Fibrations explain all you need! Apr 20, 2025

Preprint · April 2025

DOI: 10.13140/RG.2.2.11628.88967

1 author:

Gary Nan Tie

Mu Risk LLC

86 PUBLICATIONS **27** CITATIONS

SEE PROFILE

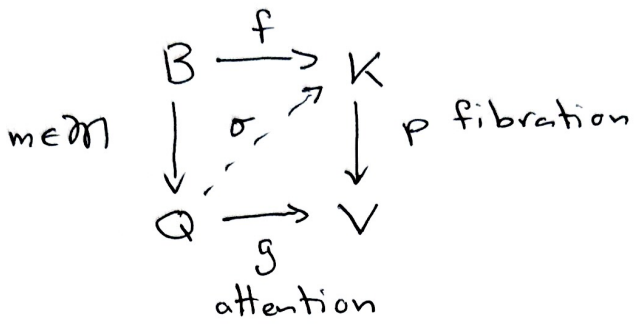
Fibration, Abduction, Attention

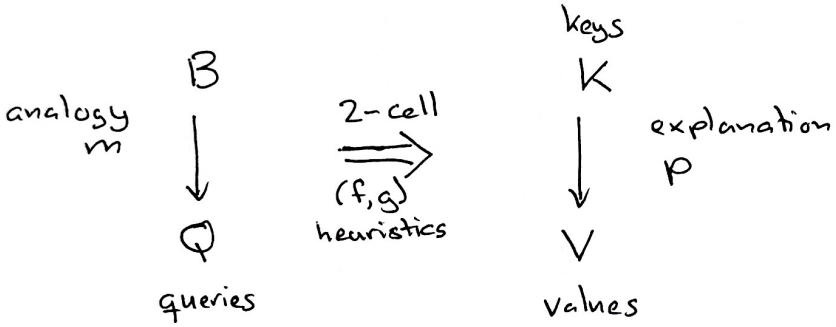
Gary Nan Tie

Easter 2025, Azores

Abstract: A Fibration in a 2-category is both a form of abduction and attention. The interchange law for horizontal and vertical composition has useful implications for the application of abductive reasoning and attention mechanisms. 'Fibrations explain all you need!'

Synopsis:





Given a query $q \in Q$, attention $g: Q \to V$ focuses on value $g(q) \in V$ and abduction $\sigma: Q \to K$ lifts g to K, that is finds a key $\sigma(q) \in K$ that explains value $g(q) \in V$ in the sense $p(\sigma(q)) = g(q)$; query q with attention g yields a key-value pair $(\sigma(q), g(q))$ that is consistent with explanation p ie $p(\sigma(q)) = g(q)$.

σ lifts g ie $p \circ \sigma = g$ and σ extends f to Q

ie $\sigma \circ m = f$ therefore $\forall b \in B \; p(\sigma(m(b))) = pf(b) = g(m(b))$.

$m(B) \subseteq Q$ so m is a weaker explanation of g than p. This is why we call $m \in \mathcal{M} \subseteq Mor(B, Q)$ an analogy of explanation p.

The interchange law for horizontal and vertical composition has useful implications for both abductive reasoning and attention mechanisms that we now explore.

Fibrational Abduction:

A fibration $p: K \to V$ with datum $p \circ f = g \circ m$ provides a lifting $\sigma(m, f, g): Q \to K$ for $m \in \mathcal{M}$.

If we think of p as a map from hypotheses to observations, the effect of hypothesis k is observation $p(k)$. Given observation $g(q)$ in V, we seek an explanation $k \in K$, that is we want to find a $k \in K$ such that the effect of hypothesis k is observation $g(q)$ ie we want $p(k) = g(q)$, denoted $k \models_p g(q)$.

Abduction is the process of deriving a hypothesis k according to effect p that explains observation $g(q)$. Lifting $\sigma: Q \to K$ performs this task since $p(\sigma(q)) = g(q)$. A fibration lifting is performing abductive reasoning with respect to the datum for $m \in \mathcal{M}$; $\sigma(q) \models_p g(q)$.

We can interpret the datum as follows:

$m \in \mathcal{M}$ is an analog of effect p, and the pair (f, g) are heuristics giving the similarity of m with p.

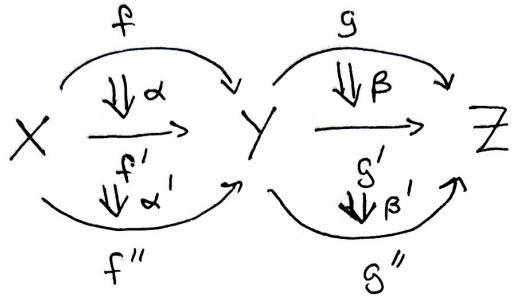
In this sense, we are using analogies to structure our abductive reasoning. Moreover, the 2-categorical compatibility conditions on liftings provides coherence; 'horizonal' for parallel explanations and 'vertical' for hierarchical explanations.

The choice of datum:

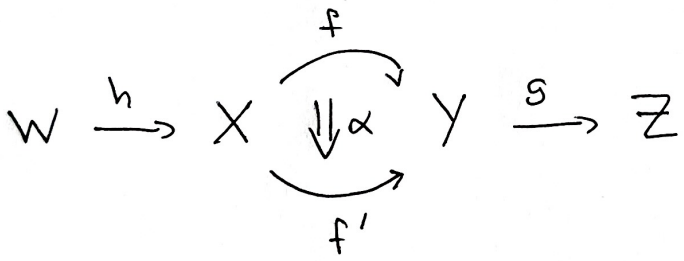
$$\begin{array}{ccc} \mathcal{B} & \xrightarrow{f} & K \\ {\scriptstyle m \in \mathcal{M}}\downarrow & \overset{\sigma}{\nearrow} & \downarrow{\scriptstyle p \text{ fibration}} \\ Q & \xrightarrow[g]{} & V \end{array}$$

will be domain knowledge specific.

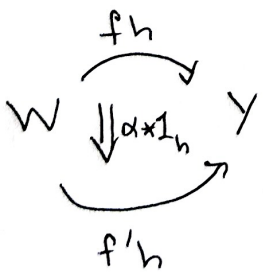
Interchange law for horizontal and vertical compositions:

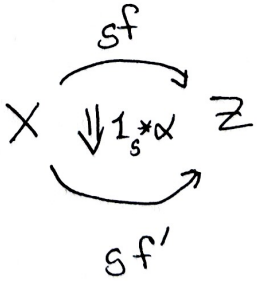


$$(\beta' \circ \beta) * (\alpha' \circ \alpha) = (\beta' * \alpha') \circ (\beta * \alpha)$$



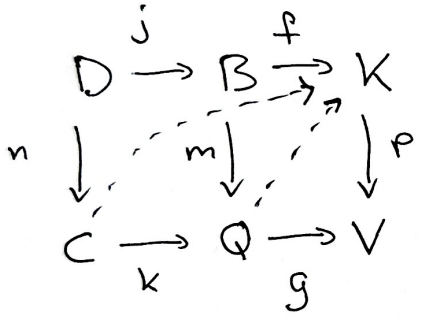
Whiskers are 2-cells:





Fibrations are 2-cells which impose horizontal and vertical compatibility of liftings, and coherence for $\models_p$.

Horizontal compatibity - decompose queries

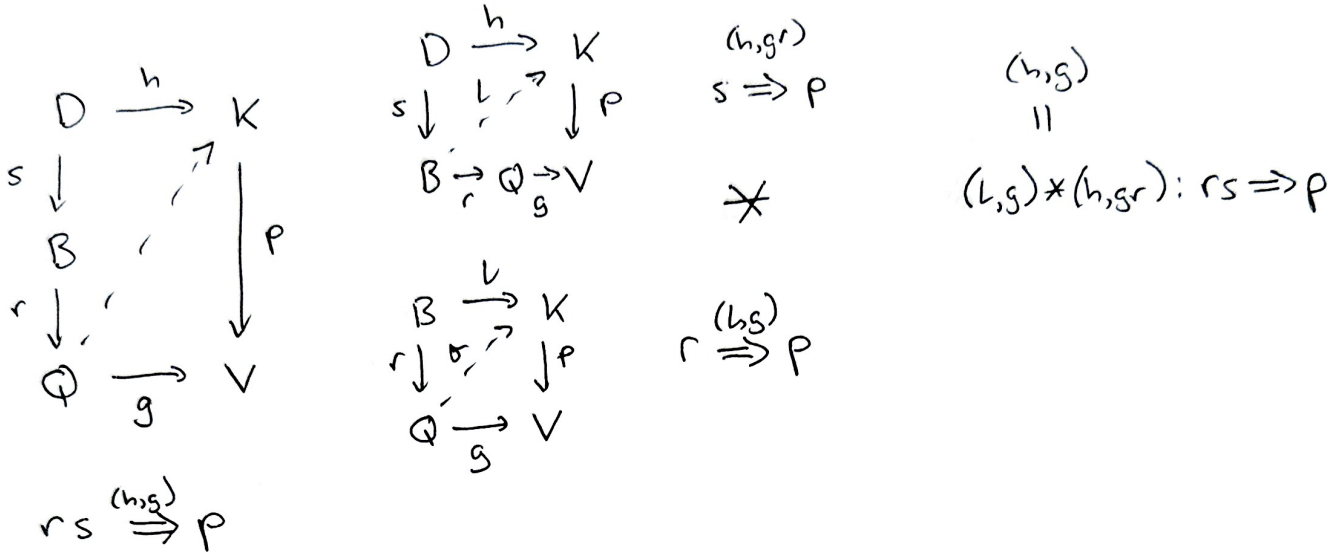


$$n \xRightarrow{(j,k)} m \xRightarrow{(f,g)} p$$

$$(f_j, g^k) = (f,g) \circ (j,k) : n \Rightarrow p$$

Require: $\sigma_{m,f,g} \circ k = \sigma_{n, f_j, g^k}$

Vertical compatibility - decompose analogies



Require: $\sigma_{rs,h,g} = \sigma_{r,L,g}$ where $L = \sigma_{s,h,gr}$

The interchange law implies the meaning of the whole is the same no matter how you slice it!

We are free to decompose queries and analogies and recombine them when reasoning abductively, ensured composition is coherent.

Fibrational Attention:

Fibrational abduction is more than a form of attention - it is a structured generalization of attention with built-in coherence guarantees. Let's analyze the implications of horizontal and vertical compatibility of liftings for fibrational attention.

Given:

- Fibration $p : K \to V$ (keys → values)
- Analogy $m : B \to Q \in \mathcal{M}$
- 2-cell $(f, g) : m \Rightarrow p$ ie $g \circ m = p \circ f$
- Lifting $\sigma_{m,f,g} : Q \to K$ such that $\sigma \circ m = f$ and $p \circ \sigma = g$.

Interpretation:

For a query q, $\sigma(q)$ is the abduced key whose value $p(\sigma(q))$ matches $g(q)$, ie $\sigma(q) \models_p g(q)$.

This is attention as hypothesis generation, not just retrieval of a key-value pair. Moreover, $(f,g): m \Rightarrow p$ being a 2-cell from m to p enforces horizontal and vertical compatibility of attention decompositions.

Horizontal compatibility - <u>Parallel Attention</u>:

Attention steps $n \Rightarrow m \Rightarrow p$ are subject to

$$\sigma(m, f, g) \circ k = \sigma(n, j \circ f, g \circ k)$$

- Parallel processing: If you split a query into sub-queries and process the separately, the results cohere when recomposed.

- Horizontal compatibility ensures that local attention (applied to sub-queries) combines consistly into global attention.

Vertical compatibility - Hierarchical Attention:

$s \Rightarrow p$

$*$ is subject to $\sigma(r \circ s, h, g) = \sigma(r, l, g)$

$r \Rightarrow p$ where $l = \sigma(s, h, g \circ r)$

- Hierarchical processing: Low-level attention (eg pixels → objects) feeds into high-level attention (eg objects → scene) without contradiction.

- Vertical compatibility ensures that coarse attention (eg whole image) refines fine attention (eg patches) coherently.

—

The interchange law enables:

- Robustness: Horizontal and vertical compatibility prevents attention from generating inconsistent hypotheses.

- Compositionality: Attention can be decomposed (parallel) and layered (hierarchical) while preserving meaning.

- Interpretability: A lifting σ makes attention explanatory, not just discriminative.

Hybrid Attention:

$$\mathrm{Attn}(q,k,v) \triangleq \mathrm{softmax}(qk^T)v \in V$$

$$\mathrm{Abd}(q) \triangleq \rho(\sigma(q)) \in V$$

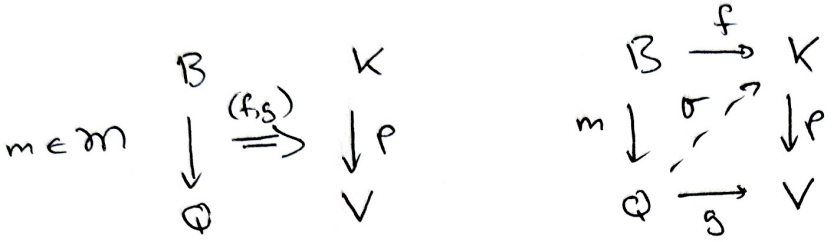
homotopy $H : [0,1] \times Q \to V$

$$H(t,q) = (1-t)\,\mathrm{Attn}(q,k,v) + t\,\mathrm{Abd}(q)$$

where t is a learned parameter dependent on q.

One can choose other forms of attention to blend with Abd.

Fibration Learning:



- Define or learn p from $\{(k_i, v_i)\}_{i=1}^{r}$
- Define or learn m from $\{(b_i, q_i)\}_{i=1}^{s}$
- Learn f and g so that $p \circ f = g \circ m$
- Learn σ so that $\sigma \circ m = f$ and $p \circ \sigma = g$
- Use fibration p for abduction and attention.

2 Note that query-analogy paths of abductive reasoning are coherent; their composition (horizontally and vertically) is the same no matter how parsed.

Summary:

Fibrations coherently unify abductive reasoning and attention mechanisms through 2-categorical compatibility. Fibrational abduction is a structured generalization of attention that coherently guarantees consistent parallel and hierarchical composability of attention components in AI systems. Fibrational attention is interpretable by design, the lifting $\sigma(q)$ explicitly represents the abduced hypothesis, $p(\sigma(q)) = g(q)$. Unlike softmax attention (which retrieves existing keys), this synthesizes explanations.

Fibrations explain all you need!