Proximal step versus gradient descent step in signal and image processing

Séminaire Français d'Optimisation – May 20th, 2021

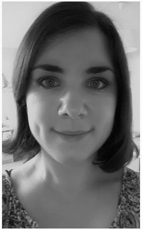
Nelly Pustelnik

CNRS, Laboratoire de Physique de l'ENS de Lyon, France

Collaborations

Patrice Abry
CNRS, Ens Lyon

Luis Briceño Arias
Univ. Técnica Federico Santa María
Chile



Barbara Pascal
Univ. Lille

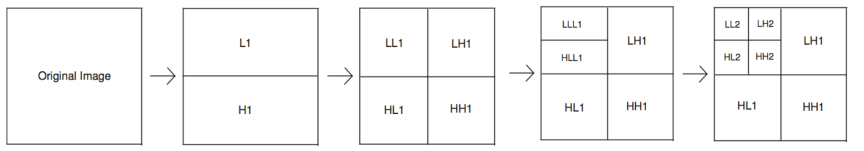
Outline

1. Proximal algorithms in signal and image processing
2. Prox versus grad for texture segmentation (numerical results)
3. Prox versus grad for piecewise constant denoising (numerical and theoretical comparisons)

Proximal algorithms in signal and image processing

From wavelet transform and sparsity to proximity operator

- Wavelets: sparse representation of most natural signals.
- Filterbank implementation of a dyadic wavelet transform: $F \in \mathbb{R}^{|\Omega| \times |\Omega|}$



$\mathbf{g} \in \mathbb{R}^{|\Omega|}$

$\zeta = F\mathbf{g}$

From wavelet transform and sparsity to proximity operator

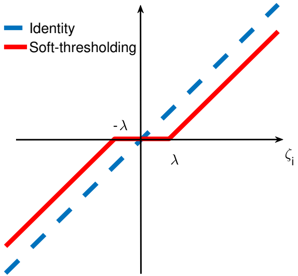
$\mathbf{g}$

$\zeta = F\mathbf{g}$

$\mathrm{soft}_\lambda(F\mathbf{g})$

$\widehat{\mathbf{u}} = F^*\mathrm{soft}_\lambda(F\mathbf{g})$

$$
\begin{aligned}
\mathrm{soft}_\lambda(\boldsymbol{\zeta}) &= \left(\max\{|\zeta_{\underline{i}}| - \lambda, 0\}\mathrm{sign}(\zeta_{\underline{i}})\right)_{\underline{i}\in\Omega} \\
&= \mathrm{prox}_{\lambda\|\cdot\|_1}(\boldsymbol{\zeta}) \\
&= \arg\min_{\boldsymbol{\nu}} \frac{1}{2}\|\boldsymbol{\nu} - \boldsymbol{\zeta}\|_2^2 + \lambda\|\boldsymbol{\nu}\|_1
\end{aligned}
$$



Proximity operator

Definition [Moreau,1965] Let $\varphi \in \Gamma_0(\mathcal{H})$ where $\mathcal{H}$ denotes a real Hilbert space. The proximity operator of φ at point $x \in \mathcal{H}$ is the unique point denoted by $\mathrm{prox}_\varphi x$ such that

$$(\forall x \in \mathcal{H}) \qquad \mathrm{prox}_\varphi x = \arg\min_{y \in \mathcal{H}} \varphi(y) + \frac{1}{2} \|x - y\|^2$$

Proximity operator

Definition [Moreau,1965] Let $\varphi \in \Gamma_0(\mathcal{H})$ where $\mathcal{H}$ denotes a real Hilbert space. The proximity operator of φ at point $x \in \mathcal{H}$ is the unique point denoted by $\mathrm{prox}_\varphi x$ such that

$$(\forall x \in \mathcal{H}) \qquad \mathrm{prox}_\varphi x = \arg\min_{y \in \mathcal{H}} \varphi(y) + \frac{1}{2}\|x - y\|^2$$

Examples: closed form expression

- $\mathrm{prox}_{\lambda\|\cdot\|_1}$: soft-thresholding with a fixed threshold $\lambda > 0$.
- $\mathrm{prox}_{\|\cdot\|_{2,1}}$ [Peyré,Fadili,2011].
- $\mathrm{prox}_{\|\cdot\|_p^p}$ with $p = \{\frac{4}{3}, \frac{3}{2}, 2, 3, 4\}$ [Chaux et al.,2005].
- $\mathrm{prox}_{D_{KL}}$ [Combettes,Pesquet,2007].
- $\mathrm{prox}_{\sum_{g \in \mathcal{G}} \|\cdot\|_q}$ with overlapping groups [Jenatton et al., 2011]
- Composition with a linear operator: $\mathrm{prox}_{\varphi \circ L}$ closed form if $LL^* = \nu\mathrm{Id}$ [Pustelnik et al., 2012]
- and many others: Prox Repository [Chierchia et al., 2016]

From wavelet transform and sparsity to proximity operator

$\mathbf{g}$

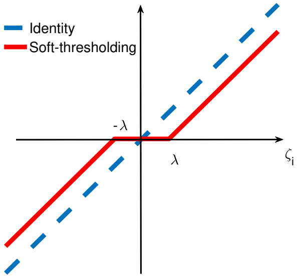
$\zeta = F\mathbf{g}$

$\mathrm{soft}_\lambda(F\mathbf{g})$

$\widehat{\mathbf{u}} = F^*\mathrm{soft}_\lambda(F\mathbf{g})$

$$\begin{aligned}\mathrm{soft}_\lambda(\zeta) &= \mathrm{prox}_{\lambda\|\cdot\|_1}(\zeta)\\ &= \arg\min_{\boldsymbol{\nu}} \frac{1}{2}\|\boldsymbol{\nu}-\zeta\|_2^2 + \lambda\|\boldsymbol{\nu}\|_1\end{aligned}$$

$$\begin{aligned}\widehat{\mathbf{u}} &= F^*\mathrm{prox}_{\lambda\|\cdot\|_1}(F\mathbf{g})\\ &= \mathrm{prox}_{\lambda\|F\cdot\|_1}(\mathbf{g})\\ &= \arg\min_{\mathbf{u}} \frac{1}{2}\|\mathbf{u}-\mathbf{g}\|_2^2 + \lambda\|F\mathbf{u}\|_1\end{aligned}$$



Example: Inverse problems

- Data: We observe data $\mathbf{g} \in \mathbb{R}^K$ being a degraded version of an original image $\bar{\mathbf{u}} \in \mathbb{R}^{|\Omega|}$ such that: $\mathbf{g} = A\bar{\mathbf{u}} + \varepsilon$
  - $A : \mathbb{R}^{K \times |\Omega|}$: denotes a linear degradation (e.g. a blur, decimation op.)
  - ε : denotes a noise (e.g. Gaussian)

- Goal: Restore the degraded image i.e., find $\widehat{\mathbf{u}}$ close to $\bar{\mathbf{u}}$:

$$\widehat{\mathbf{u}} \in \underset{\mathbf{u}\in\mathbb{R}^{|\Omega|}}{\operatorname{Argmin}} \underbrace{\frac{1}{2}\|A\mathbf{u} - \mathbf{g}\|_2^2}_{\text{Data-term}} + \lambda \underbrace{\|D\mathbf{u}\|_p^p}_{\text{Penalization}}$$

(a) Degraded Uniform blur 9×9 Gaussian noise (b) Inverse filtering Quadratic regularisation (c) $\Lambda = \mathrm{Id}$ (d) Λ Laplacian (e) Total variation

→ N. Pustelnik, A. Benazza-Benhayia, Y. Zheng, J.-C. Pesquet, Wavelet-based Image Deconvolution and Reconstruction, Wiley Encyclopedia of EEE, Feb. 2016. [PDF]

Example: Inverse problems

- Data: We observe data $\mathbf{g} \in \mathbb{R}^K$ being a degraded version of an original image $\bar{\mathbf{u}} \in \mathbb{R}^{|\Omega|}$ such that: $\mathbf{g} = A\bar{\mathbf{u}} + \varepsilon$
  - $A : \mathbb{R}^{K\times|\Omega|}$: denotes a linear degradation (e.g. a blur, decimation op.)
  - ε : denotes a noise (e.g. Gaussian)

- Goal: Restore the degraded image i.e., find $\widehat{\mathbf{u}}$ close to $\bar{\mathbf{u}}$:

$$\widehat{\mathbf{u}} \in \underset{\mathbf{u}\in\mathbb{R}^{|\Omega|}}{\mathrm{Argmin}}\, \underbrace{\frac{1}{2}\|A\mathbf{u} - \mathbf{g}\|_2^2}_{\text{Data-term}} + \lambda \underbrace{\|D\mathbf{u}\|_p^p}_{\text{Penalization}}$$

- Specificities of the data-term:
  - Data-term differentiable with $\|A\|^2$-Lipschitz gradient.
  - (Closed form) expression of the proximity operator for some $\tau > 0$,

$$\mathrm{prox}_{\frac{\tau}{2}\|A\cdot-\mathbf{g}\|^2}(\mathbf{u}) = (\tau A^* A + \mathrm{Id})^{-1}(\tau A^*\mathbf{g} + \mathbf{u})$$

  - Rarely strongly convex.

Example: COVID-19 Reproduction Number $R(t)$

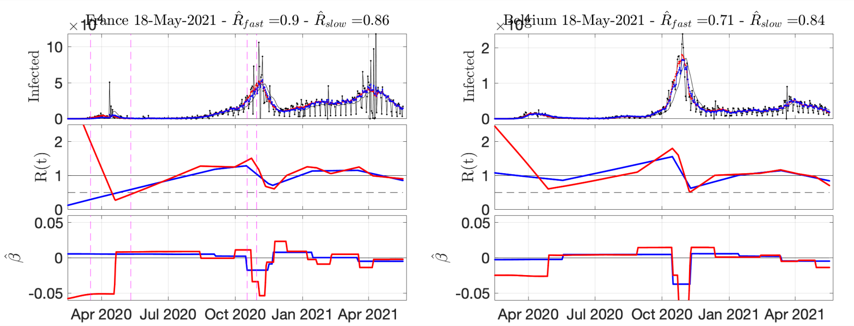
- Data/model: Poisson model to mimic the spread of an epidemic:
    - $R(t)$: propagation speed.
    - $\mathbf{g} \in \mathbb{R}^T$: Number of cases or hospitalisation for a single country or single department. Count of daily new infections $\mathbf{g} = (g_t)_{1 \le t \le T}$ modelled as Poisson random variables of parameter $p_t = R(t) \sum_{k \ge 1} \phi(k) g_{t-k}$.
    - $\sum_{k \ge 1} \phi(k) g_{t-k}$: models previous days effects.
- Goal: Estimate the reproduction number $R(t) = \widehat{\mathbf{u}}$ from the data $\mathbf{g}$:

$$\widehat{\mathbf{u}} \in \underset{\mathbf{u} \in \mathbb{R}^T}{\mathrm{Argmin}} \underbrace{\mathrm{DKL}(\mathbf{g}, \mathbf{u} \odot \Phi \mathbf{g})}_{\text{Data-term}} + \lambda \underbrace{\|D\mathbf{u}\|_1}_{\text{Penalization}}$$

with

$$\mathrm{DKL}(\mathbf{v}; \mathbf{g}) = \sum_n \psi(v_m; g_m) \quad \text{where} \quad \psi(v_m; g_m) = \begin{cases} -g_m \ln(v_m) + v_m & \text{if} \quad v_m > 0 \text{ and } g_m > 0 \\ v_m & \text{if} \quad v_m \ge 0 \text{ and } g_m = 0 \\ +\infty & \text{otherwise} \end{cases}$$

Example: COVID-19 Reproduction Number $R(t)$



→ P. Abry, N. Pustelnik, S. Roux, P. Jensen, P. Flandrin, R. Gribonval, C.-G. Lucas, E. Guichard, P. Borgnat, N. Garnier, B. Audit, Spatial and temporal regularization to estimate COVID-19 Reproduction Number R(t): Promoting piecewise smoothness via convex optimization, PLoS One, 15(8), Aug. 2020. [PDF]

→ [Daily updates]

Example: COVID-19 Reproduction Number $R(t)$

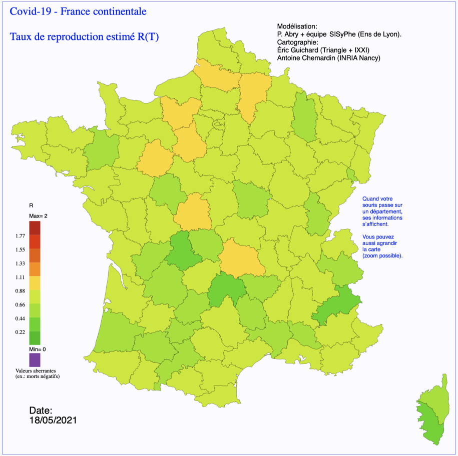
- Data/model: Poisson model to mimic the spread of an epidemic:
  - $R(t)$: propagation speed.
  - $\mathbf{g} \in \mathbb{R}^T$: Number of cases or hospitalisation for a single country or single department. Count of daily new infections $\mathbf{g} = (g_t)_{1 \leq t \leq T}$ modelled as Poisson random variables of parameter $p_t = R(t) \sum_{k \geq 1} \phi(k) g_{t-k}$.
  - $\sum_{k \geq 1} \phi(k) g_{t-k}$: models previous days effects.
- Goal: Estimate the reproduction number $R(t) = \widehat{\mathbf{u}}$ from the data $\mathbf{g}$:

$$\widehat{\mathbf{u}} = \in \underset{\mathbf{u} \in \mathbb{R}^{TN}}{\mathrm{Argmin}} \underbrace{\mathrm{DKL}(\mathbf{g}, \mathbf{u} \odot \Phi \mathbf{g})}_{\text{Data-term}} + \underbrace{\lambda_s \| G\mathbf{u} \|_1 + \lambda_t \| D\mathbf{u} \|_1}_{\text{Penalization}}$$

with

$$\mathrm{DKL}(\mathbf{v}; \mathbf{g}) = \sum_n \psi(v_m; g_m) \quad \text{where} \quad \psi(v_m; g_m) = \begin{cases} -g_m \ln(v_m) + v_m & \text{if} \quad v_m > 0 \text{ and } g_m > 0 \\ v_m & \text{if} \quad v_m \geq 0 \text{ and } g_m = 0 \\ +\infty & \text{otherwise} \end{cases}$$

Example: COVID-19 Reproduction Number $R(t)$



→ P. Abry, N. Pustelnik, S. Roux, P. Jensen, P. Flandrin, R. Gribonval, C.-G. Lucas, E. Guichard, P. Borgnat, N. Garnier, B. Audit, Spatial and temporal regularization to estimate COVID-19 Reproduction Number R(t): Promoting piecewise smoothness via convex optimization, PLoS One, 15(8), Aug. 2020. [PDF]

→ [Evolution along time and across France of R(t)]

Example: COVID-19 Reproduction Number $R(t)$

- Data:
  - $\mathbf{g} \in \mathbb{R}^T$: Number of cases or hospitalisation for a single country or single department.
  - Φ : serial interval function (the probability of secondary infections as a function of time after symptoms onset).
- Goal: Estimate the reproduction number $R(t) = \widehat{\mathbf{u}}$ from the data $\mathbf{g}$:

$$\widehat{\mathbf{u}} \in \operatorname*{Argmin}_{\mathbf{u}\in\mathbb{R}^T} \underbrace{\mathrm{DKL}(\mathbf{g}, \mathbf{u} \odot \Phi\mathbf{g})}_{\text{Data-term}} + \lambda \underbrace{\|D\mathbf{u}\|_1}_{\text{Penalization}}$$

- Specificities of the objective function:
  - Data-term differentiable but without a Lipschitz gradient.
  - Closed form expression of the proximity operator associated to $\mathrm{DKL}(\mathbf{g}, \cdot \odot \Phi\mathbf{g})$

Non-smooth optimization

- Numerous problems in signal and image processing can be modelled as a sum of convex functions composed with linear operators

For every $s \in \{1, \ldots, S\}$, $f_s \in \Gamma_0(\mathcal{G}_s)$ and $L_s \colon \mathbb{R}^{|\Omega|} \to \mathcal{G}_s$ denote a linear operator. We aim to solve:

$$\widehat{\mathbf{u}} \in \underset{\mathbf{u} \in \mathbb{R}^{|\Omega|}}{\mathrm{Argmin}} \sum_{s=1}^{S} f_s(L_s \mathbf{u})$$

- Some of them involve functions where only the proximity operator can be considered (ℓ_1-penalization, DKL,...)

- Some of them involve functions where both gradient or proximity operator can be considered (Huber function, ℓ_2^2-data-term,...)

Non-smooth optimization

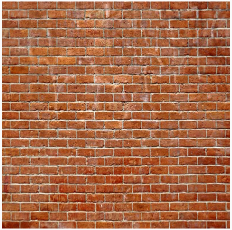
For every $s \in \{1, \ldots, S\}$, $f_s \in \Gamma_0(\mathcal{G}_s)$ and $L_s \colon \mathbb{R}^{|\Omega|} \to \mathcal{G}_s$ denote a linear operator. We aim to solve:

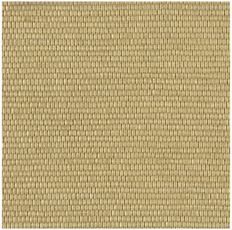
$$\widehat{\mathbf{u}} \in \operatorname*{Argmin}_{\mathbf{u} \in \mathbb{R}^{|\Omega|}} \sum_{s=1}^{S} f_s(L_s \mathbf{u})$$

- Since 2004, numerous proximal algorithms: [Bauschke-Combettes, 2017]
  - Forward-Backward $S = 2$, f_1 Lipschitz gradient and $L_2 = \mathrm{Id}$
  - ADMM Invert $\sum_{s=1}^{S} L_s L_s^*$
  - Primal-dual (Chambolle-Pock, Condat-Vũ ...)
  - ...
- Question: When both gradient step or proximal step can be performed, which type of step should we prefer ?

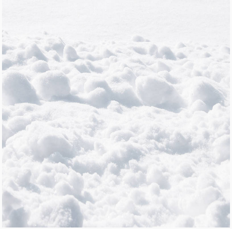
Prox versus grad for texture segmentation (strongly convex minimization problem)

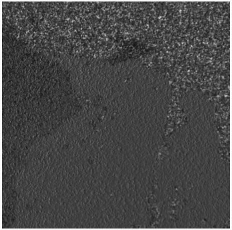
- Geometric textures → periodic





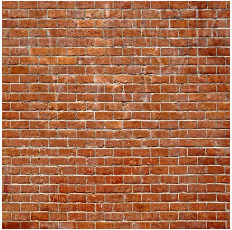
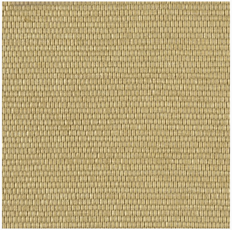
- Stochastic textures



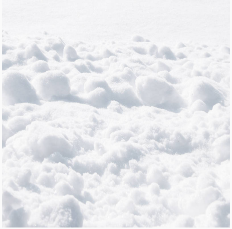
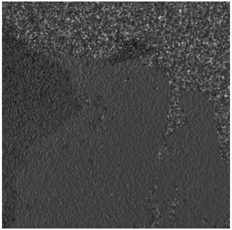


Stochastic textures

- Geometric textures → periodic

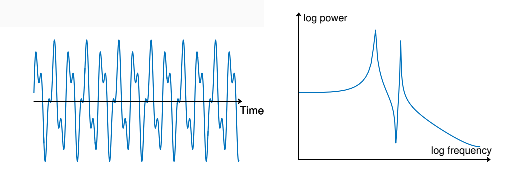




- Stochastic textures → scale-free ?

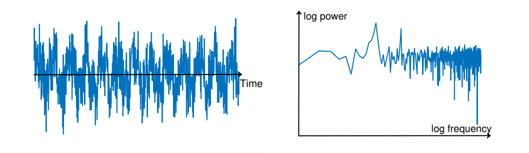




Stochastic textures

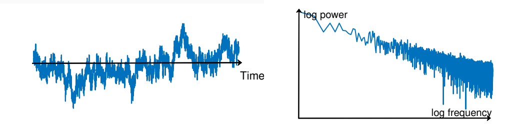
- Sinusoidal signal → periodic



- Sinusoidal signal + noise → periodic

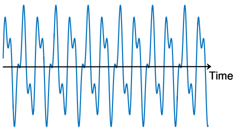


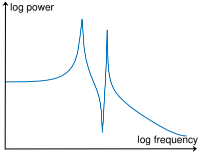
- Monofractal signal → scale-free



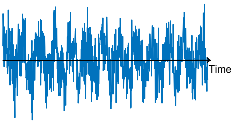
Stochastic textures

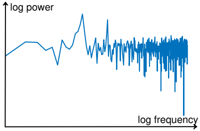
- Sinusoidal signal $\rightarrow$ periodic



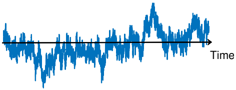


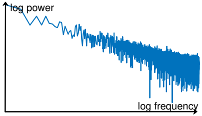
- Sinusoidal signal + noise $\rightarrow$ periodic





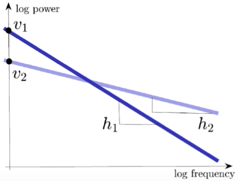
- Monofractal signal $\rightarrow$ scale-free





Texture segmentation:

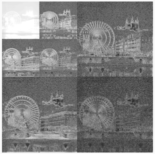
$\rightarrow$ require to compute the slope at each location



From wavelets to local regularity

- **Discrete wavelet transform**:
$$F = \left[H_{1,1}^\top, \ldots, H_{J,3}^\top, L_{J,4}^\top\right]^\top \quad \text{where} \quad \begin{cases} H_{j,m} \in \mathbb{R}^{\frac{N}{4^j} \times N} \\ L_{J,4} \in \mathbb{R}^{\frac{N}{4^J} \times N} \end{cases}$$

$\mathbf{g}$



$\zeta = F\mathbf{g}$

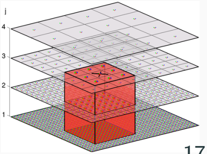
- **Wavelet coefficients** at scale $j \in \{1, \ldots, J\}$ and subband $m = \{1, 2, 3\}$:
$$\boxed{\zeta_{j,m} = H_{j,m}\mathbf{g}}$$

- **Wavelet leaders** at scale j and location $\underline{k}$

[Wendt et al., 2009]

$\rightarrow$ local supremum taken within a spatial neighborhood across all finer scales $j' \leq j$

$$\boxed{\mathcal{L}_{j,\underline{k}} = \sup_{\substack{m=\{1,2,3\} \\ \lambda_{j',\underline{k}'} \subset \Lambda_{j,\underline{k}}}} |\zeta_{j',m,\underline{k}}|} \quad \text{where} \quad \begin{cases} \lambda_{j,\underline{k}} = [\underline{k}2^j, (\underline{k}+1)2^j) \\ \Lambda_{j,\underline{k}} = \bigcup_{p \in \{-1,0,1\}^2} \lambda_{j,\underline{k}+p} \end{cases}$$

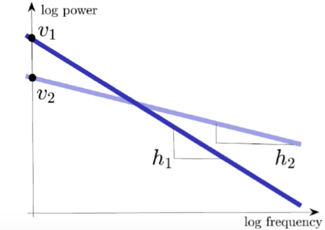


From wavelets to local regularity: joint estimation (2)

- Behavior through the scales [Jaffard, 2004]

$$\mathcal{L}_{j,\underline{n}} \simeq s_{\underline{n}} 2^{jh_{\underline{n}}} \quad \text{as} \quad 2^j \to 0$$

$$\boxed{\log_2 \mathcal{L}_{j,\underline{k}} \simeq \underbrace{\log_2 s_{\underline{n}}}_{v_{\underline{n}}} + jh_{\underline{n}} \quad \text{as} \quad 2^j \to 0.}$$



- Data-fidelity term [Pascal, Pustelnik, Abry, 2021]

$$\Phi(\boldsymbol{v}, \boldsymbol{h}; \boldsymbol{\mathcal{L}}) = \frac{1}{2} \sum_{\underline{n}} \sum_{j} (v_{\underline{n}} + jh_{\underline{n}} - \log_2 \mathcal{L}_{j,\underline{n}})^2$$

$$= \frac{1}{2} \sum_{\underline{n}} \left\| A \begin{pmatrix} v_{\underline{n}} \\ h_{\underline{n}} \end{pmatrix} - \log_2 \boldsymbol{\mathcal{L}}_{\underline{n}} \right\|_2^2 \quad \text{where} \quad A = \begin{pmatrix} 1 & 1 \\ 1 & 2 \\ \vdots & \vdots \\ 1 & J \end{pmatrix}$$

From wavelets to local regularity: joint estimation (2)

Closed form for prox_{Φ} [Pascal, Pustelnik, Abry, 2021]

For every $(\boldsymbol{v}, \boldsymbol{h}) \in \mathbb{R}^{|\Omega|} \times \mathbb{R}^{|\Omega|}$, denoting $(\boldsymbol{p}, \boldsymbol{q}) = \mathrm{prox}_{\Phi}(\boldsymbol{v}, \boldsymbol{h}) \in \mathbb{R}^{|\Omega|} \times \mathbb{R}^{|\Omega|}$ one has

$$\begin{cases} \boldsymbol{p} = \dfrac{(1+R_2)(\boldsymbol{\mathcal{S}}+\boldsymbol{v}) - R_1(\boldsymbol{\mathcal{T}}+\boldsymbol{h})}{(1+R_0)(1+R_2)-R_1^2}, \\ \boldsymbol{q} = \dfrac{(1+R_0)(\boldsymbol{\mathcal{T}}+\boldsymbol{h}) - R_1(\boldsymbol{\mathcal{S}}+\boldsymbol{v})}{(1+R_0)(1+R_2)-R_1^2}. \end{cases}$$

where $R_m = \sum_j j^m$, $\mathcal{S}_{\underline{n}} = \sum_j \log_2 \mathcal{L}_{j,\underline{n}}$, and $\mathcal{T}_{\underline{n}} = \sum_j j \log_2 \mathcal{L}_{j,\underline{n}}$.

Proof: Rely on the closed form of

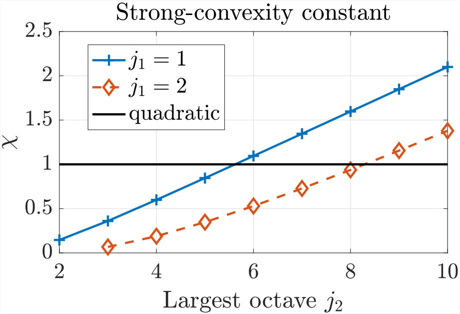
$$\begin{pmatrix} p_{\underline{n}} \\ q_{\underline{n}} \end{pmatrix} = \mathrm{prox}_{\frac{1}{2}\|A\cdot - \log_2 \boldsymbol{\mathcal{L}}_{\underline{n}}\|_2^2} \begin{pmatrix} v_{\underline{n}} \\ h_{\underline{n}} \end{pmatrix} = (A^*A + \mathrm{Id})^{-1} \left(A^* \log_2 \boldsymbol{\mathcal{L}}_{\underline{n}} + \begin{pmatrix} v_{\underline{n}} \\ h_{\underline{n}} \end{pmatrix} \right)$$

$$\text{with } A = \begin{pmatrix} 1 & 1 \\ 1 & 2 \\ \vdots & \vdots \\ 1 & J \end{pmatrix} \text{ and thus } \begin{cases} A^*A = \begin{pmatrix} R_0 & R_1 \\ R_1 & R_2 \end{pmatrix} \\ A^* \log_2 \boldsymbol{\mathcal{L}}_{\underline{n}} = \begin{pmatrix} \mathcal{S}_{\underline{n}} \\ \mathcal{T}_{\underline{n}} \end{pmatrix} \end{cases}$$

From wavelets to local regularity: joint estimation (2)

Strongly convex fidelity term Φ [Pascal, Pustelnik, Abry, 2021]

Function $\Phi(\boldsymbol{v}, \boldsymbol{h}; \mathcal{L})$ is μ-strongly convex w.r.t the variables $(\boldsymbol{v}, \boldsymbol{h})$, with $\mu = \chi$ where $\chi > 0$ is the lowest eigenvalue of the symmetric and positive definite matrix $A^*A = \begin{pmatrix} R_0 & R_1 \\ R_1 & R_2 \end{pmatrix}$ where $R_m = \sum_j j^m$.



From wavelets to local regularity: joint estimation (2)

Expression of the conjugate of Φ [Pascal, Pustelnik, Abry, 2021]

$$\Phi^*(\boldsymbol{v}, \boldsymbol{h}; \mathcal{L}) = \frac{1}{2}\langle (\boldsymbol{v}, \boldsymbol{h})^\top, \mathbf{J}^{-1}(\boldsymbol{v}, \boldsymbol{h})^\top \rangle + \langle (\mathcal{S}, \mathcal{T})^\top, \mathbf{J}^{-1}(\boldsymbol{v}, \boldsymbol{h})^\top \rangle + \mathcal{C},$$

where

$$\begin{cases} \mathcal{C} & = \frac{1}{2}\langle (\mathcal{S}, \mathcal{T})^\top, \mathbf{J}^{-1}(\mathcal{S}, \mathcal{T})^\top \rangle - \frac{1}{2}\sum_j (\log_2 \mathcal{L}_j)^2. \\ \mathcal{S} & = \sum_j \log_2 \mathcal{L}_j \\ \mathcal{T} & = \sum_j j \log_2 \mathcal{L}_j \\ \mathbf{J} & = A^* A = \begin{pmatrix} R_0 & R_1 \\ R_1 & R_2 \end{pmatrix} \quad \text{and} \quad R_m = \sum_j j^m, \end{cases}$$

From wavelets to local regularity: joint estimation (2)

By definition of the Fenchel conjugate,

$$F^*(\boldsymbol{v}, \boldsymbol{h}; \boldsymbol{\mathcal{L}}) = \sup_{\widetilde{\boldsymbol{v}} \in \mathbb{R}^{|\Omega|}, \widetilde{\boldsymbol{h}} \in \mathbb{R}^{|\Omega|}} \langle \widetilde{\boldsymbol{v}}, \boldsymbol{v} \rangle + \langle \widetilde{\boldsymbol{h}}, \boldsymbol{h} \rangle - F(\widetilde{\boldsymbol{v}}, \widetilde{\boldsymbol{h}}; \boldsymbol{\mathcal{L}}). \tag{1}$$

The supremum is obtained at $(\bar{\boldsymbol{v}}, \bar{\boldsymbol{h}})$ such that, for every $\underline{n} \in \Omega$,

$$\begin{cases} v_{\underline{n}} - \sum_j \left(\bar{v}_{\underline{n}} + j \bar{h}_{\underline{n}} - \log_2 \mathcal{L}_{j,\underline{n}} \right) = 0 \\ h_{\underline{n}} - \sum_j j \left(\bar{v}_{\underline{n}} + j \bar{h}_{\underline{n}} - \log_2 \mathcal{L}_{j,\underline{n}} \right) = 0. \end{cases} \tag{2}$$

or equivalently,

$$\begin{cases} R_0 \bar{v}_{\underline{n}} + R_1 \bar{h}_{\underline{n}} = v_{\underline{n}} + \mathcal{S}_{\underline{n}} \\ R_1 \bar{v}_{\underline{n}} + R_2 \bar{h}_{\underline{n}} = h_{\underline{n}} + \mathcal{T}_{\underline{n}} \end{cases} \tag{3}$$

that yields

$$\begin{pmatrix} \bar{v}_{\underline{n}} \\ \bar{h}_{\underline{n}} \end{pmatrix} = \mathbf{J}^{-1} \begin{pmatrix} v_{\underline{n}} + \mathcal{S}_{\underline{n}} \\ h_{\underline{n}} + \mathcal{T}_{\underline{n}} \end{pmatrix} \tag{4}$$

From wavelets to local regularity: joint estimation (2)

PLOVER: Piecewise constant LOcal VariancE and Regularity estimation
[Pascal, Pustelnik, Abry, ACHA, 2021]
Find $(\widehat{\mathbf{v}}, \widehat{\mathbf{h}}) \in \underset{\mathbf{v},\mathbf{h}}{\operatorname{Argmin}} \sum_j \| \log_2 \mathcal{L}_j - \mathbf{v} - j\mathbf{h}\|_2^2 + \lambda \underbrace{\left\| \left[\mathbf{D}\boldsymbol{v}; \alpha \mathbf{D}\boldsymbol{h}\right]^\top \right\|_{2,1}}_{\mathrm{TV}_\alpha}$

where TV_α couples spatial variations of $\boldsymbol{v}$ and $\boldsymbol{h}$ and thus favor their occurrences at same location.

- \+ Combined estimation and segmentation.
- \+ Joint estimation of the local variance and local regularity.
- \+ Strongly convex.
- \+ Closed form expression of the proximity operator associated to the data-fidelity term.
- \+ Dual formulation possible.

From wavelets to local regularity: joint estimation (2)

PLOVER: Piecewise constant LOcal VariancE and Regularity estimation
[Pascal, Pustelnik, Abry, ACHA, 2021]
Find $(\widehat{\mathbf{v}}, \widehat{\mathbf{h}}) \in \underset{\mathbf{v},\mathbf{h}}{\mathrm{Argmin}} \sum_j \| \log_2 \mathcal{L}_j - \mathbf{v} - j\mathbf{h}\|_2^2 + \lambda \underbrace{\left\| \left[\mathbf{D}\boldsymbol{v}; \alpha \mathbf{D}\boldsymbol{h}\right]^\top \right\|_{2,1}}_{\mathrm{TV}_\alpha}$

where TV_α couples spatial variations of $\boldsymbol{v}$ and $\boldsymbol{h}$ and thus favor their occurrences at same location.

Algorithmic solutions:

- Accelerated strongly convex Chambolle-Pock algorithm.
- FISTA on the dual [Chambolle-Dossal, 2015].

Algorithm 5: $\mathrm{PD_C}$: Coupled estimation (Pb. (12))

Initialization:

Set $\boldsymbol{v}^{[0]} \in \mathbb{R}^{|\Upsilon|}$, $\boldsymbol{u}^{[0]} = \mathbf{D}\boldsymbol{v}^{[0]}$, $\bar{\boldsymbol{u}}^{[0]} = \boldsymbol{u}^{[0]}$;

Set $\boldsymbol{h}^{[0]} \in \mathbb{R}^{|\Upsilon|}$, $\boldsymbol{\ell}^{[0]} = \alpha\mathbf{D}\boldsymbol{h}^{[0]}$, $\bar{\boldsymbol{\ell}}^{[0]} = \boldsymbol{\ell}^{[0]}$;

Set $\alpha > 0$ and $\lambda > 0$.

Set (δ_0, ν_0) such that $\delta_0\nu_0 \max(1,\alpha)\|\mathbf{D}\|^2 < 1$;

for $t \in \mathbb{N}^*$ **do**

Primal variable update:

$$\begin{pmatrix} \boldsymbol{v}^{[t+1]} \\ \boldsymbol{h}^{[t+1]} \end{pmatrix} = \mathrm{prox}_{\delta_t \Phi}\left(\begin{pmatrix} \boldsymbol{v}^{[t]} \\ \boldsymbol{h}^{[t]} \end{pmatrix} - \delta_t \begin{pmatrix} \mathbf{D}^* \bar{\boldsymbol{u}}^{[t]} \\ \alpha \mathbf{D}^* \bar{\boldsymbol{\ell}}^{[t]} \end{pmatrix} \right)$$

Dual variable update:

$$\begin{pmatrix} \boldsymbol{u}^{[t+1]} \\ \boldsymbol{\ell}^{[t+1]} \end{pmatrix} = \mathrm{prox}_{\nu_t (\lambda \|.\|_{2,1})^*} \begin{pmatrix} \boldsymbol{u}^{[t]} + \nu_t \mathbf{D}\boldsymbol{v}^{[t]} \\ \boldsymbol{\ell}^{[t]} + \nu_t \alpha \mathbf{D}\boldsymbol{h}^{[t]} \end{pmatrix}$$

Descent steps update:

$$\vartheta_t = (1 + 2\mu\delta_t)^{-1/2},\ \delta_{t+1} = \vartheta_t \delta_t,\ \nu_{t+1} = \nu_t / \vartheta_t$$

Auxiliary variable update:

$$\begin{pmatrix} \bar{\boldsymbol{u}}^{[t+1]} \\ \bar{\boldsymbol{\ell}}^{[t+1]} \end{pmatrix} = \begin{pmatrix} \boldsymbol{u}^{[t+1]} \\ \boldsymbol{\ell}^{[t+1]} \end{pmatrix} + \vartheta_t \left(\begin{pmatrix} \boldsymbol{u}^{[t+1]} \\ \boldsymbol{\ell}^{[t+1]} \end{pmatrix} - \begin{pmatrix} \boldsymbol{u}^{[t]} \\ \boldsymbol{\ell}^{[t]} \end{pmatrix} \right)$$

Algorithm 3: $\mathrm{FISTA}_{\mathrm{C}}$: Coupled estimation (Pb. (12))

Initialization: Set $\boldsymbol{u}^{[0]} \in \mathbb{R}^{2\times|\Upsilon|}$, $\bar{\boldsymbol{u}}^{[0]} = \boldsymbol{u}^{[0]}$;

Set $\boldsymbol{\ell}^{[0]} \in \mathbb{R}^{2\times|\Upsilon|}$, $\bar{\boldsymbol{\ell}}^{[0]} = \boldsymbol{\ell}^{[0]}$;

Let $\left(\mathcal{S}_{\underline{n}}, \mathcal{T}_{\underline{n}}\right)$ defined in (4);

Let $\mathbf{J}$ defined in (3);

Set $(\forall \underline{n})$ $\left(v_{\underline{n}}^{[0]}, h_{\underline{n}}^{[0]}\right)^{\top} = \mathbf{J}^{-1}\left(\mathcal{S}_{\underline{n}}, \mathcal{T}_{\underline{n}}\right)^{\top}$;

Set $b > 2$ and $\tau_0 = 1$;

Set $\alpha > 0$ and $\lambda > 0$;

Set $\gamma > 0$ s. t. $\gamma \max(1,\alpha)\,\|\mathbf{J}^{-1}\|\|\|\mathbf{D}\|^2 < 1$;

for $t \in \mathbb{N}$ **do**

Dual variable update:

$$\begin{pmatrix}\boldsymbol{u}^{[t+1]}\\ \boldsymbol{\ell}^{[t+1]}\end{pmatrix} = \mathrm{prox}_{\gamma(\lambda\|\cdot\|_{2,1})^*}\begin{pmatrix}\bar{\boldsymbol{u}}^{[t]} + \gamma\mathbf{D}\boldsymbol{v}^{[t]}\\ \bar{\boldsymbol{\ell}}^{[t]} + \gamma\alpha\mathbf{D}\boldsymbol{h}^{[t]}\end{pmatrix}$$

FISTA parameter update

$$\tau_{t+1} = \frac{t+b}{b}$$

Auxiliary variable update

$$\bar{\boldsymbol{u}}^{[t+1]} = \boldsymbol{u}^{[t+1]} + \frac{\tau_t - 1}{\tau_{t+1}}\left(\boldsymbol{u}^{[t+1]} - \boldsymbol{u}^{[t]}\right)$$

$$\bar{\boldsymbol{\ell}}^{[t+1]} = \boldsymbol{\ell}^{[t+1]} + \frac{\tau_t - 1}{\tau_{t+1}}\left(\boldsymbol{\ell}^{[t+1]} - \boldsymbol{\ell}^{[t]}\right)$$

Primal variable update

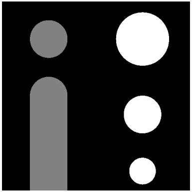
$$\begin{pmatrix}\boldsymbol{v}^{[t+1]}\\ \boldsymbol{h}^{[t+1]}\end{pmatrix} = \begin{pmatrix}\boldsymbol{v}^{[t]}\\ \boldsymbol{h}^{[t]}\end{pmatrix} - \mathbf{J}^{-1}\begin{pmatrix}\mathbf{D}^*(\boldsymbol{u}^{[t+1]} - \boldsymbol{u}^{[t]})\\ \alpha\mathbf{D}^*(\boldsymbol{\ell}^{[t+1]} - \boldsymbol{\ell}^{[t]})\end{pmatrix}$$

Two-step versus one-step texture segmentation

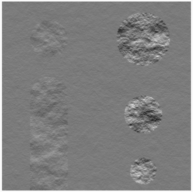
		Configuration I			Configuration III		
		T-ROF	T-*joint*	T-*coupled*	T-ROF	T-*joint*	T-*coupled*
Iterations (10^3 it.)	DFB	96 ± 48	> 250	> 250	241 ± 18	> 250	> 250
	FISTA	1.7 ± 0.4	50.2 ± 21.0	231 ± 37	$\mathbf{3.7 \pm 0.7}$	48.1 ± 3.4	> 250
	PD	31.8 ± 17.0	> 250	> 250	201 ± 69	> 250	> 250
	AcPD	$\mathbf{1.5 \pm 0.4}$	$\mathbf{31.4 \pm 4.6}$	$\mathbf{125 \pm 67}$	45.2 ± 43	$\mathbf{40.5 \pm 2.8}$	$\mathbf{121 \pm 42}$
Time (s)	DFB	$1,090 \pm 520$	$4,840 \pm 15$	$4,210 \pm 76$	$2,010 \pm 73$	$4,810 \pm 215$	$4,200 \pm 76$
	FISTA	16 ± 4	$1,030 \pm 410$	$4,800 \pm 560$	$\mathbf{30 \pm 5}$	989 ± 64	$5,110 \pm 340$
	PD	297 ± 150	$4,180 \pm 69$	$4,110 \pm 43$	$1,580 \pm 490$	$4,150 \pm 18$	$4,100 \pm 15$
	AcPD	$\mathbf{15 \pm 4}$	$\mathbf{619 \pm 96}$	$\mathbf{2,420 \pm 1,300}$	349 ± 330	$\mathbf{785 \pm 59}$	$\mathbf{2,320 \pm 790}$

Table 2: Number of iterations and computational time necessary to reach Condition (26) for the different proximal algorithms investigated, illustrated on two configurations I ($\Delta H = 0.2$, $\Delta\Sigma^2 = 0.1$) and III ($\Delta H = 0.1$, $\Delta\Sigma^2 = 0.1$). **DFB**: Dual Forward-Backward, **FISTA**: inertial acceleration of DFB, **PD**: primal-dual, **AcPD**: strong-convexity based acceleration of PD.

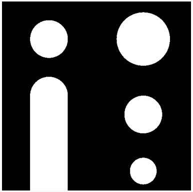
Two-step versus one-step texture segmentation



Mask

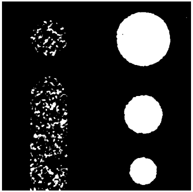


Synthetic texture



Optimal solution

T-ROF [Cai2013]

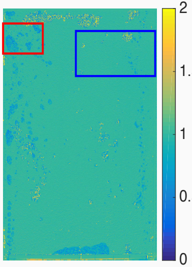


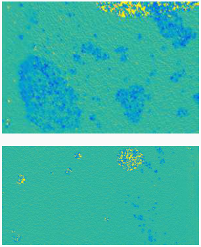
Matrix factorization [Yuan2015]

Proposed [Pascal2019]

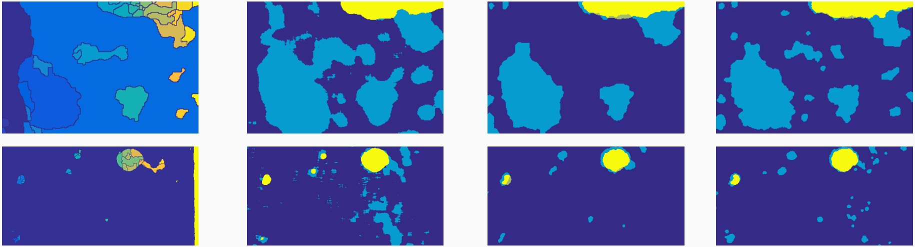
⇒ Illustration of Interface detection on a piecewise fractal textured image that mimics a multiphasic flow.

Results on multiphase flow data

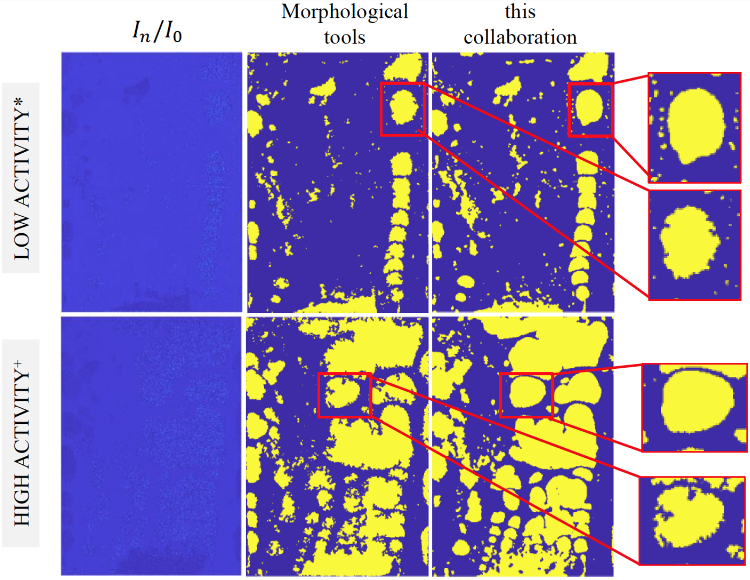




[Arbelaez et al. 2011] | [Yuan et al. 2015] | T-ROF | PLOVER



Results on multiphase flow data



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

Prox versus grad on texture segmentation: conclusions

- Joint estimation and segmentation formulated as a **strongly convex** minimization problem. $\rightarrow$ Fast algorithmic procedure. Application to large-scale problems.

- Chambolle-Pock using strong convexity faster than FISTA on the dual. $\rightarrow$ Proximal step faster than gradient step based on numerical comparisons.

- Matlab toolbox including automatic tuning of the hyperparameters : GitHub (bpascal-fr/gsugar)

Prox versus grad

Non-smooth optimization: large-scale data

$$\widehat{\mathbf{u}} \in \underset{\mathbf{u}\in\mathcal{H}}{\mathrm{Argmin}}\ f(\mathbf{u}) + g(\mathbf{u}) \qquad (5)$$

- Activating f and g via proximal steps can be advantageous numerically [Combettes, Glaudin,2019]
- The choice of the most efficient algorithm for a specific data processing problem with the form of (5) is a complicated task.
- Convergence rate is an useful tool in order to provide a theoretical comparison among algorithms.
- The theoretical behaviour of an algorithmic scheme may differ considerably from its numerical efficiency, which enlightens the importance of obtaining sharp convergence rates exploiting the properties of f and g.

Prox versus grad

$$\widehat{\mathbf{u}} \in \underset{\mathbf{u} \in \mathcal{H}}{\text{Argmin}}\ f(\mathbf{u}) + g(\mathbf{u})$$

- Sharp linear convergence rates can be obtained for several splitting algorithms under strong convexity of f and/or g. [Giselsson, Boyd,2017][Davis, Yin,2017] [Taylor,Hendrickx,Glineur,2018] [Ryu, Hannah, Yin,2019] [Ryu, Taylor, Bergeling, Gilsesson,2019]

- Sub-linear convergence rates of some first order methods depending on the KL-exponent are obtained in when $f + g$ is a KL-function. ([Attouch, Bolte, Svaiter,2013][Bolte, Daniilidis, Lewis,2006]. $\rightarrow$ KL-exponents are usually difficult to compute.

Algorithms

Gradient method Let $f \in \Gamma_0(\mathcal{H})$ and $f \in C^{1,1}_\zeta(\mathcal{H})$ (i.e. Gâteaux differentiable + ζ-Lipschitz continuous). We set, for some $\tau > 0$,

$$\Phi := \mathrm{Id} - \tau \nabla f$$

Proximal Point Algorithm (PPA) Let $f \in \Gamma_0(\mathcal{H})$. We set, for some $\tau > 0$,

$$\Phi := \mathrm{prox}_{\tau f} = (\mathrm{Id} + \tau \partial f)^{-1}.$$

Forward-backward splitting Let $f \in \Gamma_0(\mathcal{H})$ and $g \in \Gamma_0(\mathcal{H})$. Additionally, $f \in C^{1,1}_\zeta(\mathcal{H})$ (i.e. Gâteaux differentiable + ζ-Lipschitz continuous). We set, for some $\tau > 0$,

$$\Phi := \mathrm{prox}_{\tau g}(\mathrm{Id} - \tau \nabla f) = (\mathrm{Id} + \tau \partial g)^{-1}(\mathrm{Id} - \tau \nabla f)$$

Algorithms

Peaceman-Rachford splitting Let $f \in \Gamma_0(\mathcal{H})$ and $g \in \Gamma_0(\mathcal{H})$. We set, for some $\tau > 0$,

$$\Phi := (2\mathrm{prox}_{\tau g} - \mathrm{Id}) \circ (2\mathrm{prox}_{\tau f} - \mathrm{Id})$$

Douglas-Rachford splitting Let $f \in \Gamma_0(\mathcal{H})$ and $g \in \Gamma_0(\mathcal{H})$. We set, for some $\tau > 0$,

$$\Phi := \mathrm{prox}_{\tau g} \circ (2\mathrm{prox}_{\tau f} - \mathrm{Id}) + \mathrm{Id} - \mathrm{prox}_{\tau f}$$

Theoretical comparisons

Let $f \in C^{1,1}_{1/\alpha}(\mathcal{H})$ and $g \in C^{1,1}_{1/\beta}(\mathcal{H})$, for some $\alpha > 0$ and $\beta > 0$.
The problem is to

$$\underset{x \in \mathcal{H}}{\text{minimize}} \;\; f(x) + g(x),$$

under the assumption that solutions exist.

Example: **Smooth TV denoising**

$$\underset{x \in \mathbb{R}^N}{\text{minimize}} \;\; \frac{1}{2}\|x - z\|_2^2 + \chi h_\mu(Lx),$$

- $L \in \mathbb{R}^{N-1 \times N}$ denotes the first order discrete difference operator

  $$(\forall n \in \{1, \ldots, N-1\}) \quad (Lx)_n = \tfrac{1}{2}(x_n - x_{n-1})$$

- h_μ: Huber loss, the smooth approximation of the ℓ_1−norm parametrized by $\mu > 0$.

  $h_\mu \in C^{1,1}_{1/\mu}(\mathbb{R}^{N-1})$.

  Closed form expression of prox_{h_μ}.

Theoretical comparisons

Proposition (see [Briceno-Arias, Pustelnik, 2021] for detailed references)
In the context of Problem p.16, suppose that f **is $\rho-$strongly convex**, for some $\rho \in \left]0, \alpha^{-1}\right[$, and let $\tau > 0$. Then, the following holds:

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

$$r_G(\tau) := \max\left\{|1 - \tau\rho|, |1 - \tau(\beta^{-1} + \alpha^{-1})|\right\} \in]0, 1[\,. \qquad (6)$$

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

$$\tau^* = \frac{2}{\rho + \alpha^{-1} + \beta^{-1}}$$

   and

$$r_G(\tau^*) = \frac{\alpha^{-1} + \beta^{-1} - \rho}{\alpha^{-1} + \beta^{-1} + \rho}.$$

Theoretical comparisons

Proposition (see [Briceno-Arias, Pustelnik, 2021] for detailed references)
In the context of Problem p.16, suppose that f **is $\rho-$strongly convex**, for some $\rho \in \left]0, \alpha^{-1}\right[$, and let $\tau > 0$. Then, the following holds:

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

$$r_{T_1}(\tau) := \max\left\{|1 - \tau\rho|, |1 - \tau\alpha^{-1}|\right\} \in]0, 1[\,. \tag{6}$$

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

$$\tau^* = \frac{2}{\rho + \alpha^{-1}} \quad \text{and} \quad r_{T_1}(\tau^*) = \frac{\alpha^{-1} - \rho}{\alpha^{-1} + \rho}.$$

Theoretical comparisons

Proposition (see [Briceno-Arias, Pustelnik, 2021] for detailed references)
In the context of Problem p.16, suppose that f **is $\rho-$strongly convex**, for some $\rho \in \left]0, \alpha^{-1}\right[$, and let $\tau > 0$. Then, the following holds:

1. FBS Suppose that $\tau \in]0, 2\beta]$. Then $\operatorname{prox}_{\tau f}(\mathrm{Id} - \tau\nabla g)$ is $r_{T_2}(\tau)-$Lipschitz continuous, where $\mathrm{r}_{T_2}(\tau) := \dfrac{1}{1+\tau\rho} \in]0, 1[$ .In particular, the minimum in (1) is achieved at

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

Theoretical comparisons

Proposition (see [Briceno-Arias, Pustelnik, 2021] for detailed references)
In the context of Problem p.16, suppose that f **is $\rho-$strongly convex**, for some $\rho \in \left]0, \alpha^{-1}\right[$, and let $\tau > 0$. Then, the following holds:

1. 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

$$r_R(\tau) = \max\left\{\frac{1-\tau\rho}{1+\tau\rho}, \frac{\tau\alpha^{-1}-1}{\tau\alpha^{-1}+1}\right\} \in]0,1[\,. \tag{6}$$

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

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

Theoretical comparisons

Proposition (see [Briceno-Arias, Pustelnik, 2021] for detailed references)
In the context of Problem p.16, suppose that f **is $\rho-$strongly convex**, for some $\rho \in \left]0, \alpha^{-1}\right[$, and let $\tau > 0$. Then, the following holds:

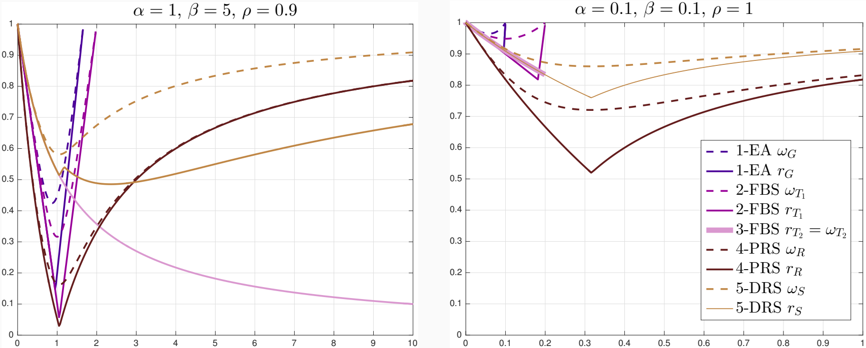
1. DRS $S_{\tau\nabla g, \tau\nabla f}$ and $S_{\tau\nabla f, \tau\nabla g}$ are $r_S(\tau)-$Lipschitz continuous, where

$$r_S(\tau) = \min\left\{\frac{1 + r_R(\tau)}{2}, \frac{\beta + \tau^2\rho}{\beta + \tau\beta\rho + \tau^2\rho}\right\} \in \left]0, 1\right[\tag{6}$$

and r_R is defined in p.16. In particular, the optimal step-size and the minimum in (1) are

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

Theoretical comparisons



Comparison of the convergence rates of EA, FBS, PRS, DRS for two choices of α, β, and ρ. Note that optimization rates are better than cocoercive rates in general.

Example: Smooth TV 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, $\alpha = 1$, and $\beta = \frac{\mu}{\chi\|L\|^2}$.

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, $\alpha = \frac{\mu}{\mu+\chi\|L_{\mathbb{I}_2}\|^2}$, and $\beta = \frac{\mu}{\chi\|L_{\mathbb{I}_1}\|^2}$.

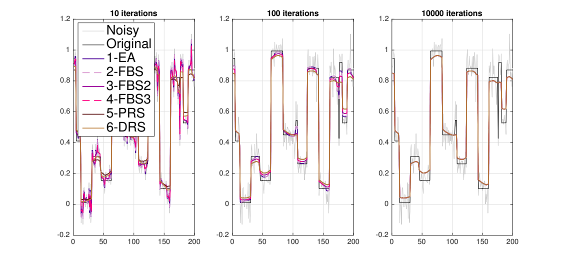
3- **FBS 2:** Use $T_{\tau\nabla\widetilde{g}, \tau\nabla\widetilde{f}}$

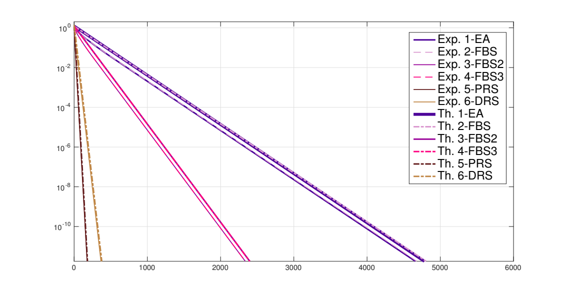
4- **FBS 3:** Use $T_{\tau\nabla\widetilde{f}, \tau\nabla\widetilde{g}}$

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

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

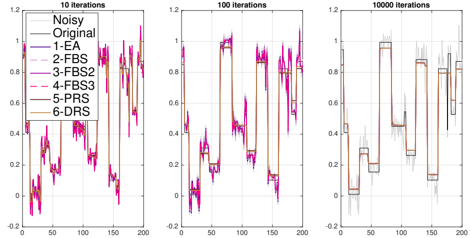
Numerical and theoretical comparisons

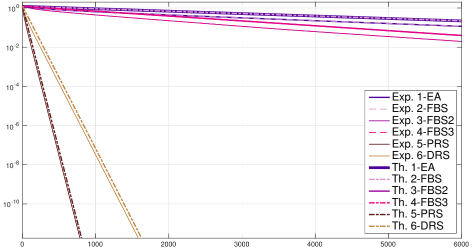




Piecewise constant denoising estimates after 10, 100, and 10000 iterations with $\chi = 0.7$ and $\mu = 0.002$ when considering gradient descent, FBS, PRS or DRS. Associated theoretical and numerical convergence rates.

Numerical and theoretical comparisons





Piecewise constant denoising estimates after 10, 100, and 10000 iterations with $\chi = 0.7$ and $\mu = 0.0001$ when considering gradient descent, FBS, PRS or DRS. Associated theoretical and numerical convergence rates.

Prox versus grad: conclusion

- Douglas-Rachford and Peaceman-Rachford better theoretical and numerical rates for piecewise constant denoising.

- Convergence rate should involve strong convexity constant but also regularization parameter and Lipschitz constant in order to integrate the different parameters having impact on signal and image processing.

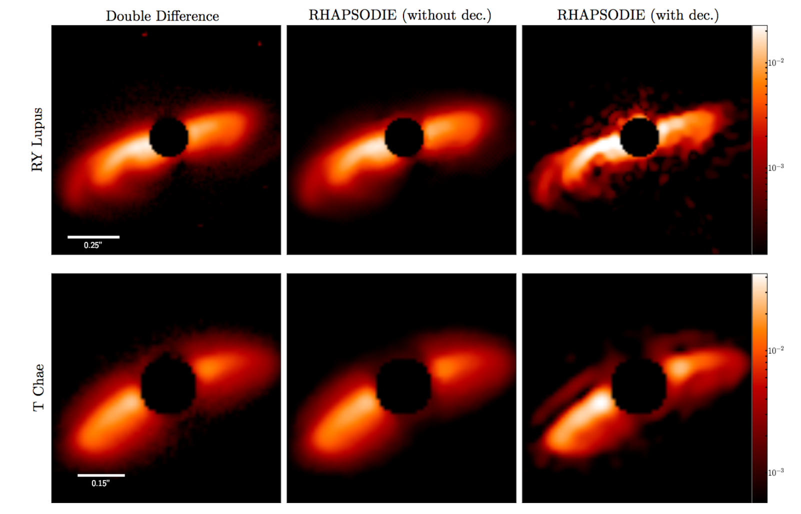
Prox versus grad: conclusion

- Signal and image processing problems with strongly convex objective functions exist. Possibility to change the constant of strong convexity when considering texture segmentation.

- Many situations where prox does not have a closed form expression:
  - Even for $\|Ax - z\|_2^2$ if $(A^*A + \mathrm{I})$ not easily invertible. In practice $A = A_1 A_2 \dots A_K$.
  - For data-term such as DKL or ℓ_1-norm

- Are the conclusions stays the same for non-strongly convex problems ?

Example: Inverse problems



→ extracted from L. Denneulin, M. Langlois, E. Thiebaut, and N. Pustelnik RHAPSODIE: Reconstruction of High-contrAst Polarized SOurces and Deconvolution for cIrcumstellar Environments, accepted to A&A, 2021.

References

- B. Pascal, N. Pustelnik, and P. Abry, Strongly Convex Optimization for Joint Fractal Feature Estimation and Texture Segmentation, Applied and Computational Harmonic Analysis, vol. 54, pp 303-322, 2021

- L. M. Briceño-Arias and N. Pustelnik, Proximal or gradient steps for cocoercive operators, submitted, 2020.

- P. Abry, N. Pustelnik, S. Roux, P. Jensen, P. Flandrin, R. Gribonval, C.-G. Lucas, E. Guichard, P. Borgnat, N. Garnier, B. Audit, Spatial and temporal regularization to estimate COVID-19 Reproduction Number R(t): Promoting piecewise smoothness via convex optimization, PLoS One, 15(8), Aug. 2020.

- L. Denneulin, M. Langlois, E. Thiebaut, and N. Pustelnik, RHAPSODIE : Reconstruction of High-contrAst Polarized SOurces and Deconvolution for cIrcumstellar Environments, accepted to A&A, 2021.

- J. Colas, N. Pustelnik, C. Oliver, P. Abry, J.-C. Geminard, V. Vidal, Nonlinear denoising for solid friction dynamics characterization, Physical Review E, 100, 032803, Sept. 2019.

- N. Pustelnik, A. Benazza-Benhayia, Y. Zheng, J.-C. Pesquet, Wavelet-based Image Deconvolution and Reconstruction, Wiley Encyclopedia of Electrical and Electronics Engineering, DOI: 10.1002/047134608X.W8294, Feb. 2016.