings. The pairwise Lancaster reduction follows the same spectral principle as Han et al. (2026a). The multi-view conclusions additionally account for aggregate competition, coverage, and a common coordinate system for the routed readouts. All optimization statements concern population global optima over the indicated square-integrable function classes.

E.1 OPERATOR STRUCTURE AND OVERLAP GEOMETRY

Let $(\mathbf{s}_1, \dots, \mathbf{s}_N)$ have a jointly specified probability law with finite second moments and nonsingular marginal covariances. Each observation map is measurable and injective on the source support, with a measurable inverse on its image. Composition with the inverse identifies observation-space and source-space L^2 encoders and preserves their joint output laws. All objectives and constraints below are unchanged by this reduction, which requires no smoothness. We do not impose Assumption 1 in this appendix. Define

$$\boldsymbol{\mu}_i := \mathbb{E}\mathbf{s}_i, \qquad \boldsymbol{\Sigma}_i := \mathrm{Cov}(\mathbf{s}_i), \qquad \mathbf{W}_i := \boldsymbol{\Sigma}_i^{-1/2}, \qquad \tilde{\mathbf{s}}_i := \mathbf{W}_i(\mathbf{s}_i - \boldsymbol{\mu}_i).$$

Write

$$\mathcal{F}_i := L_0^2(P_{\mathbf{s}_i}), \qquad \mathcal{L}_i := \{\mathbf{a}^\top \tilde{\mathbf{s}}_i : \mathbf{a} \in \mathbb{R}^{m_i}\},$$

and let $\Pi_i : \mathcal{F}_i \to \mathcal{L}_i$ be the orthogonal projection. Orthogonal complements below are taken inside the centered space $\mathcal{F}_i$. For an active edge, define

$$\mathbf{R}_{ij} := \mathbb{E}[\tilde{\mathbf{s}}_i \tilde{\mathbf{s}}_j^\top], \qquad (\mathcal{T}_{ij} f)(\mathbf{s}_i) := \mathbb{E}[f(\mathbf{s}_j) \mid \mathbf{s}_i].$$

Thus $\mathcal{T}_{ij} : \mathcal{F}_j \to \mathcal{F}_i$ is a contraction and $\mathcal{T}_{ji} = \mathcal{T}_{ij}^*$. For this general formulation, let $r_{ij} := \mathrm{rank}(\mathbf{R}_{ij}) > 0$ on each active edge, and denote its positive singular values by $t_{ij,1} \geq \cdots \geq t_{ij,r_{ij}} > 0$. Under the structured model considered below, this rank equals the latent overlap cardinality used in the main text. Define the pairwise target by

$$\mathcal{L}_{i|j} := \{\mathbf{a}^\top \tilde{\mathbf{s}}_i : \mathbf{a} \in \mathrm{col}(\mathbf{R}_{ij})\}.$$

Assumption 6 (First-order invariance)**.** For every active edge,

$$\Pi_i \mathcal{T}_{ij} = \mathcal{T}_{ij} \Pi_j.$$

This condition requires that both first-order functions and their orthogonal complements remain in the corresponding blocks under conditional expectation. Equivalently, both conditional mean regressions are linear:

$$\mathbb{E}[\tilde{\mathbf{s}}_j \mid \mathbf{s}_i] = \mathbf{R}_{ij}^\top \tilde{\mathbf{s}}_i,$$
$$\mathbb{E}[\tilde{\mathbf{s}}_i \mid \mathbf{s}_j] = \mathbf{R}_{ij} \tilde{\mathbf{s}}_j.$$

Consequently,

$$\mathcal{T}_{ij} = \Pi_i \mathcal{T}_{ij} \Pi_j + (I - \Pi_i)\mathcal{T}_{ij}(I - \Pi_j).$$

In the orthonormal bases supplied by the whitened source coordinates, the first block is represented by $\mathbf{R}_{ij}$. Define the remaining coupling strength by

$$\eta_{ij} := \left\| (I - \Pi_i)\mathcal{T}_{ij}(I - \Pi_j) \right\|_{\mathrm{op}}. \tag{31}$$

This definition includes every centered function orthogonal to $\mathcal{L}_i$ and $\mathcal{L}_j$, not only quadratic polynomials. Adjoint symmetry gives $\eta_{ji} = \eta_{ij}$. The aggregate quantities are

$$\mathbf{G}_i := \sum_{j \in \mathcal{N}_i} (\mathbf{R}_{ij} \mathbf{R}_{ij}^\top)^{1/2},$$
$$\alpha_i := \lambda_{\min}(\mathbf{G}_i), \qquad \beta_i^\perp := \sum_{j \in \mathcal{N}_i} \eta_{ij}.$$

Two geometric requirements. We distinguish linear coverage,

$$\mathrm{span}_{j\in\mathcal{N}_i}\,\mathcal{L}_{i|j}=\mathcal{L}_i,\qquad i\in[N], \tag{32}$$

from simultaneous coordinate realizability. The latter means that, for each i, there is an invertible matrix $\mathbf{T}_i$ and, for every neighbor j, a coordinate mask $\mathbf{M}_{i|j}\in\mathcal{M}(r_{ij},m_i)$ such that

$$\begin{aligned}\mathbf{q}_i &:= \mathbf{T}_i(\mathbf{s}_i-\boldsymbol{\mu}_i),\\ \mathrm{span}(\mathbf{M}_{i|j}\mathbf{q}_i) &= \mathcal{L}_{i|j}.\end{aligned} \tag{33}$$

Here the span of a vector-valued function means the linear span of its scalar coordinate functions. The same $\mathbf{q}_i$ must serve all edges incident to view i. Separate edgewise bases do not by themselves establish Equation 33.

Lemma 5 (Distribution-free covariance geometry)**.** *Retain the structural model in Equation 1. Suppose that $\mathbf{c},\boldsymbol{\epsilon}_1,\dots,\boldsymbol{\epsilon}_N$ are mutually independent, have finite second moments, and satisfy*

$$\mathrm{Cov}(\mathbf{c})=\mathbf{I}_{d_c},\qquad \mathrm{Cov}(\boldsymbol{\epsilon}_i)=\boldsymbol{\Gamma}_i\succ 0.$$

They need not be Gaussian. Then

$$\begin{aligned}\boldsymbol{\Sigma}_i &= \mathbf{A}_i\mathbf{A}_i^\top+\boldsymbol{\Gamma}_i,\\ \mathbf{R}_{ij} &= \mathbf{W}_i\mathbf{B}_i\mathbf{P}_i\mathbf{P}_j^\top\mathbf{B}_j^\top\mathbf{W}_j,\end{aligned}$$

$r_{ij}=|\mathcal{V}_i\cap\mathcal{V}_j|$, and Equation 33 holds with

$$\mathbf{q}_i=\mathbf{B}_i^\top\boldsymbol{\Sigma}_i^{-1}(\mathbf{s}_i-\boldsymbol{\mu}_i).$$

If Assumption 3 holds, then Equation 32 also holds.

Proof. The covariance identities follow from mutual independence and $\mathbf{P}_i\mathbf{P}_i^\top=\mathbf{I}_{m_i}$. Invertibility of $\mathbf{W}_i\mathbf{B}_i$ and $\mathbf{W}_j\mathbf{B}_j$ gives the rank identity. Let $\mathbf{E}_{i|j}$ select the local coordinates corresponding to $\mathcal{V}_i\cap\mathcal{V}_j$. Then

$$\mathrm{col}(\mathbf{R}_{ij})=\mathrm{col}(\mathbf{W}_i\mathbf{B}_i\mathbf{E}_{i|j}^\top).$$

Therefore,

$$\begin{aligned}\mathcal{L}_{i|j} &= \mathrm{span}\big(\mathbf{E}_{i|j}\mathbf{B}_i^\top\mathbf{W}_i\tilde{\mathbf{s}}_i\big)\\ &= \mathrm{span}(\mathbf{E}_{i|j}\mathbf{q}_i).\end{aligned}$$

The matrix $\mathbf{B}_i^\top\boldsymbol{\Sigma}_i^{-1}$ is invertible, so the coordinates of $\mathbf{q}_i$ form a basis of $\mathcal{L}_i$. View-wise coverage selects every basis coordinate on at least one incident edge, proving the final statement. □

This lemma establishes the geometry, not first-order invariance. For a non-Gaussian additive model, Assumption 6 must still be verified for its induced source-pair laws.

E.2 Finite-dimensional recovery

Proposition 4 (Pairwise recovery under a general spectral gap)**.** *Under Assumption 6, suppose that an active edge satisfies*

$$t_{ij,r_{ij}}>\eta_{ij}. \tag{34}$$

At representation dimension r_{ij} on each side, the maximum population CCA value is $\sum_{k=1}^{r_{ij}}t_{ij,k}$. Every maximizing centered, whitened pair has coordinate spans $\mathcal{L}_{i|j}$ and $\mathcal{L}_{j|i}$, respectively.

Proof. First-order invariance makes $\mathcal{T}_{ij}$ block diagonal relative to $\mathcal{L}_j\oplus\mathcal{L}_j^\perp$ and $\mathcal{L}_i\oplus\mathcal{L}_i^\perp$. Its first-order block has exactly r_{ij} positive singular values, namely $t_{ij,1},\dots,t_{ij,r_{ij}}$, and the other block has operator norm η_{ij}. Thus Equation 34 isolates these r_{ij} modes strictly above the remainder. The variational characterization of the sum of the leading r_{ij} singular values gives the stated maximum. Its strict-gap equality case forces the two feature spans to be the corresponding left and right singular subspaces, which are precisely $\mathcal{L}_{i|j}$ and $\mathcal{L}_{j|i}$. □

Theorem E.1 (Recovery beyond Gaussian sources under global first-order dominance)**.** *Assume Assumption 6 and $\alpha_i > \beta_i^{\perp}$ for every view. For centered, whitened encoders with $d_{\mathcal{Z}_i} = m_i$, the maximum of $J^{\mathcal{E}}_{\mathrm{SUMCOR}}$ is*

$$J_{\mathrm{aff}} := \sum_{\{i,j\}\in\mathcal{E}} \|\mathbf{R}_{ij}\|_* = \sum_{\{i,j\}\in\mathcal{E}} \sum_{k=1}^{r_{ij}} t_{ij,k}.$$

Every population maximizer satisfies

$$\tilde{\mathbf{h}}_i^{\star}(\mathbf{s}_i) = \mathbf{O}_i \tilde{\mathbf{s}}_i, \qquad \mathbf{O}_i \in O(m_i),$$

almost surely. This result does not require Equation 34.

Proof. Decompose an arbitrary feasible encoder into its first-order projection and orthogonal remainder:

$$\tilde{\mathbf{h}}_i = \mathbf{F}_i \tilde{\mathbf{s}}_i + \boldsymbol{\xi}_i, \qquad \mathbf{F}_i := \mathbb{E}[\tilde{\mathbf{h}}_i \tilde{\mathbf{s}}_i^{\top}].$$

Every coordinate of $\boldsymbol{\xi}_i$ belongs to $\mathcal{L}_i^{\perp}$. Whitening gives

$$\mathbf{F}_i \mathbf{F}_i^{\top} + \mathbb{E}[\boldsymbol{\xi}_i \boldsymbol{\xi}_i^{\top}] = \mathbf{I}_{m_i}, \qquad \ell_i := \mathbb{E}\|\boldsymbol{\xi}_i\|_2^2 = m_i - \|\mathbf{F}_i\|_F^2.$$

In particular, $\mathbf{D}_i := \mathbf{I}_{m_i} - \mathbf{F}_i^{\top}\mathbf{F}_i \succeq 0$ and $\mathrm{tr}(\mathbf{D}_i) = \ell_i$. First-order invariance eliminates the two mixed terms, yielding

$$\mathbb{E}[\tilde{\mathbf{h}}_i \tilde{\mathbf{h}}_j^{\top}] = \mathbf{F}_i \mathbf{R}_{ij} \mathbf{F}_j^{\top} + \mathbb{E}[\boldsymbol{\xi}_i \boldsymbol{\xi}_j^{\top}].$$

Factor a compact singular value decomposition of $\mathbf{R}_{ij}$ into two square-root factors. The inequality $\|\mathbf{X}\mathbf{Y}^{\top}\|_* \le \|\mathbf{X}\|_F \|\mathbf{Y}\|_F \le (\|\mathbf{X}\|_F^2 + \|\mathbf{Y}\|_F^2)/2$ gives

$$\begin{aligned}\|\mathbf{F}_i \mathbf{R}_{ij} \mathbf{F}_j^{\top}\|_* \le{}& \frac{1}{2}\,\mathrm{tr}\Big[\mathbf{F}_i (\mathbf{R}_{ij}\mathbf{R}_{ij}^{\top})^{1/2} \mathbf{F}_i^{\top}\Big] \\ &+ \frac{1}{2}\,\mathrm{tr}\Big[\mathbf{F}_j (\mathbf{R}_{ij}^{\top}\mathbf{R}_{ij})^{1/2} \mathbf{F}_j^{\top}\Big].\end{aligned}$$

For the remainder, let $\mathcal{U}_i : \mathbb{R}^{m_i} \to \mathcal{L}_i^{\perp}$ map $\mathbf{a}$ to $\mathbf{a}^{\top}\boldsymbol{\xi}_i$. Its squared Hilbert–Schmidt norm is ℓ_i, and

$$\mathbb{E}[\boldsymbol{\xi}_i \boldsymbol{\xi}_j^{\top}] = \mathcal{U}_i^{*}(I - \Pi_i)\mathcal{T}_{ij}(I - \Pi_j)\mathcal{U}_j.$$

The trace-norm product inequality therefore gives

$$\big\|\mathbb{E}[\boldsymbol{\xi}_i \boldsymbol{\xi}_j^{\top}]\big\|_* \le \eta_{ij}\sqrt{\ell_i \ell_j} \le \frac{\eta_{ij}}{2}(\ell_i + \ell_j).$$

Summing the two bounds over edges yields

$$\begin{aligned}J^{\mathcal{E}}_{\mathrm{SUMCOR}} &\le \frac{1}{2}\sum_i \big\{\mathrm{tr}(\mathbf{F}_i \mathbf{G}_i \mathbf{F}_i^{\top}) + \beta_i^{\perp}\ell_i\big\} \\ &\le J_{\mathrm{aff}} - \frac{1}{2}\sum_i (\alpha_i - \beta_i^{\perp})\ell_i.\end{aligned}$$

For the second inequality, use $\mathrm{tr}(\mathbf{G}_i\mathbf{D}_i) \ge \alpha_i \,\mathrm{tr}(\mathbf{D}_i)$ and $\frac{1}{2}\sum_i \mathrm{tr}(\mathbf{G}_i) = J_{\mathrm{aff}}$. The encoders $\tilde{\mathbf{h}}_i = \tilde{\mathbf{s}}_i$ attain J_{aff}. Strict global first-order dominance therefore forces $\ell_i = 0$ at every maximizer. Whitening then gives $\mathbf{F}_i\mathbf{F}_i^{\top} = \mathbf{I}_{m_i}$, so $\mathbf{F}_i$ is orthogonal, proving the conclusion. □

Theorem E.2 (Masked SUMCOR recovery beyond Gaussian sources)**.** *Assume Assumption 6, linear coverage Equation 32, simultaneous coordinate realizability Equation 33, and the pairwise gaps Equation 34 on all active edges. Use masked SUMCOR in Equation 5 with known ranks r_{ij}, raw encoder dimension $d_{\mathcal{Z}_i} = m_i$, and nonsingular encoder covariance. Then the population maximum is J_{aff}, and every maximizing encoder–mask collection satisfies*

$$\mathbf{h}_i^{\star}(\mathbf{s}_i) = \mathbf{T}_i^{\star}\mathbf{s}_i + \mathbf{b}_i, \qquad \mathbf{T}_i^{\star} \in \mathrm{GL}(m_i),$$

almost surely. No global first-order dominance condition is required.

Proof. By Proposition 4, each routed edge value is bounded above by $\sum_{k=1}^{r_{ij}} t_{ij,k}$. The common linear representations and coordinate masks from Equation 33 attain every bound simultaneously after whitening the selected blocks. They are feasible because each common representation has nonsingular covariance. Hence the maximum is J_{aff}. At any global maximizer, every edge must attain its bound. The equality case of Proposition 4 implies that the span of the centered raw encoder coordinates contains $\mathcal{L}_{i|j}$ for every neighbor j. By coverage, it contains all of $\mathcal{L}_i$. Both spaces have dimension m_i, so they coincide. Consequently the centered encoder is an invertible linear transformation of the centered source, proving the affine claim. The raw encoder is not required to be globally whitened, and the argument does not identify a unique mask support. □

E.3 Full-spectrum recovery under a trace-class condition

Theorem E.3 (Full-spectrum preservation beyond Gaussian sources)**.** *Assume Assumption 6, linear coverage Equation 32, and*

$$M_{ij} := \|\mathcal{T}_{ij}\|_* = \sum_{k\geq 1} \sigma_k(\mathcal{T}_{ij}) < \infty \qquad \textit{for every active edge.}$$

Represent each view by a finite or countable orthonormal feature family with closed span $\mathcal{H}_i \subseteq \mathcal{F}_i$. Let P_i denote the orthogonal projection onto $\mathcal{H}_i$, and maximize

$$J^{\mathcal{E}}_{\mathrm{full}} := \sum_{\{i,j\}\in\mathcal{E}} \|\mathsf{P}_i \mathcal{T}_{ij} \mathsf{P}_j\|_*.$$

The maximum is $\sum_{\{i,j\}\in\mathcal{E}} M_{ij}$. Every maximizing collection satisfies $\mathcal{L}_i \subseteq \mathcal{H}_i$ for every view, without either dominance condition. Equivalently, for an orthonormal feature family $\tilde{\mathbf{h}}_i$ spanning $\mathcal{H}_i$, there is a linear readout with square-summable rows such that

$$\mathbf{L}_i \tilde{\mathbf{h}}_i = \tilde{\mathbf{s}}_i, \qquad \mathbf{L}_i \mathbf{L}_i^\top = \mathbf{I}_{m_i},$$

where infinite coordinate expansions converge in L^2.

Proof. Compression by orthogonal projections cannot increase the trace norm, so each edge is bounded by M_{ij}. Choosing complete orthonormal bases of the centered source spaces attains all bounds simultaneously. To characterize equality, take a singular value decomposition of an active coupling,

$$\mathcal{T}_{ij} = \sum_{k:\,\sigma_k>0} \sigma_k\, u_k \langle v_k, \cdot\rangle,$$

which converges in trace norm. Then

$$\begin{aligned} \|\mathsf{P}_i \mathcal{T}_{ij} \mathsf{P}_j\|_* &\leq \sum_k \sigma_k \|\mathsf{P}_i u_k\| \|\mathsf{P}_j v_k\| \\ &\leq \sum_k \sigma_k = M_{ij}. \end{aligned}$$

If equality holds, every positive-weight product in the last sum must be one. Since orthogonal projections are contractions, $u_k \in \mathcal{H}_i$ and $v_k \in \mathcal{H}_j$ for every $\sigma_k > 0$. A global maximizer saturates every edge. Hence its feature spans contain the closures of the corresponding operator ranges. First-order invariance implies that $\mathcal{L}_{i|j} \subseteq \mathrm{ran}(\mathcal{T}_{ij})$, because it is the range of the nonzero first-order block. Coverage now yields $\mathcal{L}_i \subseteq \mathcal{H}_i$. Expanding the whitened source coordinates in an orthonormal basis of $\mathcal{H}_i$ gives the readout, and Parseval's identity gives $\mathbf{L}_i \mathbf{L}_i^\top = \mathbf{I}_{m_i}$. □

In orthonormal feature coordinates, the compressed operator has the same nonzero singular values as the feature cross-covariance operator. Thus $J^{\mathcal{E}}_{\mathrm{full}}$ extends SUMCOR to countable feature families. Allowing finite feature families covers finite-support source laws. This result preserves the first-order space inside the learned feature span. It does not make the whole feature family affine, nor establish a sufficient finite network width or convergence of trained finite-dimensional encoders.

E.4 LANCASTER SPECIALIZATION AND INTERPRETATION

A sufficient Lancaster structure. Suppose that the actual source-pair law admits a complete orthogonal-polynomial Lancaster diagonalization (Lancaster, 1958; Eagleson, 1964). In centered operator form, write

$$\mathcal{T}_{ij} f = \sum_{\nu \in \mathcal{A}_{ij}} \lambda_{ij,\nu}\, \varphi_{ij,\nu} \langle \psi_{ij,\nu}, f \rangle_{\mathcal{F}_j}. \tag{35}$$

The two displayed systems are orthonormal, the expansion converges in L^2 for every f, and they can be completed by bases of the left and right kernels. The constant mode is excluded. Require compatibility with the first-order spaces: the nonzero coupled modes can be partitioned into $\mathcal{A}_{ij}^{(1)}$ and $\mathcal{A}_{ij}^{(\perp)}$ so that

$$\operatorname{span}\{\varphi_{ij,\nu} : \nu \in \mathcal{A}_{ij}^{(1)}\} = \mathcal{L}_{i|j},$$
$$\operatorname{span}\{\psi_{ij,\nu} : \nu \in \mathcal{A}_{ij}^{(1)}\} = \mathcal{L}_{j|i},$$

while every mode indexed by $\mathcal{A}_{ij}^{(\perp)}$ lies in $\mathcal{L}_i^{\perp}$ and $\mathcal{L}_j^{\perp}$, respectively. Uncoupled first-order directions belong to the corresponding operator kernels. For polynomial systems organized by degree, this is satisfied by a degree-compatible expansion: the coupled degree-one modes span the pairwise linear targets, and the remaining coupled modes have degree at least two.

Proposition 5 (Lancaster specialization)**.** *Under the preceding diagonalization and compatibility conditions, Assumption 6 holds and*

$$\eta_{ij} = \sup_{\nu \in \mathcal{A}_{ij}^{(\perp)}} |\lambda_{ij,\nu}|, \tag{36}$$

with value zero when the index set is empty. The positive first-order singular values are the absolute coefficients indexed by $\mathcal{A}_{ij}^{(1)}$. Consequently, the finite-dimensional results above apply whenever their respective spectral and geometric conditions hold. The full-spectrum result additionally applies if

$$\sum_{\nu \in \mathcal{A}_{ij}} |\lambda_{ij,\nu}| < \infty \qquad \textit{on every active edge},$$

in which case this sum equals M_{ij}.

Proof. Projecting Equation 35 onto the first-order spaces or their orthogonal complements retains exactly the corresponding index block. Both mixed blocks vanish, which proves first-order invariance. Orthonormality makes the singular values in either block the absolute values of its diagonal coefficients, giving Equation 36 and the trace-class criterion. The recovery claims then follow from Proposition 4, Theorem E.1, Theorem E.2, and Theorem E.3. □

Gaussian specialization. For a Gaussian source pair, the Hermite–Mehler expansion gives

$$\lambda_{ij,\mathbf{n}} = \prod_{k=1}^{r_{ij}} t_{ij,k}^{n_k}, \qquad \mathbf{n} \in \mathbb{N}_0^{r_{ij}} \setminus \{\mathbf{0}\}.$$

The degree-one indices are the first-order modes. Modes involving private canonical coordinates are uncoupled. The largest higher-order coefficient is attained by squaring a leading canonical coordinate, so

$$\eta_{ij} = t_{ij,1}^2, \qquad M_{ij} = \prod_{k=1}^{r_{ij}} (1 - t_{ij,k})^{-1} - 1.$$

Together with Lemma 5, this recovers the Gaussian pairwise, global, and masked guarantees. Finite-overcomplete recovery additionally requires joint spectral compatibility (Theorem 4.3). Gaussian drift remains an existence counterexample here, without asserting drift for every non-Gaussian coupling.

Scope for non-Gaussian sources. Appropriate Lancaster couplings can use Charlier, Meixner, generalized Laguerre, or Hahn systems in place of Hermite polynomials (Han et al., 2026a). The relevant requirement is a compatible diagonalization of the *joint source-pair law*. Specifying Poisson, negative binomial, Gamma, or hypergeometric marginals alone does not verify it. In particular, a general Lancaster sequence need not obey the Gaussian identity $\eta_{ij} = t_{ij,1}^2$. Invertible affine reparameterizations of a complete source vector preserve first-order invariance: composition is a unitary map between the corresponding L^2 spaces and maps the first-order spaces onto each other. This observation does not justify adding arbitrary noise after a mixing transformation and assuming that the induced coupling retains the same structure. Finally, separately specified pairwise Lancaster laws must arise from a common multi-view joint law, and coordinate realizability must be verified independently of their pairwise diagonalizations. The following proposition verifies these conditions for the four three-dimensional non-Gaussian Setup A constructions. Recovery scores alone neither certify these population conditions nor establish a claim over all Lancaster laws.

E.5 Spectral verification for non-Gaussian Setup A

We verify the operator conditions for the four three-dimensional joint constructions in Appendix F.1. This verification uses their shared draws and conditional sampling rules, rather than their marginal distribution families alone.

Proposition 6 (Three-dimensional non-Gaussian Setup A)**.** *For each of the three-dimensional Gamma, Poisson, negative-binomial, and hypergeometric constructions in Appendix F.1, first-order invariance Assumption 6, linear coverage Equation 32, and simultaneous coordinate realizability Equation 33 hold. Every edge has rank two with $t_{ij,1} = t_{ij,2} = 1/4$, and every view has $\alpha_i = 1/4$. The exact remainder norms and aggregate gaps are given in Table 4. In particular, all four constructions satisfy*

$$t_{ij,2} > \eta_{ij}, \qquad \alpha_i > \beta_i^{\perp} = \sum_{j \in \mathcal{N}_i} \eta_{ij}.$$

Thus Theorem E.1 applies at $d_{\mathcal{Z}_i} = m_i = 3$: every centered, whitened population SUMCOR maximizer has source-space representations $\tilde{\mathbf{h}}_i^{\star}(\mathbf{s}_i) = \mathbf{O}_i \mathbf{s}_i$ almost surely, where $\mathbf{O}_i \in O(3)$.

Table 4: Exact operator constants for the specified non-Gaussian Setup A constructions. Each view has two neighbors, so $\beta_i^{\perp} = 2\eta_{ij}$.

Construction	$t_{ij,1} = t_{ij,2}$	η_{ij}	α_i	$\beta_i^{\perp}$	$\alpha_i - \beta_i^{\perp}$
Gamma	$1/4$	$1/10$	$1/4$	$1/5$	$1/20$
Poisson	$1/4$	$1/16$	$1/4$	$1/8$	$1/8$
Negative binomial	$1/4$	$1/10$	$1/4$	$1/5$	$1/20$
Hypergeometric	$1/4$	$1/16$	$1/4$	$1/8$	$1/8$

Proof. Scalar additive couplings. For one shared group in the first three constructions, write $X = C + A$ and $Y = C + B$, where C, A, B are independent and have parameters $1, 3, 3$, respectively, with the common rate or success parameter specified in Appendix F.1. The pair is exchangeable. Hence its conditional expectation operator $Tf(x) = \mathbb{E}[f(Y) \mid X = x]$ is self-adjoint on the common marginal L^2 space. For Gamma variables, $C/X \mid X \sim \mathrm{Beta}(1, 3)$. For Poisson variables, $C \mid X = x \sim \mathrm{Binom}(x, 1/4)$. For negative-binomial variables, $C \mid X = x$ is beta-binomial with trial count x and shape parameters $(1, 3)$. Their ordinary or falling-factorial conditional moments show that $\mathbb{E}[C^n \mid X = x]$ is a polynomial of degree at most n with leading coefficient

$$\lambda_n = \begin{cases} (1)_n/(4)_n, & \text{Gamma or negative binomial,} \\ 4^{-n}, & \text{Poisson,} \end{cases} \qquad n \geq 0,$$

where $(b)_n = b(b+1)\cdots(b+n-1)$ and $(b)_0 = 1$. Explicitly, the negative-binomial and Poisson conditional falling-factorial moments are $(x)_{\underline{n}}(1)_n/(4)_n$ and $(x)_{\underline{n}}4^{-n}$, respectively, where $(x)_{\underline{n}} = x(x-1)\cdots(x-n+1)$. Independence of B gives the same leading coefficient for $\mathbb{E}[(C+B)^n \mid X = x]$, since every term involving a positive power of B has conditional degree less than n. Thus T preserves every polynomial space of degree at most n.

Let p_n be its normalized degree-n orthogonal polynomial, with $p_0 = 1$. For every polynomial q of degree less than n,

$$\langle Tp_n, q\rangle = \langle p_n, Tq\rangle = 0.$$

Degree preservation and the leading coefficient calculation imply $Tp_n = \lambda_n p_n$. These polynomials are complete in the marginal L^2 space. Indeed, all three marginals have a finite positive exponential moment. If $f \in L^2$ is orthogonal to every polynomial, then $F(z) = \mathbb{E}[f(X)e^{zX}]$ is analytic on a strip containing the imaginary axis, by Cauchy–Schwarz and that exponential moment. All derivatives at zero vanish. The identity theorem and uniqueness of the Fourier transform of the finite signed measure $f\,dP_X$ yield $f = 0$ almost surely. Consequently this diagonalization controls all square-integrable functions, not only a finite polynomial truncation. The nonconstant eigenvalues are positive and decreasing, with

$$\lambda_1 = \frac{1}{4}, \qquad \lambda_2 = \begin{cases} 1/10, & \text{Gamma or negative binomial,} \\ 1/16, & \text{Poisson.} \end{cases}$$

Scalar hypergeometric coupling. Write π for $\mathrm{Hypergeom}(20, 15, 10)$ on $\{5, \dots, 10\}$ and K for the one-step kernel specified in Appendix F.1. For $a = 1 - 2^{-1/3}$, its probabilities are nonnegative:

$$p_+(x) + p_-(x) = \frac{a}{20}(2x^2 - 30x + 150) \le \frac{5a}{2} < 1.$$

The outward probabilities vanish at the two endpoints. Moreover,

$$\frac{\pi(x+1)}{\pi(x)} = \frac{(15-x)(10-x)}{(x+1)(x-4)} = \frac{p_+(x)}{p_-(x+1)}, \qquad 5 \le x < 10.$$

Thus K is reversible and preserves π. Expanding Kx^n shows that the degree-$(n+1)$ terms cancel and its degree-n coefficient is

$$\kappa_n = 1 - \frac{a}{20}n(21 - n), \qquad n = 0, \dots, 5.$$

The same self-adjointness and degree argument diagonalizes K in the orthogonal polynomials for π. Polynomials of degrees zero through five span all functions on its six-point support, so this eigenbasis is complete. The two outputs are conditionally independent three-step branches from a common $H \sim \pi$. Their pair law is

$$\begin{aligned} \mathbb{P}(X = x, Y = y) &= \sum_h \pi(h)K^3(h, x)K^3(h, y) \\ &= \pi(x)K^6(x, y), \end{aligned}$$

where the second equality uses reversibility and composition. Their conditional expectation operator is therefore K^6, with eigenvalues $\lambda_n = \kappa_n^6$. Since $1/10 < a < 1/4$, the values κ_n are positive and strictly decreasing for $n = 0, \dots, 5$. Also $\kappa_1 = 1 - a = 2^{-1/3}$, so $\lambda_1 = 1/4$. Finally,

$$\kappa_1^2 - \kappa_2 = a(a - 1/10) > 0, \qquad \lambda_2 < \lambda_1^2 = 1/16.$$

For all four scalar laws, centering and scaling by the stated population moments preserve this complete diagonalization and make p_1 the standardized coordinate.

The full three-dimensional operators. The independent groups $12, 13, 23, 0$ specify one joint law for all three views. Each view contains three independent standardized coordinates, so $\mathrm{Cov}(\mathbf{s}_i) = \mathbf{I}_3$. Each edge shares exactly two groups. Its other coordinate in each view is independent of the entire opposite view. Product orthogonal polynomials form complete bases of the three-dimensional marginal L^2 spaces. On these bases, conditional expectation is the tensor product of the two shared scalar operators and expectation over the coordinate absent from the opposite view. Every mode with positive degree in that absent coordinate is annihilated. The remaining nonconstant coefficients are exactly

$$\lambda_m\lambda_n, \qquad m, n \ge 0, \quad m + n \ge 1,$$

with $m, n \le 5$ in the hypergeometric case. The two coupled first-order modes have indices $(1, 0)$ and $(0, 1)$ and coefficient $1/4$. The third first-order direction is in the kernel. All other nonconstant

modes are orthogonal to the first-order space on both sides. This proves $\Pi_i \mathcal{T}_{ij} = \mathcal{T}_{ij} \Pi_j$. Positivity and monotonicity of the scalar coefficients give

$$\eta_{ij} = \sup_{m+n \geq 2} \lambda_m \lambda_n = \max\{\lambda_2, \lambda_1^2\}.$$

The product of two first-order functions is included in this remainder. In particular, it determines the hypergeometric remainder norm even though its scalar degree-two coefficient is smaller than $1/16$.

Coverage, routing, and the gaps. In each view's group coordinates, $(\mathbf{R}_{ij}\mathbf{R}_{ij}^\top)^{1/2}$ is $1/4$ times the coordinate projector onto the two groups shared with j. Each pair-specific group occurs on one incident edge, and group 0 occurs on both. Hence

$$\mathbf{G}_i = \mathrm{diag}(1/4, 1/4, 1/2), \qquad \alpha_i = 1/4,$$

up to coordinate ordering. Every source coordinate is covered by an incident edge. Taking $\mathbf{q}_i = \mathbf{s}_i$ and selecting the shared coordinates supplies a common representation for all masks, proving coverage and simultaneous coordinate realizability. The computed η_{ij} and the two neighbors per view give Table 4. Both gaps are strict, so Theorem E.1 proves the stated population recovery conclusion. □

This verification concerns the specified joint source laws and population global maximizers over square-integrable encoders. It does not establish that finite-sample neural training attains such a maximizer. For Gamma and negative binomial sources, $\eta_{ij} = 1/10 \neq t_{ij,1}^2 = 1/16$, so their verification specifically requires the non-Gaussian operator conditions.

F Experimental Details and Additional Results

F.1 Data-generating models

Setup A: three-dimensional sources. Write the shared groups as $(c_{12}, c_{13}, c_{23}, c_0)$. Views observe (c_{12}, c_{13}, c_0), (c_{12}, c_{23}, c_0), and (c_{13}, c_{23}, c_0), respectively. For the Gaussian experiment, these four variables are independent standard normal variables. The source model is

$$\mathbf{s}_i = \mathbf{M}_i \mathbf{P}_i \mathbf{c} + \boldsymbol{\epsilon}_i, \qquad \boldsymbol{\epsilon}_i \sim \mathcal{N}(0, \boldsymbol{\Psi}_i), \quad \boldsymbol{\Psi}_i = \mathbf{I}_3 - \mathbf{M}_i \mathbf{M}_i^\top,$$

with independent view noises and

$$\mathbf{M}_1 = \begin{pmatrix} .5 & .1 & 0 \\ 0 & .5 & 0 \\ 0 & 0 & .5 \end{pmatrix}, \quad \mathbf{M}_2 = \begin{pmatrix} .5 & .05 & 0 \\ .05 & .5 & 0 \\ 0 & 0 & .5 \end{pmatrix}, \quad \mathbf{M}_3 = \begin{pmatrix} .5 & -.25 & 0 \\ .25 & .5 & 0 \\ 0 & 0 & .5 \end{pmatrix}.$$

Each source has covariance $\mathbf{I}_3$. The noise covariances are positive definite. Applying $\boldsymbol{\Psi}_i^{-1/2}$ to the source converts this Gaussian model to isotropic noise while preserving affine recoverability under a composed observation map. The archived observation maps use three invertible linear layers with leaky-ReLU activations of negative slope 0.2 between layers. These maps are piecewise linear rather than globally smooth. They satisfy the measurable-invertibility conditions in Appendix E.1, which supply the observation-space reduction for the operator analysis.

For Gamma, Poisson, and negative binomial sources, each group has an independent shared draw with parameter one and each visible coordinate adds an independent private draw with parameter three. Gamma uses unit rate, Poisson uses these parameters as rates, and negative binomial uses success parameter 0.5 in the sampler's parameterization. The sums have mean four and variances four, four, and eight, respectively. We standardize each sum using these population moments. For the hypergeometric construction, each group starts from an independent stationary $\mathrm{Hypergeom}(20, 15, 10)$ draw. Each view runs three independent transition steps from that draw, with

$$\begin{aligned} a &= 1 - 2^{-1/3}, \\ p_+(x) &= a(10 - x)(15 - x)/20, \\ p_-(x) &= ax(x - 5)/20. \end{aligned}$$

The remaining probability leaves the state unchanged. The resulting coordinate is standardized by subtracting 7.5 and dividing by $\sqrt{75/76}$. All four constructions have unit marginal covariance and pairwise nonzero correlations $(1/4, 1/4)$. They have no additional linear mixing or Gaussian perturbation. Their observation networks have the same architecture as the Gaussian Setup A maps. The induced source-pair laws satisfy first-order invariance and the generalized pairwise and global dominance conditions by Proposition 6. This verification concerns the stated joint constructions, not their marginal distribution families.

Setups B and C. Both setups instantiate Equation 1 with $\mathbf{c} \sim \mathcal{N}(0, \mathbf{I}_4)$ and mutually independent $\boldsymbol{\epsilon}_i \sim \mathcal{N}(0, \mathbf{I}_3)$, independent of $\mathbf{c}$. The visibility sets are $\mathcal{V}_1 = \{1, 2, 3\}$, $\mathcal{V}_2 = \{1, 2, 4\}$, and $\mathcal{V}_3 = \{1, 3, 4\}$. We set $\mathbf{B}_i = \mathbf{Q}_i \mathbf{D}(\rho, \tau)$, where

$$\mathbf{D}(\rho, \tau) = \mathrm{diag}\left(\sqrt{\frac{\rho}{1 - \rho}}, \sqrt{\frac{\tau}{1 - \tau}}, \sqrt{\frac{\tau}{1 - \tau}} \right).$$

The fixed dense orthogonal matrices are

$$\mathbf{Q}_1 = \tfrac{1}{9} \begin{pmatrix} 7 & -4 & -4 \\ -4 & 1 & -8 \\ -4 & -8 & 1 \end{pmatrix}, \quad \mathbf{Q}_2 = \tfrac{1}{9} \begin{pmatrix} 1 & -4 & -8 \\ -4 & 7 & -4 \\ -8 & -4 & 1 \end{pmatrix},$$
$$\mathbf{Q}_3 = \tfrac{1}{9} \begin{pmatrix} 1 & -8 & -4 \\ -8 & 1 & -4 \\ -4 & -4 & 7 \end{pmatrix}.$$

Every edge has canonical spectrum $(\rho, \tau, 0)$ and overlap rank two. The pairwise and global first-order dominance gaps are $\tau - \rho^2$ and $\tau - 2\rho^2$, respectively. Setup B uses $(\rho, \tau) = (0.6, 0.45)$, giving gaps

0.09 and -0.27. Setup C uses $(0.6, 0.2)$, giving -0.16 and -0.52. Both retain view-wise coverage. The actual Setup A generators differ from B/C, so comparisons across A/B/C are not a sweep of these two parameters alone.

The observation map is

$$\mathbf{g}_i(\mathbf{s}) = \mathbf{V}_i\phi(\mathbf{U}_i\mathbf{s}), \qquad \phi(t) = t + 0.5\sin(t),$$

where ϕ acts coordinatewise and $\mathbf{U}_i, \mathbf{V}_i$ are seed-dependent random orthogonal matrices. Since $\phi'(t) \in [0.5, 1.5]$, these maps are smooth and invertible.

F.2 Encoders, objectives, and training protocol

The three residual encoders have hidden widths $(30, 60, 60, 60, 60)$ with batch normalization and output width $d_{\mathcal{Z}}$. The two-dimensional visualizations use widths $(20, 40, 40, 40, 40)$ and $d_{\mathcal{Z}} = 2$. SUMCOR, tSUMCOR, and masked SUMCOR use double-precision covariance arithmetic, ridge 10^{-8}, and Cholesky whitening. The SUMCOR loss is the negative average of the edgewise nuclear norms, while reported J values sum the three edge scores.

The masked SUMCOR implementation constructs a catalog of all size-r_{ij} coordinate supports at each endpoint, with $r_{ij} = 2$ in our experiments. The overlap ranks are supplied, while coordinate supports are selected. For encoder dimensions $d_{\mathcal{Z}_i}$ and $d_{\mathcal{Z}_j}$, this gives $\binom{d_{\mathcal{Z}_i}}{r_{ij}}\binom{d_{\mathcal{Z}_j}}{r_{ij}}$ support pairs per edge. Thus, equal output widths 3, 4, and 6 give 9, 36, and 225 candidates, respectively. At each training step, it centers the raw encoder outputs and extracts the within-view and cross-view covariance submatrices for every candidate. Each selected block is whitened separately using the regularized covariance, and each support pair is scored by the sum of singular values of its whitened cross-covariance. Each edge independently selects the highest-scoring pair, with ties resolved by lexicographic order. Autograd propagates through that pair's score, including its whitening, while holding the selected indices fixed. Adam updates the encoder parameters. There are no trainable mask parameters, and no continuous mask relaxation is used. Supports on different incident edges may overlap. Exhaustive enumeration solves the support-maximization subproblem exactly for the regularized empirical CCA score on each minibatch. The selected supports are shared across samples within that batch. At the population level, maximizing out the finite mask variables is equivalent to joint encoder–mask optimization (Lemma 2). This equivalence does not equate expected minibatch maxima with the population profile (Appendix C.1). Finite-batch estimation, regularized whitening, training-mode batch normalization, and nonconvex encoder optimization remain outside the population recovery guarantee. We do not claim convergence to a population optimum. In the reported evaluations, supports are reselected on every batch, so the routed objective diagnostic is a batch-profiled score. Evaluation does not include independent calibration followed by fixed-support testing. Support uniqueness is not required by the recovery theorem.

DGCCA uses output rank $d_{\mathcal{Z}}$. The rDGCCA implementation adds cross-reconstruction decoders with two hidden layers of width 30 and reconstruction coefficient 0.5. SimCLR uses squared Euclidean distance with temperature 0.1. W-MSE applies whole-batch ZCA whitening with ridge 10^{-7}, followed by sample-wise unit normalization, and minimizes the mean squared Euclidean distance between paired samples. All methods use the same encoder widths within a setup, but their objectives and, for rDGCCA, auxiliary decoders differ.

All quantitative runs use 50,000 Adam updates with learning rate 10^{-4} and fresh training batches of size 8192. Only completed runs with a final ten-batch evaluation are included. Each entry reports one trained model. The per-view tables list its training seed. Seeds for Setup B mSUMCOR at dimension four and Setup C rank-two runs at dimension six were selected using intermediate recovery diagnostics. The seed controls initialization, sampling, and observation maps. Setup A evaluates with batch-normalization batch statistics. B/C use frozen batch-normalization statistics and separate random streams for training, intermediate diagnostics, and final evaluation. Thus Setup A and B/C also differ in evaluation normalization.

F.3 Recovery metrics and aggregation

On each evaluation batch and view, an affine least-squares readout predicts the noisy source from the full learned representation. For source coordinate k, the score is

$$R^2_{i,k} = 1 - \frac{\sum_{n=1}^{4096}(s^{(n)}_{i,k} - \hat{s}^{(n)}_{i,k})^2}{\sum_{n=1}^{4096}(s^{(n)}_{i,k} - \bar{s}_{i,k})^2}.$$

The view score averages over its three source coordinates. The main tables average these scores over views and then over ten batches. The per-view standard deviations below use divisor ten. These are batch standard deviations conditional on one trained model. They are not standard errors, confidence intervals, or training-seed variation. Setup A retains final per-view means and standard deviations but not the ten individual scores. Its displayed spread is the root mean square of the three within-view batch standard deviations, $\sqrt{(\sigma_1^2 + \sigma_2^2 + \sigma_3^2)/3}$. This is not the standard deviation of the view-average score, which cannot be recovered without the per-batch scores. For B/C, the displayed spread is the standard deviation of the ten batchwise view-average scores.

For B/C, let $\mathbf{U}_s$ and $\mathbf{U}_z$ contain orthonormal bases of the centered source and full representation sample columns. The singular values of $\mathbf{U}_s^\top \mathbf{U}_z$ give the canonical correlations used in Table 3. Missing directions receive zero correlation. High average R^2 and a minimum correlation close to one jointly support retention of all source directions. The metric does not require an enlarged representation to consist entirely of first-order features. Both R^2 and these source–representation correlations use the complete encoder output, independently of the selected masks. They remain same-batch span diagnostics, not held-out readout tests. Empirical objective values need not satisfy population bounds, particularly with finite-batch covariance estimation and batchwise support maximization.

F.4 Complete per-view quantitative results

The following tables show 80 completed quantitative results (35 in Setup A, 18 in B, and 27 in C). SUMCOR uses the nuclear-norm objective in all setups. No intermediate checkpoints are included. The $d_{\mathcal{Z}} = 4$ and $d_{\mathcal{Z}} = 6$ experiments increase representation dimension while retaining a three-dimensional source. They do not vary the source dimension or establish a universal sufficient width.

Table 5: Setup A: per-view source recovery R^2 (%), mean $\pm$ population SD over ten evaluation batches of one trained model. SDs are percentage points. All sources have dimension three and all runs complete $50{,}000$ updates. Errors measure batch variability, not uncertainty across training seeds.

Distribution	Method	Seed	View 1 (%)	View 2 (%)	View 3 (%)
Gaussian	BT	2024	$43.21 \pm .35$	$36.40 \pm .29$	$34.07 \pm .31$
	VICReg	2024	$37.21 \pm .29$	$47.87 \pm .34$	35.56 ± 1.63
	W-MSE	2024	$\mathbf{99.66 \pm .02}$	$\mathbf{99.40 \pm .10}$	$\mathbf{99.47 \pm .05}$
	SimCLR	2024	$89.52 \pm .15$	$77.95 \pm .27$	$68.05 \pm .37$
	DGCCA	2024	$\mathbf{99.74 \pm .02}$	$\mathbf{99.66 \pm .07}$	$\mathbf{99.75 \pm .05}$
	rDGCCA	2024	$\mathbf{99.33 \pm .04}$	$98.40 \pm .09$	$\mathbf{99.78 \pm .02}$
	SUMCOR	2024	$\mathbf{99.71 \pm .03}$	$\mathbf{99.67 \pm .05}$	$\mathbf{99.68 \pm .03}$
Gamma	BT	2024	35.52 ± 1.33	$38.40 \pm .83$	$41.33 \pm .42$
	VICReg	2024	$35.25 \pm .77$	$47.35 \pm .37$	36.42 ± 1.61
	W-MSE	2024	$\mathbf{99.66 \pm .03}$	$\mathbf{99.40 \pm .48}$	$\mathbf{99.59 \pm .06}$
	SimCLR	2024	$83.66 \pm .23$	$70.39 \pm .33$	$65.76 \pm .25$
	DGCCA	2024	$\mathbf{99.83 \pm .02}$	$81.10 \pm .22$	$71.66 \pm .41$
	rDGCCA	2024	$96.03 \pm .25$	$94.86 \pm .51$	$\mathbf{99.72 \pm .04}$
	SUMCOR	2024	$\mathbf{99.79 \pm .02}$	$\mathbf{99.80 \pm .03}$	$\mathbf{99.77 \pm .03}$
Poisson	BT	2024	$35.21 \pm .37$	$45.66 \pm .26$	$33.73 \pm .45$
	VICReg	2024	$35.28 \pm .67$	$35.56 \pm .16$	41.01 ± 1.35
	W-MSE	2024	$\mathbf{99.39 \pm .04}$	$\mathbf{99.13 \pm .06}$	$\mathbf{99.48 \pm .03}$
	SimCLR	2024	$83.62 \pm .32$	$77.59 \pm .29$	$66.04 \pm .22$
	DGCCA	2024	$\mathbf{99.61 \pm .04}$	$83.28 \pm .26$	$81.77 \pm .37$
	rDGCCA	2024	$\mathbf{99.73 \pm .06}$	$97.34 \pm .10$	$95.67 \pm .22$
	SUMCOR	2024	$\mathbf{99.49 \pm .04}$	$\mathbf{99.14 \pm .07}$	$\mathbf{99.68 \pm .04}$

Table 6: Setup A: per-view source recovery R^2 (%), mean $\pm$ population SD over ten evaluation batches of one trained model. SDs are percentage points. All sources have dimension three and all runs complete $50{,}000$ updates. Errors measure batch variability, not uncertainty across training seeds.

Distribution	Method	Seed	View 1 (%)	View 2 (%)	View 3 (%)
Negative binomial	BT	2024	$33.57 \pm .25$	35.23 ± 1.20	$55.44 \pm .52$
	VICReg	2024	$36.75 \pm .67$	$46.30 \pm .28$	38.08 ± 1.34
	W-MSE	2024	$\mathbf{99.58 \pm .03}$	$\mathbf{99.49 \pm .05}$	$\mathbf{99.42 \pm .14}$
	SimCLR	2024	$81.81 \pm .54$	$73.99 \pm .32$	$65.40 \pm .28$
	DGCCA	2024	$\mathbf{99.85 \pm .01}$	$71.05 \pm .23$	$66.79 \pm .07$
	rDGCCA	2024	$79.78 \pm .42$	$96.48 \pm .30$	$\mathbf{99.71 \pm .04}$
	SUMCOR	2024	$\mathbf{99.78 \pm .02}$	$\mathbf{99.58 \pm .05}$	$\mathbf{99.74 \pm .04}$
Hypergeometric	BT	2024	$36.75 \pm .40$	$37.30 \pm .33$	$34.15 \pm .60$
	VICReg	2024	$33.97 \pm .39$	$33.58 \pm .09$	36.95 ± 1.19
	W-MSE	2024	$98.29 \pm .15$	$98.52 \pm .10$	$97.00 \pm .12$
	SimCLR	2024	$64.27 \pm .14$	$65.18 \pm .14$	50.53 ± 3.37
	DGCCA	2024	$\mathbf{99.55 \pm .07}$	$66.92 \pm .16$	$66.68 \pm .06$
	rDGCCA	2024	$\mathbf{99.65 \pm .04}$	$85.58 \pm .28$	$87.66 \pm .21$
	SUMCOR	2024	$\mathbf{99.42 \pm .06}$	$\mathbf{99.14 \pm .07}$	$\mathbf{99.42 \pm .06}$

Table 7: Setup B: per-view source recovery R^2 (%), mean $\pm$ population SD over ten batches of one trained model. SDs are percentage points. All sources have dimension three. All runs complete $50{,}000$ updates. Errors measure batch variability, not uncertainty across training seeds. Training details appear in Appendix F.2.

$d_{\mathcal{Z}}$	Method	Seed	View 1 (%)	View 2 (%)	View 3 (%)
3	BT	2024	$39.99 \pm .60$	$38.05 \pm .44$	$38.99 \pm .32$
	VICReg	2024	$50.95 \pm .67$	$37.63 \pm .48$	$57.56 \pm .22$
	W-MSE	2024	$70.56 \pm .55$	$67.53 \pm .35$	$67.45 \pm .22$
	SimCLR	2024	$95.19 \pm .13$	$84.61 \pm .18$	$81.34 \pm .76$
	DGCCA	2024	$42.56 \pm .49$	$69.74 \pm .56$	$67.14 \pm .25$
	rDGCCA	2024	$71.73 \pm .31$	$72.45 \pm .55$	$67.06 \pm .27$
	SUMCOR	2024	$70.64 \pm .57$	$67.40 \pm .30$	$67.37 \pm .30$
	tSUMCOR	2024	$\mathbf{99.94 \pm .00}$	$\mathbf{99.93 \pm .00}$	$\mathbf{99.93 \pm .01}$
	mSUMCOR	2024	$\mathbf{99.92 \pm .00}$	$\mathbf{99.94 \pm .00}$	$\mathbf{99.94 \pm .00}$
4	BT	2024	$43.23 \pm .50$	$45.48 \pm .53$	$38.35 \pm .25$
	VICReg	2024	$58.72 \pm .52$	$45.89 \pm .53$	$56.01 \pm .35$
	W-MSE	2024	$93.39 \pm .16$	$93.42 \pm .16$	$91.68 \pm .15$
	SimCLR	2024	$95.53 \pm .13$	$85.82 \pm .20$	$73.36 \pm .17$
	DGCCA	2024	$\mathbf{99.93 \pm .00}$	$66.98 \pm .55$	$69.78 \pm .33$
	rDGCCA	2024	$\mathbf{99.89 \pm .01}$	$68.44 \pm .46$	$93.61 \pm .20$
	SUMCOR	2024	$\mathbf{99.90 \pm .01}$	$\mathbf{99.93 \pm .00}$	$\mathbf{99.90 \pm .01}$
	tSUMCOR	2026	$\mathbf{99.96 \pm .00}$	$\mathbf{99.96 \pm .00}$	$\mathbf{99.96 \pm .00}$
	mSUMCOR	2027	$\mathbf{99.96 \pm .00}$	$\mathbf{99.96 \pm .00}$	$\mathbf{99.97 \pm .00}$

Table 8: Setup C, $d_{\mathcal{Z}} = 3$: per-view source recovery R^2 (%), mean $\pm$ population SD over ten evaluation batches of one trained model. SDs are percentage points. All sources have dimension three and all runs complete $50{,}000$ updates. Errors measure batch variability, not uncertainty across training seeds.

$d_{\mathcal{Z}}$	Method	Seed	View 1 (%)	View 2 (%)	View 3 (%)
3	BT	2024	$47.87 \pm .65$	$41.63 \pm .47$	$43.25 \pm .34$
	VICReg	2024	$54.91 \pm .63$	$41.79 \pm .56$	$55.01 \pm .22$
	W-MSE	2024	$66.62 \pm .50$	$41.17 \pm .50$	$65.21 \pm .68$
	SimCLR	2024	$90.86 \pm .15$	$78.58 \pm .33$	$75.25 \pm .31$
	DGCCA	2024	$47.24 \pm .61$	$41.15 \pm .50$	$41.28 \pm .33$
	rDGCCA	2024	$47.86 \pm .60$	$41.52 \pm .51$	$41.47 \pm .29$
	SUMCOR	2024	$47.22 \pm .61$	$41.15 \pm .49$	$41.25 \pm .32$
	tSUMCOR	2024	$47.27 \pm .62$	$41.24 \pm .45$	$41.33 \pm .32$
	mSUMCOR	2024	$47.22 \pm .62$	$42.52 \pm .45$	$41.26 \pm .33$

Table 9: Setup C, $d_{\mathcal{Z}} = 4$: per-view source recovery R^2 (%), mean $\pm$ population SD over ten evaluation batches of one trained model. SDs are percentage points. All sources have dimension three and all runs complete $50{,}000$ updates. Errors measure batch variability, not uncertainty across training seeds.

$d_{\mathcal{Z}}$	Method	Seed	View 1 (%)	View 2 (%)	View 3 (%)
4	BT	2024	$60.31 \pm .31$	$49.68 \pm .51$	$45.92 \pm .44$
	VICReg	2024	$51.59 \pm .57$	$43.92 \pm .50$	$51.55 \pm .47$
	W-MSE	2024	$82.41 \pm .33$	$81.88 \pm .79$	$78.14 \pm .90$
	SimCLR	2024	$91.25 \pm .15$	$81.35 \pm .37$	$79.25 \pm .28$
	DGCCA	2024	$47.22 \pm .62$	$41.18 \pm .50$	$41.29 \pm .33$
	rDGCCA	2024	$73.04 \pm .26$	$44.83 \pm .68$	$71.82 \pm .48$
	SUMCOR	2024	$47.23 \pm .62$	$41.21 \pm .50$	$41.26 \pm .32$
	tSUMCOR	2024	$47.26 \pm .64$	$60.63 \pm .37$	$41.37 \pm .35$
	mSUMCOR	2024	$61.02 \pm .45$	$53.84 \pm .42$	$42.07 \pm .33$

Table 10: Setup C, $d_{\mathcal{Z}} = 6$: per-view source recovery R^2 (%), mean $\pm$ population SD over ten batches of one trained model. SDs are percentage points. All sources have dimension three. All runs complete $50{,}000$ updates. Errors measure batch variability, not uncertainty across training seeds. Training details appear in Appendix F.2.

$d_{\mathcal{Z}}$	Method	Seed	View 1 (%)	View 2 (%)	View 3 (%)
6	BT	2024	$49.41 \pm .75$	$47.20 \pm .57$	$44.55 \pm .48$
	VICReg	2024	$48.58 \pm .68$	$62.46 \pm .57$	$42.68 \pm .38$
	W-MSE	2024	$94.32 \pm .19$	$93.16 \pm .15$	$93.41 \pm .18$
	SimCLR	2024	$80.98 \pm .31$	$78.20 \pm .44$	$80.94 \pm .60$
	DGCCA	2024	$\mathbf{99.45 \pm .11}$	$\mathbf{99.33 \pm .12}$	$\mathbf{99.14 \pm .27}$
	rDGCCA	2024	$\mathbf{99.90 \pm .00}$	$\mathbf{99.87 \pm .03}$	$\mathbf{99.90 \pm .00}$
	SUMCOR	2024	$\mathbf{99.75 \pm .01}$	$\mathbf{99.77 \pm .02}$	$\mathbf{99.69 \pm .02}$
	tSUMCOR	2029	51.83 ± 4.80	42.37 ± 1.60	44.20 ± 2.58
	mSUMCOR	2029	$59.13 \pm .41$	$61.31 \pm .49$	$67.87 \pm .71$

F.5 Complete two-dimensional visualizations

The two-dimensional experiments use three views with visibility sets $\{c_{12}, c_{13}\}$, $\{c_{12}, c_{23}\}$, and $\{c_{13}, c_{23}\}$. There is no globally shared coordinate, and each edge has rank one. These are separate runs from the three-dimensional numerical comparison. The four runs shown use seed 2024 and 50,000 updates. Each visualization uses 4096 samples from view 1. Hypergeometric sources and views 2 and 3 are not shown.

Figure 1 includes all four distributions and both source-coordinate colorings after ZCA whitening. Figure 2 includes the corresponding coordinate-standardized versions. Together they display 16 panels. Each pair of colorings uses identical representation points. Color indicates the corresponding raw source coordinate, scaled to the displayed range. These displays apply no fitted source alignment, and their orientation is not an estimate of a uniquely identifiable basis. All four distributions use the nuclear-norm SUMCOR objective.

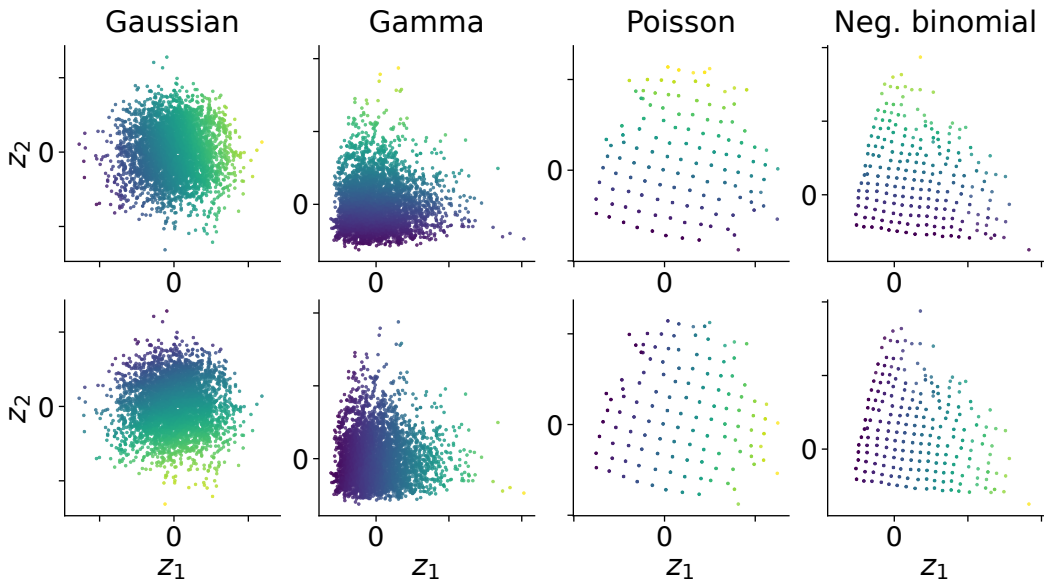


Figure 1: 2D whitened SUMCOR representations. Colors show source coordinates $s_{1,1}$ (top) and $s_{1,2}$ (bottom).

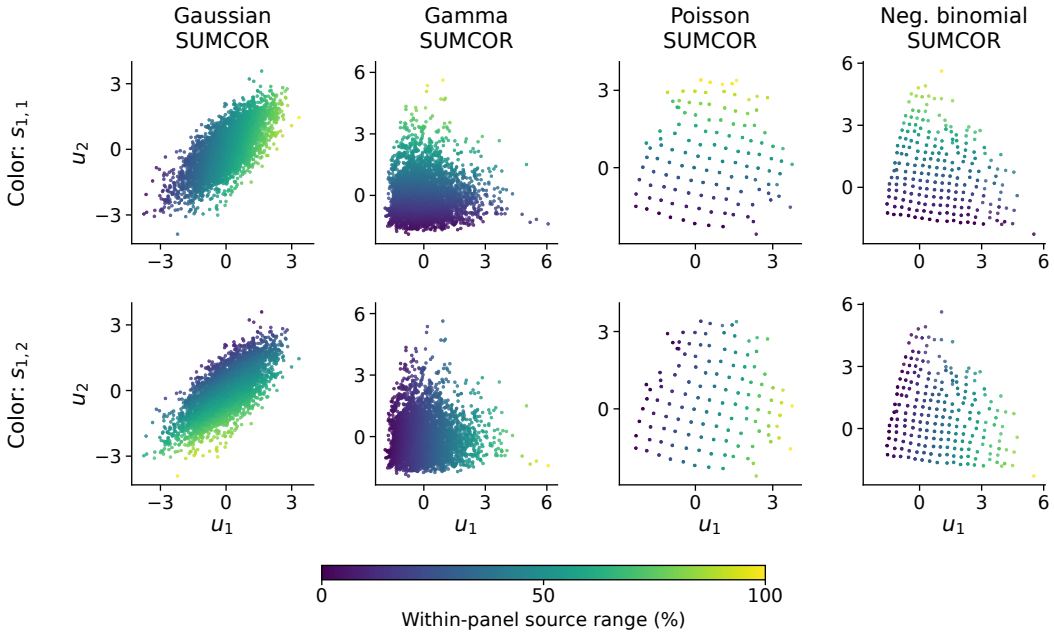


Figure 2: All coordinate-standardized counterparts of Figure 1. Each output coordinate is centered and divided by its sample standard deviation, without decorrelating the two coordinates. Both source colorings and all 4096 points are retained.