QuotientKanNet: Differentiable Kan Extensions for Functorial and Structure-Preserving Neural Computation

New York General Group, Inc.

Abstract

We introduce QuotientKanNet (QKN), a category-theoretic framework for neural representation learning based on differentiable realizations of pointwise left Kan extensions. Given a functor between finite categories and a vector-space-valued functor, the proposed construction derives a neural aggregation operator from the categorical colimit

$$(\mathrm{Lan}_J F)(d) \cong \operatorname*{colim}_{(J\downarrow d)} (F \circ \pi_d).$$

We express this colimit as the cokernel of a categorical relation operator, realize the resulting quotient through an orthogonal complement, and introduce the differentiable resolvent

$$\boxed{P_{d,\tau} = \tau(B_d B_d^* + \tau I)^{-1}, \qquad \tau > 0.}$$

The framework is developed in connection with cellular sheaf Laplacians, finite-group coinvariants, Reynolds operators, spectral diffusion, and functorial representation learning. We identify cellular sheaf diffusion and group-invariant aggregation as mathematically related special constructions while distinguishing the more general role of comma categories and arbitrary change-of-index functors.

To make the framework compatible with trainable categorical structures, we introduce a path-algebraic parameterization of generating morphisms, a cocone-constrained parameterization that preserves a nontrivial quotient subspace, and a rank-preservation principle preventing complete representation collapse. We further specify the propagation of feature vectors and morphism maps through hierarchical neural layers, separating exact functorial Kan-extension stages from nonlinear natural transformations.

We establish results concerning universal colimit factorization, spectral convergence, structural invariance, differentiability, nonexpansiveness, quotient-rank preservation, implicit differentiation, and approximation of quotient-invariant functions. A matrix-free computational formulation provides complexity bounds for sparse iterative realizations without explicitly forming the inverse operator.

The central mathematical proposition is that Kan-extension-based neural computation can be understood as functorial quotient formation equipped with a differentiable spectral approximation, while structural nondegeneracy can be guaranteed through algebraically constrained morphism parameterizations.

Keywords: Kan extension, categorical colimit, cokernel, cellular sheaf, sheaf Laplacian, group coinvariants, Reynolds operator, functorial neural networks, quotient regularization, resolvent operators, geometric deep learning.

1. Introduction

Structured representation learning concerns the construction of feature transformations that respect relationships among objects.

Conventional neural networks typically implement transformations of the form

$$h^{(\ell+1)} = \sigma_\ell(W_\ell h^{(\ell)} + b_\ell), \tag{1}$$

where W_ℓ is a parameterized linear map and σ_ℓ is a nonlinear activation.

Geometric deep learning introduces additional mathematical structure through symmetries, incidence relationships, equivariance constraints, and diffusion operators.

In group-equivariant architectures, a prescribed group action constrains the admissible transformations. In cellular sheaf neural networks, vector spaces are assigned to cells and linked through restriction maps, allowing generalizations of graph Laplacians.

Kan extensions provide a related but more general categorical perspective. They describe how a structured representation defined over one indexing category can be universally extended along a functor to another indexing category.

The present framework develops this connection at the level of finite-dimensional linear algebra.

1.1. The central construction

Let

$$J : \mathcal{C} \longrightarrow \mathcal{D}$$

be a functor, and let

$$F : \mathcal{C} \longrightarrow \mathbf{FinVect}_{\mathbb{R}}$$

describe a structured system of vector spaces and linear morphisms.

The left Kan extension of F along J assigns each d the colimit

$$(\mathrm{Lan}_J F)(d) = \operatorname*{colim}_{(J\downarrow d)} (F \circ \pi_d), \tag{2}$$

up to canonical isomorphism.

For a finite indexing category, this colimit admits a presentation

$$\operatorname{coker} B_d = V_d / \operatorname{im} B_d. \tag{3}$$

After equipping V_d with an inner product, the quotient admits an orthogonal realization through

$$P_d = I - B_d B_d^{\dagger}. \tag{4}$$

Its regularized approximation is

$$P_{d,\tau} = I - B_d (B_d^* B_d + \tau I)^{-1} B_d^*. \tag{5}$$

A neural layer can therefore be expressed as

$$\boxed{\mathcal{K}_{\theta,\tau}(x) = \phi_\theta(P_{\theta,\tau} x).} \tag{6}$$

The categorical part of the construction is determined by the relation operator, while the nonlinear map supplies task-dependent expressivity.

1.2. Mathematical scope

The framework concerns finite categories, finite-dimensional real vector spaces, differentiable matrix parameterizations, and neural maps compatible with specified categorical structure.

Its central results distinguish three objects:

- The exact left Kan extension, which satisfies a universal categorical property.
- The orthogonal quotient realization, which depends on chosen inner products.
- The regularized quotient operator, which is smooth for positive regularization but is not an exact categorical quotient.

This distinction is essential because a regularized operator is generally invertible, whereas a nontrivial categorical quotient is not.

1.3. Position among existing theories

The use of Kan extensions in learning has established precedents, including Shiebler's work on Kan extensions in data science [4]. arXiv

Similarly, cellular sheaf Laplacians and neural sheaf diffusion already provide neural constructions using matrix-valued incidence operators and associated spectral dynamics. Hansen and Ghrist developed the spectral theory of cellular sheaves, while Bodnar and collaborators introduced neural sheaf diffusion [5,7]. DOI +1

Accordingly, the mathematical interest of QKN lies in its combined formulation through arbitrary comma categories, universal colimit presentations, trainable functorial transport, and explicit rank-preserving constraints, rather than in claiming that resolvent filtering or sheaf-inspired neural computation is itself new.

2. Categorical Foundations

2.1. Vector-space-valued functors

Let $\mathcal{C}$ and $\mathcal{D}$ be finite categories.

Let

$$F : \mathcal{C} \to \mathbf{FinVect}_{\mathbb{R}} \tag{7}$$

be a covariant functor.

For each object c, write $F(c) = V_c$.

For each morphism $f : c \to c'$, let $F(f) : V_c \to V_{c'}$.

The functor axioms are

$$F(\mathrm{id}_c) = I_{V_c} \tag{8}$$

and

$$F(g \circ f) = F(g)F(f) \tag{9}$$

for composable morphisms f and g.

In particular, relations between different paths in $\mathcal{C}$ must be respected by their corresponding matrix products.

2.2. Universal characterization of left Kan extensions

Given $J : \mathcal{C} \to \mathcal{D}$, the left Kan extension of F is a functor

$$\mathrm{Lan}_J F : \mathcal{D} \to \mathbf{FinVect}_{\mathbb{R}}$$

equipped with a natural transformation

$$\eta : F \Rightarrow (\mathrm{Lan}_J F) \circ J.$$

Its universal property is

$$\boxed{\mathrm{Nat}(\mathrm{Lan}_J F, G) \cong \mathrm{Nat}(F, G \circ J)} \tag{10}$$

naturally in G.

The left Kan extension is therefore left adjoint to precomposition with J:

$$\mathrm{Lan}_J \dashv J^*. \tag{11}$$

2.3. Comma-category formulation

For an object d, define $\mathcal{J}_d = (J \downarrow d)$.

Its objects are pairs

$$(c, u), \qquad u : J(c) \to d.$$

A morphism $f : (c, u) \to (c', u')$ is a morphism satisfying

$$u' \circ J(f) = u. \tag{12}$$

Let $\pi_d : \mathcal{J}_d \to \mathcal{C}$ be the projection functor, and define

$$A_d = F \circ \pi_d. \tag{13}$$

Then

$$\boxed{(\mathrm{Lan}_J F)(d) \cong \operatorname*{colim}_{\mathcal{J}_d} A_d.} \tag{14}$$

The appearance of the comma category is important.

In general, aggregation is indexed not merely by the objects mapping directly to d, but by objects equipped with morphisms $J(c) \to d$, together with the compatible morphisms among those pairs.

Thus the aggregation structure depends on the entire functor J, not just on a fixed incidence matrix.

3. Cokernel Presentation of Categorical Colimits

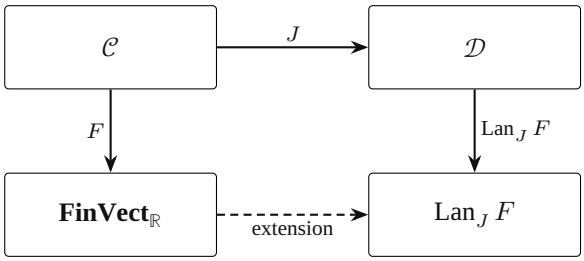
Let $\mathcal{J}_d = (J \downarrow d)$.

Define the object-feature space

$$V_d = \bigoplus_{i \in \mathrm{Ob}(\mathcal{J}_d)} A_d(i). \tag{15}$$

Define the relation-coordinate space

$$E_d = \bigoplus_{a : i \to j \in \mathrm{Mor}(\mathcal{J}_d)} A_d(i). \tag{16}$$



$$(\mathrm{Lan}_J F)(d) \cong \operatorname*{colim}_{(J\downarrow d)} (F \circ \pi_d)$$

Figure 1: Change of indexing along J and the pointwise left Kan extension. The dashed connection indicates the universal extension principle, not an additional functor.

Identity arrows may be omitted, since their contributions are zero.

Let ι_i denote the canonical direct-sum injection.

For a morphism $a : i \to j$, define

$$B_{d,a} = \iota_j A_d(a) - \iota_i. \tag{17}$$

The full relation operator is

$$B_d : E_d \to V_d, \tag{18}$$

obtained by assembling the maps $B_{d,a}$.

The associated sequence

$$E_d \xrightarrow{B_d} V_d \xrightarrow{q_d} Q_d \longrightarrow 0 \tag{19}$$

is right exact, with $Q_d = \operatorname{coker} B_d$.

Consequently,

$$\boxed{Q_d \cong (\mathrm{Lan}_J F)(d).} \tag{20}$$

Theorem 1. Universal factorization

Let W be a finite-dimensional real vector space.

Every linear map $T : V_d \to W$ satisfying

$$T B_d = 0 \tag{21}$$

factors uniquely through the quotient map q_d.

That is, there exists a unique map $\overline{T} : Q_d \to W$ such that

$$T = \overline{T} q_d. \tag{22}$$

Proof. Since $T B_d = 0$, $\operatorname{im} B_d \subseteq \ker T$. Therefore T is constant on cosets of $\operatorname{im} B_d$ and induces a unique linear transformation on the quotient $V_d / \operatorname{im} B_d$. The induced transformation is $\overline{T}$. ■

This theorem establishes the categorical meaning of the linear relations: every compatible downstream linear representation must factor through the colimit.

4. Orthogonal Quotients and Regularized Resolvents

Equip V_d and E_d with positive-definite inner products.

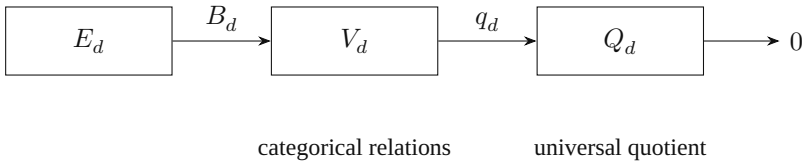


Figure 2: Cokernel presentation of the pointwise colimit.

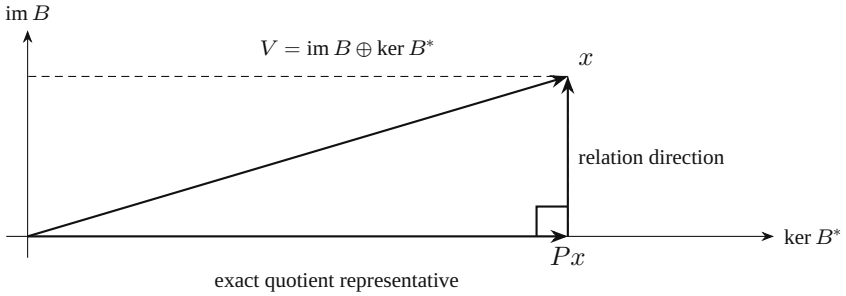


Figure 3: Orthogonal realization of the categorical quotient. $P = I - BB^\dagger$ removes components in the relation subspace and retains the unique minimum-norm representative.

Write B_d^* for the adjoint of B_d.

In finite dimensions,

$$V_d = \text{im}\, B_d \oplus \ker B_d^*. \tag{23}$$

Therefore,

$$Q_d \cong \ker B_d^*. \tag{24}$$

The isomorphism is induced by the chosen inner product and is not part of the bare categorical colimit.

4.1. Exact quotient projector

Define

$$P_d = I - B_d B_d^\dagger. \tag{25}$$

Then

$$P_d^2 = P_d, \qquad P_d^* = P_d, \tag{26}$$

and

$$\text{im}\, P_d = \ker B_d^*. \tag{27}$$

For $x \in V_d$, the vector $P_d x$ is the minimum-norm representative of its quotient class.

Equivalently,

$$P_d x = x - \operatorname*{argmin}_{r \in \text{im}\, B_d} \|x - r\|^2. \tag{28}$$

4.2. Differentiable quotient regularization

For $\tau > 0$, define

$$\boxed{P_{d,\tau} = I - B_d (B_d^* B_d + \tau I)^{-1} B_d^*.} \tag{29}$$

The resolvent identity yields

$$\boxed{P_{d,\tau} = \tau (B_d B_d^* + \tau I)^{-1}.} \tag{30}$$

This operator is positive definite and self-adjoint.

Its eigenvalues lie in $(0, 1]$.

In particular, $\|P_{d,\tau}\|_2 = 1$ whenever $\ker B_d^* \neq \{0\}$, and otherwise $\|P_{d,\tau}\|_2 < 1$.

Theorem 2. Spectral convergence

Let $B_d = U\Sigma V^*$ be a singular value decomposition, and let $\sigma_1, \dots, \sigma_r$ be the nonzero singular values.

Then, in suitable orthonormal coordinates,

$$P_{d,\tau} = U \operatorname{diag}\Big(\frac{\tau}{\sigma_1^2+\tau}, \dots, \frac{\tau}{\sigma_r^2+\tau}, 1, \dots, 1\Big)U^*. \tag{31}$$

Consequently,

$$\boxed{\lim_{\tau\to 0^+} P_{d,\tau} = P_d.} \tag{32}$$

Proof. By the singular value decomposition, $B_d B_d^* = U\Sigma\Sigma^* U^*$. Substitution into Equation (30) gives Equation (31). For every positive σ_i, $\lim_{\tau\to 0^+} \frac{\tau}{\sigma_i^2+\tau} = 0$. The zero-eigenvalue directions remain fixed with eigenvalue one. Thus the limit is precisely the orthogonal projector onto $\ker B_d^*$. ■

4.3. Variational characterization

Define the quadratic functional

$$\mathcal{E}_\tau(y; x) = \frac{1}{2}\|y - x\|^2 + \frac{1}{2\tau}\|B_d^* y\|^2. \tag{33}$$

Theorem 3. Proximal representation

For every x,

$$\boxed{P_{d,\tau}x = \operatorname*{argmin}_{y\in V_d} \mathcal{E}_\tau(y; x).} \tag{34}$$

Proof. The first-order stationarity condition is

$$y - x + \tau^{-1} B_d B_d^* y = 0.$$

Rearranging,

$$(B_d B_d^* + \tau I)y = \tau x.$$

Therefore,

$$y = \tau(B_d B_d^* + \tau I)^{-1} x.$$

The Hessian $I + \tau^{-1} B_d B_d^*$ is positive definite, so the minimizer is unique.

Thus $P_{d,\tau}$ is a proximal operator and is firmly nonexpansive.

This variational characterization connects categorical quotient construction to convex regularization without changing the underlying universal property of the exact colimit.

5. Relation to Cellular Sheaf Theory

Cellular sheaf theory provides a particularly important comparison because its operators can coincide algebraically with the relation Laplacian of QKN.

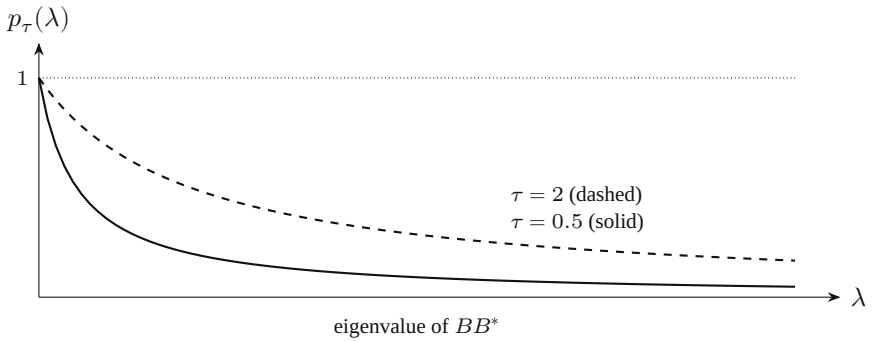


Figure 4: Resolvent response $p_\tau(\lambda) = \tau/(\lambda + \tau)$, showing exact retention at zero eigenvalue and attenuation of positive-eigenvalue directions.

5.1. Cellular sheaves on graphs

Let G be a finite graph.

A cellular sheaf associates vector spaces to vertices and edges:

$$\mathcal{F}(v), \qquad \mathcal{F}(e).$$

For each incidence $v \in e$, it assigns a restriction map

$$\rho_{v,e} : \mathcal{F}(v) \to \mathcal{F}(e).$$

Define the spaces of cochains

$$C^0(G;\mathcal{F}) = \bigoplus_{v \in V} \mathcal{F}(v), \tag{35}$$

$$C^1(G;\mathcal{F}) = \bigoplus_{e \in E} \mathcal{F}(e). \tag{36}$$

For an oriented edge $e = (u, v)$, the coboundary operator is

$$(\delta x)_e = \rho_{v,e} x_v - \rho_{u,e} x_u. \tag{37}$$

The degree-zero sheaf Laplacian is

$$\boxed{L_{\mathcal{F}} = \delta^* \delta.} \tag{38}$$

This construction generalizes the graph Laplacian to vector-valued local data and nontrivial restriction maps.

The spectral theory of cellular sheaves was developed by Hansen and Ghrist [5,6], and neural sheaf diffusion uses learned sheaf structures to define graph neural computations [7,8]. スプリンガー +1

5.2. Exact operator correspondence

Consider a QKN relation operator

$$B = \delta^*. \tag{39}$$

Then

$$\boxed{BB^* = \delta^* \delta = L_{\mathcal{F}}.} \tag{40}$$

Consequently, the regularized QKN operator becomes

$$\boxed{P_\tau = \tau (L_{\mathcal{F}} + \tau I)^{-1}.} \tag{41}$$

This is exactly a resolvent filter of the degree-zero cellular sheaf Laplacian.

Its zero-regularization limit is

$$\lim_{\tau \to 0^+} P_\tau = P_{\ker \delta}. \tag{42}$$

The kernel is the space of global sections, or degree-zero sheaf cohomology, for the relevant sheaf complex.

Hence the exact quotient realization identifies with the global-section space after choosing inner products and the adjoint identification.

5.3. Categorical realization of the sheaf operator

For a graph, form an incidence category with an object for each vertex and edge.

Use arrows $e \longrightarrow v$ for incidences, and assign each such arrow the adjoint restriction map

$$\rho^*_{v,e} : \mathcal{F}(e) \to \mathcal{F}(v).$$

This is a covariant diagram on the opposite of the usual vertex-to-edge incidence orientation.

For a fixed edge e, the corresponding span is

$$\mathcal{F}(u) \longleftarrow \mathcal{F}(e) \longrightarrow \mathcal{F}(v). \tag{43}$$

Its pushout imposes the relation

$$\rho^*_{u,e} z \sim \rho^*_{v,e} z. \tag{44}$$

After eliminating the edge-space coordinates from the full colimit presentation, these relations produce the cokernel

$$\operatorname{coker} \delta^*. \tag{45}$$

For a finite graph, this gives a categorical realization of the degree-zero sheaf quotient.

One must distinguish the reduced relation operator from the unreduced direct-sum relation matrix of the incidence diagram. They present isomorphic quotient spaces but need not produce identical regularized operators before choosing compatible metrics and eliminating auxiliary coordinates.

5.4. Scope of the generalization

The sheaf-Laplacian correspondence shows that QKN is not a completely separate spectral mechanism.

Its categorical formulation, however, permits indexing diagrams of the form $(J \downarrow d)$ for general functors J, rather than requiring a fixed graph incidence category.

These comma categories may contain multiple morphism types, compositional relations, nontrivial endomorphisms, and hierarchically organized indexing structures.

The QKN framework therefore supplies a common colimit-based language for several structured aggregation mechanisms, while cellular sheaf diffusion remains an important mathematically overlapping case.

6. Finite-Group Actions and Reynolds Averaging

A second fundamental example comes from representation theory.

Let G be a finite group, and let $\mathbf{B}G$ denote the one-object category whose endomorphisms are the elements of G.

Consider the unique functor

$$J : \mathbf{B}G \to \mathbf{1}, \tag{46}$$

where $\mathbf{1}$ is the terminal category.

Let $F : \mathbf{B}G \to \mathbf{FinVect}_{\mathbb{R}}$ correspond to a representation

$$\rho : G \to \mathrm{GL}(V). \tag{47}$$

6.1. Kan extensions as coinvariants

Since the comma category is canonically equivalent to $\mathbf{B}G$,

$$(\mathrm{Lan}_J F)(*) = \operatorname*{colim}_{\mathbf{B}G} F. \tag{48}$$

The colimit is the coinvariant space

$$\boxed{V_G = V/\operatorname{span}\{\rho(g)v - v : g \in G,\ v \in V\}.} \tag{49}$$

Let

$$R_G = \operatorname{span}\{\rho(g)v - v : g \in G,\ v \in V\}.$$

Then

$$V_G = V/R_G. \tag{50}$$

6.2. Reynolds averaging

Define the Reynolds operator

$$\boxed{\mathcal{R}_G = \frac{1}{|G|}\sum_{g\in G}\rho(g).} \tag{51}$$

It satisfies

$$\mathcal{R}_G^2 = \mathcal{R}_G, \tag{52}$$

and

$$\operatorname{im}\mathcal{R}_G = V^G, \tag{53}$$

where

$$V^G = \{v \in V : \rho(g)v = v\ \forall g \in G\}$$

is the invariant subspace.

Moreover,

$$\ker\mathcal{R}_G = R_G. \tag{54}$$

Therefore,

$$\boxed{V_G \cong V^G.} \tag{55}$$

This isomorphism holds over $\mathbb{R}$ for finite groups because averaging is available.

For a representation whose matrices are orthogonal with respect to the chosen inner product,

$$\rho(g)^* = \rho(g)^{-1},$$

the Reynolds operator is self-adjoint.

Hence

$$\boxed{P_{R_G^\perp} = \mathcal{R}_G.} \tag{56}$$

For a nonorthogonal representation equipped with an arbitrary Euclidean inner product, Reynolds averaging need not equal the orthogonal quotient projector. An invariant inner product, obtained by averaging an initial inner product over the group, restores this equality.

The Reynolds operator is a classical construction in invariant theory [12]. Wolfram MathWorld

Theorem 4. Group averaging as a Kan quotient projection

Let G be a finite group acting orthogonally on a finite-dimensional real inner-product space V.

Let B be a relation operator whose image is

$$\text{span}\{\rho(g)v - v : g \in G,\ v \in V\}.$$

Then

$$\boxed{I - BB^\dagger = \frac{1}{|G|}\sum_{g\in G}\rho(g).} \tag{57}$$

Proof. The Reynolds operator is idempotent and self-adjoint.

Its image is V^G, and its kernel is R_G.

Therefore $\mathcal{R}_G$ is the orthogonal projector onto $R_G^\perp$.

The operator $I - BB^\dagger$ is the unique orthogonal projector onto that same subspace.

Hence the two operators coincide. ■

This gives a direct categorical interpretation of invariant aggregation for finite orthogonal group actions.

Group-equivariant networks are broader than this invariant quotient construction: equivariant layers may retain nontrivial representation types, whereas this exact quotient retains only the trivial representation component.

Thus invariant pooling is a special case of QKN, but general group equivariance is not equivalent to quotient projection.

7. Functorial Parameterization of Learnable Morphisms

An important distinction arises when the categorical morphism maps are trainable.

A family of arbitrary learned matrices does not necessarily define a functor. In particular, the identities

$$F_\theta(g \circ f) = F_\theta(g)F_\theta(f) \tag{58}$$

must remain valid for all composable morphisms.

We therefore construct trainable feature functors through categorical presentations.

7.1. Free categories generated by quivers

Let $\mathsf{Q} = (Q_0, Q_1, s, t)$ be a finite directed quiver, where Q_0 is the set of vertices and Q_1 is the set of generating arrows.

Let $\mathcal{C} = \text{Free}(\mathsf{Q})$ be the free category generated by Q.

For each vertex c, choose a vector space $V_c = \mathbb{R}^{n_c}$.

For each generating arrow $a : c \to c'$, introduce a trainable matrix

$$M_a(\theta) \in \mathbb{R}^{n_{c'} \times n_c}. \tag{59}$$

For a composite path $p = a_k \circ \cdots \circ a_1$, define

$$\boxed{F_\theta(p) = M_{a_k}(\theta) \cdots M_{a_1}(\theta).} \tag{60}$$

For identity morphisms, define

$$F_\theta(\text{id}_c) = I_{n_c}. \tag{61}$$

Theorem 5. Exact functoriality of path-generated representations

The assignment defined by Equations (59)–(61) determines a functor

$$F_\theta : \text{Free}(\mathsf{Q}) \to \mathbf{FinVect}_{\mathbb{R}}$$

for every parameter value θ.

Proof. Identity morphisms are assigned identity matrices by construction.

Let p and q be composable paths.

The composite is their concatenation.

By the definition of the path representation,

$$F_\theta(q \circ p) = F_\theta(q) F_\theta(p).$$

Thus both functor axioms hold identically. ■

This construction ensures that parameter updates do not violate functoriality.

It does not require a composition-consistency penalty.

7.2. Categories with relations between paths

A general finitely presented category may have a presentation

$$\mathcal{C} = \text{Free}(\mathsf{Q})/\!\sim, \tag{62}$$

where $\sim$ identifies certain paths.

Suppose the defining relations include $p_i \sim q_i$.

The learned generating matrices must satisfy

$$\boxed{M_{p_i}(\theta) = M_{q_i}(\theta).} \tag{63}$$

A path-generated assignment descends to a functor on $\mathcal{C}$ if and only if it respects these defining relations.

One may therefore define the admissible parameter space

$$\Theta_{\mathcal{C}} = \left\{\theta : M_{p_i}(\theta) = M_{q_i}(\theta) \text{ for all defining relations}\right\}. \tag{64}$$

For exact categorical semantics, parameterization must remain in $\Theta_{\mathcal{C}}$.

There are several mathematical possibilities: solve the defining relations symbolically, use a constrained family of representations, or parameterize matrices through structures that satisfy the relations identically.

A penalty for violating Equation (63) is only a relaxation and does not guarantee an exact functor at nonzero penalty.

7.3. Acyclicity and finiteness

A finite quiver can generate an infinite category if it contains directed cycles.

For finite-dimensional construction over explicitly enumerable finite comma categories, a finite acyclic quiver is particularly convenient because it has finitely many directed paths.

For cyclic or infinite categories, additional finiteness assumptions, finite presentations, or exact finite-dimensional reductions are needed before constructing a finite relation matrix.

8. Preventing Trivial Representation Collapse

The quotient dimension is determined by the rank of the relation operator.

Let $N = \dim V_d$, $r = \operatorname{rank} B_d$.

Then

$$\boxed{\dim Q_d = N - r.} \tag{65}$$

If $r = N$, the quotient is the zero vector space.

Equivalently, $\ker B_d^* = \{0\}$.

In this case,

$$\lim_{\tau \to 0^+} P_{d,\tau} = 0. \tag{66}$$

Thus unconstrained learned relations can eliminate every representational degree of freedom.

8.1. Cocone-constrained nondegeneracy

We introduce a structural method for preserving a nonzero quotient.

Let $\mathcal{J}_d$ be the relevant comma category, and choose a vector space

$$U_d \cong \mathbb{R}^{k_d}, \qquad k_d > 0.$$

For each object i, choose a linear map

$$T_i : A_d(i) \to U_d. \tag{67}$$

Require that these maps form a cocone:

$$\boxed{T_j A_d(a) = T_i} \tag{68}$$

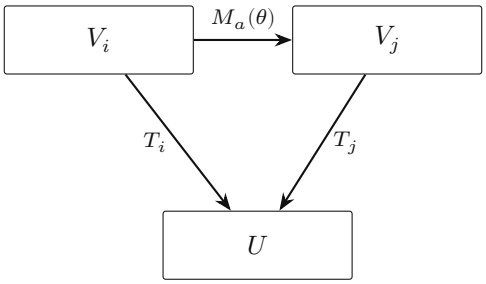


Figure 5: Cocone-constrained morphism learning. The commutative triangle fixes a common anchor space U and ensures that the relation operator is annihilated by the assembled anchor map.

for every morphism $a : i \to j$.

Assemble them into

$$T_d : V_d \to U_d \tag{69}$$

using the direct-sum universal property.

The cocone equations imply

$$\boxed{T_d B_d = 0.} \tag{70}$$

Theorem 6. Quotient-rank preservation

Suppose T_d is surjective and satisfies $T_d B_d = 0$.

Then

$$\boxed{\dim \operatorname{coker} B_d \geq k_d.} \tag{71}$$

Equivalently,

$$\boxed{\operatorname{rank} B_d \leq N - k_d.} \tag{72}$$

Proof. By the universal property of the cokernel, T_d factors uniquely through q_d:

$$T_d = \overline{T}_d q_d.$$

Because T_d and q_d are surjective, $\overline{T}_d$ is surjective.

Consequently, $\dim Q_d \geq \dim U_d = k_d$.

The rank-nullity theorem gives $\dim Q_d = N - \operatorname{rank} B_d$.

Therefore, $\operatorname{rank} B_d \leq N - k_d$. ■

This gives an exact algebraic guarantee against complete quotient collapse.

The result does not require monitoring singular values or explicitly computing the rank during learning.

8.2. Direct preservation of an anchor subspace

The cocone condition yields a stronger operator-theoretic statement.

Since $T_d B_d = 0$, we have $B_d^* T_d^* = 0$.

Therefore,

$$\operatorname{im} T_d^* \subseteq \ker B_d^*. \tag{73}$$

It follows that

$$\boxed{P_{d,\tau} T_d^* u = T_d^* u} \tag{74}$$

for every $u \in U_d$ and every $\tau > 0$.

Thus the regularized operator preserves the entire anchor subspace exactly, independently of the regularization parameter.

This is stronger than a lower bound on the quotient dimension: it identifies an explicit family of representation directions that cannot be attenuated by the quotient operator.

8.3. A constructive morphism parameterization

The cocone conditions can be enforced directly on generating arrows.

Suppose every anchor map $T_j : V_j \to U$ is surjective.

Choose a right inverse $S_j : U \to V_j$ such that

$$T_j S_j = I_U. \tag{75}$$

For a generating arrow $a : i \to j$, define

$$\boxed{M_a(\theta) = S_j T_i + (I_{V_j} - S_j T_j) K_a(\theta),} \tag{76}$$

where $K_a(\theta) : V_i \to V_j$ is an arbitrary trainable linear map.

Then

$$\begin{aligned} T_j M_a(\theta) &= T_j S_j T_i \\ &\quad + T_j (I_{V_j} - S_j T_j) K_a(\theta) \\ &= T_i. \end{aligned}$$

Hence

$$\boxed{T_j M_a(\theta) = T_i} \tag{77}$$

for every parameter value.

When the source category is freely generated by the quiver, all composite morphisms inherit this compatibility automatically.

For categories with path relations, those relations must additionally be enforced; the cocone condition alone does not imply that distinct paths representing the same morphism have equal matrix products.

Theorem 7. Simultaneous functoriality and anchor preservation

Let $\mathcal{C}$ be the free category generated by a finite acyclic quiver.

Equip its objects with vector spaces V_i, and choose surjective maps $T_i : V_i \to U$ to a fixed nonzero vector space U.

Define each generating-arrow matrix by Equation (76), and define composite matrices by path multiplication.

Then, for every parameter value θ:

- F_θ is a functor.
- $T_j F_\theta(a) = T_i$ for every morphism $a : i \to j$.

- The induced relation operator satisfies $T_d B_d = 0$.
- If T_d is surjective, the resulting quotient has dimension at least $\dim U$.

Proof. The first statement follows from Theorem 5.

Equation (77) establishes the second statement on generating arrows. Its extension to composite paths follows by induction on path length.

The third statement follows by applying the cocone identity to each relation column.

The fourth statement follows from Theorem 6. ■

This construction supplies a concrete family of trainable, exactly functorial, nontrivially quotient-preserving representations.

8.4. Limits of anchor preservation

Anchor preservation prevents complete collapse but does not guarantee that every non-anchor direction remains informative.

For a vector v in a positive-eigenvalue eigenspace of $L_d = B_d B_d^*$ with eigenvalue λ,

$$P_{d,\tau} v = \frac{\tau}{\lambda + \tau} v. \tag{78}$$

If $\lambda \gg \tau$, this component is strongly attenuated.

Accordingly, quotient dimension and finite-regularization signal retention are distinct properties.

A useful spectral condition is

$$\|B_d\|_2 \le M_d. \tag{79}$$

Under this assumption,

$$\boxed{\frac{\tau}{M_d^2 + \tau} I \preceq P_{d,\tau} \preceq I.} \tag{80}$$

Thus the smallest singular value of the regularized operator is bounded away from zero for fixed positive τ and bounded relation-operator norm.

Anchor constraints provide exact rank preservation, while spectral constraints provide quantitative control of finite-regularization attenuation.

9. Computational Complexity and Matrix-Free Operator Theory

The regularized quotient operator is defined by

$$P_\tau x = \tau (B B^* + \tau I)^{-1} x. \tag{81}$$

Let $B \in \mathbb{R}^{N \times M}$.

Explicitly constructing and inverting the matrix generally requires cubic arithmetic complexity and quadratic storage.

However, the definition of P_τ does not require explicit matrix inversion.

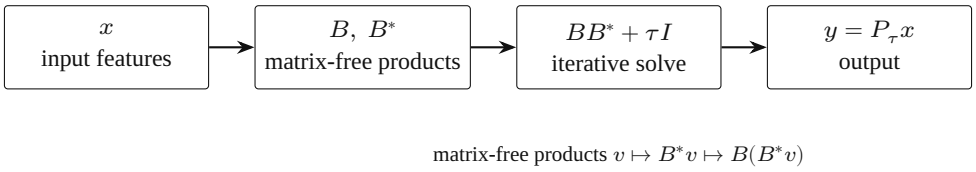


Figure 6: Matrix-free resolvent computation. Krylov iterations use applications of B and B^* rather than assembling or inverting a dense matrix. The adjoint product is shown as an available primitive rather than a required preprocessing step.

9.1. Equivalent linear systems

The desired representation $y = P_\tau x$ is the unique solution of

$$\boxed{(BB^* + \tau I)y = \tau x.} \tag{82}$$

Alternatively, solve

$$(B^*B + \tau I)z = B^*x \tag{83}$$

and recover

$$\boxed{y = x - Bz.} \tag{84}$$

Both expressions define the same operator in exact arithmetic.

The smaller of the two systems can be selected according to the dimensions N and M.

9.2. Sparse conjugate-gradient realization

Because $A_\tau = BB^* + \tau I$ is symmetric positive definite for $\tau > 0$, conjugate gradients can be applied.

Its condition number satisfies

$$\begin{aligned} \kappa(A_\tau) &= \frac{\lambda_{\max}(BB^*) + \tau}{\lambda_{\min}(BB^*) + \tau} \\ &\le 1 + \frac{\|B\|_2^2}{\tau}. \end{aligned} \tag{85}$$

For conjugate gradients initialized at y_0, the standard energy-norm estimate is

$$\frac{\|y_k - y^\star\|_{A_\tau}}{\|y_0 - y^\star\|_{A_\tau}} \le 2\left(\frac{\sqrt{\kappa} - 1}{\sqrt{\kappa} + 1}\right)^k, \tag{86}$$

where $y^\star$ is the exact solution and $\kappa = \kappa(A_\tau)$.

Hence a sufficient asymptotic iteration bound is

$$k = O\left(\sqrt{\kappa}\log\frac{1}{\varepsilon}\right) \tag{87}$$

for a prescribed relative energy-norm accuracy ε.

If B is sparse, an application of BB^* can be implemented as successive products

$$v \mapsto B^*v \mapsto B(B^*v).$$

The arithmetic cost of each iteration is then $O(\mathrm{nnz}(B))$, where $\mathrm{nnz}(B)$ is the number of nonzero matrix entries.

Thus the approximate solve has complexity

$$\boxed{O\left(\mathrm{nnz}(B)\sqrt{\kappa}\log\frac{1}{\varepsilon}\right),} \tag{88}$$

under the standard conjugate-gradient convergence bound.

This excludes the costs of constructing the categorical diagram, generating matrix entries, and applying the downstream neural map.

9.3. Regularization and numerical conditioning

Equation (85) identifies a fundamental tradeoff.

Decreasing τ improves approximation to the exact quotient projector, but can increase the condition number of the associated linear system.

For fixed nonzero singular values, the quotient approximation error satisfies

$$\|P_\tau - P\|_2 = \frac{\tau}{(\sigma^+_{\min}(B))^2 + \tau}. \tag{89}$$

Therefore, small τ improves categorical quotient accuracy while potentially increasing computational difficulty.

This tradeoff is intrinsic to the resolvent representation and cannot be removed merely by replacing a dense inverse with an iterative solver.

Preconditioning may reduce the effective condition number, although its effectiveness depends on the structure of the categorical relation operator.

9.4. Implicit differentiation

Let

$$y = \tau A_\tau^{-1} x, \qquad A_\tau = BB^* + \tau I. \tag{90}$$

For a scalar objective $\mathcal{L}$, define $g = \nabla_y \mathcal{L}$.

The gradient with respect to the input is

$$\boxed{\nabla_x \mathcal{L} = \tau A_\tau^{-1} g.} \tag{91}$$

This requires a linear solve with the same symmetric positive-definite operator used in the forward computation.

For the gradient with respect to B, let $z = A_\tau^{-1} x$ and $w = \tau A_\tau^{-1} g$.

Then

$$dy = -\tau A_\tau^{-1}\big((dB)B^* + B(dB)^*\big)z.$$

Using the Frobenius inner product, for real matrices we obtain

$$\boxed{\nabla_B \mathcal{L} = -w(B^* z)^* - z(B^* w)^*.} \tag{92}$$

Consequently, derivatives can be evaluated through matrix-vector products and linear solves without differentiating through an explicitly stored matrix inverse.

For inexact iterative solutions, the resulting gradient is approximate; its accuracy depends on the forward and adjoint solve residuals.

10. Hierarchical Feature Propagation

We now specify a multilayer categorical architecture in which both feature vectors and morphism maps propagate between levels.

Consider categories

$$\mathcal{C}_0 \xrightarrow{J_0} \mathcal{C}_1 \xrightarrow{J_1} \cdots \xrightarrow{J_{L-1}} \mathcal{C}_L. \tag{93}$$

At level ℓ, assume a vector-space-valued functor

$$F_\ell : \mathcal{C}_\ell \to \mathbf{FinVect}_{\mathbb{R}}. \tag{94}$$

For every object c, let

$$h_c^{(\ell)} \in F_\ell(c) \tag{95}$$

be its current feature vector.

The collection of feature vectors need not itself be a natural transformation from the constant functor. In particular, it is not necessary that

$$F_\ell(f)h_c^{(\ell)} = h_{c'}^{(\ell)}$$

for every morphism $f : c \to c'$. Such a condition would impose additional compatibility on the data.

10.1. Explicit construction of the next-layer input

Fix $d \in \mathcal{C}_{\ell+1}$, and define

$$\mathcal{I}_{\ell,d} = (J_\ell \downarrow d).$$

The corresponding direct-sum space is

$$V_{\ell,d} = \bigoplus_{(c,u)\in \mathrm{Ob}(\mathcal{I}_{\ell,d})} F_\ell(c). \tag{96}$$

The next-layer input is the vector

$$\boxed{x_d^{(\ell)} = \bigoplus_{(c,u)\in \mathrm{Ob}(\mathcal{I}_{\ell,d})} h_c^{(\ell)}.} \tag{97}$$

If a source object appears in several distinct comma-category objects (c, u), its feature vector is included once for each occurrence.

Define the corresponding relation operator $B_{\ell,d} : E_{\ell,d} \to V_{\ell,d}$ from the morphism maps of F_ℓ.

The regularized aggregation is

$$\boxed{z_d^{(\ell)} = P_{\ell,d,\tau_\ell} x_d^{(\ell)}.} \tag{98}$$

This equation explicitly determines the categorical feature input at each layer.

10.2. Propagation of morphism maps

The exact left Kan extension produces a functor

$$Q_\ell = \mathrm{Lan}_{J_\ell} F_\ell. \tag{99}$$

For every object d, $Q_\ell(d) \cong \mathrm{coker}\, B_{\ell,d}$.

For each morphism $g : d \to d'$ in $\mathcal{C}_{\ell+1}$, postcomposition induces a map between the relevant comma categories.

This induces a linear map

$$Q_\ell(g) : Q_\ell(d) \to Q_\ell(d'). \tag{100}$$

These maps satisfy

$$Q_\ell(\mathrm{id}_d) = I, \qquad Q_\ell(g' \circ g) = Q_\ell(g')Q_\ell(g). \tag{101}$$

Therefore the exact quotient construction propagates not only objectwise vector spaces but also the full functorial system of morphism maps.

10.3. Nonlinear transformation of quotient features

To construct a nonlinear layer, choose a target functor

$$U_\ell : \mathcal{C}_{\ell+1} \to \mathbf{FinVect}_{\mathbb{R}}. \tag{102}$$

For every object d, choose a nonlinear transformation

$$\phi_{\ell,d} : Q_\ell(d) \to U_\ell(d). \tag{103}$$

The transformed feature is

$$\boxed{h_d^{(\ell+1)} = \phi_{\ell,d}\left(q_{\ell,d} x_d^{(\ell)}\right),} \tag{104}$$

where $q_{\ell,d}$ is the exact quotient map.

Using an orthogonal realization, the same computation can be represented as

$$h_d^{(\ell+1)} = \hat{\phi}_{\ell,d}\left(P_{\ell,d} x_d^{(\ell)}\right). \tag{105}$$

A regularized implementation substitutes $P_{\ell,d,\tau}$ for $P_{\ell,d}$, producing an approximate quotient-based neural layer.

The next feature functor is then defined by

$$\boxed{F_{\ell+1} = U_\ell.} \tag{106}$$

Its morphism maps are therefore

$$\boxed{F_{\ell+1}(g) = U_\ell(g),} \tag{107}$$

not an unspecified consequence of the nonlinear activation.

This closes the mathematical definition of the multilayer architecture.

10.4. Naturality of nonlinear transformations

For the nonlinear maps to be compatible with target morphisms, require

$$\boxed{U_\ell(g) \circ \phi_{\ell,d} = \phi_{\ell,d'} \circ Q_\ell(g).} \tag{108}$$

Because ϕ may be nonlinear, Equation (108) is a naturality condition after passing to the underlying set-valued functors, rather than a natural transformation in the category of finite-dimensional vector spaces and linear maps.

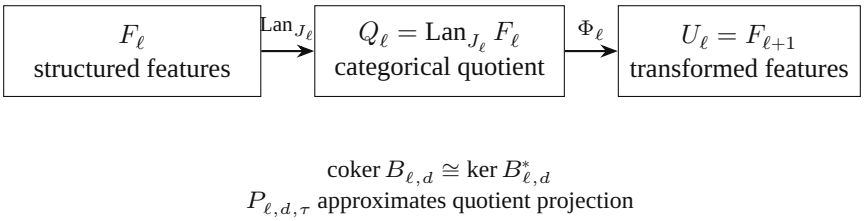


Figure 7: Hierarchical QKN computation separates exact functorial propagation from the nonlinear transformation.

For example, a componentwise nonlinearity commutes with permutation matrices:

$$\sigma(\Pi x) = \Pi\sigma(x). \tag{109}$$

Thus, when the relevant morphism actions are coordinate permutations, componentwise nonlinearities provide a simple class of compatible transformations.

For general linear morphism actions, such commutation is not automatic.

10.5. Naturality of the regularized operator

The exact Kan quotient is functorial, but the regularized orthogonal realization need not commute with the induced target morphism maps.

Suppose there are linear maps $S_g : V_{\ell,d} \to V_{\ell,d'}$ that implement the action of a target morphism on the pre-quotient representations.

The sufficient condition

$$S_g L_{\ell,d} = L_{\ell,d'} S_g, \tag{110}$$

where $L_{\ell,d} = B_{\ell,d} B^*_{\ell,d}$, implies

$$\boxed{S_g P_{\ell,d,\tau} = P_{\ell,d',\tau} S_g} \tag{111}$$

for a common positive regularization parameter τ.

This follows from the resolvent identity and the fact that the relevant operators are invertible.

Equation (110) is a sufficient compatibility condition, not a property automatically satisfied by every categorical diagram.

Hence the architecture distinguishes exact functorial quotient propagation from additional metric and spectral compatibility required for strictly natural regularized layers.

11. Structural Invariance and Approximation Theory

Let $R = \operatorname{im} B$, and let $P = P_{\ker B^*}$.

Theorem 8. Exact relation invariance

Let $f(x) = \phi(Px)$ for an arbitrary function ϕ on the orthogonal quotient realization.

Then

$$\boxed{f(x + r) = f(x)} \tag{112}$$

for every $r \in R$.

Proof. Since $Pr = 0$, $P(x + r) = Px$.

Applying ϕ gives the assertion. ■

Theorem 9. Approximate invariance under regularization

Suppose ϕ is L_ϕ-Lipschitz and $r \in R$.

Let $\sigma^+_{\min}(B)$ denote the smallest positive singular value of B.

Then, for $\tau > 0$,

$$\boxed{\begin{aligned}&\|\phi(P_\tau(x + r)) - \phi(P_\tau x)\| \\ &\quad \le L_\phi \frac{\tau}{(\sigma^+_{\min}(B))^2 + \tau}\|r\|.\end{aligned}} \tag{113}$$

Proof. Lipschitz continuity gives

$$\|\phi(P_\tau(x + r)) - \phi(P_\tau x)\| \le L_\phi \|P_\tau r\|.$$

On R, the largest eigenvalue of P_τ is $\frac{\tau}{(\sigma^+_{\min}(B))^2+\tau}$.

The bound follows. ■

11.1. Universal approximation on quotient spaces

Let $K \subseteq V$ be compact, and suppose $f^\star : K \to \mathbb{R}^m$ is continuous and constant on fibers of P.

Then there exists a continuous function $\overline{f} : P(K) \to \mathbb{R}^m$ such that

$$f^\star = \overline{f} \circ P|_K. \tag{114}$$

If a downstream neural class is dense in the continuous functions on $P(K)$, then, for every $\varepsilon > 0$, there exists a neural function satisfying

$$\boxed{\sup_{x \in K} \|N_\theta(Px) - f^\star(x)\| < \varepsilon.} \tag{115}$$

The proof follows from the quotient-map property of the continuous surjection and standard approximation on the compact image $P(K)$.

This theorem identifies the natural expressive domain of the exact quotient architecture.

Exact quotient layers cannot distinguish features lying in the same equivalence class. Their expressivity is therefore universal only within the class of functions compatible with the prescribed relations.

12. Composition of Kan-Extension Layers

Let

$$\mathcal{C}_0 \xrightarrow{J_0} \mathcal{C}_1 \xrightarrow{J_1} \mathcal{C}_2 \tag{116}$$

be functors between finite categories.

Given $F_0 : \mathcal{C}_0 \to \mathbf{FinVect}_{\mathbb{R}}$, the universal properties of left Kan extensions imply

$$\boxed{\mathrm{Lan}_{J_1}\left(\mathrm{Lan}_{J_0} F_0\right) \cong \mathrm{Lan}_{J_1 \circ J_0} F_0.} \tag{117}$$

Theorem 10. Composition of exact categorical aggregation

Assume the relevant left Kan extensions exist.

Then successive exact Kan extensions are naturally isomorphic to a single left Kan extension along the composite indexing functor.

Proof. For every functor $G : \mathcal{C}_2 \to \mathbf{FinVect}_{\mathbb{R}}$, the successive adjunctions give

$$\begin{aligned} &\mathrm{Nat}\left(\mathrm{Lan}_{J_1}\,\mathrm{Lan}_{J_0}\,F_0, G\right) \\ &\cong \mathrm{Nat}\left(\mathrm{Lan}_{J_0}\,F_0, G \circ J_1\right) \\ &\cong \mathrm{Nat}\left(F_0, G \circ J_1 \circ J_0\right). \end{aligned}$$

The last expression characterizes the left Kan extension along $J_1 \circ J_0$.

Uniqueness of left adjoints up to natural isomorphism proves the result. ■

12.1. Limits of compositional equivalence

Equation (117) concerns the exact categorical extension operators.

A multilayer neural network may interleave them with nonlinear transformations:

$$F_0 \xrightarrow{\mathrm{Lan}_{J_0}} Q_1 \xrightarrow{\Phi_1} F_1 \xrightarrow{\mathrm{Lan}_{J_1}} Q_2. \tag{118}$$

In general,

$$\mathrm{Lan}_{J_1}\left(\Phi_1(\mathrm{Lan}_{J_0}\,F_0)\right)$$

cannot be identified with a single left Kan extension of F_0.

Likewise, compositions of regularized quotient operators need not equal the regularized quotient for the composite category.

These distinctions preserve the separation between universal categorical operations and nonlinear neural computations.

13. Unified Mathematical Interpretation

The QKN framework connects several mathematical constructions through a common relation operator. The core mathematical chain is

$$\boxed{\begin{aligned}(\mathrm{Lan}_J\,F)(d) &\cong \operatorname*{colim}_{(J\downarrow d)}(F \circ \pi_d) \\ &\cong \mathrm{coker}\,B_d \\ &\cong \ker B_d^*.\end{aligned}} \tag{119}$$

The final identification depends on the choice of inner products.

The differentiable realization is

$$\boxed{P_{d,\tau} = \tau(B_d B_d^* + \tau I)^{-1}.} \tag{120}$$

For trainable morphisms, exact functoriality can be maintained through path-generated representations, while compatible cocones provide quotient-rank guarantees.

Theoretical framework	**Mathematical object**	**Role in QKN**
Category theory	$\mathrm{Lan}_J F$	Universal extension of structured representations
Homological algebra	$\mathrm{coker}\, B_d$	Quotient by categorical relation vectors
Inner-product geometry	$\ker B_d^*$	Orthogonal realization of the quotient
Operator theory	$P_{d,\tau}$	Differentiable quotient approximation
Sheaf theory	$L_{\mathcal{F}}$	Special relation-Laplacian construction
Representation theory	V_G	Coinvariants of finite-group actions
Invariant theory	$\mathcal{R}_G$	Exact group averaging under suitable inner products
Convex analysis	$\mathcal{E}_\tau$	Variational realization of regularized aggregation
Path algebra	F_θ	Functor-preserving parameterization
Linear algebra	$T_d B_d = 0$	Nondegeneracy through a compatible cocone

Table 1: Unified mathematical interpretation of QuotientKanNet.

These constructions produce three distinct but complementary mathematical constraints:

$$\boxed{\begin{aligned} \text{Functoriality:} \quad & F_\theta(g \circ f) = F_\theta(g) F_\theta(f), \\ \text{Quotient preservation:} \quad & T_d B_{d,\theta} = 0, \\ \text{Spectral regularity:} \quad & P_{d,\tau} = \tau (B_{d,\theta} B_{d,\theta}^* + \tau I)^{-1}. \end{aligned}} \tag{121}$$

Taken together, these equations define a constrained family of differentiable categorical neural operators.

14. Conclusion

QuotientKanNet provides a mathematical framework in which neural aggregation is derived from pointwise left Kan extensions.

The construction begins with a category of structured objects and a vector-space-valued functor. For each target object, the corresponding comma category determines a system of linear relations. Their cokernel defines the categorical quotient, which can be realized orthogonally after choosing inner products.

A regularized resolvent then supplies a smooth operator suitable for differentiable representation learning:

$$\boxed{\mathcal{K}_{\theta,\tau}(x) = \phi_\theta \left(\tau (B_\theta B_\theta^* + \tau I)^{-1} x \right).} \tag{122}$$

The framework establishes exact invariance in the quotient limit, quantitative approximate invariance at positive regularization, a proximal variational formulation, and stability properties for the learned relation operator.

The categorical formulation also clarifies its relationship to established constructions. Finite-group coinvariants yield Reynolds averaging under invariant inner products, while cellular sheaf Laplacians arise from corresponding adjoint relation operators. These cases demonstrate that QKN is connected to existing geometric and algebraic approaches rather than being an unrelated spectral mechanism.

Two structural principles strengthen the mathematical formulation.

First, functorial parameterization ensures that learned morphism maps continue to satisfy the categorical composition laws. Second, cocone-constrained nondegeneracy guarantees the survival of a prescribed quotient subspace, excluding complete representation collapse.

The hierarchical construction specifies both the feature-vector assembly rule and the propagation of morphism maps, providing a complete categorical interpretation of successive neural aggregation stages.

The resulting theoretical perspective may be summarized as follows:

$$\boxed{\begin{array}{c}\text{Universal categorical structure} \longrightarrow \\ \text{quotient representation} \longrightarrow \\ \text{regularized spectral operator} \longrightarrow \\ \text{structure-preserving neural computation.}\end{array}}$$

The distinctive mathematical aim of QuotientKanNet is to treat categorical equivalence as a first-class architectural constraint while retaining a differentiable representation through resolvent regularization.

References

1. Mac Lane, S. (1998). *Categories for the Working Mathematician*, 2nd ed. Springer.
2. Riehl, E. (2016). *Category Theory in Context*. Dover Publications.
3. Borceux, F. (1994). *Handbook of Categorical Algebra 1: Basic Category Theory*. Cambridge University Press.
4. Shiebler, D. (2022). Kan Extensions in Data Science and Machine Learning. arXiv:2203.09018.
5. Hansen, J., and Ghrist, R. (2019). Toward a Spectral Theory of Cellular Sheaves. *Journal of Applied and Computational Topology*, 3, 315–358. DOI.
6. Hansen, J. (2020). *Laplacians of Cellular Sheaves: Theory and Applications*. Doctoral dissertation, University of Pennsylvania. Dissertation.
7. Bodnar, C., Di Giovanni, F., Chamberlain, B., Lió, P., and Bronstein, M. (2022). Neural Sheaf Diffusion: A Topological Perspective on Heterophily and Oversmoothing in GNNs. *Advances in Neural Information Processing Systems* 35. Proceedings.
8. Barbero, F., Bodnar, C., Sáez de Ocáriz Borde, H., Bronstein, M., Veličković, P., and Liò, P. (2022). Sheaf Neural Networks with Connection Laplacians. *Proceedings of Topological, Algebraic, and Geometric Learning Workshops*, PMLR 196, 28–36. Proceedings.
9. Ben-Israel, A., and Greville, T. N. E. (2003). *Generalized Inverses: Theory and Applications*, 2nd ed. Springer.
10. Rockafellar, R. T. (1970). *Convex Analysis*. Princeton University Press.
11. Saad, Y. (2003). *Iterative Methods for Sparse Linear Systems*, 2nd ed. SIAM.
12. Serre, J.-P. (1977). *Linear Representations of Finite Groups*. Springer.