BRIGEABLE

ANR Chair in AI

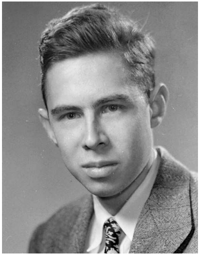
Jean-Christophe Pesquet

Center for Visual Computing, OPIS Inria group, CentraleSupélec, University Paris-Saclay

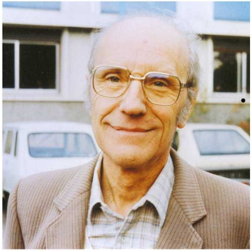
DATAIA - September 2020

Motivation

BRIDinG thE gAp Between iterative proximaL methods and nEural networks



Frank Rosenblatt (1928–1971)



Jean-Jacques Moreau (1923–2014)

Gradient descent

✓ Basic optimization problem

$$\underset{x\in C}{\text{minimize}} \ \frac{1}{2}\|Hx - y\|^2$$

where C nonempty closed convex subset of $\mathbb{R}^N$, $y \in \mathbb{R}^M$, and $H \in \mathbb{R}^{M\times N}$.

✓ Projected gradient algorithm

$$(\forall n \in \mathbb{N}\setminus\{0\}) \quad x_n = \mathrm{proj}_C\big(x_{n-1} - \gamma_n H^\top(Hx_{n-1} - y)\big)$$

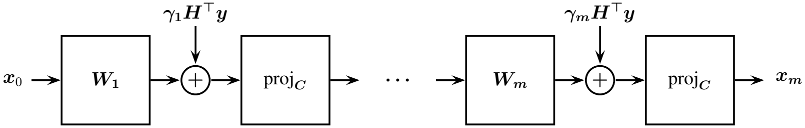
where $\gamma_n > 0$ is the step-size

Gradient descent

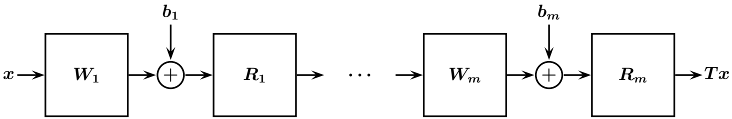
✓ Projected gradient algorithm

$$(\forall n \in \mathbb{N} \setminus \{0\}) \quad x_n = \mathrm{proj}_C\big(x_{n-1} - \gamma_n H^\top (H x_{n-1} - y)\big) = \mathrm{proj}_C(W_n x_{n-1} + \gamma_n H^\top y)$$

where $\gamma_n > 0$ is the step-size and $W_n = \mathrm{Id} - \gamma_n H^\top H$.



Feedforward NNs



Neural network model

$$T = T_m \circ \cdots \circ T_1$$

where $(\forall i \in \{1, \ldots, m\}) \quad T_i \colon \mathbb{R}^{N_{i-1}} \to \mathbb{R}^{N_i} \colon x \mapsto R_i(W_i x + b_i)$,
$W_i \in \mathbb{R}^{N_i \times N_{i-1}}$ is a weight matrix,
b_i is a bias vector in $\mathbb{R}^{N_i}$,
and $R_i \colon \mathbb{R}^{N_i} \to \mathbb{R}^{N_i}$ is an activation operator.

Remark $(W_i)_{1 \leqslant i \leqslant m}$ can be convolutive operators

Link

✓ Proximity operator [Moreau, 1962]

Let $f\colon \mathbb{R}^N \to \,]-\infty, +\infty]$ be a lower-semicontinuous convex function. For every $x \in \mathbb{R}^N$,

$$\operatorname{prox}_f(x) = \underset{z \in \mathbb{R}^N}{\operatorname{argmin}} \ \frac{1}{2}\|z - x\|^2 + f(z).$$

If f is the indicator function of C, then $\operatorname{prox}_f = \operatorname{proj}_C$.

projected gradient algorithm $\rightsquigarrow$ proximal gradient algorithm

Link

✓ Proximity operator [Moreau, 1962]
Let $f\colon \mathbb{R}^N \to]-\infty, +\infty]$ be a lower-semicontinuous convex function. For every $x \in \mathbb{R}^N$,

$$\operatorname{prox}_f(x) = \underset{z \in \mathbb{R}^N}{\operatorname{argmin}} \ \frac{1}{2}\|z - x\|^2 + f(z).$$

If f is the indicator function of C, then $\operatorname{prox}_f = \operatorname{proj}_C$.
projected gradient algorithm $\rightsquigarrow$ proximal gradient algorithm

✓ Most of the activation operators are proximity operators

Link

✓ Most of the activation operators are proximity operators

Example of the squashing function used in capsnets

$$(\forall x \in \mathbb{R}^N) \quad Rx = \frac{\mu\|x\|}{1+\|x\|^2}x = \operatorname{prox}_{\phi\circ\|\cdot\|}x, \quad \mu = \frac{8}{3\sqrt{3}},$$

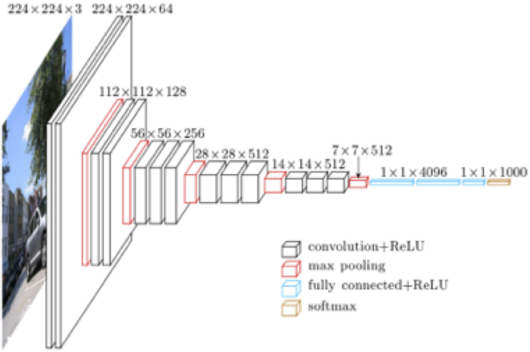
where

$$\phi\colon \xi \mapsto \begin{cases} \mu \arctan\sqrt{\dfrac{|\xi|}{\mu-|\xi|}} - \sqrt{|\xi|(\mu-|\xi|)} - \dfrac{\xi^2}{2}, & \text{if } |\xi| < \mu; \\ \dfrac{\mu(\pi-\mu)}{2}, & \text{if } |\xi| = \mu; \\ +\infty, & \text{otherwise.} \end{cases}$$

Link

✓ Most of the activation operators are proximity operators

✓ Difficulty



Objective

BETTER UNDERSTANDING OF NEURAL NETWORKS

EXPLAINABILITY

Under some assumptions, NNs are shown to solve variational inequalities [Combettes, Pesquet, 2020]

Objective

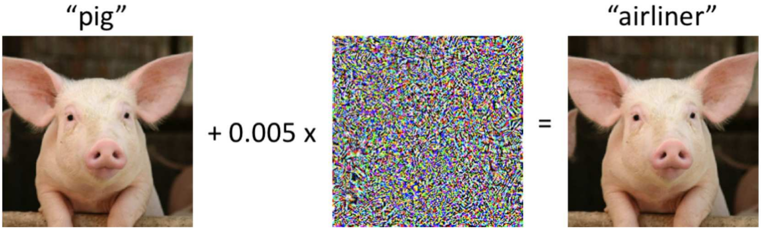
Better understanding of neural networks

Explainability

Under some assumptions, NNs are shown to solve variational inequalities [Combettes, Pesquet, 2020]

Robustness

Sensitivity to adversarial perturbations [Szegedy et al., 2013]



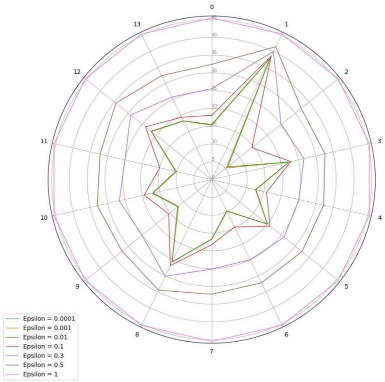
Robustness issues

- ✓ Certifiability requirement for NNs in critically safe environments
- ✓ Deriving sharp Lipschitz constant estimates

Robustness issues

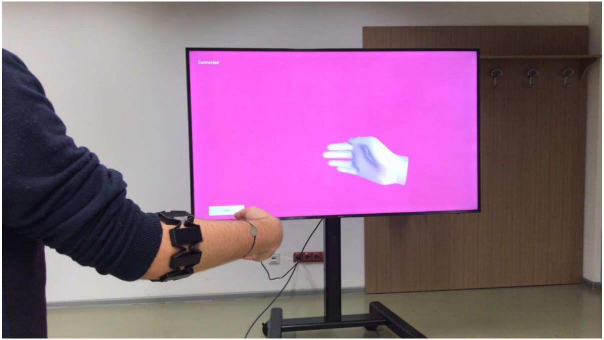
Example of a NN for Air Traffic Management developed by Thales (CIFRE PhD thesis of K. Gupta)

LIPSCHITZ STAR



Robustness issues

Example of Automatic Gesture Recognition based on surface Electromyographic signals (PhD thesis of A. Neacsu in collaboration with Polithenica University of Bucharest)



✓ standard training
accuracy = 99.78 %, but Lipschitz constant $> 10^{12}$

Robustness issues

Example of Automatic Gesture Recognition based on surface Electromyographic signals (PhD thesis of A. Neacsu in collaboration with Polithenica University of Bucharest)

- ✓ standard training
  accuracy = 99.78 %, but Lipschitz constant $> 10^{12}$
- ✓ proximal algorithm for training the network subject to a Lispchitz bound constraint

Accuracy	75 %	80 %	85 %	90 %	95 %
Lipschitz constant	0.36	0.46	0.82	2.68	**3.38**

Workplan

✓ WP1: Design of robust networks
generalization of existing results, constrained training,...

✓ WP2: Proposition of new fixed point strategies
link with plug and play methods, fixed point training,...

✓ WP3: Proximal view of Deep Dictionary Learning
change of metrics, theoretical analysis,...

Workplan

✓ WP1: Design of robust networks
generalization of existing results, constrained training,...

✓ WP2: Proposition of new fixed point strategies
link with plug and play methods, fixed point training,...

✓ WP3: Proximal view of Deep Dictionary Learning
change of metrics, theoretical analysis,...

... September 2020 $\rightarrow$ August 2024

Partners

✓ Industrial

- Schneider Electric (WP 1)
- GE Healthcare (WP 2)
- IFPEN (WP 3)
- Additional collaborations with Thales and Essilor

Partners

✓ Industrial

- Schneider Electric (WP 1)
- GE Healthcare (WP 2)
- IFPEN (WP 3)
- Additional collaborations with Thales and Essilor

✓ Academic

- P. Combettes, NCSU (WP 1)
- A. Repetti and Y. Wiaux, Heriot Watt University (WP 2)
- H. Krim, NCSU (WP 3)
- M. Kaaniche, Univ. Sorbonne Paris Nord (WP 3).

Some references

P. L. Combettes and J.-C. Pesquet
Proximal splitting methods in signal processing
in *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*, H. H. Bauschke, R. Burachik, P. L. Combettes, V. Elser, D. R. Luke, and H. Wolkowicz editors. Springer-Verlag, New York, pp. 185-212, 2011.

C. Bertocchi, E. Chouzenoux, M.-C. Corbineau,J.-C. Pesquet, M. Prato
Deep unfolding of a proximal interior point method for image restoration
Inverse Problems, vol. 36, no 3, pp. 034005, Feb. 2020.

P. L. Combettes and J.-C. Pesquet
Lipschitz certificates for layered network structures driven by averaged activation operators
SIAM Journal on Mathematics of Data Science, vol. 2, no. 2, pp. 529–557, June 2020.

P. L. Combettes and J.-C. Pesquet
Deep neural network structures solving variational inequalities
Set-Valued and Variational Analysis, vol. 28, pp. 491–518, Sept. 2020.

P. L. Combettes and J.-C. Pesquet
Fixed point strategies in data science
https://arxiv.org/abs/2008.02260, 2020.