Fixed Point Strategies in Image Processing

Jean-Christophe Pesquet

https://jc.pesquet.eu

Centre de Vision Numérique – Inria OPIS
CentraleSupélec, Univ. Paris Saclay, France

SSVM 2023

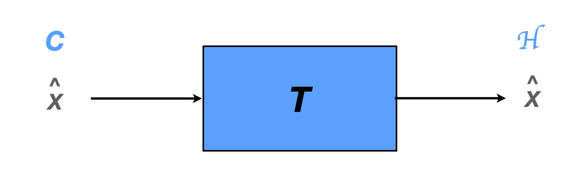
Motivations

Problem

Let T be a (possibly nonlinear) system with

- output space $\mathcal{H}$
- input set $C \subset \mathcal{H}$.

We want to find a fixed point $\widehat{x}$ of T:



$\operatorname{Fix} T \equiv$ set of fixed points of T.

Example: If T is a feedforward neural network, $\operatorname{Fix} T$ is defined by a system of subdifferential inclusions [Combettes, Pesquet - 2020]

Fixed point theorem

(Emile Picard, 1856-1941)

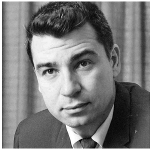
If $C = \mathcal{H}$ is a Hilbert space and

T is a Banach contraction, i.e. there exists $\rho \in [0, 1[$ such that

$$\big(\forall (x, x') \in \mathcal{H}^2\big) \qquad \|Tx - Tx'\| \leqslant \rho \|x - x'\|.$$

Then T has a unique fixed point $\widehat{x}$.
The sequence $(x_n)_{n\in\mathbb{N}}$ defined as $(\forall n \in \mathbb{N})\ x_{n+1} = Tx_n$ with $x_0 \in \mathcal{H}$, converges linearly to $\widehat{x}$.

Feasibility problems

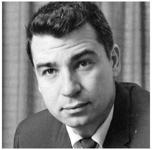
(John Von Neumann, 1903-1957 – Dante C. Youla, 1925-2021)

Problem

Let S_1 and S_2 be two closed convex subsets of $\mathcal{H}$ such $S_1 \cap S_2 \neq \varnothing$. We want to

$$\text{Find } \widehat{x} \in S_1 \cap S_2.$$

Feasibility problems



(John Von Neumann, 1903-1957 – Dante C. Youla, 1925-2021)

Problem

Let S_1 and S_2 be two closed convex subsets of $\mathcal{H}$ such $S_1 \cap S_2 \neq \varnothing$. We want to

$$\begin{aligned} & \text{Find } \widehat{x} \in S_1 \cap S_2 \\ \Leftrightarrow \quad & \text{Find } \widehat{x} = T(\widehat{x}) \text{ with } T = \mathrm{proj}_{S_1} \circ \mathrm{proj}_{S_2}. \end{aligned}$$

Feasibility problems

Problem

Let S_1 and S_2 be two closed convex subsets of $\mathcal{H}$ such $S_1 \cap S_2 \neq \varnothing$. We want to

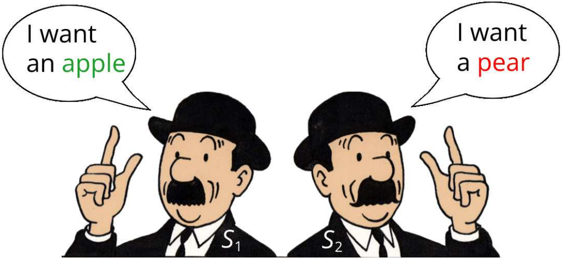
$$\begin{aligned} &\text{Find } \ \widehat{x} \in S_1 \cap S_2 \\ \Leftrightarrow \quad &\text{Find } \ \widehat{x} = T(\widehat{x}) \ \text{ with } \ T = \mathrm{proj}_{S_1} \circ \mathrm{proj}_{S_2}. \end{aligned}$$

POCS algorithm

$$\begin{aligned} &\text{Set } x_0 \in \mathcal{H} \\ &\text{For } n = 0, \ldots \\ &\lfloor \ x_{n+1} = \mathrm{proj}_{S_1}(\mathrm{proj}_{S_2} x_n). \end{aligned}$$

Convergence properties although T is not a Banach contraction

Possibly infeasible problems



Problem

Let S_1 and S_2 be two closed convex subsets of $\mathcal{H}$.
We want to

$$\underset{x \in S_1}{\text{minimize}} \quad \underbrace{d^2_{S_2}(x)}_{\|x - \text{proj}_{S_2} x\|^2}$$

Possibly infeasible problems

Problem

Let S_1 and S_2 be two closed convex subsets of $\mathcal{H}$.
We want to

$$\underset{x\in S_1}{\text{minimize}} \qquad \underbrace{d^2_{S_2}(x)}_{\|x-\text{proj}_{S_2}x\|^2}$$

Projected gradient algorithm

Set $x_0 \in \mathcal{H}$

For $n = 0, \ldots$

$\lfloor\; x_{n+1} = \text{proj}_{S_1}\left(x_n - \frac{1}{2}\nabla d^2_{S_2}(x_n)\right)$

Possibly infeasible problems

Problem

Let S_1 and S_2 be two closed convex subsets of $\mathcal{H}$.
We want to

$$\underset{x\in S_1}{\text{minimize}} \qquad \underbrace{d^2_{S_2}(x)}_{\|x-\text{proj}_{S_2}x\|^2}$$

Projected gradient algorithm

Set $x_0 \in \mathcal{H}$

For $n = 0, \ldots$

$$\left\lfloor \begin{aligned} x_{n+1} &= \text{proj}_{S_1}\big(x_n - \tfrac{1}{2}\nabla d^2_{S_2}(x_n)\big) \\ &= \text{proj}_{S_1}(\text{proj}_{S_2} x_n) \end{aligned} \right.$$

POCS with 2 sets solves a minimization problem

More than 2 sets

POCS algorithm

$(S_i)_{1\leqslant i\leqslant m}$ closed convex subsets of $\mathcal{H}$

Set $x_0 \in \mathcal{H}$

For $n = 0, \ldots$

$$x_{n+1} = \underbrace{\mathrm{proj}_{S_1} \circ \cdots \circ \mathrm{proj}_{S_m}}_{T} x_n.$$

More than 2 sets

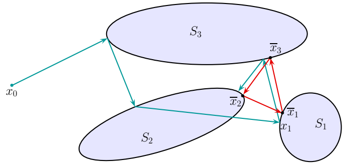
POCS algorithm

$(S_i)_{1\leqslant i\leqslant m}$ closed convex subsets of $\mathcal{H}$

$$\begin{aligned}&\text{Set } x_0 \in \mathcal{H}\\ &\text{For } n = 0, \dots\\ &\quad x_{n+1} = \underbrace{\mathrm{proj}_{S_1} \circ \cdots \circ \mathrm{proj}_{S_m}}_{T} x_n.\end{aligned}$$

- If $\bigcap_{i=1}^m S_i \neq \varnothing$, (weak) convergence to a point in the intersection
- Otherwise, generates a limit cycle $(\overline{x}_1, \dots, \overline{x}_m)$ such that

$$\begin{cases}\overline{x}_1 = \mathrm{proj}_{S_1}\overline{x}_2\\ \overline{x}_2 = \mathrm{proj}_{S_2}\overline{x}_3\\ \quad\vdots\\ \overline{x}_{m-1} = \mathrm{proj}_{S_{m-1}}\overline{x}_m\\ \overline{x}_m = \mathrm{proj}_{S_m}\overline{x}_1\end{cases}$$



$(\overline{x}_1, \dots, \overline{x}_m)$ with $m \geqslant 3$ is not a solution to an optimization problem [Baillon, Combettes, Cominetti - 2012]

Mathematical tools

Some vocabulary

An operator $T\colon \mathcal{H} \to \mathcal{H}$ is

- ρ-Lipschitz with $\rho \in\,]0, +\infty[$ if

$$(\forall (x,y) \in \mathcal{H}^2) \quad \|Tx - Ty\| \leqslant \rho\|x - y\|$$

- nonexpansive if T is 1-Lispchitz

- α-averaged with $\alpha \in]0,1]$ if $T = (1-\alpha)\mathrm{Id} + \alpha Q$ where Q is nonexpansive

- firmly nonexpansive if it is $1/2$-averaged

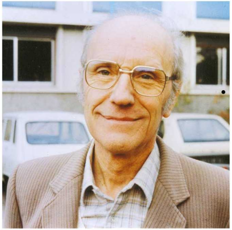
- β-cocoercive with $\beta \in\,]0, +\infty[$ if βT is firmly nonexpansive.

Examples

- Firmly nonexpansive operators
  - projection onto a closed convex set
  - proximity operator of a function $f \in \Gamma_0(\mathcal{H})$

    $\Gamma_0(\mathcal{H})$: class of lower-semicontinuous convex function from $\mathcal{H}$ to $]-\infty, +\infty]$ which are proper (i.e. finite at least at one point)

$$\operatorname{prox}_f : \mathcal{H} \to \mathcal{H} \colon x \mapsto \underset{y \in \mathcal{H}}{\operatorname{argmin}} \; f(y) + \frac{1}{2}\|y - x\|^2.$$

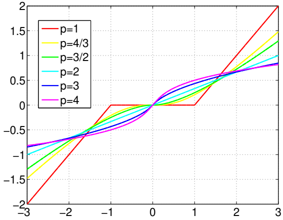


(Jean-Jacques Moreau, 1923–2014)

Examples

- Firmly nonexpansive operators
  - projection onto a closed convex set
  - proximity operator of a function $f \in \Gamma_0(\mathcal{H})$

$$\mathrm{prox}_f : \mathcal{H} \to \mathcal{H} : x \mapsto \underset{y \in \mathcal{H}}{\operatorname{argmin}} \; f(y) + \frac{1}{2}\|y - x\|^2.$$



$\mathrm{prox}_{|\cdot|^p}$

Remark: If $C \subset \mathcal{H}$ nonempty, closed, and convex set, then $\mathrm{proj}_C = \mathrm{prox}_{\iota_C}$ where ι_C is the indicator function of C:

$$(\forall x \in \mathcal{H}) \quad \iota_C(x) = \begin{cases} 0 & \text{if } x \in C \\ +\infty & \text{otherwise.} \end{cases}$$

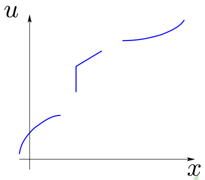
Examples

- Firmly nonexpansive operators
  - projection onto a closed convex set
  - proximity operator of a function $f \in \Gamma_0(\mathcal{H})$
  - resolvent of a maximally monotone operator (MMO) A:

$$J_A = (\mathrm{Id} + A)^{-1} : \mathcal{H} \to \mathcal{H}$$

A multivalued operator $A : \mathcal{H} \to 2^{\mathcal{H}}$ is monotone if

$$(\forall (x_1, x_2) \in \mathcal{H}^2)(\forall (u_1, u_2) \in Ax_1 \times Ax_2) \ \langle u_1 - u_2 \mid x_1 - x_2 \rangle \geqslant 0.$$



Examples

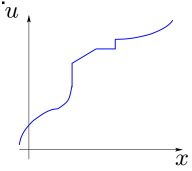
- FIRMLY NONEXPANSIVE OPERATORS
  - projection onto a closed convex set
  - proximity operator of a function $f \in \Gamma_0(\mathcal{H})$
  - resolvent of a maximally monotone operator (MMO) A:

$$J_A = (\mathrm{Id} + A)^{-1} : \mathcal{H} \to \mathcal{H}$$

A multivalued operator $A : \mathcal{H} \to 2^{\mathcal{H}}$ is monotone if

$$(\forall (x_1, x_2) \in \mathcal{H}^2)(\forall (u_1, u_2) \in Ax_1 \times Ax_2) \ \langle u_1 - u_2 \mid x_1 - x_2 \rangle \geqslant 0 .$$

It is maximally monotone if its graph is maximal in the sense of the inclusion relation.



Examples

- FIRMLY NONEXPANSIVE OPERATORS
  - projection onto a closed convex set
  - proximity operator of a function $f \in \Gamma_0(\mathcal{H})$
  - resolvent of a maximally monotone operator (MMO) A:

$$J_A = (\mathrm{Id} + A)^{-1} \colon \mathcal{H} \to \mathcal{H}$$

A multivalued operator $A : \mathcal{H} \to 2^{\mathcal{H}}$ is monotone if

$$(\forall (x_1, x_2) \in \mathcal{H}^2)(\forall (u_1, u_2) \in Ax_1 \times Ax_2) \ \langle u_1 - u_2 \mid x_1 - x_2 \rangle \geqslant 0.$$

It is maximally monotone if its graph is maximal in the sense of the inclusion relation.

Remark:

★ If $f \colon \mathcal{H}_1 \times \mathcal{H}_2 \to [-\infty, +\infty]$ where $\mathcal{H}_1$ and $\mathcal{H}_2$ are two real Hilbert spaces and, for every $(x_1, x_2) \in \mathcal{H}_1 \times \mathcal{H}_1$, $f(\cdot, x_2) \in \Gamma_0(\mathcal{H}_1)$ and $-f(x_1, \cdot) \in \Gamma_0(\mathcal{H}_2)$, then

$$A \colon (x_1, x_2) \mapsto \partial f(\cdot, x_2) \times (-\partial f(x_1, \cdot))$$

is maximally monotone.

Examples

- Firmly nonexpansive operators
  - projection onto a closed convex set
  - proximity operator of a function $f \in \Gamma_0(\mathcal{H})$
  - resolvent of a maximally monotone operator (MMO) A:

$$J_A = (\mathrm{Id} + A)^{-1} : \mathcal{H} \to \mathcal{H}$$

A multivalued operator $A : \mathcal{H} \to 2^{\mathcal{H}}$ is monotone if

$$(\forall (x_1, x_2) \in \mathcal{H}^2)(\forall (u_1, u_2) \in Ax_1 \times Ax_2) \ \langle u_1 - u_2 \mid x_1 - x_2 \rangle \geqslant 0$$

It is maximally monotone if its graph is maximal in the sense of the inclusion relation.

Remark:

⋆ If $f \in \Gamma_0(\mathcal{H})$, ∂f is maximally monotone and $\mathrm{prox}_f = J_{\partial f}$.

Examples

- β-COCOERCIVE OPERATORS
  - gradient ∇f of a differentiable convex function f if ∇f is $1/\beta$-Lipschitzian
  - if $(T_i)_{1\leqslant i\leqslant m}$ are β_i-cocoercive and $(L_i)_{1\leqslant i\leqslant m}$ are linear bounded operators with adjoints $(L_i^*)_{1\leqslant i\leqslant m}$ defined on Hilbert spaces, then $x\mapsto\sum_{i=1}^m L_i^*(T_i(L_ix))$ is cocoercive.

Examples

- β-COCOERCIVE OPERATORS
  - gradient ∇f of a differentiable convex function f if ∇f is $1/\beta$-Lipschitzian
  - if $(T_i)_{1\leqslant i\leqslant m}$ are β_i-cocoercive and $(L_i)_{1\leqslant i\leqslant m}$ are linear bounded operators with adjoints $(L_i^*)_{1\leqslant i\leqslant m}$ defined on Hilbert spaces, then $x \mapsto \sum_{i=1}^m L_i^*(T_i(L_i x))$ is cocoercive.
- α-AVERAGED OPERATORS
  - Banach contractions
  - if T is β-cocoercive, then $\mathrm{Id} - \gamma T$ is $\gamma/(2\beta)$-averaged when $\gamma \in]0, 2\beta[$

    Remark: If $T = \nabla f$, $\mathrm{Id} - \gamma T$: gradient descent operator

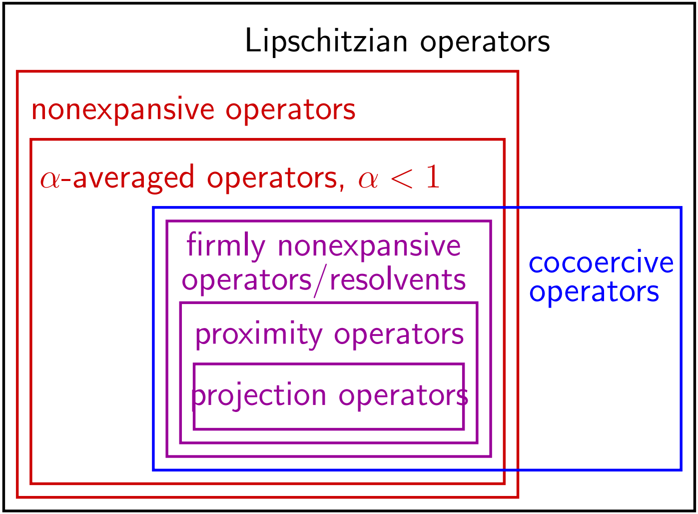
Examples

- β-COCOERCIVE OPERATORS
  - gradient ∇f of a differentiable convex function f if ∇f is $1/\beta$-Lipschitzian
  - if $(T_i)_{1\leqslant i\leqslant m}$ are β_i-cocoercive and $(L_i)_{1\leqslant i\leqslant m}$ are linear bounded operators with adjoints $(L_i^*)_{1\leqslant i\leqslant m}$ defined on Hilbert spaces, then $x \mapsto \sum_{i=1}^m L_i^*(T_i(L_i x))$ is cocoercive.

- α-AVERAGED OPERATORS
  - Banach contractions
  - if T is β-cocoercive, then $\mathrm{Id} - \gamma T$ is $\gamma/(2\beta)$-averaged when $\gamma \in]0, 2\beta[$

    Remark: If $T = \nabla f$, $\mathrm{Id} - \gamma T$: gradient descent operator
  - a convex combination or a composition of averaged operators is an averaged operator

Map of operator world



Fixed point algorithms

Krasnosel'skii-Mann-like algorithm

(Mark O. Krasnosel'skii, 1920-1997 —- William R. Mann, 1920-2006)

Let $T\colon \mathcal{H} \to \mathcal{H}$ be an α-averaged operator with $\alpha \in]0,1]$ such that $\operatorname{Fix} T \neq \varnothing$.
Let $(\lambda_n)_{n\in\mathbb{N}}$ be a sequence in $[0, 1/\alpha]$ such that

$$\sum_{n\in\mathbb{N}} \lambda_n(1-\alpha\lambda_n) = +\infty.$$

Let $x_0 \in \mathcal{H}$ and $(\forall n \in \mathbb{N})$ $x_{n+1} = x_n + \lambda_n(Tx_n - x_n)$.
Then $(x_n)_{n\in\mathbb{N}}$ converges (weakly) to a point in $\operatorname{Fix} T$.

Fixed point algorithms

Krasnosel'skii-Mann-like algorithm

Let $T\colon \mathcal{H} \to \mathcal{H}$ be an α-averaged operator with $\alpha \in]0,1]$ such that $\operatorname{Fix} T \neq \varnothing$.
Let $(\lambda_n)_{n\in\mathbb{N}}$ be a sequence in $[0, 1/\alpha]$ such that

$$\sum_{n\in\mathbb{N}} \lambda_n(1-\alpha\lambda_n) = +\infty.$$

Let $x_0 \in \mathcal{H}$ and $(\forall n \in \mathbb{N})$ $x_{n+1} = x_n + \lambda_n(Tx_n - x_n)$.
Then $(x_n)_{n\in\mathbb{N}}$ converges (weakly) to a point in $\operatorname{Fix} T$.

Remark: if $\alpha < 1$, one can choose $(\forall n \in \mathbb{N})$ $\lambda_n = 1$, that is

$$(\forall n \in \mathbb{N}) \quad x_{n+1} = Tx_n.$$

Fixed point algorithms

Krasnosel'skii-Mann-like algorithm: stochastic variant

Same assumptions on T and $(\lambda_n)_{n\in\mathbb{N}}$.
Let x_0 and $(e_n)_{n\in\mathbb{N}}$ be $\mathcal{H}$-valued random variables and

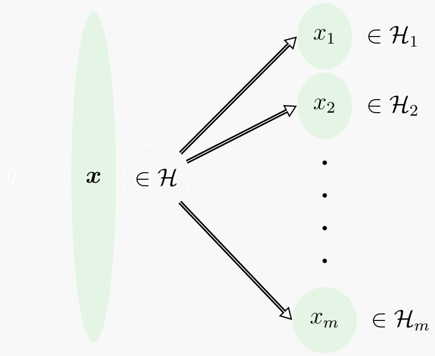
$$(\forall n \in \mathbb{N}) \quad x_{n+1} = x_n + \lambda_n(Tx_n + e_n - x_n).$$

Suppose that $\sum_{n\in\mathbb{N}} \lambda_n \sqrt{\mathsf{E}(\|e_n\|^2 \,|\, \mathcal{X}_n)} < +\infty$ a.s., where $\mathcal{X}_n$ is the σ-algebra generated by $(x_0, \ldots, x_n)$.
Then $(x_n)_{n\in\mathbb{N}}$ converges (weakly) a.s. to a $(\operatorname{Fix} T)$-valued random variable.

Fixed point algorithms

Krasnosel'skii-Mann-like algorithm: random block-coordinate variant

- Variable splitting



$\mathcal{H} = \times_{i=1}^m \mathcal{H}_i$

$\mathcal{H}_1, \ldots, \mathcal{H}_m$ are separable real Hilbert spaces

- Block decomposition of $T : x \mapsto (T_1 x, \ldots, T_m x)$

Fixed point algorithms

Krasnosel'skii-Mann-like algorithm: random block-coordinate variant

- Variable splitting
- Block decomposition of $T\colon x \mapsto (T_1 x, \ldots, T_m x)$
- Update of selected coordinates

$$(\forall i \in \{1,\ldots,m\})(\forall n \in \mathbb{N}) \quad x_{i,n+1} = x_{i,n} + \varepsilon_{i,n}\lambda_n(T_i x_n - x_{i,n})$$

where $\varepsilon_{i,n} \in \{0,1\}$ random activation variable.

- Assumptions
  - $\inf_{n\in\mathbb{N}} \lambda_n > 0$ and $\sup_{n\in\mathbb{N}} \lambda_n < 1/\alpha$
  - $(\varepsilon_n)_{n\in\mathbb{N}}$ are identically distributed
  - For every $n \in \mathbb{N}$, ε_n and $(x_0,\ldots,x_n)$ are mutually independent.
  - $(\forall i \in \{1,\ldots,m\})$ $\mathsf{P}[\varepsilon_{i,0} = 1] > 0$.

Then $(x_n)_{n\in\mathbb{N}}$ converges (weakly) a.s. to a $(\operatorname{Fix} T)$-valued random variable.

Optimization

Fixed point formulation

- Problem
  Let $f\colon \mathcal{H} \to \left]-\infty, +\infty\right]$. We want to

$$\text{Find} \quad \widehat{x} \in \operatorname{Argmin} f$$

- Reformulation

$$\text{Find} \quad \widehat{x} = T(\widehat{x})$$

  where $T\colon \mathcal{H} \to \mathcal{H}$.

Fixed point formulation

- Problem
  Let $f\colon \mathcal{H} \to \left]-\infty, +\infty\right]$. We want to

$$\text{Find} \quad \widehat{x} \in \operatorname{Argmin} f$$

- More general reformulation

$$\text{Find} \quad \begin{cases} \widehat{x} = \Phi(\widehat{z}) \\ \widehat{z} = T(\widehat{z}) \end{cases}$$

where $\Phi\colon \mathcal{K} \to \mathcal{H}$ and $T\colon \mathcal{K} \to \mathcal{K}$.

Fixed point formulation

- Problem

  Let $f\colon \mathcal{H} \to \left]-\infty, +\infty\right]$. We want to

$$\text{Find} \quad \widehat{x} \in \operatorname{Argmin} f$$

- More general reformulation

$$\text{Find} \quad \begin{cases} \widehat{x} = \Phi(\widehat{z}) \\ \widehat{z} = T(\widehat{z}) \end{cases}$$

  where $\Phi\colon \mathcal{K} \to \mathcal{H}$ and $T\colon \mathcal{K} \to \mathcal{K}$.

- Primal-dual methods

$$\widehat{z} = \begin{bmatrix} \widehat{x} \\ \widehat{v} \end{bmatrix}$$

  where $\widehat{v}$ solution to the dual optimization problem.

Algorithms

algorithm	function	Φ	T
gradient descent	ℓ	Id	$\mathrm{Id} - \gamma\nabla\ell$
proximal point	g	Id	$\mathrm{prox}_{\gamma g}$
proximal gradient/ forward-backward (FB)	$g+\ell$	Id	$\mathrm{prox}_{\gamma g} \circ (\mathrm{Id} - \gamma\nabla\ell)$
Tseng/ FBF	$g+\ell$	Id	$(\mathrm{Id} - \gamma\nabla\ell) \circ \mathrm{prox}_{\gamma g} \circ (\mathrm{Id} - \gamma\nabla\ell) + \gamma\nabla\ell$
Dual FB/ dual ascent	$g + h \circ L$ $+\frac{1}{2}\|\cdot - \widetilde{x}\|^2$	$\mathrm{prox}_g(\widetilde{x} - L^*\cdot)$	$\mathrm{prox}_{\gamma h^*} \circ \big(\mathrm{Id} + \gamma L \mathrm{prox}_g(\widetilde{x} - L^*\cdot)\big)$
Douglas-Rachford	$g+h$	$\mathrm{prox}_{\gamma h}$	$(2\mathrm{prox}_{\gamma g} - \mathrm{Id}) \circ (2\mathrm{prox}_{\gamma h} - \mathrm{Id})$
3 operator splitting	$g+h+\ell$	$\mathrm{prox}_{\gamma h}$	$\mathrm{prox}_{\gamma g} \circ \big(2\mathrm{prox}_{\gamma h} - \mathrm{Id} - \gamma\nabla\ell \circ \mathrm{prox}_{\gamma h}\big) + \mathrm{Id} - \mathrm{prox}_{\gamma h}$
ADMM	$g + h \circ L$	$(x,y,\lambda) \mapsto x$	$(x,y,\lambda) \mapsto \Big((\gamma^{-1}\partial f + L^*L)^{-1}L^*(y-\lambda), \mathrm{prox}_{\gamma^{-1}g}(Lx+\lambda), \lambda + Lx - y\Big)$
Condat-Vũ $\ell = 0$: Chambolle-Pock	$g + h \circ L + \ell$	$(x,v) \mapsto x$	$J_{MA} \circ (\mathrm{Id} - MB)$ $A(x,v) = \begin{bmatrix} \partial g & L^* \\ -L & \partial h^* \end{bmatrix} \begin{bmatrix} x \\ v \end{bmatrix}$, $B(x,v) = \begin{bmatrix} \nabla\ell(x) \\ 0 \end{bmatrix}$, $M = \begin{bmatrix} \tau^{-1}\mathrm{Id} & -L^* \\ -L & \sigma^{-1}\mathrm{Id} \end{bmatrix}^{-1}$
Loris-Verhoeven	$h \circ L + \ell$	$(x,v) \mapsto x$	$J_{MA} \circ (\mathrm{Id} - MB)$ $A(x,v) = \begin{bmatrix} 0 & L^* \\ -L & \partial h^* \end{bmatrix} \begin{bmatrix} x \\ v \end{bmatrix}$, $B(x,v) = \begin{bmatrix} \nabla\ell(x) \\ 0 \end{bmatrix}$, $M = \begin{bmatrix} \tau^{-1}\mathrm{Id} & 0 \\ 0 & \sigma^{-1}\mathrm{Id} - \tau L^* L \end{bmatrix}^{-1}$
Combettes-Pesquet $\ell = 0$: Briceño Arias-Combettes	$g + h \circ L + \ell$	$(x,v) \mapsto x$	$(\mathrm{Id} - \gamma B) \circ J_{\gamma A} \circ (\mathrm{Id} - \gamma B) + \gamma B$ $A(x,v) = \begin{bmatrix} \partial g(x) \\ \partial h^*(v) \end{bmatrix}$, $B(x,v) = \begin{bmatrix} \nabla\ell & L^* \\ -L & 0 \end{bmatrix} \begin{bmatrix} x \\ v \end{bmatrix}$

g and h proper lower semi-continuous convex functions
h^* Fenchel-Legendre conjugate of h
ℓ convex and differentiable function
L linear bounded operator with adjoint L^*
$\gamma \in]0,+\infty[$, $(\tau,\sigma) \in]0,+\infty[^2$ such that $\tau\sigma\|L\|^2 < 1$

Adjoint mismatch

E. Chouzenoux, J.-C. Pesquet, C. Riddell, M. Savanier, and Y. Trousset,
Convergence of proximal gradient algorithm in the presence of adjoint mismatch,
Inverse Problems, June 2021.

Problem

Penalized least squares

$$\underset{x\in\mathcal{H}}{\text{minimize}}\ \ \frac{1}{2}\|Hx-z\|^2+g(x)+\frac{\kappa}{2}\|x\|^2,$$

where

- $\mathcal{H}$ and $\mathcal{G}$ real Hilbert spaces
- $z\in\mathcal{G}$ and $H\colon\mathcal{H}\to\mathcal{G}$ bounded linear operator (i.e. projector in tomography)
- elastic net-like penalty: $g\in\Gamma_0(\mathcal{H})$ and $\kappa\in[0,+\infty[$

Forward-backward algorithm

$$(\forall n\in\mathbb{N})\ \ x_{n+1}=\operatorname{prox}_{\gamma g}\big((1-\gamma\kappa)x_n-\gamma H^*(Hx_n-z)\big),\quad \gamma>0$$

Difficulty: H^* may be hard to compute exactly
(i.e. backprojector in tomography)

Mismatched algorithm

Form

$$(\forall n \in \mathbb{N}) \quad x_{n+1} = \text{prox}_{\gamma g}\big((1-\gamma\kappa)x_n - \gamma K_n(Hx_n - z)\big).$$

where

- $\sum_{n\in\mathbb{N}} \|K_n - \overline{K}\| < +\infty$
- $(K_n)_{n\in\mathbb{N}}$ and $\overline{K}$ bounded linear operator from $\mathcal{G}$ to $\mathcal{H}$.

Key assumptions

- $L = \overline{K}H + \kappa\text{Id}$ is β- cocoercive .
  If no mismatch ($\overline{K} = H^*$), $\beta^{-1} = \|H\|^2 + \kappa$.

Mismatched algorithm

Form

$$(\forall n \in \mathbb{N}) \quad x_{n+1} = \mathrm{prox}_{\gamma g}\big((1-\gamma\kappa)x_n - \gamma K_n(Hx_n - z)\big).$$

where

- $\sum_{n\in\mathbb{N}} \|K_n - \overline{K}\| < +\infty$
- $(K_n)_{n\in\mathbb{N}}$ and $\overline{K}$ bounded linear operator from $\mathcal{G}$ to $\mathcal{H}$.

Key assumptions

- $L = \overline{K}H + \kappa \mathrm{Id}$ is β- cocoercive .

  Sufficient condition: $\lambda_{\min} > 0$, $\beta^{-1} = \left(\sqrt{\lambda_{\max}} + \frac{\|L-L^*\|}{2\sqrt{\lambda_{\min}}}\right)^2$,
  where $\lambda_{\min}$ (resp. $\lambda_{\max}$) minimum (resp. maximum) spectral value of $(L+L^*)/2$.

Mismatched algorithm

Form

$$(\forall n \in \mathbb{N}) \quad x_{n+1} = \operatorname{prox}_{\gamma g}\big((1-\gamma\kappa)x_n - \gamma K_n(Hx_n - z)\big).$$

where

- $\sum_{n\in\mathbb{N}} \|K_n - \overline{K}\| < +\infty$
- $(K_n)_{n\in\mathbb{N}}$ and $\overline{K}$ bounded linear operator from $\mathcal{G}$ to $\mathcal{H}$.

Key assumptions

- $L = \overline{K}H + \kappa\mathrm{Id}$ is β- cocoercive .
- $\mathcal{F} = \{x \in \mathcal{H} \mid 0 \in Lx - \overline{K}z + \partial g(x)\} \neq \varnothing$.
  If no mismatch, $\mathcal{F}$ is the set of minimizers.

Mismatched algorithm

Form

$$(\forall n \in \mathbb{N}) \quad x_{n+1} = \operatorname{prox}_{\gamma g}\big((1-\gamma\kappa)x_n - \gamma K_n(Hx_n - z)\big).$$

where

- $\sum_{n\in\mathbb{N}} \|K_n - \overline{K}\| < +\infty$
- $(K_n)_{n\in\mathbb{N}}$ and $\overline{K}$ bounded linear operator from $\mathcal{G}$ to $\mathcal{H}$.

Key assumptions

- $L = \overline{K}H + \kappa \mathrm{Id}$ is β- cocoercive .
- $\mathcal{F} = \big\{x \in \mathcal{H} \;\big|\; 0 \in Lx - \overline{K}z + \partial g(x)\big\} \neq \varnothing$.
  Sufficient condition: $\operatorname{dom}\partial g = \mathcal{H}$ and $\lim_{\|x\|\to+\infty} \frac{1}{2}\langle x \mid Lx\rangle + g(x) = +\infty$.

Main results

Convergence

Let $\gamma \in]0, 2\beta[$. Then the sequence $(x_n)_{n\in\mathbb{N}}$ generated by the mismatched algorithm converges weakly to a point $\widetilde{x} \in \mathcal{F}$.
In addition, if $\lambda_{\min} > 0$ and, for every $n \in \mathbb{N}$, $K_n = \overline{K}$, then $(x_n)_{n\in\mathbb{N}}$ converges linearly.

Error bound

Assume that the following hold.

1. Let $\mu \in [0, +\infty[$ be the strong convexity modulus of g. Either $\mu > 0$ or $\lambda_{\min} \neq 0$.
2. $\widehat{x}$ is a solution to the minimization problem.

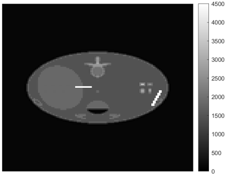
Then there exists a unique solution $\widetilde{x} \in \mathcal{F}$ and

$$\|\widetilde{x} - \widehat{x}\| \leqslant \chi \, \|(H^* - \overline{K})(H\widehat{x} - z)\|$$

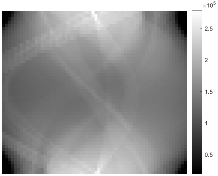
where $\chi \leqslant 1/(\mu + 2\lambda_{\min})$.

Reconstruction example

abdomen phantom in fan beam geometry
180° using 50 angles, 62 bin detector
$\overline{K}$ pixel-driven backprojector
$g \propto \|W \cdot\|_1$ with W orthogonal Symlet wavelet decomposition



original



sinogram

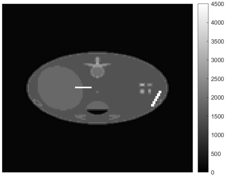
Reconstruction example

abdomen phantom in fan beam geometry

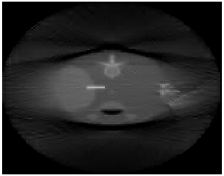
180° using 50 angles, 62 bin detector

$\overline{K}$ pixel-driven backprojector

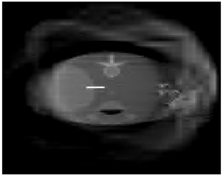
$g \propto \|W\cdot\|_1$ with W orthogonal Symlet wavelet decomposition



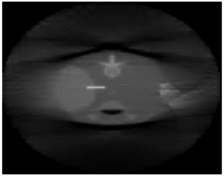
original



matched



mismatched divergent



mismatched convergent

Extensions

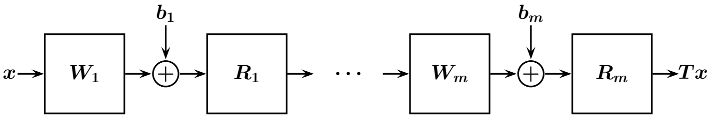
- Primal-dual formulations
  [Chouzenoux, Contreras, Pesquet, Savanier - 2023]

- Unmatched preconditioning
  [Chouzenoux, Savanier, Pesquet, Riddel - 2022]

Neural network compression

S. Verma and J.-C. Pesquet,
Sparsifying networks via subdifferential inclusion,
International Machine Learning Conference, Jul., 2021.

Feedforward NNs



Neural network model

Let $m \geqslant 1$ be an integer.

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

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 operator
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 a nonlinear (activation) operator.

REMARK More generally, $(W_i)_{1 \leqslant i \leqslant m}$ can be MIMO convolutive operators

Standard activation operators

Most of them are proximity operators [Combettes, Pesquet - 2020]

- Rectified linear unit (ReLU)

$$\varrho\colon \mathbb{R} \to \mathbb{R}\colon \xi \mapsto \begin{cases} \xi, & \text{if } \xi > 0; \\ 0, & \text{if } \xi \leqslant 0. \end{cases}$$

  Then, $\varrho = \mathrm{proj}_{[0,+\infty[}$.

- Parametric ReLU

$$\varrho\colon \mathbb{R} \to \mathbb{R}\colon \xi \mapsto \begin{cases} \xi, & \text{if } \xi > 0; \\ \alpha\xi, & \text{if } \xi \leqslant 0 \end{cases}, \qquad \alpha \in]0,1].$$

  Then $\varrho = \mathrm{prox}_\phi$ where

$$\phi\colon \mathbb{R} \to \mathbb{R}\colon \xi \mapsto \begin{cases} 0, & \text{if } \xi > 0; \\ (1/\alpha - 1)\xi^2/2, & \text{if } \xi \leqslant 0. \end{cases}$$

Standard activation operators

Most of them are proximity operators [Combettes, Pesquet - 2020]

- Unimodal sigmoid

$$\varrho\colon \mathbb{R} \to \mathbb{R}\colon \xi \mapsto \frac{1}{1+e^{-\xi}} - \frac{1}{2}$$

Then $\varrho = \operatorname{prox}_\phi$ where

$$\phi\colon \xi \mapsto \begin{cases} (\xi+1/2)\ln(\xi+1/2) + (1/2-\xi)\ln(1/2-\xi) - \frac{1}{2}(\xi^2+1/4) & \text{if } |\xi| < 1/2; \\ -1/4, & \text{if } |\xi| = 1/2; \\ +\infty, & \text{if } |\xi| > 1/2. \end{cases}$$

- Elliot function

$$\varrho\colon \mathbb{R} \to \mathbb{R}\colon \xi \mapsto \frac{\xi}{1+|\xi|}.$$

We have $\varrho = \operatorname{prox}_\phi$, where

$$\phi\colon \mathbb{R} \to]-\infty,+\infty]\colon \xi \mapsto \begin{cases} -|\xi| - \ln(1-|\xi|) - \frac{\xi^2}{2}, & \text{if } |\xi| < 1; \\ +\infty, & \text{if } |\xi| \geqslant 1. \end{cases}$$

Standard activation operators

Most of them are proximity operators [Combettes, Pesquet - 2020]

- Softmax

$$R\colon \mathbb{R}^N \to \mathbb{R}^N \colon (\xi_k)_{1\leqslant k\leqslant N} \mapsto \left(\exp(\xi_k) \Big/ \sum_{j=1}^{N} \exp(\xi_j)\right)_{1\leqslant k\leqslant N} - u,$$

where $u = (1, \ldots, 1)/N \in \mathbb{R}^N$.
Then $R = \operatorname{prox}_\varphi$ where $\varphi = \psi(\cdot + u) + \langle \cdot \mid u\rangle$ and

$$\psi\colon \mathbb{R}^N \to]-\infty, +\infty]$$

$$(\xi_k)_{1\leqslant k\leqslant N} \mapsto \begin{cases} \displaystyle\sum_{k=1}^{N}\left(\xi_k \ln \xi_k - \frac{\xi_k^2}{2}\right), & \text{if } (\xi_k)_{1\leqslant i\leqslant N} \in [0,1]^N \\ & \text{and } \displaystyle\sum_{k=1}^{N} \xi_k = 1; \\ +\infty, & \text{otherwise.} \end{cases}$$

Standard activation operators

Most of them are proximity operators [Combettes, Pesquet - 2020]

- 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

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

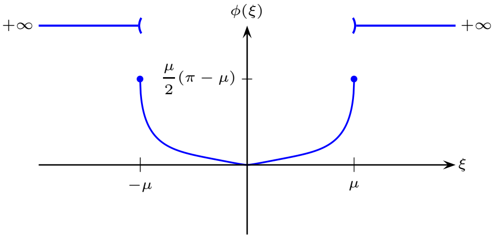
Standard activation operators

Most of them are proximity operators [Combettes, Pesquet - 2020]

- Squashing function used in capsnets

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

where



FB model of NNs

Let $m \geqslant 1$ be an integer.

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

where $(\forall i \in \{1, \ldots, m\}) \quad T_i \colon \mathbb{R}^{N_{i-1}} \to \mathbb{R}^{N_i} : x \mapsto \mathrm{prox}_{f_i}(W_i x + b_i)$,
$W_i \in \mathbb{R}^{N_i \times N_{i-1}}$ is a weight operator
b_i is a (bias) vector in $\mathbb{R}^{N_i}$,
and $f_i \in \Gamma_0(\mathbb{R}^{N_i})$

Subdifferential inclusion

$$\underbrace{x_i}_{\text{output of } i\text{-th layer}} = \mathrm{prox}_{f_i}(W_i \underbrace{x_{i-1}}_{\text{input of } i\text{-th layer}} + b_i)$$

$$\Leftrightarrow \quad W_i x_{i-1} + b_i - x_i \in \partial f_i(x_i)$$

$$\Leftrightarrow \quad d_{\partial f_i(x_i)}(W_i x_{i-1} + b_i - x_i) = 0$$

Convex formulation of NN compression

- Data decomposition: P mini-batches $(\mathbb{B}_j)_{1 \leqslant j \leqslant P}$

- Minimization problem

$$\underset{(W_i, b_i) \in \bigcap_{j=1}^{P} C_{i,j}}{\text{minimize}} \quad \|W_i\|_1$$

where, for every $j \in \{1, \ldots, P\}$,

$$C_{i,j} = \Big\{(W_i, b_i) \mid \sum_{t \in \mathbb{B}_j} d^2_{\partial f_i(x_{i,t})}(W_i x_{i-1,t} + b_i - x_{i,t}) = 0\Big\}$$

Convex formulation of NN compression

- Data decomposition: P mini-batches $(\mathbb{B}_j)_{1\leqslant j\leqslant P}$

- Minimization problem

$$\underset{(W_i,b_i)\in\bigcap_{j=1}^P C_{i,j}}{\text{minimize}} \quad \|W_i\|_1$$

where, for every $j \in \{1,\ldots,P\}$,

$$C_{i,j} = \Big\{(W_i,b_i) \;\Big|\; \sum_{t\in\mathbb{B}_j} d^2_{\partial f_i(x_{i,t})}(W_i x_{i-1,t} + b_i - x_{i,t}) \leqslant |\mathbb{B}_j|\eta\Big\}$$

$\eta > 0$: accuracy tolerance

Convex formulation of NN compression

- Data decomposition: P mini-batches $(\mathbb{B}_j)_{1\leqslant j\leqslant P}$
- Minimization problem

$$\underset{(W_i,b_i)\in\bigcap_{j=1}^P C_{i,j}}{\text{minimize}} \quad \|W_i\|_1$$

where, for every $j \in \{1,\ldots,P\}$,

$$C_{i,j} = \Big\{(W_i,b_i) \;\Big|\; \sum_{t\in\mathbb{B}_j} d^2_{\partial f_i(x_{i,t})}(W_i x_{i-1,t} + b_i - x_{i,t}) \leqslant |\mathbb{B}_j|\eta\Big\}$$

$\eta > 0$: accuracy tolerance

- SIS algorithm
Douglas-Rachford iterations
$\rightsquigarrow \operatorname{proj}_{\bigcap_{j=1}^P C_{i,j}}$ computed by block-iterative subgradient projection algorithm

Numerical results

Accuracy

Dataset	CIFAR-10			CIFAR-100		
Pruning ratio	90%	95%	98%	90%	95%	98%
ResNet50 (Baseline)	94.62	-	-	77.39	-	-
SNIP	92.65	90.86	87.21	73.14	69.25	58.43
GraSP	92.47	91.32	88.77	73.28	70.29	62.12
SynFlow	92.49	91.22	88.82	73.37	70.37	62.17
STR	92.59	91.35	88.75	73.45	70.45	62.34
FORCE	92.56	91.46	88.88	73.54	70.37	62.39
LRR	92.62	91.27	89.11	**74.13**	70.38	62.47
RigL	92.55	91.42	89.03	73.77	70.49	62.33
SIS (Ours)	**92.81**	**91.69**	**90.11**	73.81	**70.62**	**62.75**

Infererence FLOPS at 90% sparsity level on ImageNet

ResNet50 (Baseline)	SNIP	GraSP	SynFlow	STR	FORCE	SIS (Ours)
4.14G	409M	470M	465M	341M	455M	**298M**

Maximally Monotone Regularization

J.-C. Pesquet, A. Repetti, M. Terris, and Y. Wiaux,
Learning maximally monotone operators for image recovery,
SIAM Journal on Imaging Sciences, Aug. 2021.

Variational formulations of inverse problems

- Optimization problem

$$\text{Find} \quad \hat{x} \in \underset{x \in \mathbb{R}^N}{\operatorname{Argmin}} \quad \underbrace{\ell(x)}_{\text{Data fidelity term}} \quad + \quad \underbrace{g(x)}_{\text{Regularization term}}$$

MAIN CHALLENGE:
Choose the right form of regularizer and its right parameters

Variational formulations of inverse problems

- Optimization problem

$$\text{Find} \quad \hat{x} \in \underset{x \in \mathbb{R}^N}{\operatorname{Argmin}} \quad \underbrace{\ell(x)}_{\text{Data fidelity term}} \quad + \quad \underbrace{g(x)}_{\text{Regularization term}}$$

- Equivalent variational inclusion problem
If $\ell \in \Gamma_0(\mathbb{R}^N)$ and $g \in \Gamma_0(\mathbb{R}^N)$ (+ qualification condition), then

$$0 \in \partial \ell(\widehat{x}) + \partial g(\widehat{x})$$

Variational formulations of inverse problems

- Optimization problem

$$\text{Find} \quad \hat{x} \in \underset{x \in \mathbb{R}^N}{\text{Argmin}} \; \underbrace{\ell(x)}_{\text{Data fidelity term}} + \underbrace{g(x)}_{\text{Regularization term}}$$

- Extension to monotone inclusion problem

If $\ell \in \Gamma_0(\mathbb{R}^N)$, then

$$0 \in \partial\ell(\widehat{x}) + A(\widehat{x})$$

where A is a MMO

$\rightsquigarrow$ new regularization paradigm

$\rightsquigarrow$ more general

$\rightsquigarrow$ easier to learn

PnP approach

- Assumption: ℓ is differentiable with a $1/\beta$-Lipschitz gradient, $\beta \in]0, +\infty[$
- Forward-backward algorithm

$$(\forall n \in \mathbb{N}) \quad x_{n+1} = \underbrace{J_{\gamma A}}_{\text{Denoiser}} \big(x_n - \gamma \nabla \ell(x_n)\big)$$

with $\gamma \in]0, 2\beta[$.

- Objectives
  - Learn the best denoiser $\sim$ recent PnP approaches
  - With guaranteed convergence conditions
  - By characterizing the limit point.

Learning strategy

- Minty's theorem

$$J_{\gamma A} = \frac{\mathrm{Id} + Q}{2}$$

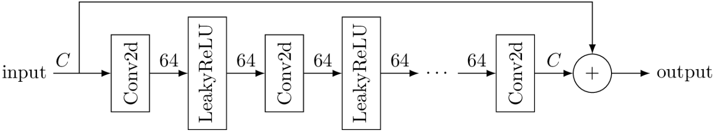
  where Q nonexpansive
- Q modelling by nonexpansive neural network
  $\rightsquigarrow$ universal approximation theorem to MMOs using nonexpansive feedforward NNs
- Nonexpansiveness condition

$$(\forall x \in \mathbb{R}^N) \quad \|\nabla Q_\theta(x)\| \leqslant 1$$

  θ: parameters of the NN
  $\rightsquigarrow$ penalization $\lambda \sum_{t=1}^{T} \max\{\|\nabla Q_\theta(x_t)\|^2, 1-\epsilon\}$
  $\epsilon \in [0,1[$, $\lambda \in]0,+\infty[$, $(x_t)_{1 \leqslant t \leqslant T}$: training sequence
- Learning θ
  image denoising task in the presence of zero-mean white Gaussian noise based on penalized MSE loss

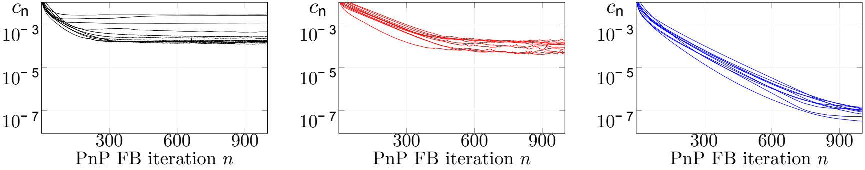
Image restoration results

- $Q_\theta \equiv$ DnCNN – 20 layers



- Convergence of PnP-FB

  ☛ Evaluate $c_n = \|x_n - x_{n-1}\|/\|x_0\|$, for generated sequence $(x_n)_{n\in\mathbb{N}}$



(a) BM3D (b) RealSN (c) Proposed

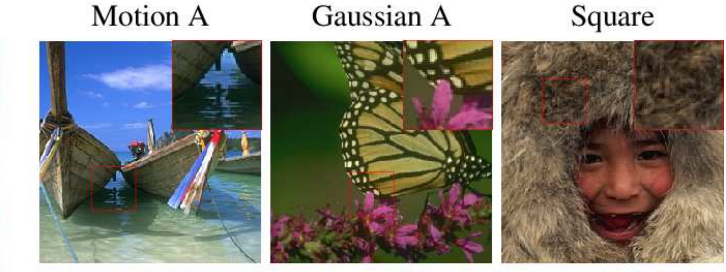
Image restoration results

- $Q_\theta \equiv$ DnCNN – 20 layers
- PSNR for grayscale images (BSD68)

denoiser	kernel							
	(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)
Observation	23.36	22.93	23.43	19.49	23.84	19.85	20.75	20.67
RealSN	26.24	26.25	26.34	25.89	25.08	25.84	24.81	23.92
$\mathrm{prox}_{\mu_{\ell_1}\|\Psi^\dagger_{\mathrm{Wav}}\cdot\|_1}$	29.44	29.20	29.31	28.87	30.90	30.81	29.40	29.06
$\mathrm{prox}_{\mu_{\mathrm{TV}}\|\cdot\|_{\mathrm{TV}}}$	29.70	29.35	29.43	29.15	30.67	30.62	29.61	29.23
DnCNN	29.82	29.24	29.26	28.88	30.84	30.95	29.54	29.17
BM3D	30.05	29.53	29.93	29.10	31.08	30.78	29.56	29.41
Proposed	**30.86**	**30.33**	**30.31**	**30.14**	**31.72**	**31.69**	**30.42**	**30.09**

Visual results on color images

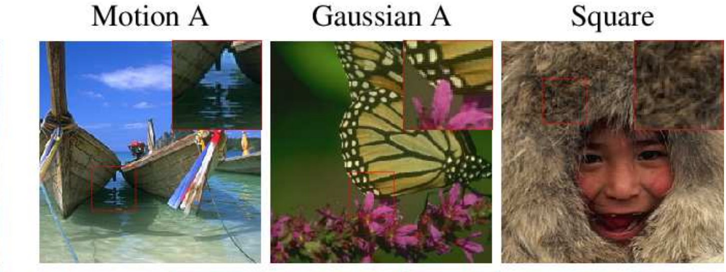
BSD500 test set



Observed

Visual results on color images

BSD500 test set



VAR

Visual results on color images

BSD500 test set

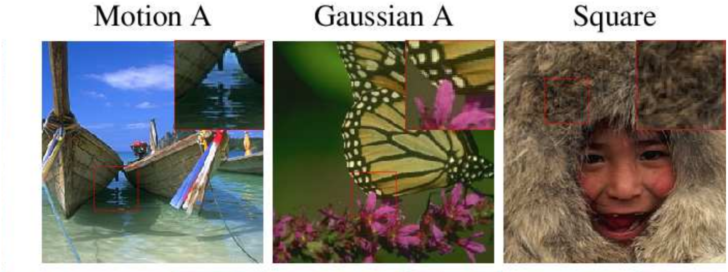
Motion A Gaussian A Square

BM3D

(29.73, 0.834) (29.32, 0.891) (26.97, 0.611)

Visual results on color images

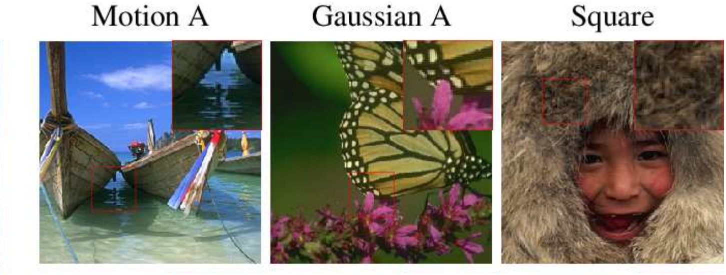
BSD500 test set



DnCNN

Visual results on color images

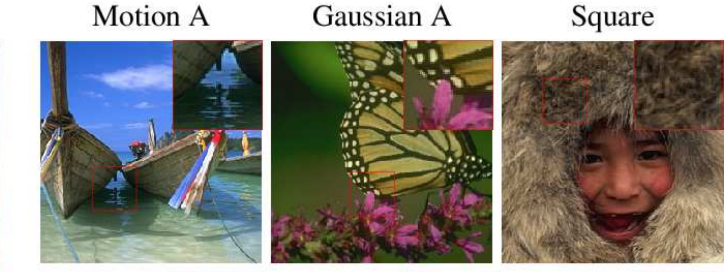
BSD500 test set



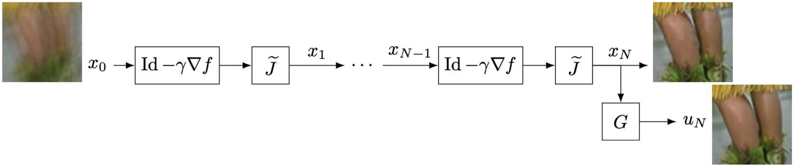
Proposed

Visual results on color images

BSD500 test set



Further improvement: U-Net postprocessing



Conclusion

- Fixed point theory: backbone of optimization methods
- General framework for analyzing approaches which go beyond optimization
- Wide number of applications
- Many developments skipped: parallel splitting, primal-dual formulations, Bregman divergences, game theory,...

P. L. Combettes and J.-C. Pesquet,
Fixed point strategies in data science,
IEEE Transactions on Signal Processing, March 2021.

Thank you for your attention!

https://jc.pesquet.eu