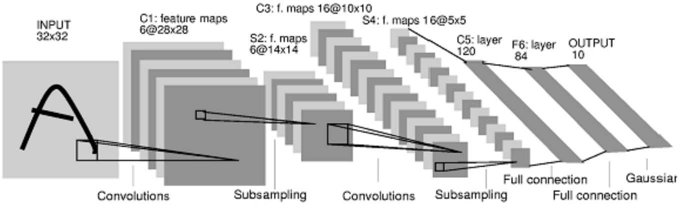
Invariance and Stability of Deep Convolutional Representations - NIPS 2017
Alberto Bietti, Julien Mairal

MSDMA Journal Club
Nicolas Thome
Prenom.Nom@cnam.fr
`http://cedric.cnam.fr/~thomen/`

January, 12th 2018

Stability: Context

- Deep ConvNet: impressive results in various contexts, especially image classification
- Deep ConvNet:
  - Model local appearance at multiple scales
  - Gain some invariance by pooling operations
- **BUT:** exact nature of this invariance and the characteristics of functional spaces where convolutional neural networks live poorly understood



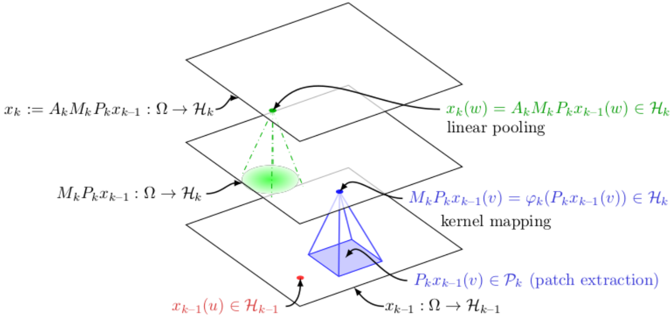
Context

- Understanding geometry of ConvNet functional spaces: fundamental question
  - Regularization: providing ways to control the variations of prediction functions in a principled manner
  - Small deformations of natural signals often preserve class info, but these deformations much richer class of transfo than translations
  - Representations stable to small deformations $\Rightarrow$ more robust models, improved sample complexity
    - $=$ better generalization performances for a given number of training samples OR
    - $\sim$ generalization performances with fewer training samples
  - Previsous stability studies for convolutional multilayer architectures based on wavelets [Mal12, BM13], BUT
    - Non learned parameters (filters)
    - Architecture significantly different from state-of-the-art ConvNets

Context

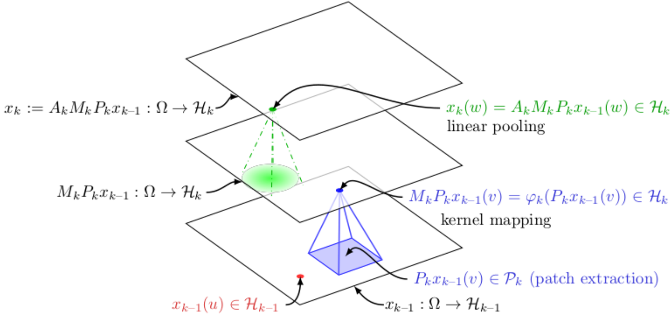
- This paper:
  - Analysis of stability properties of functional space derived from a kernel for multi-dimensional signals, which admits a multilayer and convolutional structure that generalizes the construction of convolutional kernel networks (CKNs) [MKHS14, Mai16]
    - Generalization to continuous inputs
  - Analysis of translation-invariance properties of the kernel representation and its stability to the action of diffeomorphisms
    - Using the framework of Mallat [Mal12], similar stability results obtained WHILE preserving signal information
  - Stability results can be translated to predictive models by controlling their norm in the functional space
    - RKHS norm controls both stability and generalization, so that stability may lead to improved sample complexity

Multilayer Convolutional Kernel



- Generalization of [Mai16] to continuous inputs
  - Patch extraction operator
  - Kernel mapping operator
  - Pooling operator

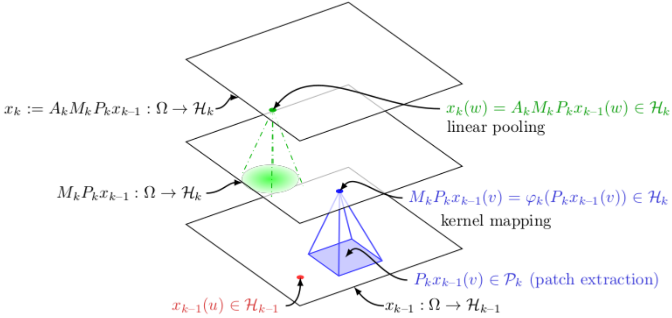
Multilayer Convolutional Kernel



- Patch extraction operator:

$$P_k x_{k-1}(u) = (v \mapsto x_{k-1}(u+v))_{v \in S_k} \in \mathcal{P}_k$$

Multilayer Convolutional Kernel



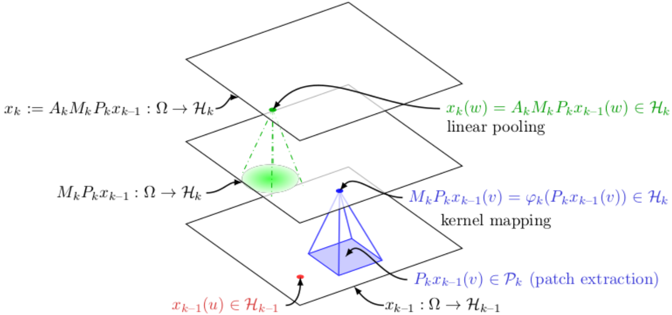
- Kernel mapping operator:

$$M_k P_k x_{k-1}(u) := \varphi_k(P_k x_{k-1}(u)) \in \mathcal{H}_k.$$

  - Homogeneous dot-product kernels:

$$K_k(z, z') = \|z\| \|z'\| \kappa_k \left(\frac{\langle z, z' \rangle}{\|z\| \|z'\|} \right) \quad \text{with} \quad \kappa_k(1) = 1$$

Multilayer Convolutional Kernel



- Pooling operator:

$$x_k(u) = A_k M_k P_k x_{k-1}(u) = \int_{\mathbb{R}^d} h_{\sigma_k}(u-v) M_k P_k x_{k-1}(v) dv \in \mathcal{H}_k$$

Multilayer Convolutional Kernel

- Multilayer construction:

$$\Phi_n(x_0) := x_n = A_n M_n P_n A_{n-1} M_{n-1} P_{n-1} \cdots A_1 M_1 P_1 x_0 \in L^2(\Omega, \mathcal{H}_n)$$

Lemma 1 (Signal preservation). *Assume that $\mathcal{H}_k$ contains linear functions $\langle w, \cdot \rangle$ with w in $\mathcal{P}_k$ (this is true for all kernels K_k described in Appendix B), then the signal x_{k-1} can be recovered from a sampling of $x_k = A_k M_k P_k x_{k-1}$ at discrete locations as soon as the union of patches centered at these points covers all of Ω. It follows that x_k can be reconstructed from such a sampling.*

Stability on Multilayer Convolutional Kernel

- Stability to diffeomorphisms: $L_\tau x(u) = x(u - \tau(u))$
  - Closely follwoing stability analysis of scattering [Mal12]
- Assumptions for each layer k:

(A1) **Norm preservation**: $\|\varphi_k(x)\| = \|x\|$ for all x in $\mathcal{P}_k$;
(A2) **Non-expansiveness**: $\|\varphi_k(x) - \varphi_k(x')\| \leq \|x - x'\|$ for all x, x' in $\mathcal{P}_k$;
(A3) **Patch sizes**: there exists $\kappa > 0$ such that at any layer k we have

$$\sup_{c \in S_k} |c| \leq \kappa \sigma_{k-1}.$$

- Stability bound (proof apprendix C):

$$\|\Phi(L_\tau x) - \Phi(x)\| \leq \left(\sum_{k=1}^{n} \|[P_k A_{k-1}, L_\tau]\| + \|[A_n, L_\tau]\| + \|L_\tau A_n - A_n\| \right) \|x\| \qquad (5)$$

  - With $[A, B] = AB - BA$ commutator

Stability on Multilayer Convolutional Kernel

$$\|\Phi(L_\tau x) - \Phi(x)\| \leq \left(\sum_{k=1}^{n} \|[P_k A_{k-1}, L_\tau]\| + \|[A_n, L_\tau]\| + \|L_\tau A_n - A_n\| \right) \|x\| \qquad (5)$$

- Bound on $\|[P_k A_{k-1}, L_\tau]\|$:

$$\|[P_k A_{k-1}, L_\tau]\| \leq C_1 \|\nabla \tau\|_\infty$$

- Bound on $\|L_\tau A_n - A_n\|$:

$$\|L_\tau A_\sigma - A_\sigma\| \leq \frac{C_2}{\sigma} \|\tau\|_\infty$$

- Leading to

$$\|\Phi(L_\tau x) - \Phi(x)\| \leq \left(C_1 (1+n) \|\nabla \tau\|_\infty + \frac{C_2}{\sigma_n} \|\tau\|_\infty \right) \|x\|$$

Stability on Multilayer Convolutional Kernel

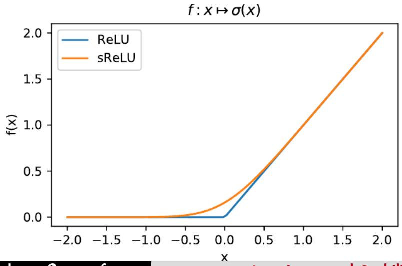
- Stability bound valid for the kernel approximation (to gite finite dimensional maps)
- Extension for designing global invariance to group actions

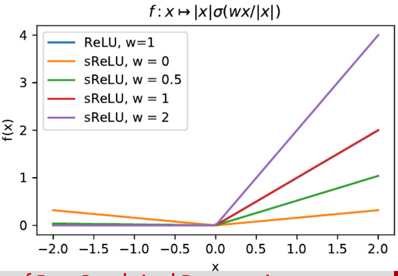
Link to ConvNets

- Multilayer Convolutional Kernel (MKN) contains a set of CNNs on continuous domains
  - With certain type of smooth homogeneous activation functions
- CNN map construction:

$$\tilde{z}_k^i(u) = n_k(u)\sigma\Big(\langle w_k^i, P_k z_{k-1}(u)\rangle / n_k(u)\Big)$$

  - $n_k(u)$ unsual (homogeneization, next), otherwise very standard (except linear pooling!)
- Homogeneous activation: contains smoothed version of popular RelU





Stability & Generalization

- recap stability bound:

$$\|\Phi(L_\tau x) - \Phi(x)\| \leq \left(C_1 \left(1+n\right) \|\nabla \tau\|_\infty + \frac{C_2}{\sigma_n} \|\tau\|_\infty \right) \|x\|$$

- Bound on $||f||$:

$$||f_\sigma||^2 \leq p_n \sum_{i=1}^{p_n} ||w_{n+1}^i||_2^2 B_{n,i}$$

- Cauchy-Schawrtz:

$$|f(z) - f(z')| \leq ||f||_{\mathcal{H}} ||\varphi(z) - \varphi(z')||_{\mathcal{H}}$$

  - x, x' close but different labels $\Rightarrow$ large $||f||$ generalization harder
  - **Link between stabilty & generalization**

Conclusion

- Linear pooling: what about max pooling
- Stability to diffeomorphisms:
  - Beyond diffeomorphisms ? (Small deformations of natural signals often preserve class info, but these deformations much richer class of transfo than translations)
  - adversarial examples ?

References I

Joan Bruna and Stephane Mallat, *Invariant scattering convolution networks*, IEEE Trans. Pattern Anal. Mach. Intell. 35 (2013), no. 8, 1872–1886.

Julien Mairal, *End-to-End Kernel Learning with Supervised Convolutional Kernel Networks*, Advances in Neural Information Processing Systems (NIPS) (Barcelona, France), December 2016.

Stéphane Mallat, *Group invariant scattering*, Communications on Pure and Applied Mathematics 65 (2012), no. 10, 1331–1398.

Julien Mairal, Piotr Koniusz, Zaïd Harchaoui, and Cordelia Schmid, *Convolutional kernel networks*, Advances in Neural Information Processing Systems 27: Annual Conference on Neural Information Processing Systems 2014, December 8-13 2014, Montreal, Quebec, Canada, 2014, pp. 2627–2635.