# On the Geometry of Rectifier Convolutional Neural Networks


Matteo Gamba, Stefan Carlsson, Hossein Azizpour, Mårten Björkman

KTH Royal Institute of Technology


## General Problem

- What is the **inductive bias** of state of the art Convolutional Neural Networks (CNNs)?
- For a fixed convolutional architecture, what family of functions are **attained in practice** when training the model on natural data?
- What is the relevant notion of "**complexity**"?

## Motivation

- Each convolutional layer defines **hyperplane arrangements** in its preactivation space, which in turn induce **classification regions** in the input space.
- Studying the **preimage of convolutional layers** might reveal the inductive bias of SGD on *natural data*.

## Research Question

How to describe and **characterize hyperplane arrangements** for pairs of stacked convolutional layers?

## Convolutional Layers

Cross-correlation between an input tensor $\mathcal{X} \in \mathbb{R}^{C\times H\times W}$ and a tensor $\mathcal{W} \in \mathbb{R}^{n_{\text{out}}\times n_{\text{in}}\times k\times k}$:

$$\tilde{\mathcal{O}}(o,i,j) = b_o + \sum_{c=0}^{n_{\text{in}}-1}\sum_{m=0}^{k-1}\sum_{n=0}^{k-1} \mathcal{X}(c,i+m,j+n)\cdot \mathcal{W}_o(c,m,n)$$

$$i = 0,\dots,r-1 \quad \text{and} \quad j = 0,\dots,r-1 \quad (1)$$

for each $\mathcal{W}_o := \mathcal{W}[o,:,:,:]$, $o = 1,\dots,n_{\text{out}}$.

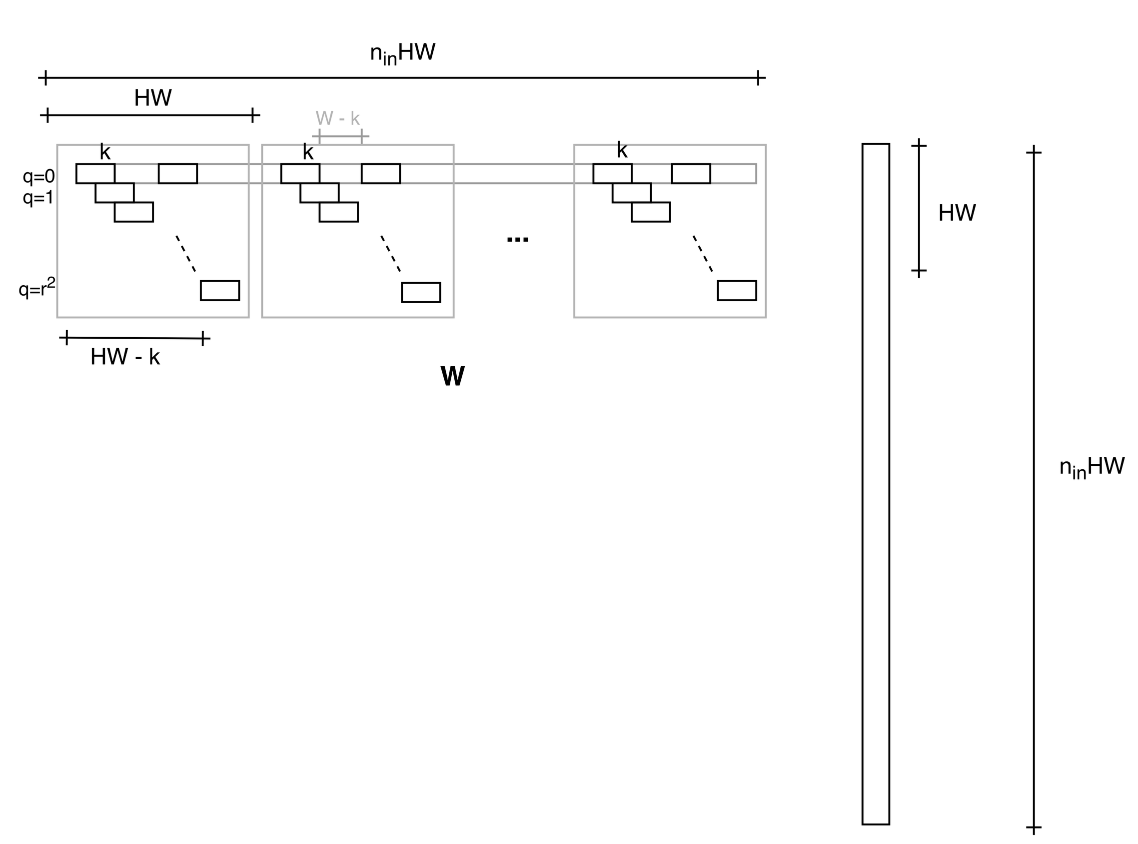


Figure: Tensor vectorization for a convolutional filter.

## Hyperplane Arrangements

For each node in the network, a forward pass computes:

$$\varphi(\mathbf{W}_q^T\mathbf{x} + b) =: O_q \in \mathbb{R} \quad (2)$$

ReLU induces two **affine halfspaces** in the preactivation space of the layer:

$$X_q^+ = \{\mathbf{x} \in \mathbb{R}^{n_{\text{in}}HW} \mid \mathbf{W}_q^T\mathbf{x} + b \geq 0\}$$
$$X_q^- = \{\mathbf{x} \in \mathbb{R}^{n_{\text{in}}HW} \mid \mathbf{W}_q^T\mathbf{x} + b < 0\} \quad (3)$$

For stride $s = 1$, $\mathbf{W}$ is a **Toepliz matrix** identifying $r^2$ hyperplanes in $\mathbb{R}^D$, with $D = n_{\text{in}}HW$:

$$\begin{cases} w_0x_0 + w_1x_1 + \ldots + w_{D-1}x_{D-1} + b \geq 0 \\ w_{D-1}x_0 + w_0x_1 + \ldots + w_{D-2}x_{D-1} + b \geq 0 \\ \vdots \\ w_{D-r^2+1}x_0 + \ldots + w_{D-r^2}x_{D-1} + b \geq 0 \end{cases} \quad (4)$$

Each convolutional filter defines a **polytope** in its **preactivation space**.

## Polyhedral Cones

- Each channel of each filter defines a **polyhedral cone** with *apex on the identity line*.
- Data is mapped to a **subspace of lower or equal dimension**.

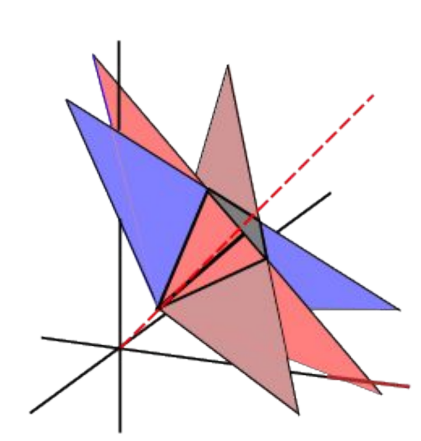

Figure: Polyhedral cone in three dimensions.

## Summary Statistics

Pairwise offsets between **apex position**, **opening angle** and **rotation angles** are used to define four discrete states:

1. `OUT_FULL_IN`: second cone *fully included* in first.
2. `OUT_PARTIAL_IN`: second cone *partially included* in first.
3. `IN_FULL_OUT`: first cone *fully included* in second.
4. `IN_PARTIAL_OUT`: first cone *partially included* in second.

## Evaluation

- For two consecutive layers $l$ and $l+1$
- For each cone $C_i^{l+1}$ compute mutual arrangement w.r.t. combination of cones $\{C_p^l\}_{p=0}^{n_l-1}$

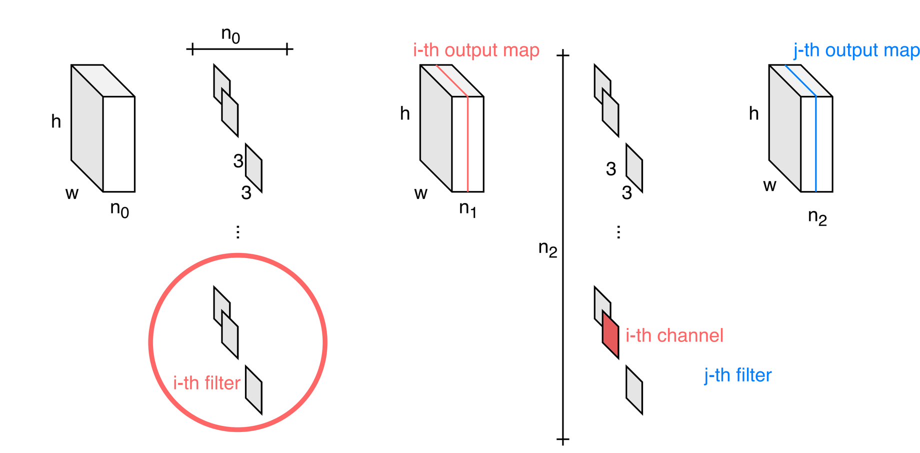


Figure: Computing statistics for pairs of stacked layers.

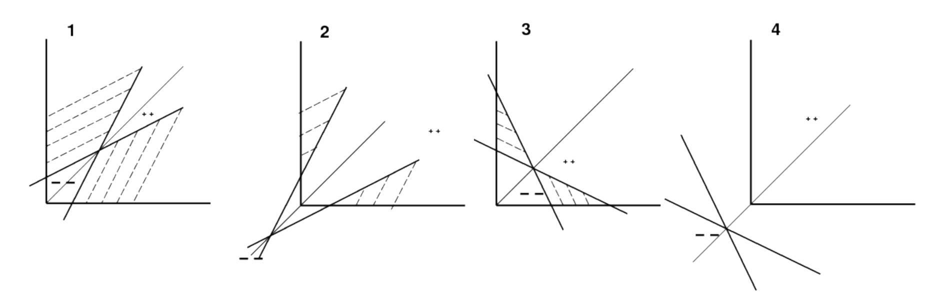


Figure: Discrete states in two dimensions.

## Results

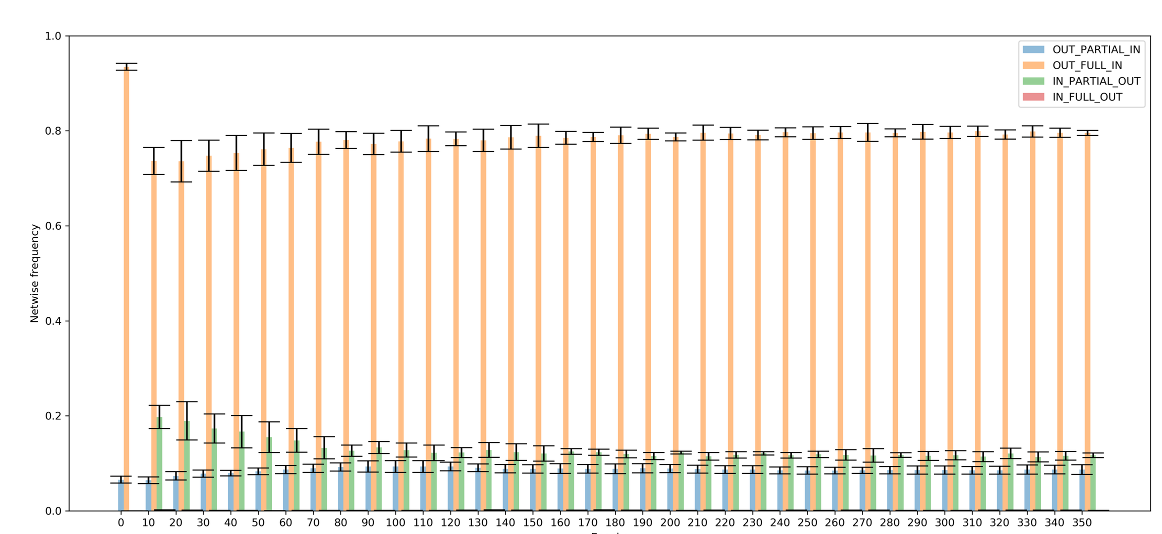


Figure: Net-wise distribution for LeNet9.

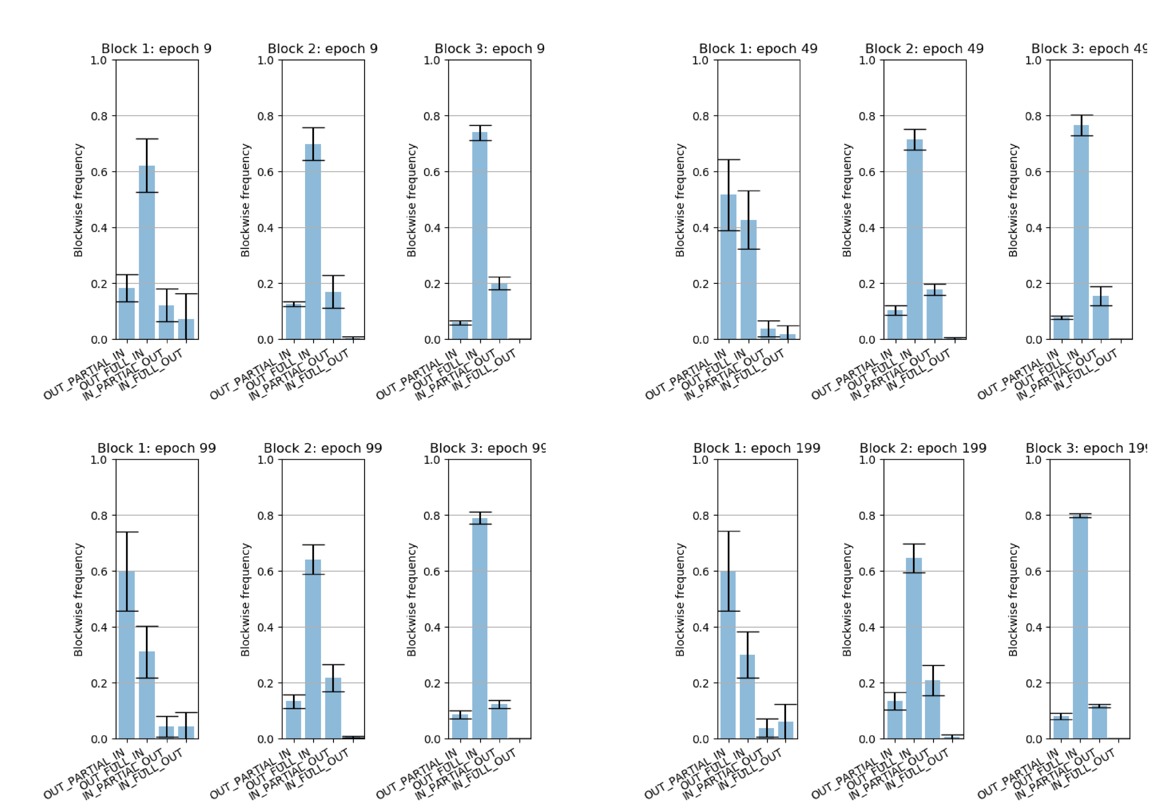


Figure: Block-wise distribution for LeNet9.

## Conclusion

- Low variance for higher blocks.
- Distribution "stabilizes" during training.
- How strongly does this reflect the behaviour of large CNNs?

## Models and Datasets

Experiments performed on 9-layer versions of VGG and LeNet trained on CIFAR-10.

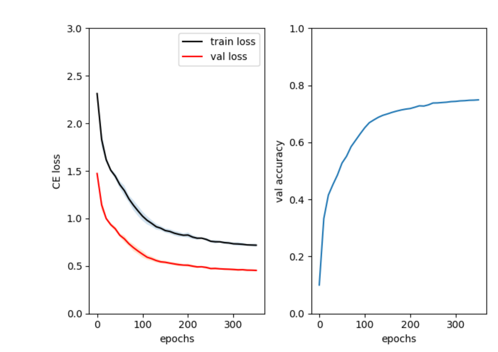

Figure: Train and test loss and accuracy for LeNet9.

## References


[1] S. Carlsson, H. Azizpour, and A. Razavian, "The preimage of rectifier network activities," 2016.

[2] S. Carlsson, "Geometry of deep convolutional networks," *CoRR*, 2019.

[3] G. Montufar, R. Pascanu, K. Cho, and Y. Bengio, "On the number of linear regions of deep neural networks," in *Adv. in NIPS*, 2014.

[4] R. Pascanu, G. Montufar, and Y. Bengio, "On the number of response regions of deep feed forward networks with piece-wise linear activations," 2013.


## Acknowledgements


This work is partially funded by the Swedish Research Council project 2017-04609 and Wallenberg AI, Autonomous Systems and Software Program (WASP).


## Contact Information


- Matteo Gamba
- ✉ mgamba@kth.se