On strong convexity for the understanding and design of (unfolded) algorithms

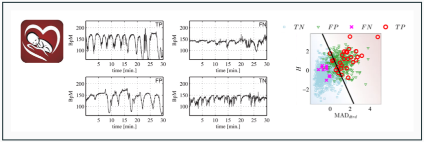
Nelly Pustelnik

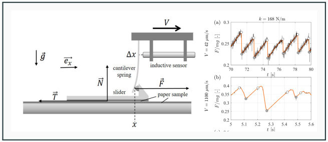
Challenges and advances in modern variational analysis
Limoges – March 15th, 2023

ENS DE LYON

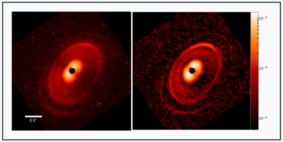
Context

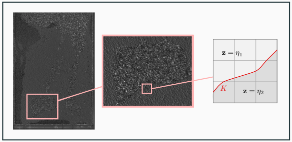
- **Signal processing**



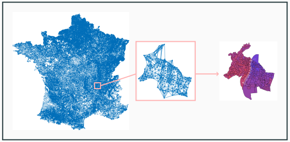


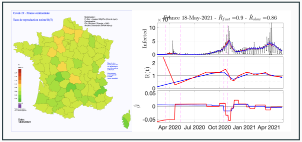
- **Image processing**



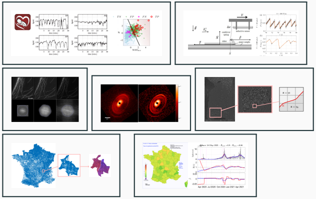


- **Graph processing**





Context



➞ **Three quantities of interest**:

- $\mathrm{z} \in \mathbb{R}^{\overline{N}}$: Data/measures.
- $\overline{\mathrm{x}} \in \mathbb{R}^{N}$: True (unknown) parameters.
- $\widehat{\mathrm{x}} \in \mathbb{R}^{N}$: Estimated parameters.

Context

Data: z

Direct/acquisition model

Design a cost function

Minimization algorithm

Hyperparameter tuning

Estimated parameters: $\widehat{\mathrm{x}}(\mathrm{z}; \widehat{\lambda})$

$$\mathrm{z} = \mathcal{D}(\mathrm{A}\overline{\mathrm{x}})$$

$$\widehat{\mathrm{x}}(\mathrm{z}; \lambda) \in \underset{\mathrm{x}}{\mathrm{Argmin}}\ \varphi(\mathrm{x}; \mathrm{z}) + \lambda \psi(\mathrm{x})$$

Sequence $\mathrm{x}^{[k+1]} = \boldsymbol{\Phi} \mathrm{x}^{[k]}$

$$\widehat{\lambda} \in \underset{\lambda}{\mathrm{Argmin}} \|\overline{\mathrm{x}} - \widehat{\mathrm{x}}(\mathrm{z}; \lambda)\|_2^2$$

On strong convexity for the understanding and design of algorithms

Context

Data: z

Direct/acquisition model

Design a cost function

Minimization algorithm

Hyperparameter tuning

Estimated parameters: $\widehat{\mathrm{x}}(\mathrm{z}; \widehat{\lambda})$

$$\mathrm{z} = \mathcal{D}(\mathrm{A}\overline{\mathrm{x}})$$

$$\widehat{\mathrm{x}}(\mathrm{z}; \lambda) \in \underset{\mathrm{x}}{\mathrm{Argmin}}\ \varphi(\mathrm{x}; \mathrm{z}) + \lambda \psi(\mathrm{x})$$

Sequence $\mathrm{x}^{[k+1]} = \boldsymbol{\Phi} \mathrm{x}^{[k]}$

$$\widehat{\lambda} \in \underset{\lambda}{\mathrm{Argmin}} \|\overline{\mathrm{x}} - \widehat{\mathrm{x}}(\mathrm{z}; \lambda)\|_2^2$$

Iterative scheme

➡ **Minimization problem** :

$$\widehat{\mathrm{x}} \in \underset{\mathrm{x}}{\mathrm{Argmin}}\ f(\mathrm{x}) + g(\mathrm{x})$$

➡ **Design of a recursive sequence of the form**

$$(\forall k \in \mathbb{N}) \qquad \mathrm{x}^{[k+1]} = \boldsymbol{\Phi}\mathrm{x}^{[k]},$$

Gradient descent	$\boldsymbol{\Phi} = \mathrm{Id} - \tau(\nabla f + \nabla g)$
Proximal point algorithm	$\boldsymbol{\Phi} = \mathrm{prox}_{\tau(f+g)}$
Forward-Backward	$\boldsymbol{\Phi} = \mathrm{prox}_{\tau g}(\mathrm{Id} - \tau\nabla f)$
Peaceman-Rachford	$\boldsymbol{\Phi} = (2\mathrm{prox}_{\tau g} - \mathrm{Id}) \circ (2\mathrm{prox}_{\tau f} - \mathrm{Id})$
Douglas-Rachford	$\boldsymbol{\Phi} = \mathrm{prox}_{\tau g}(2\mathrm{prox}_{\tau f} - \mathrm{Id}) + \mathrm{Id} - \mathrm{prox}_{\tau f}$

Iterative scheme

➡ **Minimization problem** :

$$\widehat{x} \in \underset{x}{\operatorname{Argmin}}\ f(x) + h(Dx)$$

- **Require the computation of** $\operatorname{prox}_{h(D\cdot)}$. **Few closed form.**
- **Reformulation in the dual**: $\min_{w \in \mathcal{G}} f^*(-D^* w) + h^*(w)$,
- **Primal-dual algorithms**: $\min_{x} f(x) + \widetilde{f}(x) + h(Dx)$,

  $\rightarrow$ f has a ν-Lipschitz gradient.

Hyperparameters setting: $\tau > 0$, $\gamma > 0$, such that $\frac{1}{\tau} - \gamma \|D\|^2 > \frac{\nu}{2}$

For $k = 0, 1, \ldots$

$$\left\lfloor \begin{array}{l} w^{[k+1]} = \operatorname{prox}_{\tau \widetilde{f}}\big(w^{[k]} - \tau \nabla f(w^{[k]}) - \tau D^* x^{[k]}\big) \\ x^{[k+1]} = \operatorname{prox}_{\gamma h^*}\big(x^{[k]} + \gamma D(2w^{[k+1]} - w^{[k]})\big) \end{array} \right.$$

Iterative scheme

➡ **Minimization problem** :

$$\widehat{\mathrm{x}} \in \underset{\mathrm{x}}{\mathrm{Argmin}}\ f(\mathrm{x}) + h(\mathrm{D}\mathrm{x})$$

- **Require the computation of** $\mathrm{prox}_{h(\mathrm{D}\cdot)}$. **Few closed form.**
- **Reformulation in the dual**: $\min_{\mathrm{w}\in\mathcal{G}} f^*(-\mathrm{D}^*\mathrm{w}) + h^*(\mathrm{w})$,
- **Primal-dual algorithms**: $\min_{\mathrm{x}} f(\mathrm{x}) + \widetilde{f}(\mathrm{x}) + h(\mathrm{D}\mathrm{x})$,

  $\rightarrow$ Acceleration when $\widetilde{f}$ **strongly convex**.

Hyperparameters setting: $\tau > 0$, $\gamma > 0$, such that $\frac{1}{\tau} - \gamma\|\mathrm{D}\|^2 > \frac{\nu}{2}$

For $k = 0, 1, \ldots$

$$\left\lfloor \begin{array}{l} \mathrm{w}^{[k+1]} = \mathrm{prox}_{\tau\widetilde{f}}\left(\mathrm{w}^{[k]} - \tau\nabla f(\mathrm{w}^{[k]}) - \tau\mathrm{D}^*\mathrm{x}^{[k]}\right) \\ \mathrm{x}^{[k+1]} = \mathrm{prox}_{\gamma h^*}\left(\mathrm{x}^{[k]} + \gamma\mathrm{D}(2\mathrm{w}^{[k+1]} - \mathrm{w}^{[k]})\right) \end{array} \right.$$

Preliminary remarks

➡ **Minimization problem** :

$$\widehat{x} \in \underset{x}{\operatorname{Argmin}} \; f(x) + g(x)$$

- Smooth and strongly convex.
- Focus on (linear) convergence of the iterates, i.e.

$$(\forall k \in \mathbb{N}) \qquad \|x^{[k]} - \widehat{x}\| \leq r^k \|x^{[0]} - \widehat{x}\|.$$

Questions:

- Proximal step or gradient step ?
- Design efficiency region diagram ?

Notations: $C_L^{1,1}$ models the class of differentiable functions having a L-Lipschitz gradient.

Theoretical comparisons

Proposition (see [**Briceño-Arias, Pustelnik, 2021**] for detailed references)
In the context of $\min f + g$ where $f, g \in \Gamma_0(\mathcal{H})$, $f \in C^{1,1}_{L_f}(\mathcal{H})$, f is ρ-strongly convex, and $g \in C^{1,1}_{L_g}(\mathcal{H})$, for some $\rho \in \left]0, L_f\right[$, and let $\tau > 0$. Then, the following holds:

- **Gradient descent** Suppose that $\tau \in \left]0, 2L_f^{-1}L_g^{-1}/(L_g^{-1} + L_f^{-1})\right[$. Then, $\mathrm{Id} - \tau(\nabla g + \nabla f)$ is $r_G(\tau)-$Lipschitz continuous, where

$$r_G(\tau) := \max\left\{|1 - \tau\rho|, |1 - \tau(L_f + L_g)|\right\} \in \left]0, 1\right[.$$

Theoretical comparisons

Proposition (see [**Briceño-Arias, Pustelnik, 2021**] for detailed references)
In the context of $\min f + g$ where $f, g \in \Gamma_0(\mathcal{H})$, $f \in C^{1,1}_{L_f}(\mathcal{H})$, f is ρ-strongly convex, and $g \in C^{1,1}_{L_g}(\mathcal{H})$, for some $\rho \in]0, L_f[$, and let $\tau > 0$. Then, the following holds:

- **Gradient descent** Suppose that $\tau \in \left]0, 2L_f^{-1}L_g^{-1}/(L_g^{-1} + L_f^{-1})\right[$. Then, $\mathrm{Id} - \tau(\nabla g + \nabla f)$ is $r_G(\tau)-$Lipschitz continuous, where

$$r_G(\tau) := \max\left\{|1 - \tau\rho|, |1 - \tau(L_f + L_g)|\right\} \in]0, 1[\,.$$

In particular, the minimum is achieved at

$$\tau^* = \frac{2}{\rho + L_f + L_g}$$

and

$$r_G(\tau^*) = \frac{L_f + L_g - \rho}{L_f + L_g + \rho}.$$

Theoretical comparisons

Proposition (see [Bricẽno-Arias, Pustelnik, 2021] for detailed references)
In the context of $\min f + g$ where $f, g \in \Gamma_0(\mathcal{H})$, $f \in C^{1,1}_{L_f}(\mathcal{H})$, f is ρ-strongly convex, and $g \in C^{1,1}_{L_g}(\mathcal{H})$, for some $\rho \in \,]0, L_f[$, and let $\tau > 0$.

- **FBS** Suppose that $\tau \in \left]0, 2L_f^{-1}\right[$. Then $\operatorname{prox}_{\tau g}(\mathrm{Id} - \tau\nabla f)$ is $r_{T_1}(\tau)-$Lipschitz continuous, where

$$r_{T_1}(\tau) := \max\left\{|1 - \tau\rho|, |1 - \tau L_f|\right\} \in \,]0, 1[\,.$$

Theoretical comparisons

Proposition (see [Bricěno-Arias, Pustelnik, 2021] for detailed references)
In the context of $\min f + g$ where $f, g \in \Gamma_0(\mathcal{H})$, $f \in C^{1,1}_{L_f}(\mathcal{H})$, f is ρ-strongly convex, and $g \in C^{1,1}_{L_g}(\mathcal{H})$, for some $\rho \in]0, L_f[$, and let $\tau > 0$.

- **FBS** Suppose that $\tau \in \left]0, 2L_f^{-1}\right[$. Then $\mathrm{prox}_{\tau g}(\mathrm{Id} - \tau \nabla f)$ is $r_{T_1}(\tau)-$Lipschitz continuous, where

$$r_{T_1}(\tau) := \max\left\{|1 - \tau\rho|, |1 - \tau L_f|\right\} \in]0, 1[\,.$$

In particular, the minimum in (1) is achieved at

$$\tau^* = \frac{2}{\rho + L_f} \quad \text{and} \quad r_{T_1}(\tau^*) = \frac{L_f - \rho}{L_f + \rho}.$$

Theoretical comparisons

Proposition (see [**Briceño-Arias, Pustelnik, 2021**] for detailed references)
In the context of $\min f + g$ where $f, g \in \Gamma_0(\mathcal{H})$, $f \in C^{1,1}_{L_f}(\mathcal{H})$, f is ρ-strongly convex, and $g \in C^{1,1}_{L_g}(\mathcal{H})$, for some $\rho \in]0, L_f[$, and let $\tau > 0$.

- **FBS (v2)** Suppose that $\tau \in \left]0, 2L_g^{-1}\right]$. Then $\mathrm{prox}_{\tau f}(\mathrm{Id} - \tau\nabla g)$ is $r_{T_2}(\tau)-$Lipschitz continuous, where

$$r_{T_2}(\tau) := \frac{1}{1 + \tau\rho} \in]0, 1[\,.$$

Theoretical comparisons

Proposition (see [**Briceño-Arias, Pustelnik, 2021**] for detailed references)
In the context of $\min f + g$ where $f, g \in \Gamma_0(\mathcal{H})$, $f \in C^{1,1}_{L_f}(\mathcal{H})$, f is ρ-strongly convex, and $g \in C^{1,1}_{L_g}(\mathcal{H})$, for some $\rho \in \,]0, L_f[$, and let $\tau > 0$.

- **FBS (v2)** Suppose that $\tau \in \left]0, 2L_g^{-1}\right]$. Then $\mathrm{prox}_{\tau f}(\mathrm{Id} - \tau\nabla g)$ is $r_{T_2}(\tau)-$Lipschitz continuous, where

$$r_{T_2}(\tau) := \frac{1}{1+\tau\rho} \in \,]0,1[\,.$$

In particular, the minimum is achieved at

$$\tau^* = 2L_g^{-1} \quad \text{and} \quad r_{T_2}(\tau^*) = \frac{1}{1+2L_g^{-1}\rho}.$$

Theoretical comparisons

Proposition (see [Briceño-Arias, Pustelnik, 2021] for detailed references)
In the context of $\min f + g$ where $f, g \in \Gamma_0(\mathcal{H})$, $f \in C^{1,1}_{L_f}(\mathcal{H})$, f is ρ-strongly convex, and $g \in C^{1,1}_{L_g}(\mathcal{H})$, for some $\rho \in \,]0, L_f[$, and let $\tau > 0$.

•**PRS** $(2\mathrm{prox}_{\tau g} - \mathrm{Id}) \circ (2\mathrm{prox}_{\tau f} - \mathrm{Id})$ and $(2\mathrm{prox}_{\tau f} - \mathrm{Id}) \circ (2\mathrm{prox}_{\tau g} - \mathrm{Id})$ are $r_R(\tau)-$Lipschitz continuous, where

$$r_R(\tau) = \max\left\{\frac{1-\tau\rho}{1+\tau\rho}, \frac{\tau L_f - 1}{\tau L_f + 1}\right\} \in \,]0,1[\,.$$

Theoretical comparisons

Proposition (see [**Briceño-Arias, Pustelnik, 2021**] for detailed references)
In the context of $\min f + g$ where $f, g \in \Gamma_0(\mathcal{H})$, $f \in C^{1,1}_{L_f}(\mathcal{H})$, f is ρ-strongly convex, and $g \in C^{1,1}_{L_g}(\mathcal{H})$, for some $\rho \in]0, L_f[$, and let $\tau > 0$.

•**PRS** $(2\mathrm{prox}_{\tau g} - \mathrm{Id}) \circ (2\mathrm{prox}_{\tau f} - \mathrm{Id})$ and $(2\mathrm{prox}_{\tau f} - \mathrm{Id}) \circ (2\mathrm{prox}_{\tau g} - \mathrm{Id})$ are $r_R(\tau)-$Lipschitz continuous, where

$$r_R(\tau) = \max\left\{\frac{1-\tau\rho}{1+\tau\rho}, \frac{\tau L_f - 1}{\tau L_f + 1}\right\} \in]0,1[\,.$$

In particular, the minimum is achieved at

$$\tau^* = \sqrt{\frac{1}{\rho L_f}} \quad \text{and} \quad r_R(\tau^*) = \frac{1-\sqrt{L_f^{-1}\rho}}{1+\sqrt{L_f^{-1}\rho}}.$$

Theoretical comparisons

Proposition (see [Briceño-Arias, Pustelnik, 2021] for detailed references)
In the context of $\min f + g$ where $f, g \in \Gamma_0(\mathcal{H})$, $f \in C^{1,1}_{L_f}(\mathcal{H})$ and $g \in C^{1,1}_{L_g}(\mathcal{H})$, suppose that f is $\rho-$strongly convex, for some $\rho \in]0, L_f[$, and let $\tau > 0$. Then, the following holds:

- **DRS** $S_{\tau\nabla g, \tau\nabla f}$ and $S_{\tau\nabla f, \tau\nabla g}$ are $r_S(\tau)-$Lipschitz continuous, where

$$r_S(\tau) = \min\left\{\frac{1 + r_R(\tau)}{2}, \frac{L_g^{-1} + \tau^2\rho}{L_g^{-1} + \tau L_g^{-1}\rho + \tau^2\rho}\right\} \in]0, 1[$$

and r_R is defined in p.21.

Theoretical comparisons

Proposition (see [Briceño-Arias, Pustelnik, 2021] for detailed references)
In the context of $\min f + g$ where $f, g \in \Gamma_0(\mathcal{H})$, $f \in C^{1,1}_{L_f}(\mathcal{H})$ and $g \in C^{1,1}_{L_g}(\mathcal{H})$, suppose that f is $\rho-$strongly convex, for some $\rho \in]0, L_f[$, and let $\tau > 0$. Then, the following holds:

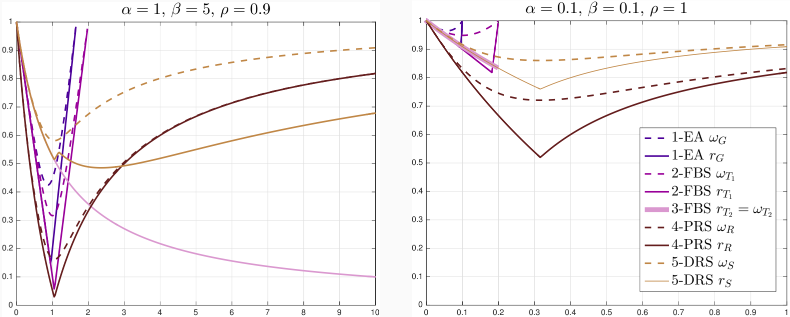
- **DRS** $S_{\tau\nabla g,\tau\nabla f}$ and $S_{\tau\nabla f,\tau\nabla g}$ are $r_S(\tau)-$Lipschitz continuous, where

$$r_S(\tau) = \min\left\{\frac{1 + r_R(\tau)}{2}, \frac{L_g^{-1} + \tau^2\rho}{L_g^{-1} + \tau L_g^{-1}\rho + \tau^2\rho}\right\} \in]0, 1[$$

and r_R is defined in p.21.
In particular, the optimal step-size and the minimum in (1) are

$$(\tau^*, r_S(\tau^*)) = \begin{cases} \left(\sqrt{\frac{1}{\rho L_f}}, \frac{1}{1+\sqrt{L_f^{-1}\rho}}\right), & \text{if } L_f \leq 4L_g; \\ \left(\sqrt{\frac{1}{\rho L_g}}, \frac{2}{2+\sqrt{L_g^{-1}\rho}}\right), & \text{otherwise.} \end{cases}$$

Theoretical comparisons



Comparison of the convergence rates of EA, FBS, PRS, DRS for two choices of $\alpha = L_f^{-1}$, $\beta = L_g^{-1}$, and ρ. Note that optimization rates are better than cocoercive rates in general.

Theoretical comparisons

Proposition [Briceño-Arias, Pustelnik, 2021]
Let $(L_g, \rho) \in \,]0, +\infty[\, \times \,]0, 1[$. Then $r^*_G(L_g, \rho) > r^*_{T_1}(\rho) > r^*_R(\rho)$.

➡ The linear convergence rate of PRS is always smaller than those of algorithms governed by operators EA (gradient descent) and FBS (forward-backward splitting).

Theoretical comparisons

Proposition [Briceño-Arias, Pustelnik, 2021]
Let $(L_g, \rho) \in \ ]0, +\infty[\times]0, 1[$ and

$$\eta(L_g) = \frac{1 - \sqrt{1 - 4L_g}}{1 + \sqrt{1 - 4L_g}} \in]0, 1[\,.$$

. Then, the following holds:

- Suppose that $L_g < \frac{1}{4}$ and that $\rho \in I(L_g)$, where $I(\beta) = \left[L_g \max\{1/16, \eta(L_g)\}, \frac{L_g}{\eta(L_g)} \right]$.

Then

$$r^*_{T_2}(L_g, \rho) \leq \min\{r^*_S(L_g, \rho), r^*_R(\rho)\}.$$

- Suppose that $L_g < \frac{1}{16}$ and that $\rho < \chi(L_g)$, where $\chi(L_g) = \min\left\{ \frac{L_g}{16}, 1 - 8L_g(\sqrt{L_g^{-1}} - 2) \right\}$.

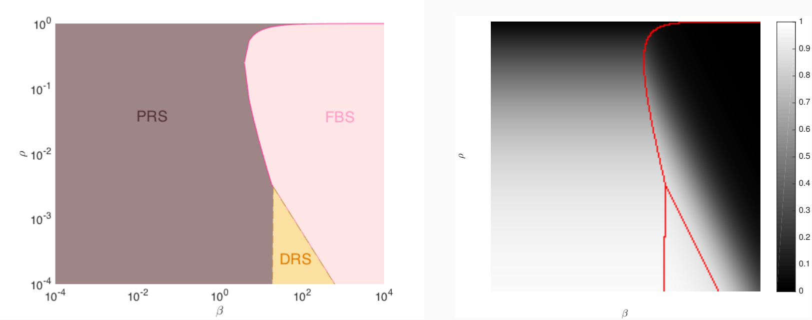
Then

$$r^*_S(L_g, \rho) < \min\{r^*_{T_2}(L_g, \rho), r^*_R(\rho)\}.$$

In any other case, we have

$$r^*_R(\rho) \leq \min\{r^*_{T_2}(L_g, \rho), r^*_S(L_g, \rho)\}.$$

Theoretical comparisons



Comparison of the convergence rates of EA, FBS, PRS, DRS for two choices of $\alpha = L_f^{-1}$, $\beta = L_g^{-1}$, and ρ.

Numerical comparisons: Smooth TV1D denoising

First formulation: $\underset{x\in\mathbb{R}^N}{\text{minimize}} \quad \underbrace{\frac{1}{2}\|x-z\|_2^2}_{f(x)} + \underbrace{\chi h(Lx)}_{g(x)}$

$\rightarrow f$ is $\rho = 1$ strongly convex, $L_f = 1$, and $L_g = \frac{\chi\|L\|^2}{\mu}$.

1- **EA:** Use $G_{\tau(\nabla g + \nabla f)}$

2- **FBS:** Use $T_{\tau\nabla f, \tau\nabla g}$

Second formulation: $\min_{x\in\mathcal{H}} \underbrace{\frac{1}{2}\|x-z\|_2^2 + \chi h_{\mathbb{I}_1}(L_{\mathbb{I}_1}x)}_{\tilde{f}(x)} + \underbrace{\chi h_{\mathbb{I}_2}(L_{\mathbb{I}_2}x)}_{\tilde{g}(x)}$

$\rightarrow \widetilde{f}$ is $\rho = 1$ strongly convex, $L_{\tilde{f}} = \frac{\mu + \chi\|L_{\mathbb{I}_2}\|^2}{\mu}$ and $L_{\tilde{g}} = \frac{\chi\|L_{\mathbb{I}_1}\|^2}{\mu}$

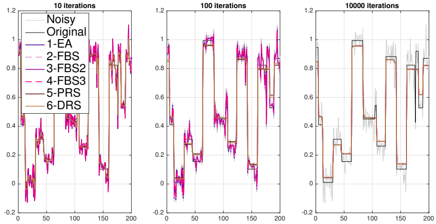
3- **FBS 2:** Use $T_{\tau\nabla\widetilde{g}, \tau\nabla\widetilde{f}}$

4- **FBS 3:** Use $T_{\tau\nabla\widetilde{f}, \tau\nabla\widetilde{g}}$

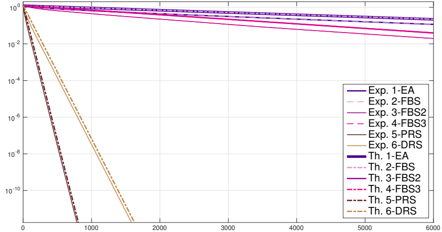
5- **PRS:** Use $R_{\tau\nabla\widetilde{f}, \tau\nabla\widetilde{g}}$

6- **DRS:** Use $S_{\tau\nabla\widetilde{f}, \tau\nabla\widetilde{g}}$

Numerical comparisons: Smooth TV1D denoising



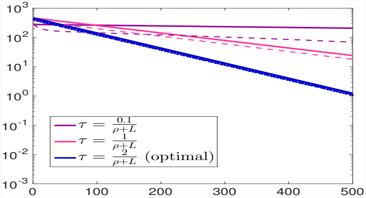
Original/degraded/reconstructed signals



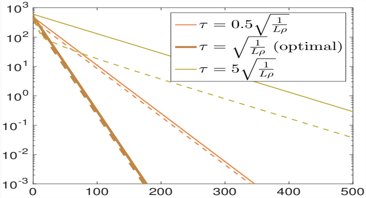
Errors vs Iterations

Numerical comparisons: $\min_{\mathrm{v,h}} \sum_j \| \log_2 \mathcal{L}_j - \mathrm{v} - j\mathrm{h}\|_2^2 + \lambda_v \|\mathrm{Dv}\|_1 + \lambda_h \|\mathrm{Dh}\|_1$

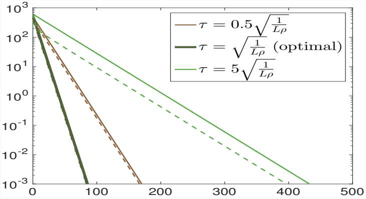
FBS



DRS



PRS



Problem solved: segmentation problem over the range $j \in \{2,3,4\}$, $\lambda_v = 0.1$, and $\lambda_h = 200$.

Display: Comparisons of the theoretical upper bound (i.e., $r_\Phi(\tau)^k \|x_0 - x_\infty\|_2$) versus the numerical error (i.e., $\|x_k - x_\infty\|_2$) w.r.t. the number of iterations.

Supervised deep image analysis: Strong convexity to design supervised deep proximal architectures

Context

Data: z

Direct/acquisition model

Design a cost function

Minimization algorithm

Hyperparameter tuning

Estimated parameters: $\widehat{\mathrm{x}}(\mathrm{z}; \widehat{\lambda})$

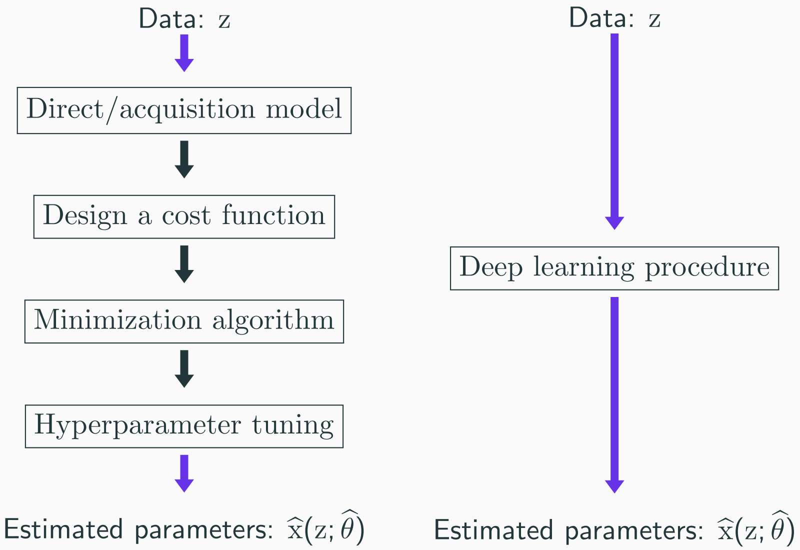
$$\mathrm{z} = \mathcal{D}(\mathrm{A}\overline{\mathrm{x}})$$

$$\widehat{\mathrm{x}}(\mathrm{z}; \lambda) \in \underset{\mathrm{x}}{\mathrm{Argmin}}\; \varphi(\mathrm{x}; \mathrm{z}) + \lambda \psi(\mathrm{x})$$

Sequence $\mathrm{x}^{[k+1]} = \boldsymbol{\Phi} \mathrm{x}^{[k]}$

$$\widehat{\lambda} \in \underset{\lambda}{\mathrm{Argmin}} \|\overline{\mathrm{x}} - \widehat{\mathrm{x}}(\mathrm{z}; \lambda)\|_2^2$$

Standard learning and deep learning



Standard learning and deep learning

➡ **Create a database** $\mathcal{S} = \left\{ (\overline{\mathrm{x}}_\ell, \mathrm{z}_\ell) \in \mathbb{R}^N \times \mathbb{R}^{\overline{N}} \mid \ell \in \{1, \ldots, L\} \right\}$

➡ **Learn a prediction function** d_Θ

$$\widehat{\Theta} \in \underset{\Theta}{\mathrm{Argmin}}\ \mathrm{E}(\Theta) := \frac{1}{L} \sum_{\ell=1}^{L} f_1\big(\mathrm{d}_\Theta(\mathrm{z}_\ell), \overline{\mathrm{x}}_\ell\big) + f_2(\Theta)$$

- **Linear model**: $\mathrm{d}_\Theta(\mathrm{z}_\ell) = \Theta^\top \mathrm{z}_\ell$

Standard learning and deep learning

➡ **Create a database** $\mathcal{S} = \{(\overline{\mathrm{x}}_\ell, \mathrm{z}_\ell) \in \mathbb{R}^N \times \mathbb{R}^{\overline{N}} \mid \ell \in \{1, \ldots, L\}\}$

➡ **Learn a prediction function** d_Θ

$$\widehat{\Theta} \in \underset{\Theta}{\mathrm{Argmin}}\ \mathrm{E}(\Theta) := \frac{1}{L}\sum_{\ell=1}^{L} f_1\big(\mathrm{d}_\Theta(\mathrm{z}_\ell), \overline{\mathrm{x}}_\ell\big) + f_2(\Theta)$$

- **Linear model**: $\mathrm{d}_\Theta(\mathrm{z}_\ell) = \Theta^\top \mathrm{z}_\ell$

- **Non-linear model**:

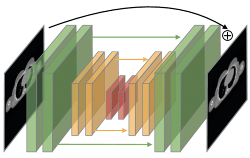
$$\boxed{\mathrm{d}_\Theta(\mathrm{z}_\ell) = \eta^{[K]}\big(W^{[K]} \ldots \eta^{[1]}(W^{[1]}\mathrm{z}_\ell + b^{[1]}) \ldots + b^{[K]}\big)}$$

where $\Theta = \{W^{[k]}, b^{[k]}\}_{1 \le k \le K}$ with

$W^{[k]}$ denotes a weight matrix,

$b^{[k]}$ is a bias vector,

$\eta^{[k]}$ is the nonlinear activation function.



Synthesis formulation and proximal gradient descent: LISTA

➡ **Synthesis formulation:** $\boxed{\min_{\mathrm{x}} \frac{1}{2}\|\mathrm{Hx} - \mathrm{z}\|_2^2 + \lambda\|\mathrm{x}\|_1}$ where $\mathrm{H} \in \mathbb{R}^{\overline{N}\times N}$

➡ **Forward-backward iterations:**

$$\mathrm{x}^{[k+1]} = \mathrm{prox}_{\tau\lambda\|\cdot\|_1}(\mathrm{x}^{[k]} - \tau \mathrm{H}^*(\mathrm{Hx}^{[k]} - \mathrm{z}))$$

➡ **Reformulation:**

$$\mathrm{x}^{[k+1]} = \mathrm{prox}_{\tau\lambda\|\cdot\|_1}((\mathrm{Id} - \tau \mathrm{H}^*\mathrm{H})\mathrm{x}^{[k]} + \tau \mathrm{H}^* \mathrm{z}))$$

➡ **Layer network:**

$$\mathrm{x}^{[k+1]} = \underbrace{\boxed{\mathrm{prox}_{\tau\lambda\|\cdot\|_1}}}_{\eta^{[k]}} \Big(\underbrace{\boxed{\mathrm{Id} - \tau \mathrm{H}^*\mathrm{H}}}_{\mathrm{W}^{[k]}} \mathrm{x}^{[k]} + \underbrace{\boxed{\tau \mathrm{H}^*\mathrm{z}}}_{\mathrm{b}^{[k]}} \Big)$$

[Gregor, LeCun, 2010]

Standard activation functions

➡ **Preliminary remarks** [Combettes, Pesquet, 2020]

- **Most of activation functions are proximity operator** :
  ReLU, Unimodal sigmoid, Softmax ...

- For $W^{[k]}$ bounded linear operators and η_k proximity operators, d_Θ model allows to derive tight Lipschitz bounds for feedforward neural networks in order to evaluate **robustness**.

Iterative scheme

➡ **Minimization problem** : $\widehat{x} = \arg\min_{x} \frac{1}{2}\|x - z\|^2 + \|Dx\|_1$

➡ **Dual reformulation**: $\widehat{w} \in \underset{w \in \mathcal{G}}{\mathrm{Argmin}}\ \frac{1}{2}\|z - D^\top w\|^2 + \iota_{\|\cdot\|_\infty \leq 1}(w)$

- Primal solution: $\widehat{x} = z - D^\top \widehat{w}$.
- Solution obtained with proximal gradient based procedure.
- Accelerated schemes (e.g., FISTA).

➡ **Primal-dual algorithms**:

- Resolution with Chambolle-Pock iterations.

For $k = 0, 1, \ldots$

$$\left\lfloor \begin{array}{l} x^{[k+1]} = \mathrm{prox}_{\frac{\tau}{2}\|\cdot - z\|_2^2}\left(x^{[k]} - \tau D^\top w^{[k]}\right) \\ w^{[k+1]} = \mathrm{prox}_{\iota_{\|\cdot\|_\infty \leq 1}}\left(w^{[k]} + \gamma D(2x^{[k+1]} - x^{[k]})\right) \end{array} \right.$$

- Acceleration when the data-term is **strongly convex**.

(F)ISTA in the dual

➡ **Minimization problem**: $\widehat{\mathrm{x}} = \arg\min_{\mathrm{x}} \frac{1}{2}\|\mathrm{x}-\mathrm{z}\|^2 + \|\mathrm{D}\mathrm{x}\|_1$

➡ **(F)ISTA to solve dual reformulation**:

Set $\mathrm{w}_1 \in \mathbb{R}^{|\mathbb{F}|}$, and $\mathrm{y}_1 \in \mathbb{R}^{|\mathbb{F}|}$. For every iteration k,

$$\left\lfloor \begin{array}{ll} \mathrm{w}_{k+1} & = \mathrm{prox}_{\iota_{\|\cdot\|_\infty \leq 1}}\Big((\mathrm{Id} - \tau_k \mathrm{D}\mathrm{D}^\top)\mathrm{y}_k + \tau_k \mathrm{D}\mathrm{z} \Big) \\ \mathrm{y}_{k+1} & = (1+\alpha_k)\mathrm{w}_{k+1} - \alpha_k \mathrm{w}_k \end{array} \right.$$

➡ **Preliminary remarks**:

- FISTA: $(\mathrm{w}_k)_{k\in\mathbb{N}}$ converges to $\widehat{\mathrm{w}}$ when $\alpha_k = \frac{t_k - 1}{t_{k+1}}$ and $t_{k+1} = \frac{k+a-1}{a}$, $a > 2$, $\tau < \frac{1}{\|\mathrm{D}\|^2}$ and $\widetilde{F}(\mathrm{w}_k) - \widetilde{F}(\widehat{\mathrm{w}}) \leq \frac{\zeta}{k^2}$.
- ISTA: When $\alpha_k \equiv 0$, $(\mathrm{w}_k)_{k\in\mathbb{N}}$ converges to $\widehat{\mathrm{w}}$ when $\tau < \frac{2}{\|\mathrm{D}\|^2}$ for this limit case, and $\widetilde{F}(\mathrm{w}_k) - \widetilde{F}(\widehat{\mathrm{w}}) \leq \frac{\zeta}{k}$.
- (F)ISTA: $\widehat{\mathrm{x}} = \mathrm{z} - \mathrm{D}^\top \widehat{\mathrm{w}}$

(F)ISTA in the dual

➡ **Minimization problem**: $\widehat{x} = \arg\min_{x} \frac{1}{2}\|x - z\|^2 + \|Dx\|_1$

➡ **(F)ISTA to solve dual reformulation**:
Set $w_1 \in \mathbb{R}^{|\mathbb{F}|}$, and $y_1 \in \mathbb{R}^{|\mathbb{F}|}$. For every iteration k,

$$\left\lfloor \begin{array}{ll} w_{k+1} & = \mathrm{prox}_{\iota_{\|\cdot\|_\infty \leq 1}}\Big((\mathrm{Id} - \tau_k DD^\top) y_k + \tau_k Dz\Big) \\ y_{k+1} & = (1 + \alpha_k) w_{k+1} - \alpha_k w_k \end{array} \right.$$

Proposition : The proximity operator of the conjugate of the ℓ_1-norm scaled by parameter $\lambda > 0$ fits the HardTanh activation function,:

$$(\forall x = (x_i)_{1 \leq i \leq N}) \qquad P_{\|\cdot\|_\infty \leq \lambda}(x) = \mathrm{HardTanh}_\lambda(x) = (p_i)_{1 \leq i \leq N}$$

where

$$p_i = \begin{cases} -\lambda & \text{if} \quad p_i < -\lambda, \\ \lambda & \text{if} \quad p_i > \lambda, \\ p_i & \text{otherwise.} \end{cases}$$

(F)ISTA in the dual

➡ **Minimization problem**: $\widehat{\mathrm{x}} = \arg\min_{\mathrm{x}} \frac{1}{2}\|\mathrm{x}-\mathrm{z}\|^2 + \|\mathrm{Dx}\|_1$

➡ **(F)ISTA to solve dual reformulation**:
Set $\mathrm{w}_1 \in \mathbb{R}^{|\mathbb{F}|}$, and $\mathrm{y}_1 \in \mathbb{R}^{|\mathbb{F}|}$. For every iteration k,

$$\left\lfloor \begin{aligned} \mathrm{w}_{k+1} &= \mathrm{HardTanh}_1\Big((\mathrm{Id} - \tau_k \mathrm{DD}^\top)\mathrm{y}_k + \tau_k \mathrm{Dz}\Big) \\ \mathrm{y}_{k+1} &= (1+\alpha_k)\mathrm{w}_{k+1} - \alpha_k \mathrm{w}_k \end{aligned} \right.$$

Proposition : The proximity operator of the conjugate of the ℓ_1-norm scaled by parameter $\lambda > 0$ fits the HardTanh activation function,:

$$(\forall \mathrm{x} = (\mathrm{x}_i)_{1\le i\le N}) \qquad \mathrm{P}_{\|\cdot\|_\infty \le \lambda}(\mathrm{x}) = \mathrm{HardTanh}_\lambda(\mathrm{x}) = (\mathrm{p}_i)_{1\le i\le N}$$

where

$$\mathrm{p}_i = \begin{cases} -\lambda & \text{if} \quad \mathrm{p}_i < -\lambda, \\ \lambda & \text{if} \quad \mathrm{p}_i > \lambda, \\ \mathrm{p}_i & \text{otherwise.} \end{cases}$$

(F)ISTA in the dual

➡ **Minimization problem**: $\widehat{x} = \arg\min_{x} \frac{1}{2}\|x - z\|^2 + \|Dx\|_1$

➡ **(F)ISTA to solve dual reformulation**:
Set $w_1 \in \mathbb{R}^{|\mathbb{F}|}$, and $y_1 \in \mathbb{R}^{|\mathbb{F}|}$. For every iteration k,

$$\begin{cases} w_{k+1} &= \mathrm{HardTanh}_1\Big((\mathrm{Id} - \tau_k D D^\top) y_k + \tau_k D z\Big) \\ y_{k+1} &= (1+\alpha_k) w_{k+1} - \alpha_k w_k \end{cases}$$

➡ **Unfolded (F)ISTA**:

$$\begin{bmatrix} w^{[k]} \\ w^{[k+1]} \end{bmatrix} = \underbrace{\begin{Bmatrix} \mathrm{Id}_{|\mathbb{F}|} \\ \mathrm{HardTanh}_1 \end{Bmatrix}}_{\eta^{[k]}} \left(\underbrace{\begin{bmatrix} 0 & \mathrm{Id}_{|\mathbb{F}|} \\ -\alpha_{k-1}(\mathrm{Id}_{|\mathbb{F}|} - D_1^{[k]} D_2^{[k]}) & (1+\alpha_{k-1})(\mathrm{Id}_{|\mathbb{F}|} - D_1^{[k]} D_2^{[k]}) \end{bmatrix}}_{W^{[k]}} \begin{bmatrix} w^{[k-1]} \\ w^{[k]} \end{bmatrix} + \underbrace{\begin{bmatrix} 0 \\ D_1^{[k]} \mathbf{z}_l \end{bmatrix}}_{b^{[k]}} \right)$$

Network Deep-(F)ISTA-GD

➡ **Network**: For every layer $k \in \{2, \dots, K-1\}$:

$$\begin{cases} W^{[1]} = \begin{bmatrix} \mathrm{D}_1^{[1]} \\ (\mathrm{Id}_{|\mathbb{F}|} - \mathrm{D}_1^{[1]}\mathrm{D}_2^{[1]})\mathrm{D}_1^{[1]} \end{bmatrix}, \\ b^{[1]} = \begin{bmatrix} 0 \\ \mathrm{D}_1^{[1]}\mathbf{z}_I \end{bmatrix}, \eta^{[1]} = \begin{Bmatrix} \mathrm{Id}_{|\mathbb{F}|} \\ \mathrm{HardTanh}_\lambda \end{Bmatrix}, \\ W^{[k]} = \begin{bmatrix} 0 & \mathrm{Id}_{|\mathbb{F}|} \\ -\alpha_{k-1}(\mathrm{Id}_{|\mathbb{F}|} - \mathrm{D}_1^{[k]}\mathrm{D}_2^{[k]}) & (1+\alpha_{k-1})(\mathrm{Id}_{|\mathbb{F}|} - \mathrm{D}_1^{[k]}\mathrm{D}_2^{[k]}) \end{bmatrix}, \\ b^{[k]} = \begin{bmatrix} 0 \\ \mathrm{D}_1^{[k]}\mathbf{z}_I \end{bmatrix}, \eta^{[k]} = \begin{Bmatrix} \mathrm{Id}_{|\mathbb{F}|} \\ \mathrm{HardTanh}_\lambda \end{Bmatrix}, \\ W^{[K]} = \begin{bmatrix} 0 & -\mathrm{D}_2^{[k]} \end{bmatrix}, b^{[K]} = \mathbf{z}_I, \eta^{[K]} = \mathrm{Id}_N. \end{cases}$$

➡ **Proposition**: If $\mathrm{D}_1^{[k]} = \tau_k \mathrm{D}$ and $\mathrm{D}_2^{[k]} = \mathrm{D}^\top$, then `Deep-(F)ISTA-GD` network fits the generic `(F)ISTA` scheme.

(Sc)CP

➡ **Minimization problem**: $\widehat{\mathrm{x}} = \arg\min_{\mathrm{x}} \frac{1}{2}\|\mathrm{x}-\mathrm{z}\|^2 + \|\mathrm{Dx}\|_1$

➡ **(Sc)CP to solve the minimization problem**:

Set $\mathrm{w}_1 \in \mathbb{R}^{|\mathbb{F}|}$, and $\mathrm{x}_1 = \mathrm{x}_0 \in \mathbb{R}^{|\mathbb{F}|}$. For every iteration k,

$$\left\lfloor \begin{array}{ll} \mathrm{w}_{k+1} & = \mathrm{prox}_{\iota_{\|\cdot\|_\infty \leq 1}}\Big(\mathrm{w}_k + \tau_k \mathrm{D}\Big((1+\alpha_k)\,\mathrm{x}_k - \alpha_k \mathrm{x}_{k-1}\Big)\Big) \\ \mathrm{x}_{k+1} & = \mathrm{prox}_{\frac{\sigma_k}{2}\|\cdot - \mathrm{z}\|_2^2}\Big(\mathrm{x}_k - \sigma_k \mathrm{D}^\top \mathrm{w}_{k+1}\Big) \end{array} \right.$$

➡ **Remarks** :

- ScCP: $\alpha_k = \frac{1}{\sqrt{1+2\gamma\sigma_k}}$, $\sigma_{k+1} = \alpha_k \sigma_k$, $\tau_{k+1} = \frac{\tau_k}{\alpha_k}$.
- CP: $\gamma = 0$, $\sigma_k \equiv \sigma$, $\tau_k \equiv \tau$ and assuming $\sigma\tau\|D\|^2 < 1$.
- $(\mathrm{x}_k)_{k\in\mathbb{N}}$ converges to $\widehat{\mathrm{x}}$.
- Convergence rate $O(1/k)$ for CP and $O(1/k^2)$ for ScCP.

(Sc)CP

➡ **Minimization problem**: $\widehat{\mathrm{x}} = \arg\min_{\mathrm{x}} \frac{1}{2}\|\mathrm{x}-\mathrm{z}\|^2 + \|\mathrm{D}\mathrm{x}\|_1$

➡ **(Sc)CP to solve the minimization problem**:

Set $\mathrm{w}_1 \in \mathbb{R}^{|\mathbb{F}|}$, and $\mathrm{x}_1 = \mathrm{x}_0 \in \mathbb{R}^{|\mathbb{F}|}$. For every iteration k,

$$\left\lfloor \begin{array}{ll} \mathrm{w}_{k+1} & = \mathrm{HardTanh}_1\Big(\mathrm{w}_k + \tau_k \mathrm{D}\Big((1+\alpha_k)\,\mathrm{x}_k - \alpha_k \mathrm{x}_{k-1}\Big)\Big) \\ \mathrm{x}_{k+1} & = \frac{\sigma_k}{1+\sigma_k}\mathrm{z} + \frac{1}{1+\sigma_k}\mathrm{x}_k - \frac{\sigma_k}{1+\sigma_k}\mathrm{D}^\top \mathrm{w}_{k+1} \end{array} \right.$$

➡ **Remarks** :

- ScCP: $\alpha_k = \frac{1}{\sqrt{1+2\gamma\sigma_k}}$, $\sigma_{k+1} = \alpha_k \sigma_k$, $\tau_{k+1} = \frac{\tau_k}{\alpha_k}$.
- CP: $\gamma = 0$, $\sigma_k \equiv \sigma$, $\tau_k \equiv \tau$ and assuming $\sigma\tau\|D\|^2 < 1$.
- $(\mathrm{x}_k)_{k\in\mathbb{N}}$ converges to $\widehat{\mathrm{x}}$.
- Convergence rate $O(1/k)$ for CP and $O(1/k^2)$ for ScCP.

(Sc)CP

➡ **Minimization problem**: $\widehat{\mathrm{x}} = \arg\min_{\mathrm{x}} \frac{1}{2}\|\mathrm{x}-\mathrm{z}\|^2 + \|\mathrm{D}\mathrm{x}\|_1$

➡ **(Sc)CP to solve the minimization problem**:

Set $\mathrm{w}_1 \in \mathbb{R}^{|\mathbb{F}|}$, and $\mathrm{x}_1 = \mathrm{x}_0 \in \mathbb{R}^{|\mathbb{F}|}$. For every iteration k,

$$\left\lfloor \begin{array}{ll} \mathrm{w}_{k+1} & = \mathrm{HardTanh}_1\Big(\mathrm{w}_k + \tau_k \mathrm{D}\big((1+\alpha_k)\mathrm{x}_k - \alpha_k \mathrm{x}_{k-1}\big)\Big) \\ \mathrm{x}_{k+1} & = \frac{\sigma_k}{1+\sigma_k}\mathrm{z} + \frac{1}{1+\sigma_k}\mathrm{x}_k - \frac{\sigma_k}{1+\sigma_k}\mathrm{D}^\top \mathrm{w}_{k+1} \end{array}\right.$$

➡ **Unfolded (Sc)CP**:

$$\begin{bmatrix} \mathrm{x}_k \\ \mathrm{w}_{k+1} \end{bmatrix} = \underbrace{\begin{Bmatrix} \mathrm{Id}_N \\ \mathrm{HardTanh}_1 \end{Bmatrix}}_{\eta^{[k]}} \left(\underbrace{\begin{bmatrix} \frac{1}{1+\sigma_{k-1}} & -\frac{\sigma_{k-1}}{1+\sigma_{k-1}}\mathrm{D}_2^{[k-1]} \\ \frac{1+\alpha_k}{1+\sigma_{k-1}}\mathrm{D}_1^{[k]} - \alpha_k \mathrm{D}_1^{[k]} & \mathrm{Id}_{|\mathbb{F}|} - \frac{(1+\alpha_k)\sigma_{k-1}}{1+\sigma_{k-1}}\mathrm{D}_1^{[k]}\mathrm{D}_2^{[k-1]} \end{bmatrix}}_{W^{[k]}} \begin{bmatrix} \mathrm{x}_{k-1} \\ \mathrm{w}_k \end{bmatrix} + \underbrace{\begin{bmatrix} \frac{\sigma_{k-1}}{1+\sigma_{k-1}}\mathbf{z} \\ \frac{(1+\alpha_k)\sigma_{k-1}}{1+\sigma_{k-1}}\mathrm{D}_1^{[k]}\mathbf{z} \end{bmatrix}}_{b^{[k]}} \right)$$

Network Deep-(Sc)CP-GD

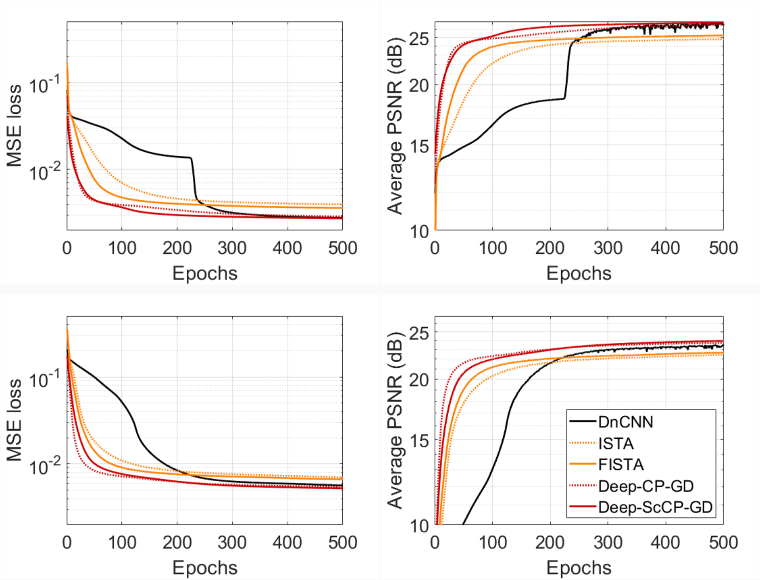
➡ **Network**: For every layer $k \in \{2, \ldots, K-1\}$:

$$\begin{cases} W^{[1]} = \begin{bmatrix} \mathrm{Id}_N \\ 2D_1^{[1]} \end{bmatrix}, b^{[1]} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \eta^{[1]} = \begin{Bmatrix} \mathrm{Id}_N \\ \mathrm{HardTanh}_\lambda \end{Bmatrix}, \\ W^{[k]} = \begin{bmatrix} \frac{1}{1+\sigma_{k-1}} & -\frac{\sigma_{k-1}}{1+\sigma_{k-1}}{}_2^{[k-1]} \\ \frac{1+\alpha_k}{1+\sigma_{k-1}} D_1^{[k]} - \alpha_k {}_1^{[k]} & \mathrm{Id}_{|\mathbb{F}|} - \frac{(1+\alpha_k)\sigma_{k-1}}{1+\sigma_{k-1}}{}_1^{[k]}{}_2^{[k-1]} \end{bmatrix}, \\ b^{[k]} = \begin{bmatrix} \frac{\sigma_{k-1}}{1+\sigma_{k-1}} \mathbf{z} \\ \frac{(1+\alpha_k)\sigma_{k-1}}{1+\sigma_{k-1}}{}_1^{[k]} \mathbf{z} \end{bmatrix}, \eta^{[k]} = \begin{Bmatrix} \mathrm{Id}_N \\ \mathrm{HardTanh}_\lambda \end{Bmatrix}, \\ W^{[K]} = \begin{bmatrix} \mathrm{Id}_N & 0 \end{bmatrix}, b^{[K]} = 0, \eta^{[K]} = \mathrm{Id}_N. \end{cases}$$

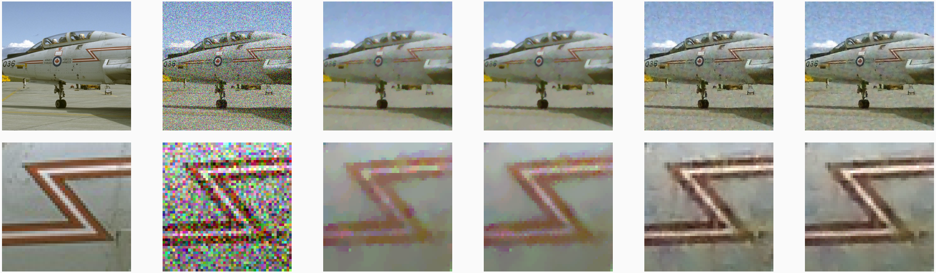
➡ **Proposition**: If $\mathrm{D}_1^{[k]} = \tau_k \mathrm{D}$ and $\mathrm{D}_2^{[k]} = \mathrm{D}^\top$, then the `Deep-(Sc)CP-GD` network fits the generic `(Sc)CP` scheme.

Performance Gaussian image denoising

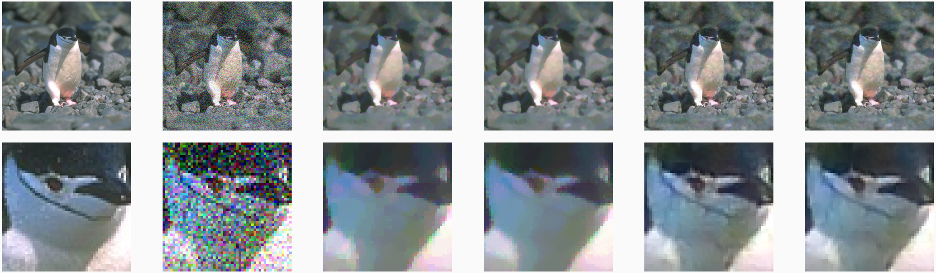
Training loss | PSNR on test dataset



Original	Noisy	TV	NL-TV	DnCNN	Proposed

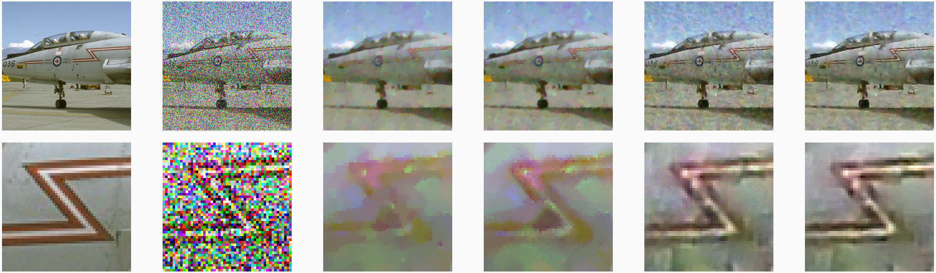


PSNR/SSIM	14.1/0.25	26.0/0.84	26.6/0.85	27.9/0.86	**28.2/0.87**

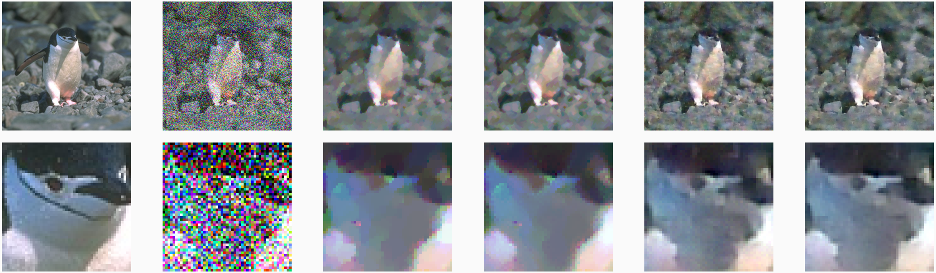


PSNR/SSIM	14.1/0.13	26.0/0.76	27.7/0.79	28.5/0.79	**28.8/0.81**

Original	Noisy	TV	NL-TV	DnCNN	Proposed



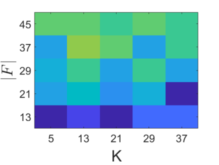
PSNR/SSIM	8.13/0.09	23.6/0.76	24.0/0.76	24.4/0.76	**25.2/0.80**



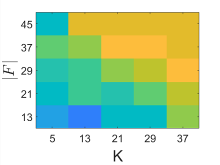
PSNR/SSIM	8.14/0.043	24.5/0.64	25.1/0.65	25.4/0.65	**25.9/0.70**

- SNR

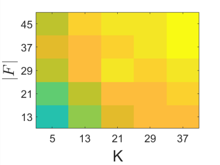
Deep-ISTA-GD



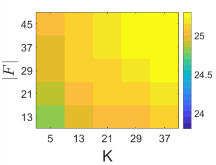
Deep-FISTA-GD



Deep-CP-GD

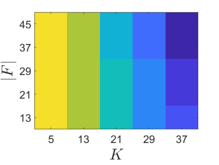


Deep-ScCP-GD

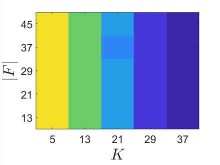


- Robustness: $\|f_\Theta(\mathbf{z}+\epsilon) - f_\Theta(\mathbf{z})\| \leq \chi\|\epsilon\|$.

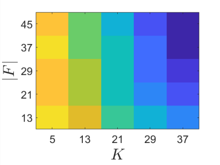
Deep-ISTA-GD



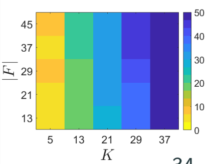
Deep-FISTA-GD



Deep-CP-GD

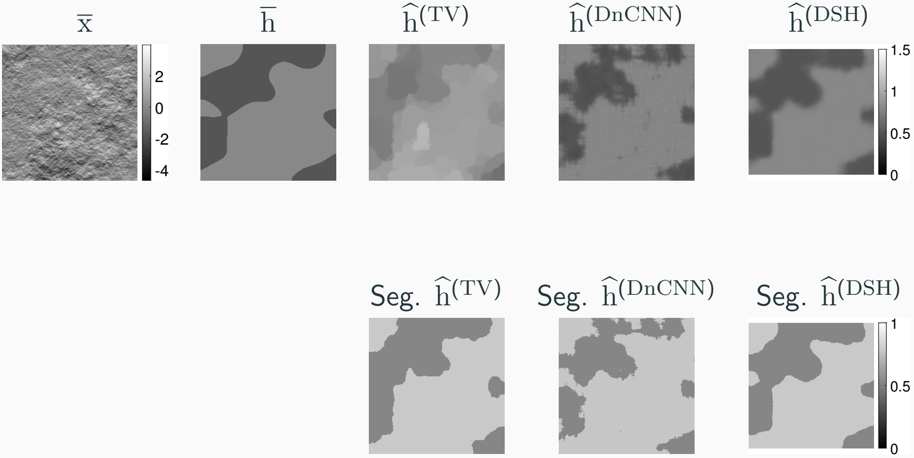


Deep-ScCP-GD

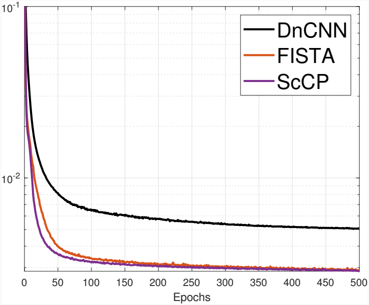


Performance texture segmentation:

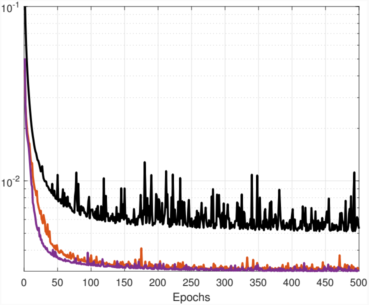
➡ **Minimization problem**: $\min_{\mathrm{x}} \frac{1}{2}\|\mathrm{x} - \sum_j w_{j,n} \log_2 \mathcal{L}_{j,n}\|^2 + \|\mathrm{Dx}\|_1$



Performance texture segmentation



(i)



(ii)

I- **Model**: Strong convexity in unsupervised image processing.

II- **Algorithmic**: Efficiency regions of strong convexity and Lipschitz parameters to identifying the most efficient first-order algorithm for image analysis.

III- **Deep learning**: On strong convexity in the design of unfolded deep learning architectures.

I- **Model**: Strong convexity in unsupervised image processing.
→ Denoising, texture segmentation.
→ **Require particular care for modeling.**

II- **Algorithmic**: Efficiency regions of strong convexity and Lipschitz parameters to identifying the most efficient first-order algorithm for image analysis.

III- **Deep learning**: On strong convexity in the design of unfolded deep learning architectures.

I- **Model**: Strong convexity in unsupervised image processing.

→ Denoising, texture segmentation.

→ **Require particular care for modeling.**

II- **Algorithmic**: Efficiency regions of strong convexity and Lipschitz parameters to identifying the most efficient first-order algorithm for image analysis.

→ Efficiency diagram for first order schemes.

→ **Extension to other schemes.**

III- **Deep learning**: On strong convexity in the design of unfolded deep learning architectures.

Outline → Conclusions and perspectives

I- **Model**: Strong convexity in unsupervised image processing.
→ Denoising, texture segmentation.
→ **Require particular care for modeling.**

II- **Algorithmic**: Efficiency regions of strong convexity and Lipschitz parameters to identifying the most efficient first-order algorithm for image analysis.
→ Efficiency diagram for first order schemes.
→ **Extension to other schemes.**

III- **Deep learning**: On strong convexity in the design of unfolded deep learning architectures.
→ Deep-(F)ISTA and Deep-(Sc)CP in between standard image analysis and deep learning.
→ **Compute tight Lipschitz bounds.**

References

Model:
- J. Colas, N. Pustelnik, C. Oliver, J.-C. Geminard, V. Vidal, Nonlinear denoising for solid friction dynamics characterization, Physical Review E, 100, 032803, Sept. 2019.
- B. Pascal, N. Pustelnik, and P. Abry, Strongly Convex Optimization for Joint Fractal Feature Estimation and Texture Segmentation, ACHA, vol. 54, pp 303-322, 2021.

Convergence rate:
- L. M. Briceño-Arias and N. Pustelnik, Proximal or gradient steps for cocoercive operators, accepted to Signal Processing, 2022.

Deep unfolded schemes:
- M. Jiu and N. Pustelnik, A deep primal-dual proximal network for image restoration, IEEE JSTSP Special Issue on Deep Learning for Image/Video Restoration and Compression, vol. 15, no. 2,pp. 190–203, Feb. 2021.
- H.T.V. Le, N. Pustelnik, M. Foare, The faster proximal algorithmm, the better unfolded deep learning architecture ? The study case of image denoising, EUSIPCO, Belgrade, Serbia, 2022.

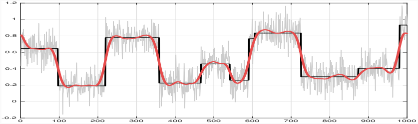
Strong convexity in signal/image analysis

Piecewise constant denoising: $z = \bar{x} + n$ with $n = \mathcal{N}(0, \sigma^2 \mathrm{Id})$

➡ **Minimization problem**:

$$\widehat{x}(z; \widehat{\lambda}) = \arg \min_{x \in \mathbb{R}^N} \frac{1}{2} \|x - z\|_2^2 + \lambda \|Dx\|_\bullet \quad \text{where} \quad \begin{cases} Dx = d * x \\ \lambda > 0 \end{cases}$$

➡ **Linear denoising**



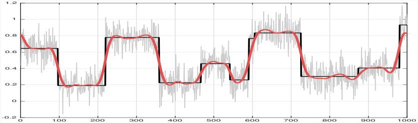
$$d = \begin{bmatrix} 1 & -1 \end{bmatrix}; \quad \|\cdot\|_\bullet = \|\cdot\|_2^2$$

Piecewise constant denoising: $z = \bar{x} + n$ with $n = \mathcal{N}(0, \sigma^2 \mathrm{Id})$

➡ **Minimization problem**:

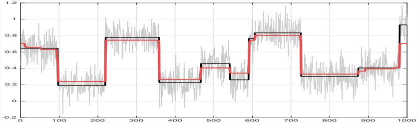
$$\widehat{x}(z; \widehat{\lambda}) = \arg\min_{x \in \mathbb{R}^N} \frac{1}{2}\|x - z\|_2^2 + \lambda \|Dx\|_\bullet \quad \text{where} \quad \begin{cases} Dx = d * x \\ \lambda > 0 \end{cases}$$

➡ **Linear denoising**

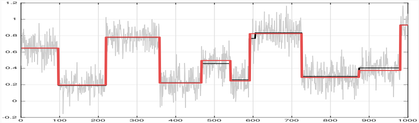


$$d = \begin{bmatrix} 1 & -1 \end{bmatrix}; \quad \|\cdot\|_\bullet = \|\cdot\|_2^2$$

➡ **Non-linear denoising**.



$$d = \begin{bmatrix} 1 & -1 \end{bmatrix} \text{ and } \|\cdot\|_\bullet = \|\cdot\|_1$$



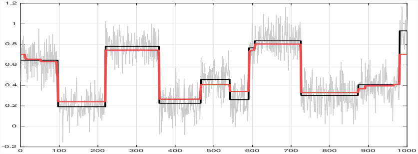
$$d = \begin{bmatrix} 1 & -1 \end{bmatrix} \text{ and } \|\cdot\|_\bullet = \|\cdot\|_0$$

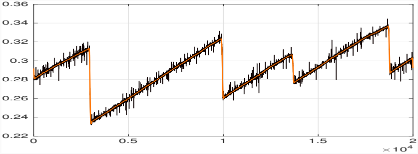
Piecewise linear denoising: $z = \bar{x} + n$ with $n = \mathcal{N}(0, \sigma^2 \mathrm{Id})$

➡ Minimization problem

$$\widehat{x}(z; \widehat{\lambda}) = \arg\min_{x \in \mathbb{R}^N} \frac{1}{2}\|x - z\|_2^2 + \lambda \| D x \|_\bullet \quad \text{where} \quad \begin{cases} Dx = d * x \\ \lambda > 0 \end{cases}$$

➡ Non-linear denoising: piecewise constant/linear

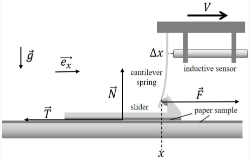


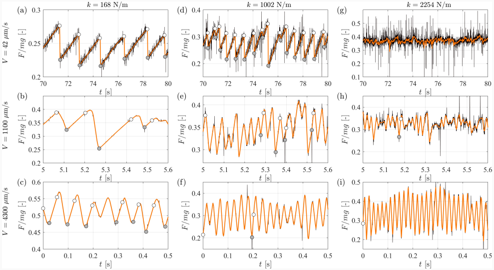


$d = \begin{bmatrix} 1 & -1 \end{bmatrix}$ and $\|\cdot\|_\bullet = \|\cdot\|_1$

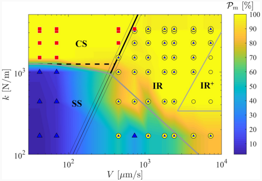
$d = \begin{bmatrix} 1 & -2 & 1 \end{bmatrix}$ and $\|\cdot\|_\bullet = \|\cdot\|_1$

➡ **Solid friction**, LPENSL experiment [Colas, et al., 2019]



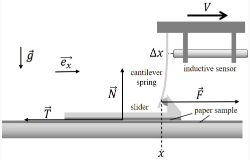


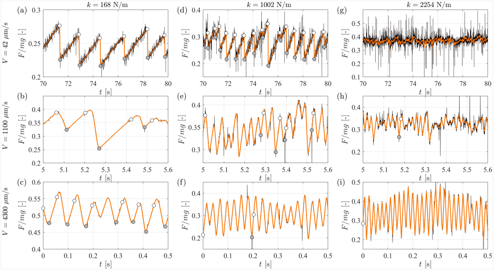
➡ **Phase diagram**:



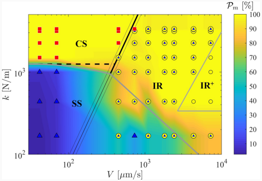
Piecewise linear denoising: Stick-Slip

➡ **Solid friction**, LPENSL experiment [Colas, et al., 2019]





➡ **Phase diagram**:



Limitations:

→ Large dimensionality of each signal.

→ Evaluation for different λ values.

→ Large amount of signal to analyze.

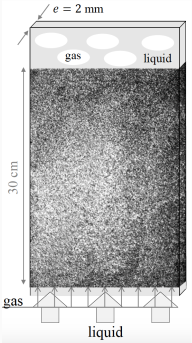
Advantages:

→ Strongly convex formulation.

Texture segmentation: Multiphase flow

➡ **Gas/liquid flow in porous medium**: LPENSL experiment

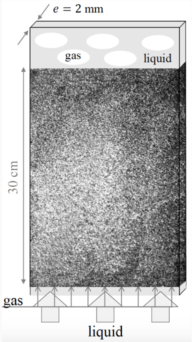
- Segment gas/liquid + accurate estimation of the interface.
- Large-scale data.



Texture segmentation: Multiphase flow

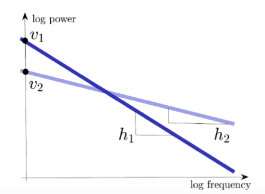
➡ **Gas/liquid flow in porous medium**: LPENSL experiment

- Segment gas/liquid + accurate estimation of the interface.
- Large-scale data.



➡ **Texture segmentation**:

- Scale-free descriptors.
- Require to compute the slope at each location.

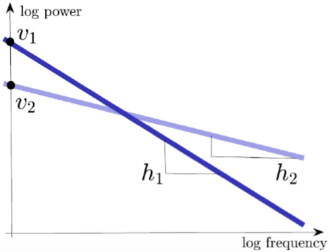


Texture segmentation: Multiphase flow

➡ **PLOVER:** [Pascal, Pustelnik, Abry, ACHA, 2021]

$$(\widehat{\mathrm{v}}, \widehat{\mathrm{h}}) \in \underset{\mathrm{v},\mathrm{h}}{\mathrm{Argmin}} \sum_j \| \log_2 \mathcal{L}_j - \mathrm{v} - j\mathrm{h}\|_2^2 + \lambda \left\| \left[\mathrm{Dv}; \alpha \mathrm{Dh}\right]^\top \right\|_{2,1}$$

- Behavior through the scales $\mathcal{L}_{j,n} \simeq s_n 2^{jh_n}$ when $2^j \to 0$
- Wavelet coefficients/leaders $\zeta_j = \mathrm{D}_j \mathrm{z}$ and $\mathcal{L}_{j,n} = \sup_{\lambda_{j',n'} \subset \Lambda_{j,n}} |\zeta_{j',n'}|$

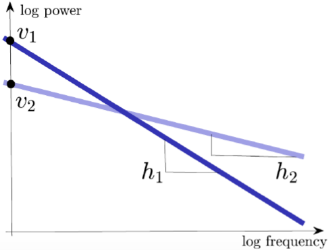


Texture segmentation: Multiphase flow

➡ **PLOVER:** [Pascal, Pustelnik, Abry, ACHA, 2021]

$$(\widehat{\mathrm{v}}, \widehat{\mathrm{h}}) \in \operatorname*{Argmin}_{\mathrm{v},\mathrm{h}} \sum_j \| \log_2 \mathcal{L}_j - \mathrm{v} - j\mathrm{h} \|_2^2 + \lambda \left\| \left[\mathrm{Dv}; \alpha \mathrm{Dh} \right]^\top \right\|_{2,1}$$

- Behavior through the scales $\mathcal{L}_{j,n} \simeq s_n 2^{jh_n}$ when $2^j \to 0$
- Wavelet coefficients/leaders $\zeta_j = \mathrm{D}_j \mathrm{z}$ and $\mathcal{L}_{j,n} = \sup_{\lambda_{j',n'} \subset \Lambda_{j,n}} |\zeta_{j',n'}|$



Limitations:

→ Large dimensionality of each image.

→ Evaluation for different λ values.

→ Large amount of signal to analyze.

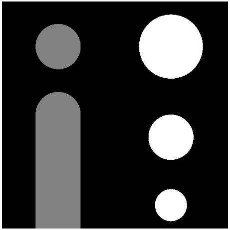
Advantages:

→ Combined estimation and segmentation.

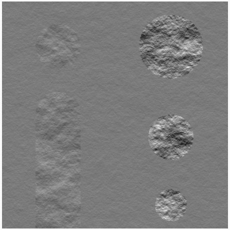
→ Joint estimation local variance/regularity.

→ Strongly convex, closed form proximity operator for data-fidelity term, dual formulation possible.

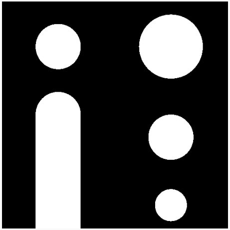
Texture segmentation: Multiphase flow



Mask



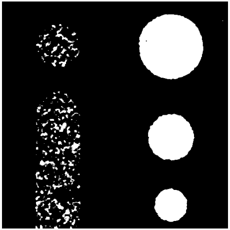
Synthetic texture



Optimal solution

T-ROF

[Cai2013]



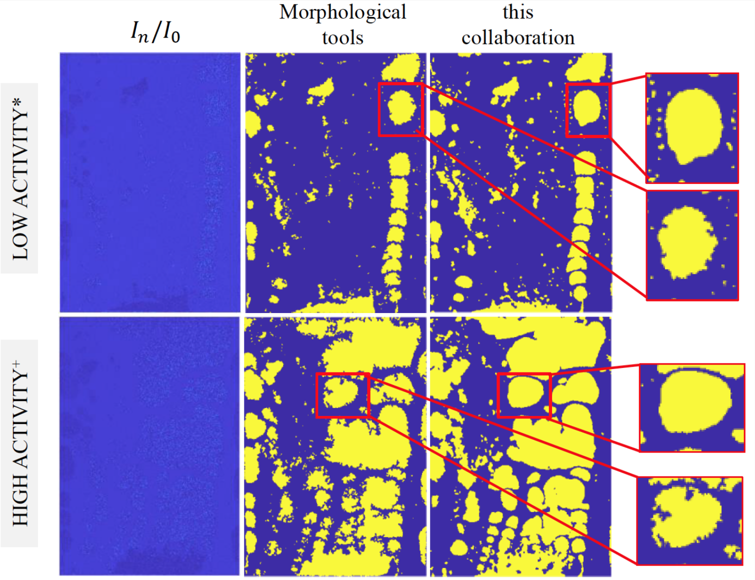
Matrix factorization

[Yuan2015]

Proposed

[Pascal2019]

Results on multiphase flow data



$^*(Q_G, Q_L) = (300{,}300)$ mL/min $\quad ^+(Q_G, Q_L) = (1200{,}300)$ mL/min