Neural network layers as parametric spans

mattiagbergomi@gmail.com, pietro.vertechi@protonmail.com

Preprint available at arXiv:2208.00809.

Structure

- Context:
  - neural networks,
  - the *zoo* of linear neural network layers,
  - commonalities among different linear layers.

Structure

- Context:
  - neural networks,
  - the *zoo* of linear neural network layers,
  - commonalities among different linear layers.

- General definition of linear layer:
  - Frobenius integration theory,
  - parametric span.

Structure

- Context:
  - neural networks,
  - the *zoo* of linear neural network layers,
  - commonalities among different linear layers.

- General definition of linear layer:
  - Frobenius integration theory,
  - parametric span.

- Interpret classical layers in the light of this general definition:
  - dense layer,
  - convolutional layer,
  - geometric deep learning.

Neural networks: stack "simple" layers to approximate complex functions

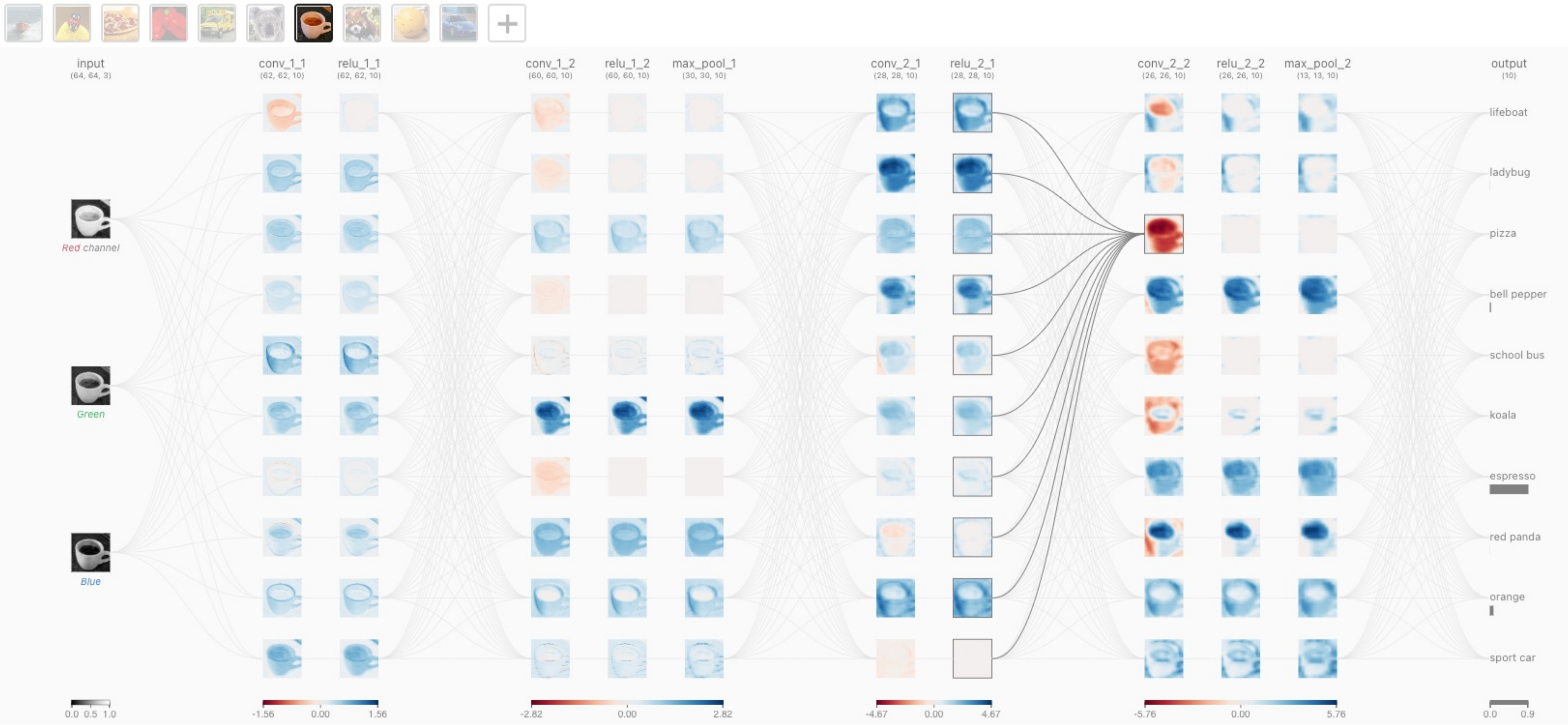
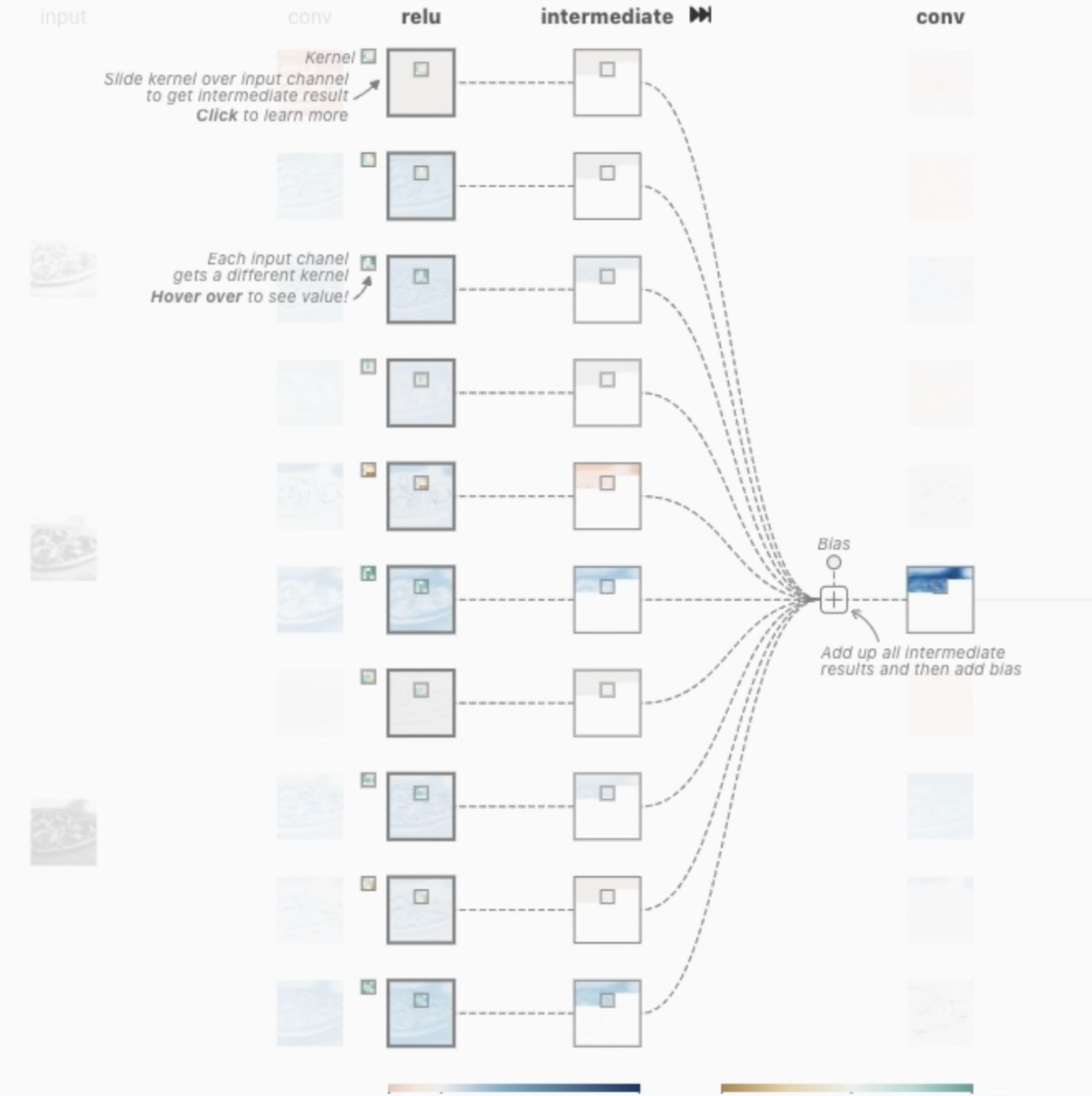


Image credits: https://github.com/poloclub/cnn-explainer



See supplementary video (credits: https://github.com/poloclub/cnn-explainer).

How does one choose what layers to use?

Many linear layers exist:

- dense layer,
- planar convolution,
- transposed convolution,
- group-equivariant convolution,
- graph convolution,
- diffusion convolution,
- geodesic convolution,
- anisotropic convolution,
- ...

How does one choose what layers to use?

Many linear layers exist:

- dense layer,
- planar convolution,
- transposed convolution,
- group-equivariant convolution,
- graph convolution,
- diffusion convolution,
- geodesic convolution,
- anisotropic convolution,
- ...

Each with their own set of hyperparameters.

How does one choose what layers to use?

Many linear layers exist:

- dense layer,
- planar convolution,
- transposed convolution,
- group-equivariant convolution,
- graph convolution,
- diffusion convolution,
- geodesic convolution,
- anisotropic convolution,
- ...

Each with their own set of hyperparameters.

Question 1. What are the *defining features* of a linear layer?

How does one choose what layers to use?

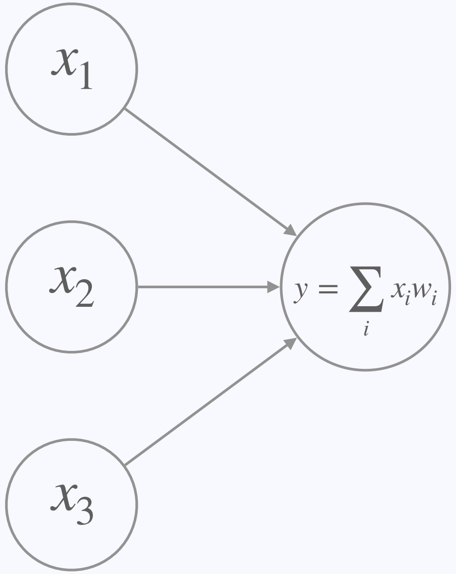
Many linear layers exist:

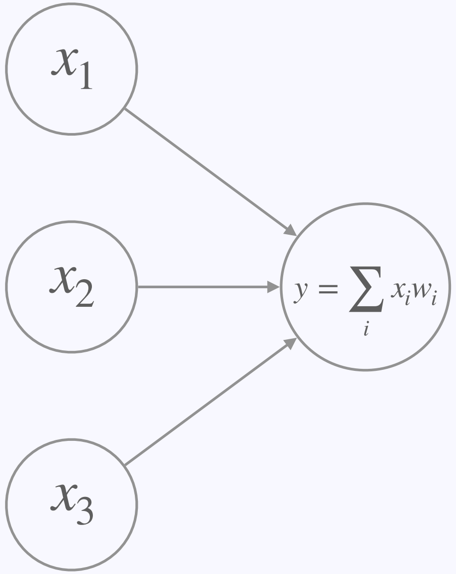
- dense layer,
- planar convolution,
- transposed convolution,
- group-equivariant convolution,
- graph convolution,
- diffusion convolution,
- geodesic convolution,
- anisotropic convolution,
- ...

Each with their own set of hyperparameters.

Question 1. What are the *defining features* of a linear layer?

Question 2. Is there a space of *all* linear layers?

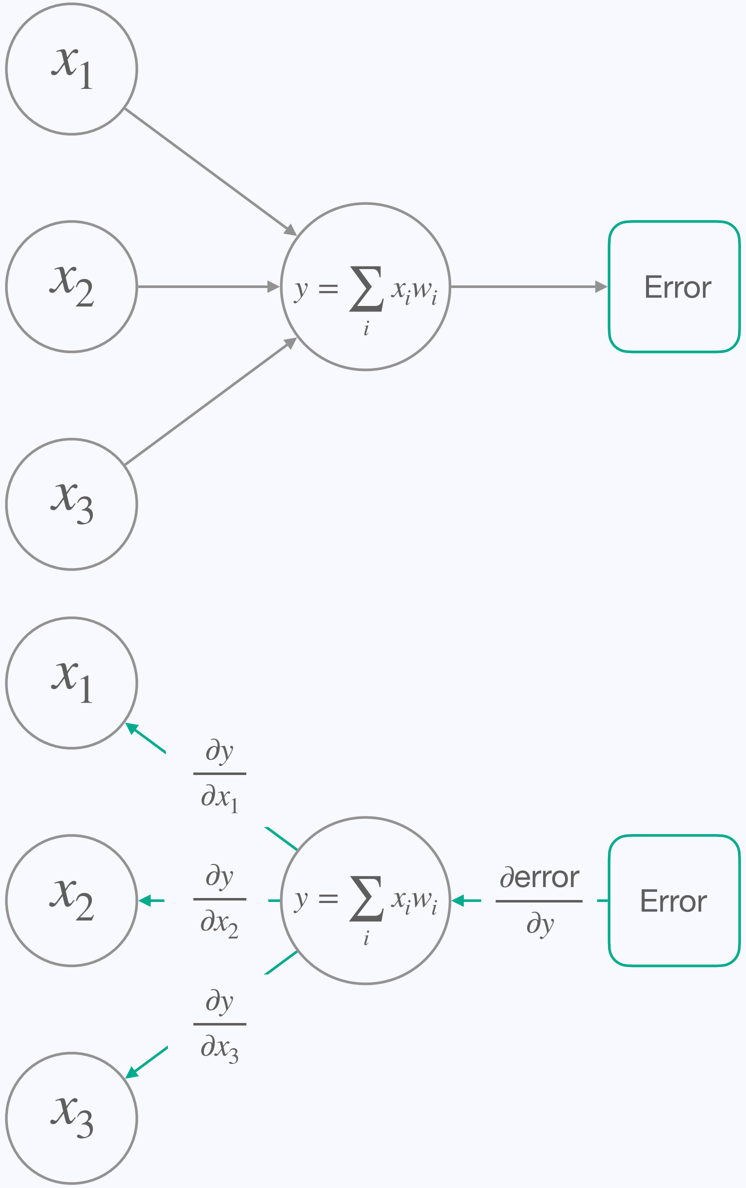

x_1
x_2
x_3
$y = \sum_i x_i w_i$



General requirements

Bilinearity.

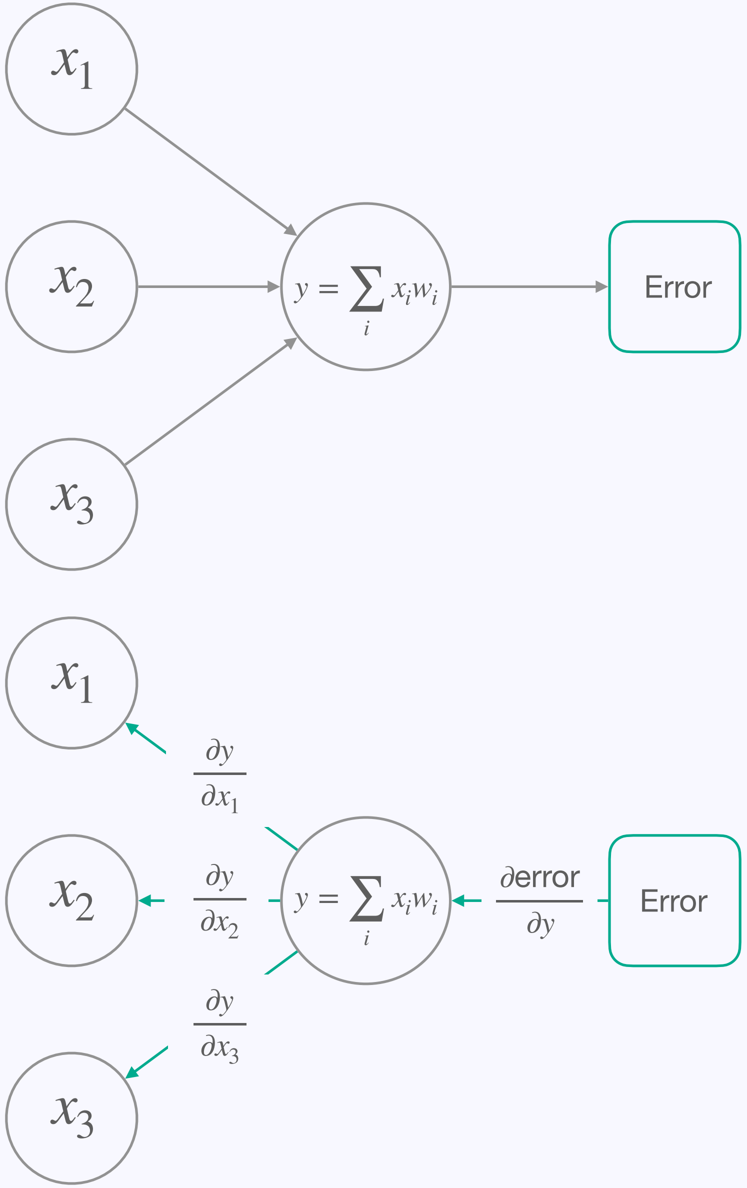
The output value is separately linear in the input value and in the weights.



General requirements

Bilinearity.

The output value is separately linear in the input value and in the weights.



General requirements

Bilinearity.

The output value is separately linear in the input value and in the weights.

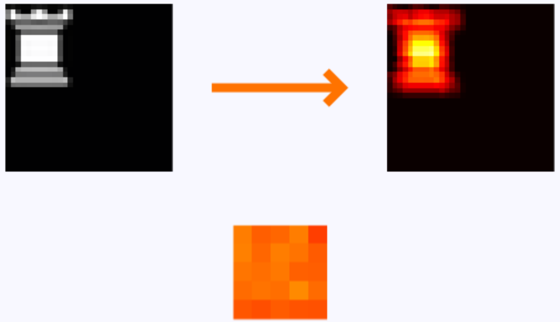
Duality.

The dual (also known as adjoint, or backward pass) exists and is again a linear layer.

Domain-specific requirements

Equivariance.

Convolutional layers owe their success to the notion of equivariance (weight sharing).

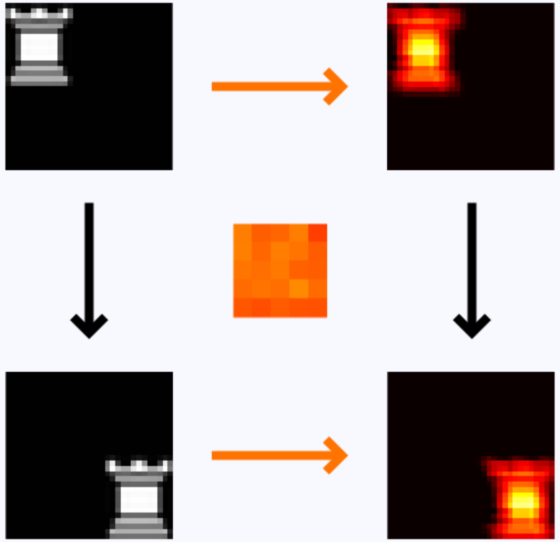


Adapted from Kayan & Gemert (2020).

Domain-specific requirements

Equivariance.

Convolutional layers owe their success to the notion of equivariance (weight sharing).

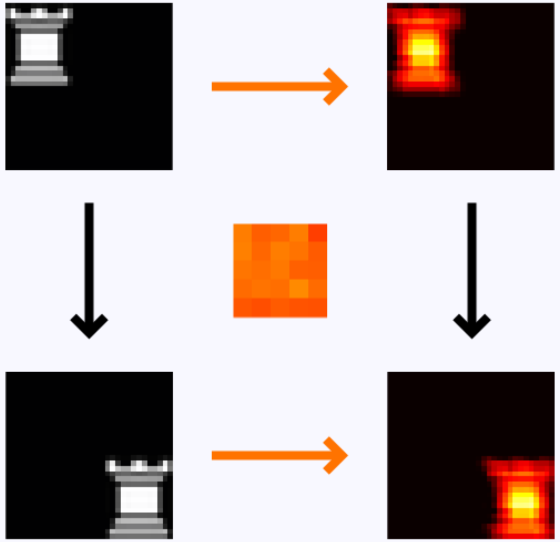


Adapted from Kayan & Gemert (2020).

Domain-specific requirements

Equivariance.

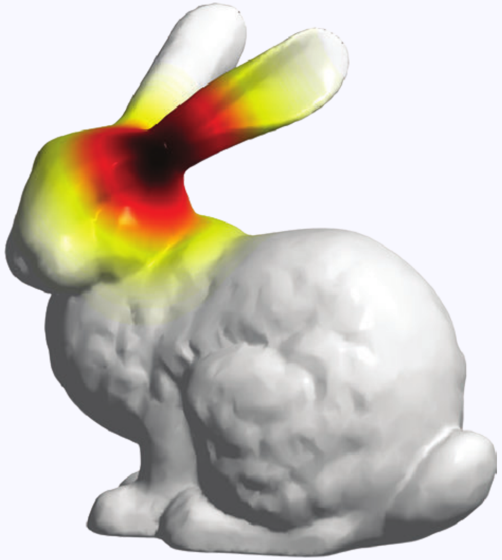
Convolutional layers owe their success to the notion of equivariance (weight sharing).



Adapted from Kayan & Gemert (2020).

Locality.

If there is spatial structure, the inputs of a given output should be localized in space.



Adapted from Bronstein et al. (2017).

Action plan

Give a general unifying definition of linear layer that

- respects bilinearity and duality (easily computable backward pass),

Action plan

Give a general unifying definition of linear layer that

- respects bilinearity and duality (easily computable backward pass),
- incorporates local connectivity and equivariance constraints (weight sharing),

Action plan

Give a general unifying definition of linear layer that

- respects bilinearity and duality (easily computable backward pass),
- incorporates local connectivity and equivariance constraints (weight sharing),
- works on a variety of different spaces, both discrete and continuous.

Action plan

Give a general unifying definition of linear layer that

- respects bilinearity and duality (easily computable backward pass),
- incorporates local connectivity and equivariance constraints (weight sharing),
- works on a variety of different spaces, both discrete and continuous.

Key ingredients.

- Frobenius integration theories
    - formalize (via category theory) the interplay between functions and measures,
    - naturally lead to bilinearity and duality,
    - can be applied to smooth manifolds (*continuous* layers) or finite sets (*discrete* layers).

Action plan

Give a general unifying definition of linear layer that

- respects bilinearity and duality (easily computable backward pass),
- incorporates local connectivity and equivariance constraints (weight sharing),
- works on a variety of different spaces, both discrete and continuous.

Key ingredients.

- Frobenius integration theories
  - formalize (via category theory) the interplay between functions and measures,
  - naturally lead to bilinearity and duality,
  - can be applied to smooth manifolds (*continuous* layers) or finite sets (*discrete* layers).
- Parametric spans
  - formalize locality and weight sharing,
  - recover classical linear neural network layers, both discrete and continuous.

Frobenius integration theories

Let X be a smooth manifold and let us denote

- $\mathcal{F}(X)$ the space of smooth real-valued functions on X,

Frobenius integration theories

Let X be a smooth manifold and let us denote

- $\mathcal{F}(X)$ the space of smooth real-valued functions on X,
- $\mathcal{M}(X)$ the space of smooth densities (of compact support) on X,

Frobenius integration theories

Let X be a smooth manifold and let us denote

- $\mathcal{F}(X)$ the space of smooth real-valued functions on X,
- $\mathcal{M}(X)$ the space of smooth densities (of compact support) on X,
- $\int_X$ the integral.

Frobenius integration theories

Let X be a smooth manifold and let us denote

- $\mathcal{F}(X)$ the space of smooth real-valued functions on X,
- $\mathcal{M}(X)$ the space of smooth densities (of compact support) on X,
- $\int_X$ the integral.

Structure.

- $\mathcal{F}(X)$ is a commutative $\mathbb{R}$-algebra:
    - linear combinations of smooth functions,
    - pointwise multiplication of smooth functions.

Frobenius integration theories

Let X be a smooth manifold and let us denote

- $\mathcal{F}(X)$ the space of smooth real-valued functions on X,
- $\mathcal{M}(X)$ the space of smooth densities (of compact support) on X,
- $\int_X$ the integral.

Structure.

- $\mathcal{F}(X)$ is a commutative $\mathbb{R}$-algebra:
  - linear combinations of smooth functions,
  - pointwise multiplication of smooth functions.
- $\mathcal{M}(X)$ is an $\mathcal{F}(X)$-module:
  - linear combinations of smooth densities,
  - pointwise multiplication of a smooth density by a smooth function.

Frobenius integration theories

Let X be a smooth manifold and let us denote

- $\mathcal{F}(X)$ the space of smooth real-valued functions on X,
- $\mathcal{M}(X)$ the space of smooth densities (of compact support) on X,
- $\int_X$ the integral.

Structure.

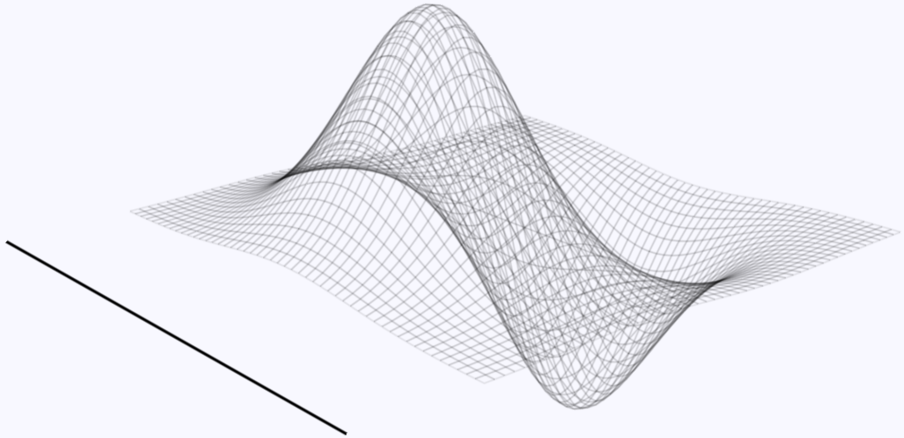
- $\mathcal{F}(X)$ is a commutative $\mathbb{R}$-algebra:
  - linear combinations of smooth functions,
  - pointwise multiplication of smooth functions.
- $\mathcal{M}(X)$ is an $\mathcal{F}(X)$-module:
  - linear combinations of smooth densities,
  - pointwise multiplication of a smooth density by a smooth function.
- $\int_X$ is an $\mathbb{R}$-linear functional on $\mathcal{M}(X)$:
  - integrating a smooth density yields a real number.

Frobenius integration theories

Definition. We say that a smooth map

$$f\colon X \to Y$$

is a *submersion* if its differential is, at every point, surjective.



Visualization powered by Makie and Beautiful Makie

Frobenius integration theories

Definition. We say that a smooth map

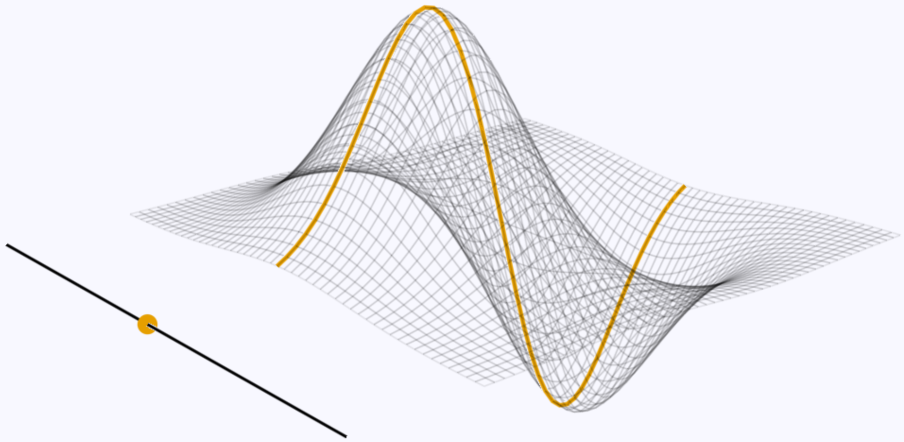
$$f\colon X \to Y$$

is a *submersion* if its differential is, at every point, surjective.

Inverse function theorem. For each point $p \in Y$, the *fiber*

$$X_p := f^{-1}(\{p\})$$

is a submanifold of X.



Visualization powered by Makie and Beautiful Makie

Frobenius integration theories

Definition. We say that a smooth map

$$f\colon X \to Y$$

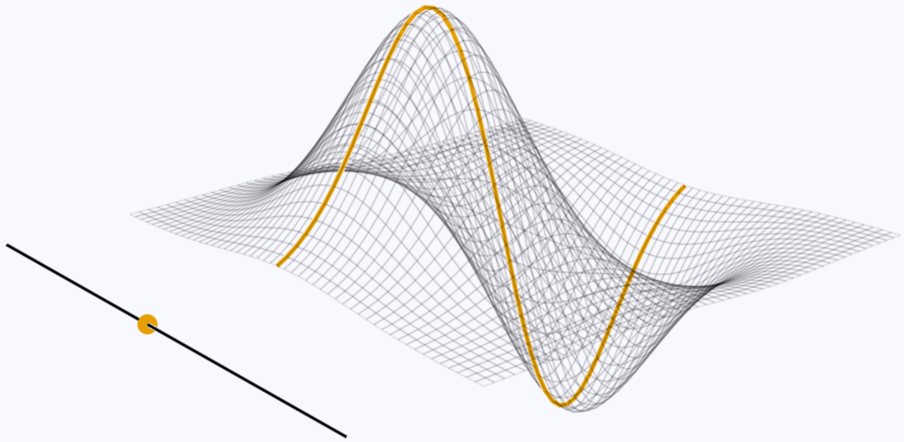
is a *submersion* if its differential is, at every point, surjective.

Inverse function theorem. For each point $p \in Y$, the *fiber*

$$X_p := f^{-1}(\{p\})$$

is a submanifold of X.

Integration along fibers. Transform a quantity on X into a quantity on Y (linear pooling).



Visualization powered by Makie and Beautiful Makie

Frobenius integration theories

Structure.

$\mathcal{F}$ and $\mathcal{M}$ are functors of opposite variance.

A smooth submersion $f\colon X \to Y$ induces

- an algebra homomorphism $f^*\colon \mathcal{F}(Y) \to \mathcal{F}(X)$ (function pullback, given by precomposition),
- an $\mathbb{R}$-linear map $f_*\colon \mathcal{M}(X) \to \mathcal{M}(Y)$ (density pushforward, given by integration along fibers).

Frobenius integration theories

Structure.

$\mathcal{F}$ and $\mathcal{M}$ are functors of opposite variance.

A smooth submersion $f\colon X \to Y$ induces

- an algebra homomorphism $f^*\colon \mathcal{F}(Y) \to \mathcal{F}(X)$ (function pullback, given by precomposition),
- an $\mathbb{R}$-linear map $f_*\colon \mathcal{M}(X) \to \mathcal{M}(Y)$ (density pushforward, given by integration along fibers).

Properties.

Furthermore, for all $\mu \in \mathcal{M}(X)$ and $y \in \mathcal{F}(Y)$,

- $f_*(f^*y \cdot \mu) = y \cdot f_*\mu$ (Frobenius reciprocity),

Frobenius integration theories

Structure.

$\mathcal{F}$ and $\mathcal{M}$ are functors of opposite variance.

A smooth submersion $f\colon X \to Y$ induces

- an algebra homomorphism $f^*\colon \mathcal{F}(Y) \to \mathcal{F}(X)$ (function pullback, given by precomposition),
- an $\mathbb{R}$-linear map $f_*\colon \mathcal{M}(X) \to \mathcal{M}(Y)$ (density pushforward, given by integration along fibers).

Properties.

Furthermore, for all $\mu \in \mathcal{M}(X)$ and $y \in \mathcal{F}(Y)$,

- $f_*(f^*y \cdot \mu) = y \cdot f_*\mu$ (Frobenius reciprocity),
- $\int_X \mu = \int_Y f_*\mu$ (Fubini's theorem).

Frobenius integration theories

Structure.

$\mathcal{F}$ and $\mathcal{M}$ are functors of opposite variance.

A smooth submersion $f\colon X \to Y$ induces

- an algebra homomorphism $f^*\colon \mathcal{F}(Y) \to \mathcal{F}(X)$ (function pullback, given by precomposition),
- an $\mathbb{R}$-linear map $f_*\colon \mathcal{M}(X) \to \mathcal{M}(Y)$ (density pushforward, given by integration along fibers).

Properties.

Furthermore, for all $\mu \in \mathcal{M}(X)$ and $y \in \mathcal{F}(Y)$,

- $f_*(f^*y \cdot \mu) = y \cdot f_*\mu$ (Frobenius reciprocity),
- $\int_X \mu = \int_Y f_*\mu$ (Fubini's theorem).

As a consequence, f_* and f^* are "adjoint" operators: $\int_X f^*y \cdot \mu = \int_Y y \cdot f_*\mu$.

Frobenius integration theories

Structure.

$\mathcal{F}$ and $\mathcal{M}$ are functors of opposite variance.

A smooth submersion $f\colon X \to Y$ induces

- an algebra homomorphism $f^*\colon \mathcal{F}(Y) \to \mathcal{F}(X)$ (function pullback, given by precomposition),
- an $\mathbb{R}$-linear map $f_*\colon \mathcal{M}(X) \to \mathcal{M}(Y)$ (density pushforward, given by integration along fibers).

Properties.

Furthermore, for all $\mu \in \mathcal{M}(X)$ and $y \in \mathcal{F}(Y)$,

- $f_*(f^*y \cdot \mu) = y \cdot f_*\mu$ (Frobenius reciprocity),
- $\int_X \mu = \int_Y f_*\mu$ (Fubini's theorem).

As a consequence, f_* and f^* are "adjoint" operators: $\int_X f^*y \cdot \mu = \int_Y y \cdot f_*\mu$.

Punch line. f_* is the backward pass of f^* and vice versa.

Frobenius integration theories

Propositions 1 and 2. *All* the structures and properties defined above can be succinctly described as a functor

$$\mathbf{Subm} \to \mathbf{Gr}(\mathbf{Mod}/\mathbb{R}),$$

where $\mathbf{Subm}$ is the category of smooth manifolds and submersions.

Frobenius integration theories

Propositions 1 and 2. *All* the structures and properties defined above can be succinctly described as a functor

$$\mathbf{Subm} \to \mathbf{Gr}(\mathbf{Mod}/\mathbb{R}),$$

where $\mathbf{Subm}$ is the category of smooth manifolds and submersions.

Intuition on $\mathbf{Gr}(\mathbf{Mod}/\mathbb{R})$.

- For a commutative $\mathbb{R}$-algebra A, we construct $\mathbf{Mod}_A/\mathbb{R}$, the category of A-modules with an $\mathbb{R}$-linear functional.

Frobenius integration theories

Propositions 1 and 2. *All* the structures and properties defined above can be succinctly described as a functor

$$\mathbf{Subm} \to \mathbf{Gr}(\mathbf{Mod}/\mathbb{R}),$$

where $\mathbf{Subm}$ is the category of smooth manifolds and submersions.

Intuition on $\mathbf{Gr}(\mathbf{Mod}/\mathbb{R})$.

- For a commutative $\mathbb{R}$-algebra A, we construct $\mathbf{Mod}_A/\mathbb{R}$, the category of A-modules with an $\mathbb{R}$-linear functional.
- These categories form a functor $\mathbf{Mod}/\mathbb{R}\colon \mathbf{CAlg}_{\mathbb{R}}^{\mathrm{op}} \to \mathbf{Cat}$.

Frobenius integration theories

Propositions 1 and 2. *All* the structures and properties defined above can be succinctly described as a functor

$$\mathbf{Subm} \to \mathbf{Gr}(\mathbf{Mod}/\mathbb{R}),$$

where $\mathbf{Subm}$ is the category of smooth manifolds and submersions.

Intuition on $\mathbf{Gr}(\mathbf{Mod}/\mathbb{R})$.

- For a commutative $\mathbb{R}$-algebra A, we construct $\mathbf{Mod}_A/\mathbb{R}$, the category of A-modules with an $\mathbb{R}$-linear functional.
- These categories form a functor $\mathbf{Mod}/\mathbb{R} \colon \mathbf{CAlg}_{\mathbb{R}}^{\mathrm{op}} \to \mathbf{Cat}$.
- We glue all these categories together by means of the *covariant Grothendieck construction* [1].

[1] Spivak, D. I. (2019). Generalized Lens Categories via functors $\mathcal{C}^{\mathrm{op}} \to \mathbf{Cat}$. arXiv preprint arXiv:1908.02202.

Frobenius integration theories

Propositions 1 and 2. *All* the structures and properties defined above can be succinctly described as a functor

$$\mathbf{Subm} \to \mathbf{Gr}(\mathbf{Mod}/\mathbb{R}),$$

where $\mathbf{Subm}$ is the category of smooth manifolds and submersions.

Definition. A functor from some category $\mathcal{C}$ to $\mathbf{Gr}(\mathbf{Mod}/\mathbb{R})$ is a *Frobenius integration theory* on $\mathcal{C}$.

Frobenius integration theories

Propositions 1 and 2. *All* the structures and properties defined above can be succinctly described as a functor

$$\mathbf{Subm} \to \mathbf{Gr}(\mathbf{Mod}/\mathbb{R}),$$

where $\mathbf{Subm}$ is the category of smooth manifolds and submersions.

Definition. A functor from some category $\mathcal{C}$ to $\mathbf{Gr}(\mathbf{Mod}/\mathbb{R})$ is a *Frobenius integration theory* on $\mathcal{C}$.

Examples.

There exist Frobenius integration theories based on

- smooth manifolds and submersions,

Frobenius integration theories

Propositions 1 and 2. *All* the structures and properties defined above can be succinctly described as a functor

$$\mathbf{Subm} \to \mathbf{Gr}(\mathbf{Mod}/\mathbb{R}),$$

where $\mathbf{Subm}$ is the category of smooth manifolds and submersions.

Definition. A functor from some category $\mathcal{C}$ to $\mathbf{Gr}(\mathbf{Mod}/\mathbb{R})$ is a *Frobenius integration theory* on $\mathcal{C}$.

Examples.

There exist Frobenius integration theories based on

- smooth manifolds and submersions,
- measurable spaces and nullset-preserving measurable functions (see manuscript),

Frobenius integration theories

Propositions 1 and 2. *All* the structures and properties defined above can be succinctly described as a functor

$$\mathbf{Subm} \to \mathbf{Gr}(\mathbf{Mod}/\mathbb{R}),$$

where $\mathbf{Subm}$ is the category of smooth manifolds and submersions.

Definition. A functor from some category $\mathcal{C}$ to $\mathbf{Gr}(\mathbf{Mod}/\mathbb{R})$ is a *Frobenius integration theory* on $\mathcal{C}$.

Examples.

There exist Frobenius integration theories based on

- smooth manifolds and submersions,
- measurable spaces and nullset-preserving measurable functions (see manuscript),
- finite sets and functions (exercise).

Take-home message

- The Grothendieck construction yields a global category $\mathbf{Gr}(\mathbf{Mod}/\mathbb{R})$ that axiomatizes the behavior of
    - functions,
    - densities,
    - integrals.

Take-home message

- The Grothendieck construction yields a global category $\mathbf{Gr}(\mathbf{Mod}/\mathbb{R})$ that axiomatizes the behavior of
  - functions,
  - densities,
  - integrals.
- A *Frobenius integration theory* is a functor to $\mathbf{Gr}(\mathbf{Mod}/\mathbb{R})$.

Take-home message

- The Grothendieck construction yields a global category $\mathbf{Gr}(\mathbf{Mod}/\mathbb{R})$ that axiomatizes the behavior of
  - functions,
  - densities,
  - integrals.
- A *Frobenius integration theory* is a functor to $\mathbf{Gr}(\mathbf{Mod}/\mathbb{R})$.
- There is a natural Frobenius integration theory

$$\mathbf{Subm} \to \mathbf{Gr}(\mathbf{Mod}/\mathbb{R})$$

  on the category of smooth manifolds and submersions.

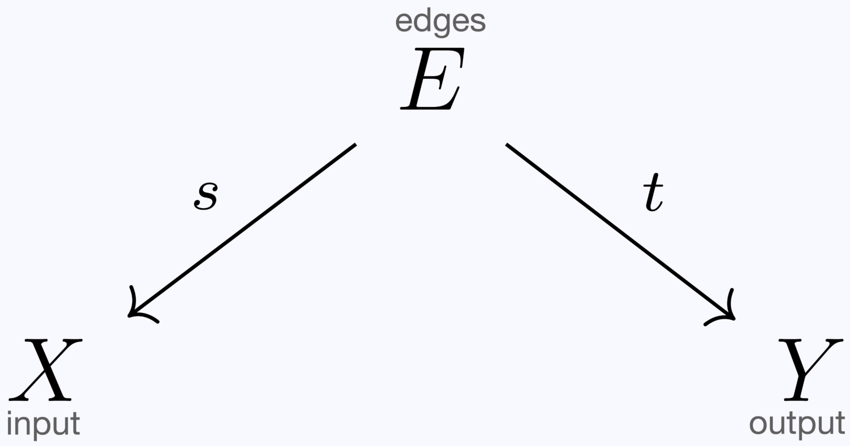
Take-home message

- The Grothendieck construction yields a global category $\mathbf{Gr}(\mathbf{Mod}/\mathbb{R})$ that axiomatizes the behavior of
  - functions,
  - densities,
  - integrals.
- A *Frobenius integration theory* is a functor to $\mathbf{Gr}(\mathbf{Mod}/\mathbb{R})$.
- There is a natural Frobenius integration theory

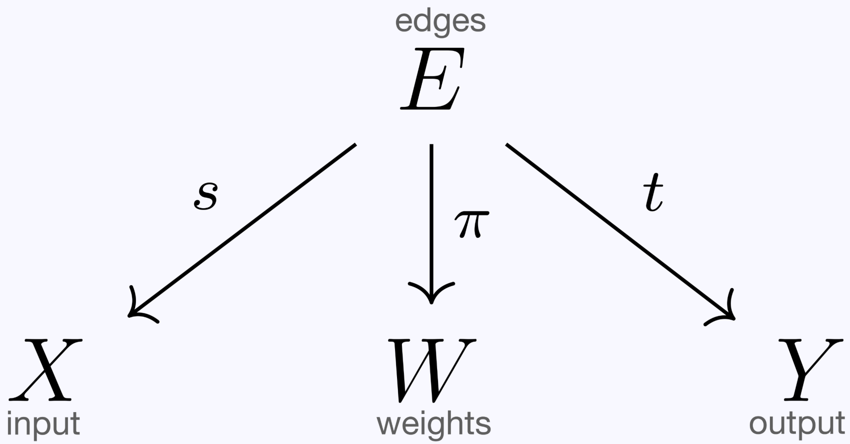
$$\mathbf{Subm} \to \mathbf{Gr}(\mathbf{Mod}/\mathbb{R})$$

  on the category of smooth manifolds and submersions.
- Frobenius integration theories naturally lead to dualizable bilinear operators.

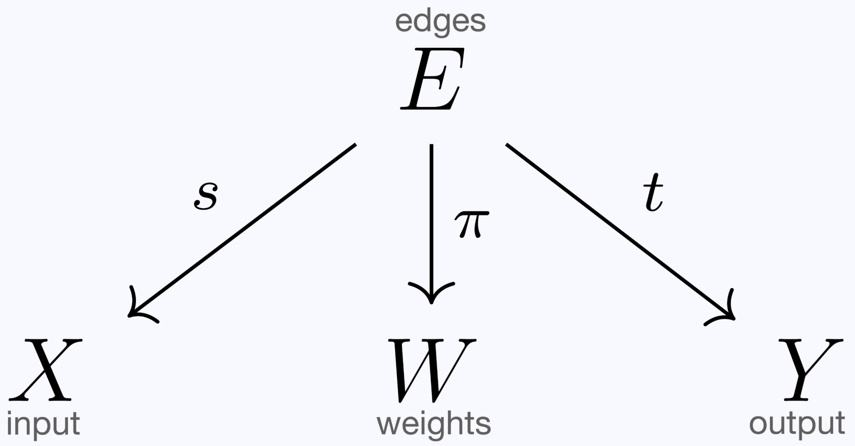
Spans

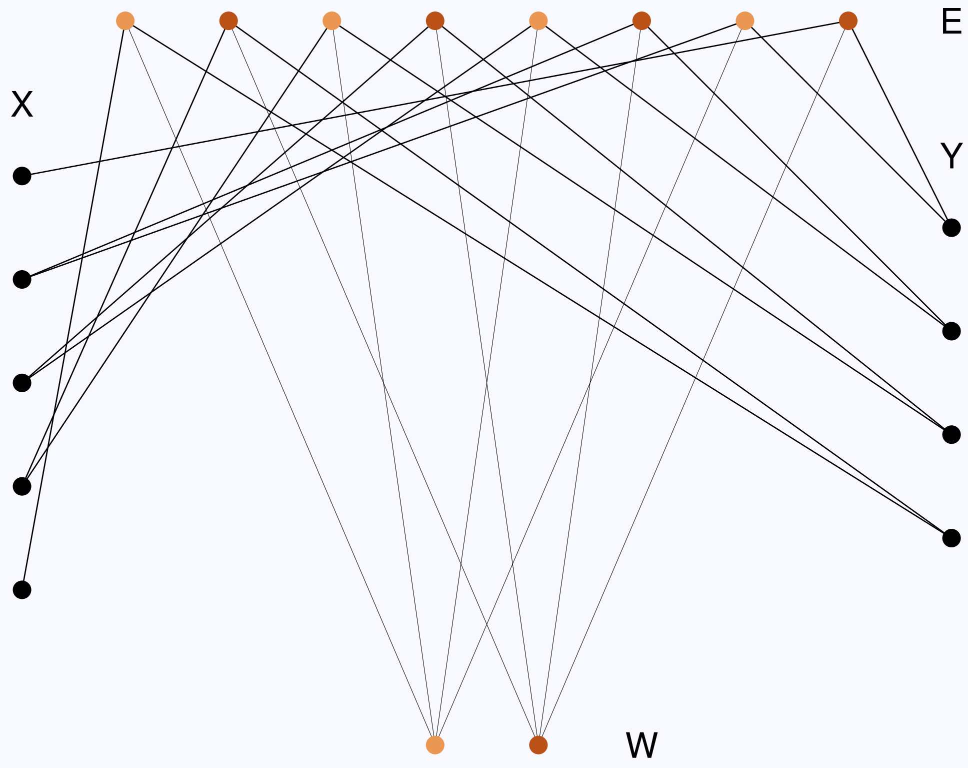


Parametric spans



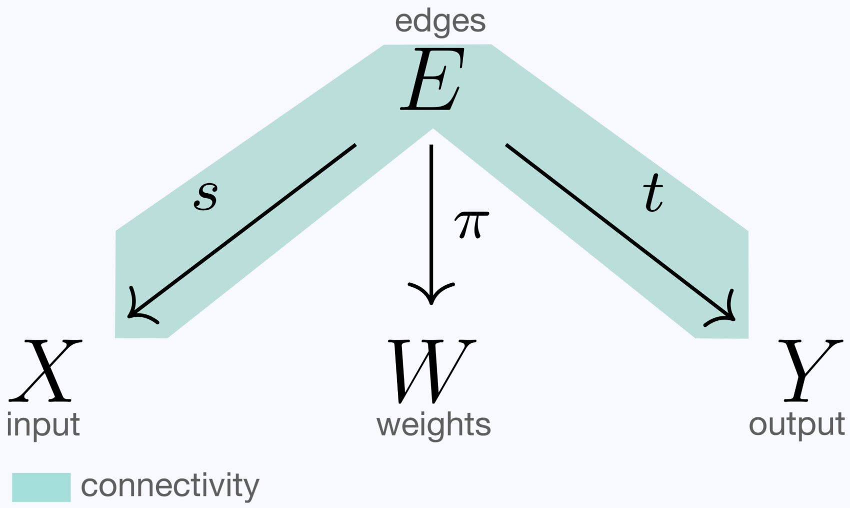
Parametric spans



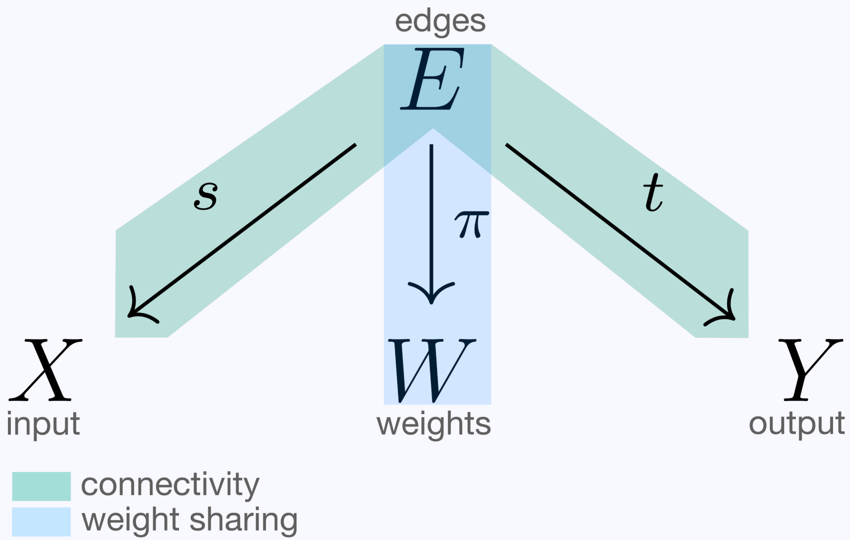


See supplementary video.

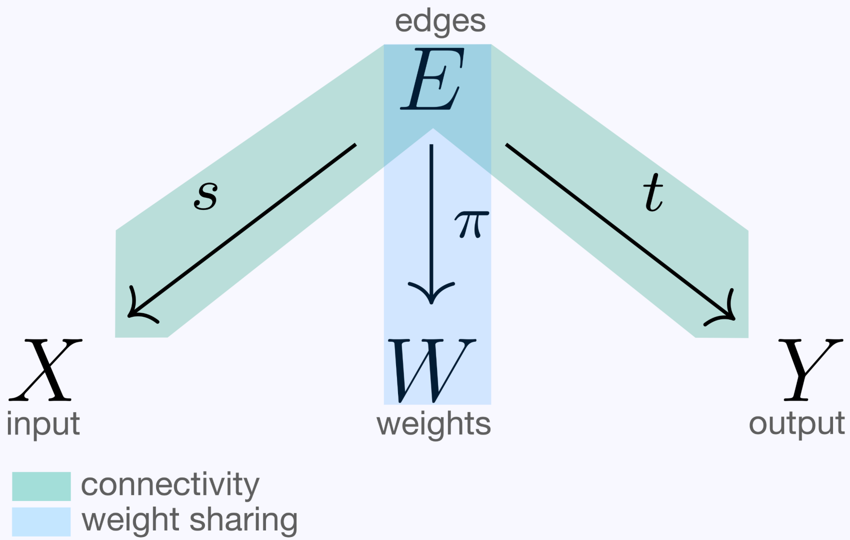
Parametric spans



Parametric spans



Parametric spans



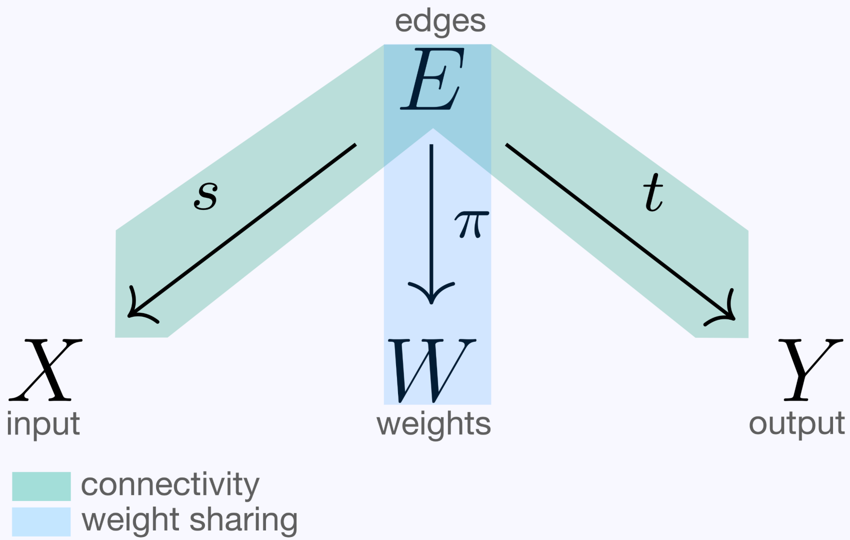
Assumptions.

Frobenius integration theory on $\mathcal{C}$.

E, X, W, Y are objects in $\mathcal{C}$.

s, π, t are morphisms in $\mathcal{C}$.

Parametric spans



Assumptions.

Frobenius integration theory on $\mathcal{C}$.

E, X, W, Y are objects in $\mathcal{C}$.

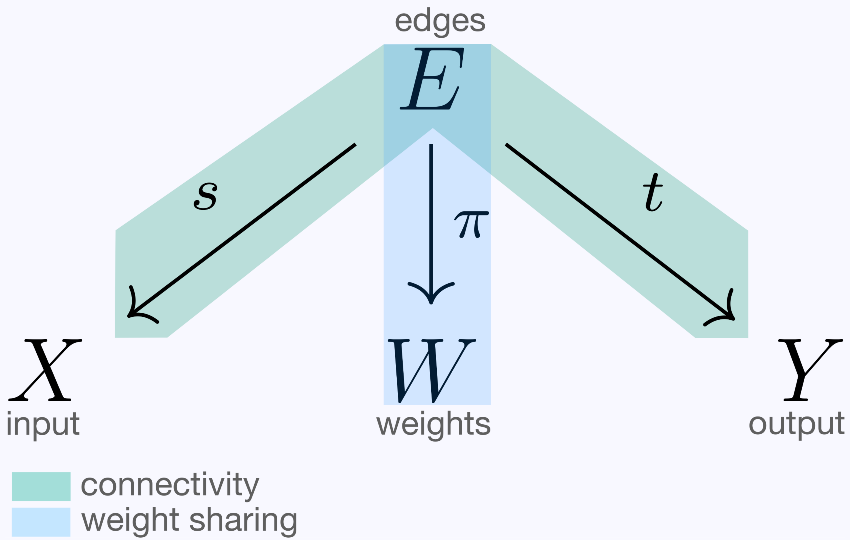
s, π, t are morphisms in $\mathcal{C}$.

Proposition 3. A parametric span and $\mu \in \mathcal{M}(E)$ induce a layer (separately $\mathbb{R}$-linear in x and w)

$$\mathcal{F}(X) \times \mathcal{F}(W) \to \mathcal{M}(Y)$$

$$(x, w) \mapsto t_*(s^*x \cdot \pi^*w \cdot \mu).$$

Parametric spans



Proposition 3. A parametric span and $\mu \in \mathcal{M}(E)$ induce a layer (separately $\mathbb{R}$-linear in x and w)

$$\mathcal{F}(X) \times \mathcal{F}(W) \to \mathcal{M}(Y)$$

$$(x, w) \mapsto t_*(s^*x \cdot \pi^*w \cdot \mu).$$

Proposition 4. The dual (backward pass) can be computed by permuting the legs of the parametric span.

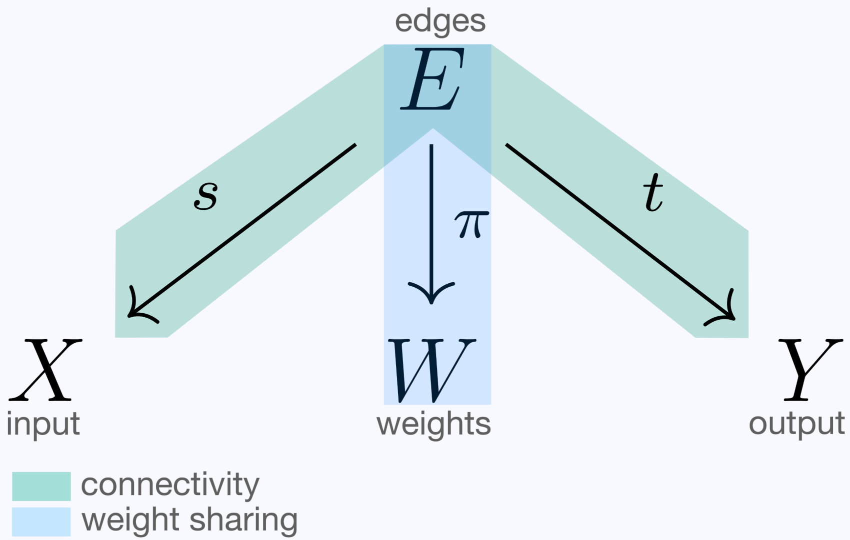
Assumptions.

Frobenius integration theory on $\mathcal{C}$.

E, X, W, Y are objects in $\mathcal{C}$.

s, π, t are morphisms in $\mathcal{C}$.

Parametric spans



Proposition 3. A parametric span and $\mu \in \mathcal{M}(E)$ induce a layer (separately $\mathbb{R}$-linear in x and w)

$$\mathcal{F}(X) \times \mathcal{F}(W) \to \mathcal{M}(Y)$$

$$(x, w) \mapsto t_*(s^* x \cdot \pi^* w \cdot \mu).$$

Proposition 4. The dual (backward pass) can be computed by permuting the legs of the parametric span.

Assumptions.

Frobenius integration theory on $\mathcal{C}$.

E, X, W, Y are objects in $\mathcal{C}$.

s, π, t are morphisms in $\mathcal{C}$.

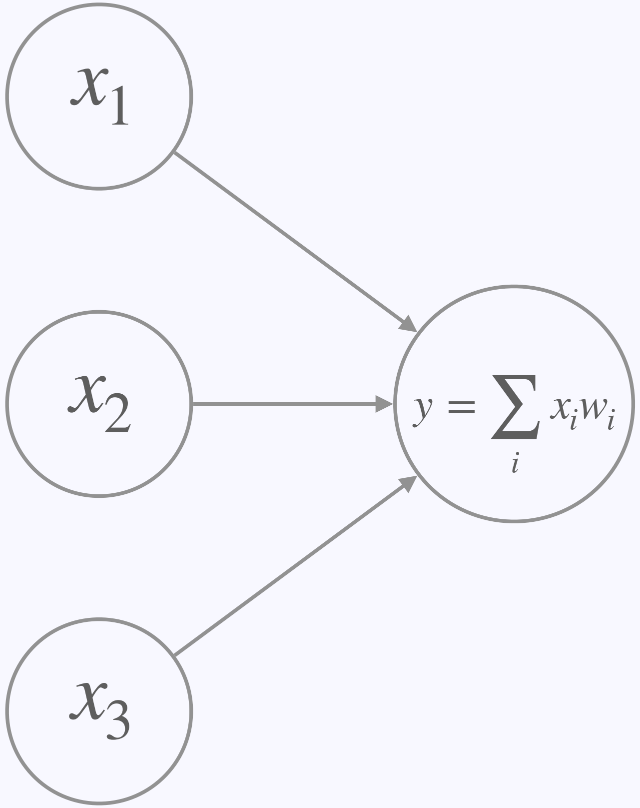
Punch line.

Parametric spans can be used to define linear layers with

- local connectivity,
- weight sharing,
- computable backward pass.

Classical architectures

Dense layer



Classical architectures

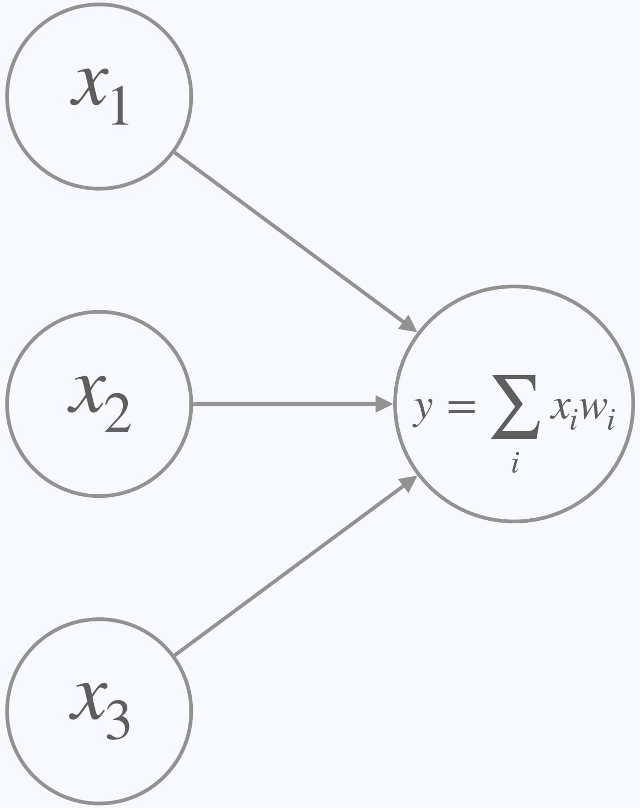
Dense layer

Features

- Domain: Discrete
- Symmetry: No symmetry

Parametric Span

$$\begin{array}{ccccc} & & n_i \times n_o & & \\ & \swarrow & \downarrow & \searrow & \\ n_i & & n_i \times n_o & & n_o \end{array}$$



Classical architectures

Convolutional layer

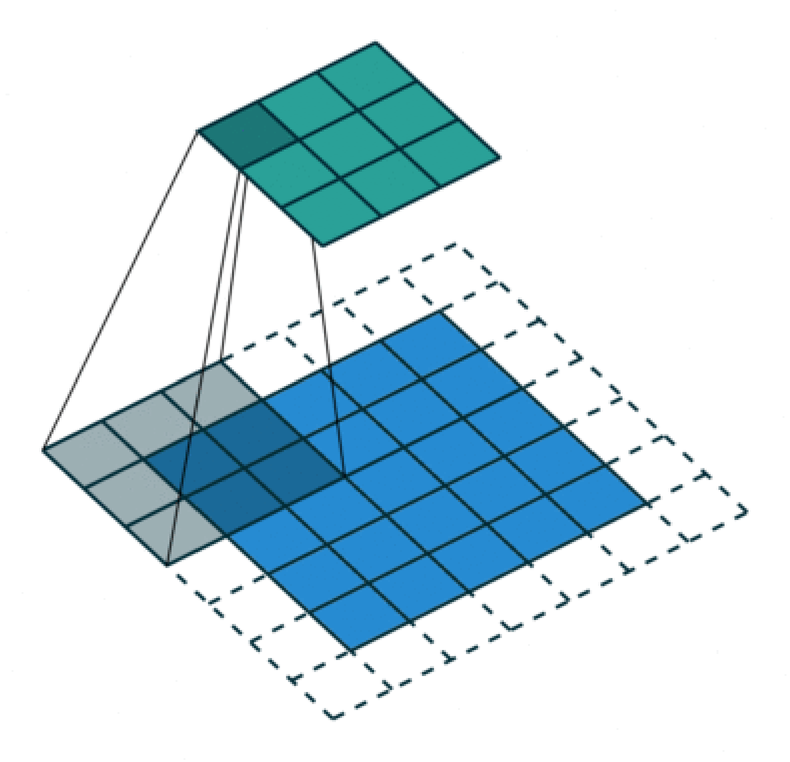


Image credits: Đặng Hà Thế Hiển

Classical architectures

Convolutional layer

Features

- Domain: Discrete & continuous
- Symmetry: Translation

Parametric Span

$$\begin{array}{ccc} & F \times S_o & \\ \swarrow & \downarrow & \searrow \\ S_i & F & S_o \end{array}$$

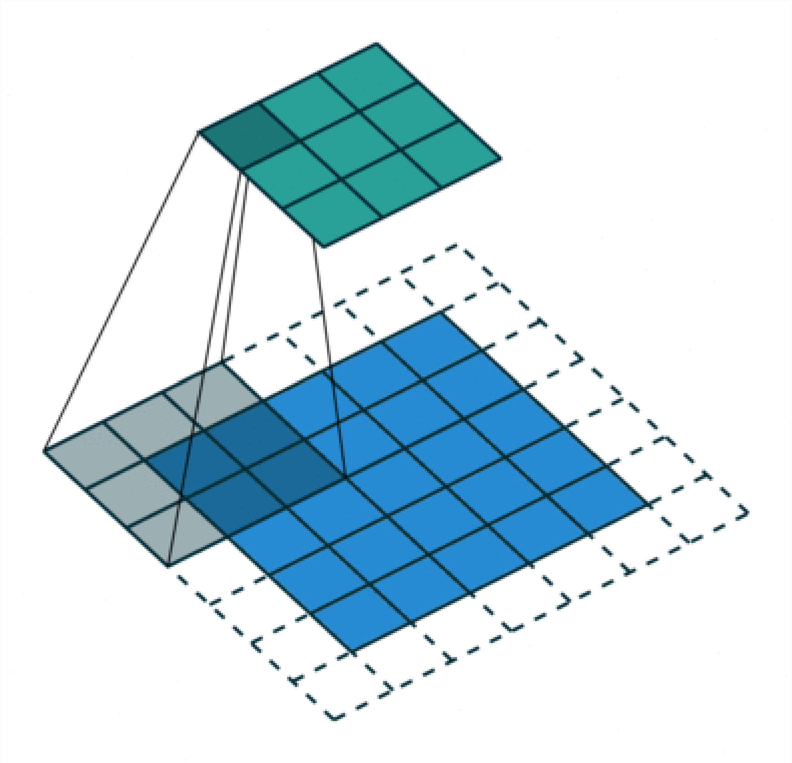


Image credits: Đặng Hà Thế Hiển

Classical architectures

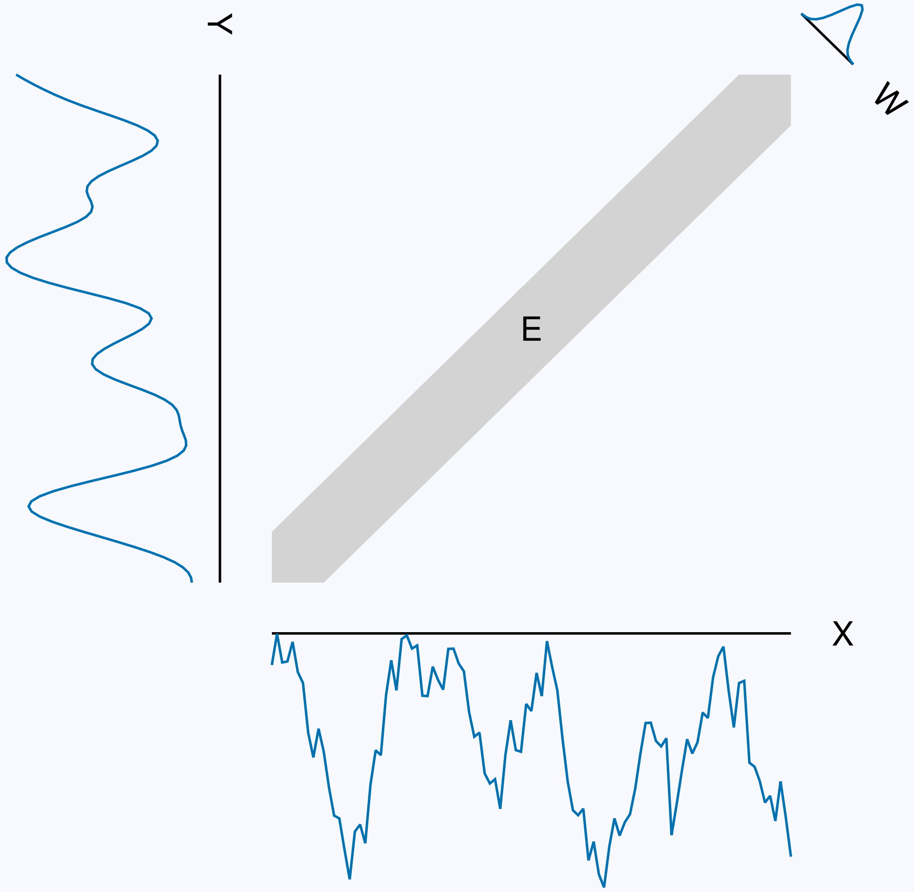
Convolutional layer

Features

- Domain: Discrete & continuous
- Symmetry: Translation

Parametric Span

$$\begin{array}{ccccc} & & F \times S_o & & \\ & \swarrow & \downarrow & \searrow & \\ S_i & & F & & S_o \end{array}$$



Classical architectures

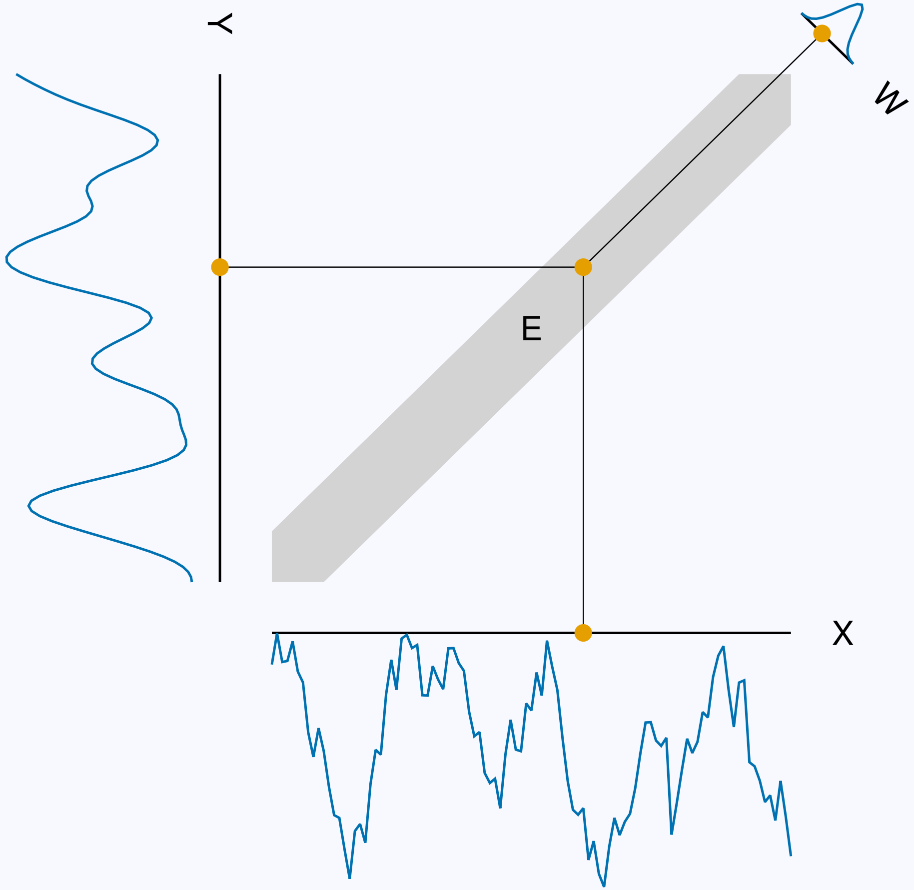
Convolutional layer

Features

- Domain: Discrete & continuous
- Symmetry: Translation

Parametric Span

$$\begin{array}{ccccc} & & F \times S_o & & \\ & \swarrow & \downarrow & \searrow & \\ S_i & & F & & S_o \end{array}$$



Classical architectures

Convolutional layer

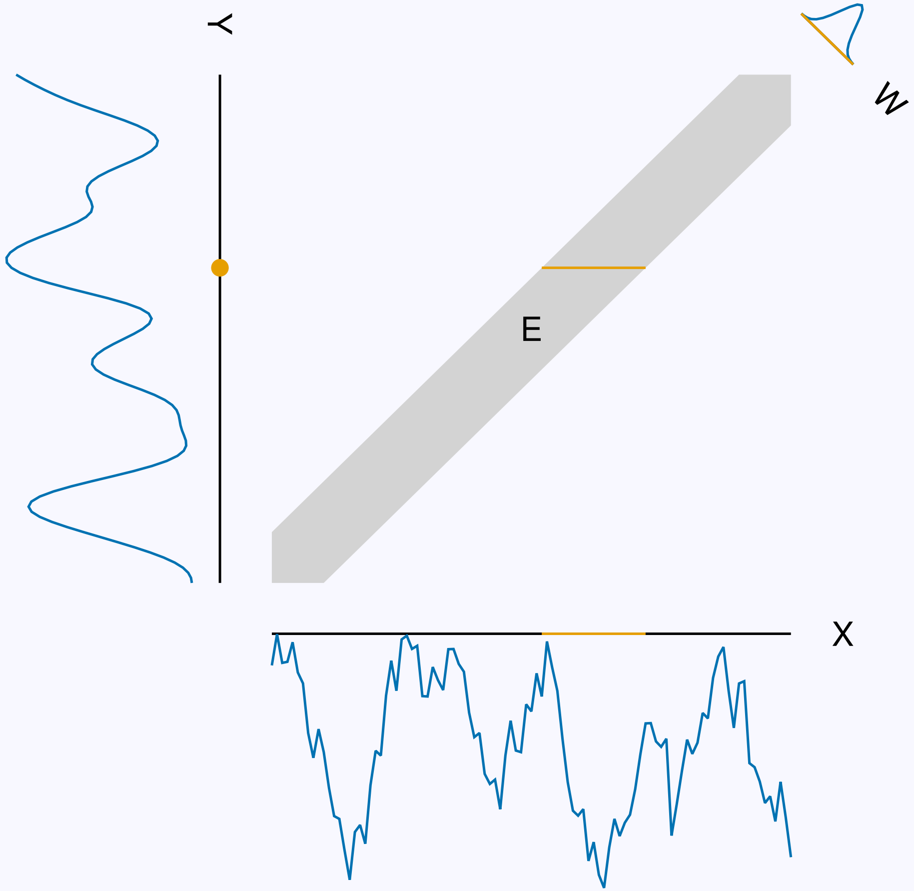
Features

- Domain: Discrete & continuous
- Symmetry: Translation

Parametric Span

$$F \times S_o$$

$$S_i \leftarrow F \times S_o \rightarrow S_o, \quad F \times S_o \rightarrow F$$



Classical architectures

Geometric deep learning

Polar coordinates ρ, θ

Adapted from Monti et al. (2017).

Classical architectures

Geometric deep learning

Features

- Domain: Discrete & continuous
- Symmetry: Learned

Parametric Span

$$\begin{array}{ccccc} & & \{ (q,p) \mid p \in \mathcal{N}(q) \} & & \\ & \swarrow & \downarrow \mathbf{u} & \searrow & \\ X & & \mathbb{R}^d & & X \end{array}$$

Polar coordinates ρ, θ

Adapted from Monti et al. (2017).

Conclusions and future directions

- Parametric spans induce linear layers with local connectivity and weight sharing.

Conclusions and future directions

- Parametric spans induce linear layers with local connectivity and weight sharing.
- The backward pass of such layers can be computed by permuting the legs of the parametric span.

Conclusions and future directions

- Parametric spans induce linear layers with local connectivity and weight sharing.
- The backward pass of such layers can be computed by permuting the legs of the parametric span.
- Parametric spans in the category of manifolds and submersions encompass
  - dense layers,
  - convolutional layers and variations thereof,
  - many geometric deep learning layers.

Conclusions and future directions

- Parametric spans induce linear layers with local connectivity and weight sharing.
- The backward pass of such layers can be computed by permuting the legs of the parametric span.
- Parametric spans in the category of manifolds and submersions encompass
  - dense layers,
  - convolutional layers and variations thereof,
  - many geometric deep learning layers.
- Thus, we can define the *microstructure* of a single linear layer in categorical terms.

Conclusions and future directions

- Parametric spans induce linear layers with local connectivity and weight sharing.
- The backward pass of such layers can be computed by permuting the legs of the parametric span.
- Parametric spans in the category of manifolds and submersions encompass
    - dense layers,
    - convolutional layers and variations thereof,
    - many geometric deep learning layers.
- Thus, we can define the *microstructure* of a single linear layer in categorical terms.
- In the future, we plan to
    - incorporate nonlinearities,
    - encode *global neural architectures* (not just single layers).

Conclusions and future directions

- Parametric spans induce linear layers with local connectivity and weight sharing.
- The backward pass of such layers can be computed by permuting the legs of the parametric span.
- Parametric spans in the category of manifolds and submersions encompass
  - dense layers,
  - convolutional layers and variations thereof,
  - many geometric deep learning layers.
- Thus, we can define the *microstructure* of a single linear layer in categorical terms.
- In the future, we plan to
  - incorporate nonlinearities,
  - encode *global neural architectures* (not just single layers).
- Our overarching aim is to create a framework for neural architectures with the following properties:
  - modularity and composability [1],
  - existence and computability of duals for reverse-mode differentiation [2].

[1] Vertechi, P., Frosini, P., & Bergomi, M. G. (2020). Parametric machines: a fresh approach to architecture search. arXiv preprint arXiv:2007.02777.
[2] Vertechi, P., & Bergomi, M. G. (2022). Machines of finite depth: towards a formalization of neural networks. arXiv preprint arXiv:2204.12786.