Semi-Linearized Proximal Alternating Minimization for a Discrete Mumford-Shah Model

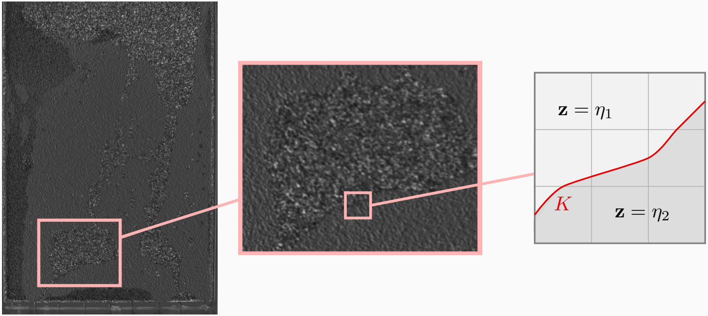
Nelly Pustelnik

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

Context

Motivation

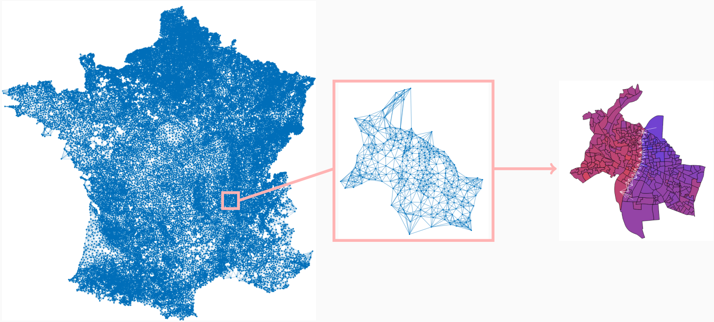
Defi Imag'IN CNRS SIROCCO (2017-2018): Multiphasic flow experiment modeling gas and liquid in a porous medium.



- Goal: classify gas/liquid + accurate estimation of the perimeter.
- Datasize: Image composed with 2.10^7 pixels. Analysis to be performed on a sequence of images.

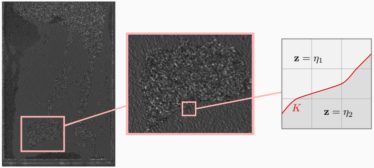
Motivation

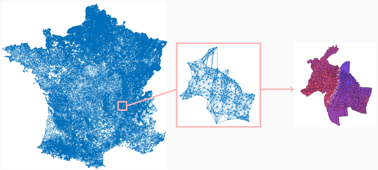
Defi Infinity CNRS OptMatGeo (2018): vote transfer matrix estimation between two elections



- Goal: clustering the areas with similar transfer matrix (i.e. similar electoral behaviour) + sharp transitions.
- Datasize: $\sim 2.10^7$

Motivation





- Structured data (e.g. images, graphs).
- Difficulty to label data.
- For Physics or societal applications, the truth is not as sharp as clustering but it is closer from regression (smooth behaviour) and possibly sharp transition.
- Huge amount of data.

Goal: Design algorithmic solutions with convergence guarantees to extract piecewise smooth behaviour.

Collaborations

Marion Foare
CPE and ENS de Lyon

Laurent Condat
CNRS, Gipsa-lab, France
& KAUST, Arabie Saoudite

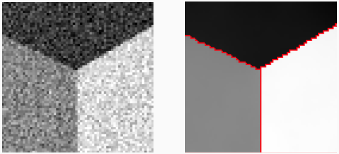
Mumford-Shah

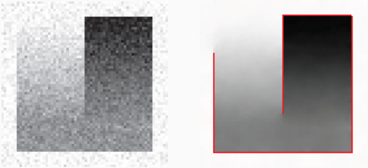
Mumford-Shah (1989)

$$\underset{\mathbf{u},K}{\text{minimize}} \; \underbrace{\frac{1}{2}\int_{\Omega}(\mathbf{u}-\mathbf{z})^2 dxdy}_{\text{fidelity}} + \beta \underbrace{\int_{\Omega\setminus K}|\nabla \mathbf{u}|^2 dxdy}_{\text{smoothness}} + \underbrace{\lambda\mathcal{H}^1(K\cap\Omega)}_{\text{length}}$$

- Ω: image domain,
- $\mathbf{z} \in L^\infty(\Omega)$: data,
- $\boxed{\mathbf{u}} \in W^{1,2}(\Omega)$: piecewise smooth approximation of $\mathbf{z}$,

  $W^{1,2}(\Omega) = \left\{u \in L^2(\Omega)\ \partial u \in L^2(\Omega)\right\}$ where ∂ weak derivative operator
- $\boxed{K}$: set of discontinuities,
- $\mathcal{H}^1$: Hausdorff measure.





Discrete Mumford-Shah like models

$$\underset{\mathbf{u},K}{\text{minimize}}\ \frac{1}{2}\int_\Omega (\mathbf{u}-\mathbf{z})^2 dxdy + \beta \int_{\Omega\setminus K} |\nabla \mathbf{u}|^2 dxdy + \lambda \mathcal{H}^1(K\cap\Omega)$$

- Potts model (1952):
  [Rudin et al., 1992] [Cai, Steidl, 2013] [Storath, Weinmann, 2014]
  $$\underset{\mathbf{u}}{\text{minimize}}\ \frac{1}{2}\|\mathbf{u}-\mathbf{z}\|_2^2 + \gamma\|D\mathbf{u}\|_0$$
- Blake-Zisserman problem (1987):
  [Strekalovskiy, Cremers, 2014] [Hohm et al., 2015]
  $$\underset{\mathbf{u}}{\text{minimize}}\ \frac{1}{2}\|\mathbf{u}-\mathbf{z}\|_2^2 + \gamma\sum_i \min(|(D\mathbf{u})_i|^p, \alpha^p)$$
- Ambrosio-Tortorelli (1990)
  [Foare et al., 2016]
  $$\underset{\mathbf{u},\mathbf{e}}{\text{minimize}}\ \frac{1}{2}\|\mathbf{u}-\mathbf{z}\|_2^2 + \beta\|(1-\mathbf{e})\odot D\mathbf{u}\|^2 + \lambda\left(\varepsilon\|\tilde{D}\mathbf{e}\|_2^2 + \frac{1}{4\varepsilon}\|\mathbf{e}\|_2^2\right)$$

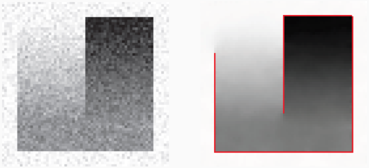
Discrete Mumford-Shah like models

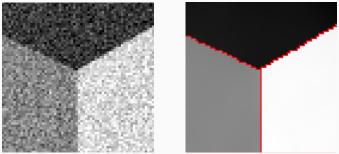
$$\underset{\mathbf{u},K}{\text{minimize}}\ \frac{1}{2}\int_\Omega (\mathbf{u}-\mathbf{z})^2 dxdy + \beta\int_{\Omega\setminus K} |\nabla \mathbf{u}|^2 dxdy + \lambda \mathcal{H}^1(K\cap\Omega)$$

- Potts model (1952): $\Rightarrow$ TV denoising (ROF) (1992)
  [Rudin et al., 1992] [Cai, Steidl, 2013] [Storath, Weinmann, 2014]
  $$\underset{\mathbf{u}}{\text{minimize}}\ \frac{1}{2}\|\mathbf{u}-\mathbf{z}\|_2^2 + \gamma \|D\mathbf{u}\|_1$$
- Blake-Zisserman problem (1987):
  [Strekalovskiy, Cremers, 2014] [Hohm et al., 2015]
  $$\underset{\mathbf{u}}{\text{minimize}}\ \frac{1}{2}\|\mathbf{u}-\mathbf{z}\|_2^2 + \gamma \sum_i \min(|(D\mathbf{u})_i|^p, \alpha^p)$$
- Ambrosio-Tortorelli (1990)
  [Foare et al., 2016]
  $$\underset{\mathbf{u},\mathbf{e}}{\text{minimize}}\ \frac{1}{2}\|\mathbf{u}-\mathbf{z}\|_2^2 + \beta\|(1-\mathbf{e})\odot D\mathbf{u}\|^2 + \lambda\Big(\varepsilon\|\tilde{D}\mathbf{e}\|_2^2 + \frac{1}{4\varepsilon}\|\mathbf{e}\|_2^2\Big)$$

Mumford-Shah like models: summary

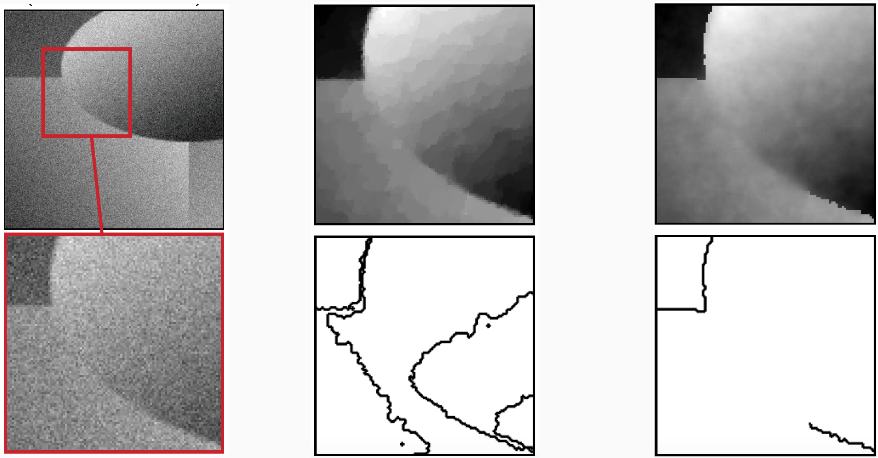
	Potts	Blake-Zisserman	Ambrosio-Tortorelli
Smooth estimate	X	V	V
Data-term flexibility	X	X	X
Convergence	V	X	X
Large scale dataset	V	V	X
Open contours	X	V	V





⇒ Revisit Ambrosio-Tortorelli model in order to provide a large-scale flexible convergent discrete MS like model.

Mumford-Shah versus ROF



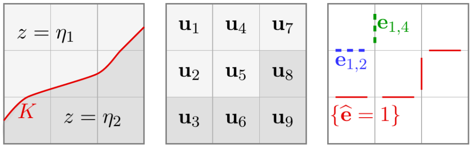
Noisy data | Estimation and contour detection obtained with ROF [Cai et al. 2019] | Estimation and contour detection obtained with the proposed MS model

D-MS for images

Proposed Discrete Mumford-Shah like (D-MS) model

$$\underset{\mathbf{u},\mathbf{e}}{\text{minimize}}\ \Psi(\mathbf{u},\mathbf{e}) := \mathcal{L}(\mathbf{u};\mathbf{z}) + \beta\|(1-\mathbf{e})\odot D\mathbf{u}\|^2 + \lambda\mathcal{R}(\mathbf{e})$$

- $\Omega = \{1,\ldots,N_1\}\times\{1,\ldots,N_2\}$;
- $\mathbf{z}\in\mathbb{R}^{|\tilde{\Omega}|}$: input data = image/graph (e.g. $\mathbf{z} = A\bar{\mathbf{u}} + \epsilon$);
- $\mathbf{u}\in\mathbb{R}^{|\Omega|}$: piecewise smooth approximation of $\mathbf{z}$;
- $\mathcal{L}$: data fidelity term convex, l.s.c., proper;
- $D\in\mathbb{R}^{|\mathbb{E}|\times|\Omega|}$: models a finite difference operator;
- $\mathbf{e}\in\mathbb{R}^{|\mathbb{E}|}$: edges between nodes whose value is 1 when a contour change is detected and 0 otherwise;
- $\mathcal{R}$: favors sparse solution (i.e. "short $|K|$"), convex, l.s.c., proper.



Proposed Discrete Mumford-Shah like (D-MS) model

$$\underset{\mathbf{u},\mathbf{e}}{\text{minimize}}\ \Psi(\mathbf{u},\mathbf{e}) := \mathcal{L}(\mathbf{u};\mathbf{z}) + \beta\|(1-\mathbf{e})\odot D\mathbf{u}\|^2 + \lambda\mathcal{R}(\mathbf{e})$$

- $\Omega = \{1,\ldots,N_1\}\times\{1,\ldots,N_2\}$;
- $\mathbf{z}\in\mathbb{R}^{|\tilde{\Omega}|}$: input data = image/graph (e.g. $\mathbf{z} = A\bar{\mathbf{u}} + \epsilon$);
- $\mathbf{u}\in\mathbb{R}^{|\Omega|}$: piecewise smooth approximation of $\mathbf{z}$;
- $\mathcal{L}$: data fidelity term convex, l.s.c., proper;
- $D\in\mathbb{R}^{|\mathbb{E}|\times|\Omega|}$: models a finite difference operator;
- $\mathbf{e}\in\mathbb{R}^{|\mathbb{E}|}$: edges between nodes whose value is 1 when a contour change is detected and 0 otherwise;
- $\mathcal{R}$: favors sparse solution (i.e. "short $|K|$"), convex, l.s.c., proper.

$\Rightarrow$ Identify the assumptions on $\mathcal{L}$ and $\mathcal{R}$ to design an algorithmic scheme with convergence guarantees (confidence in the solution).

D-MS: Gauss-Seidel iterations

$$\underset{\mathbf{u},\mathbf{e}}{\text{minimize}}\ \Psi(\mathbf{u},\mathbf{e}) := \mathcal{L}(\mathbf{u};\mathbf{z}) + \underbrace{\beta\|(1-\mathbf{e})\odot D\mathbf{u}\|^2}_{\mathcal{S}(D\mathbf{u},\mathbf{e})} + \lambda\mathcal{R}(\mathbf{e})$$

- Gauss-Seidel scheme = coordinate descent

  Set $\mathbf{e}^{[0]} \in \mathbb{R}^{|\mathbb{E}|}$.

  For $k \in \mathbb{N}$

  $\mathbf{u}^{[k+1]} \in \operatorname{Arg}\min_{\mathbf{u}} \Psi(\mathbf{u}, \mathbf{e}^{[k]})$

  $\mathbf{e}^{[k+1]} \in \operatorname{Arg}\min_{\mathbf{e}} \Psi(\mathbf{u}^{[k+1]}, \mathbf{e})$

- Under technical assumptions, convergence of the sequence $(\mathbf{u}^{[k]}, \mathbf{e}^{[k]})_{\ell\in\mathbb{N}}$ to a critical point $(\mathbf{u}^*, \mathbf{e}^*)$ of Ψ.

  Technical assumptions = minimum is attained at each iteration, e.g. by assuming strict convexity w.r.t one argument. (Auslender1976, Bertsekas1999)

D-MS: Gauss-Seidel iterations

$$\underset{\mathbf{u},\mathbf{e}}{\text{minimize}}\ \Psi(\mathbf{u},\mathbf{e}) := \mathcal{L}(\mathbf{u};\mathbf{z}) + \underbrace{\beta\|(1-\mathbf{e})\odot D\mathbf{u}\|^2}_{\mathcal{S}(D\mathbf{u},\mathbf{e})} + \lambda\mathcal{R}(\mathbf{e})$$

- PAM [Attouch et al. 2010]

  Set $\mathbf{e}^{[0]} \in \mathbb{R}^{|\mathbb{E}|}$ and $\mathbf{u}^{[0]} \in \mathbb{R}^{|\Omega|}$.

  For $k \in \mathbb{N}$

  $$\left\lfloor \begin{array}{l} \mathbf{u}^{[k+1]} = \arg\min_{\mathbf{u}} \Psi(\mathbf{u},\mathbf{e}^{[k]}) + \frac{c_k}{2}\|\mathbf{u}-\mathbf{u}^{[k]}\|^2 \\ \mathbf{e}^{[k+1]} = \arg\min_{\mathbf{e}} \Psi(\mathbf{u}^{[k+1]},\mathbf{e}) + \frac{d_k}{2}\|\mathbf{e}-\mathbf{e}^{[k]}\|^2 \end{array} \right.$$

- Under technical assumptions, the sequence $(\mathbf{u}^{[k]},\mathbf{e}^{[k]})_{\ell\in\mathbb{N}}$ converges to a critical point $(\mathbf{u}^*,\mathbf{e}^*)$ of Ψ.

  Technical assumptions = closed form of the proximity operators.

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

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

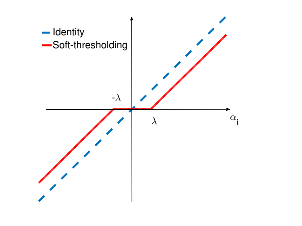
Proximity operator

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 $\lambda > 0$.



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 $\operatorname{prox}_\varphi x$ such that

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

Examples: closed form expression

- $\operatorname{prox}_{\lambda\|\cdot\|_1}$: soft-thresholding with a fixed threshold $\lambda > 0$.
- $\operatorname{prox}_{\|\cdot\|_{1,2}}$ [Peyré,Fadili,2011].
- $\operatorname{prox}_{\|\|_p^p}$ with $p = \{\frac{4}{3}, \frac{3}{2}, 2, 3, 4\}$ [Chaux et al.,2005].
- $\operatorname{prox}_{D_{KL}}$ [Combettes,Pesquet,2007].

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 $\lambda > 0$.
- $\mathrm{prox}_{\|\cdot\|_{1,2}}$[Peyré,Fadili,2011].
- $\mathrm{prox}_{\|\|_p^p}$ 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]

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 $\lambda > 0$.
- $\mathrm{prox}_{\|\cdot\|_{1,2}}$[Peyré,Fadili,2011].
- $\mathrm{prox}_{\|\|_p^p}$ 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]

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 $\lambda > 0$.
- $\mathrm{prox}_{\|\cdot\|_{1,2}}$[Peyré,Fadili,2011].
- $\mathrm{prox}_{\|\|_p^p}$ 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: http://proximity-operator.net

D-MS: PAM iterations

$$\underset{\mathbf{u},\mathbf{e}}{\text{minimize}}\ \Psi(\mathbf{u},\mathbf{e}) := \mathcal{L}(\mathbf{u};\mathbf{z}) + \beta \underbrace{\|(1-\mathbf{e})\odot D\mathbf{u}\|^2}_{\mathcal{S}(D\mathbf{u},\mathbf{e})} + \lambda\mathcal{R}(\mathbf{e})$$

- PAM [Attouch et al. 2010]

  Set $\mathbf{e}^{[0]} \in \mathbb{R}^{|\mathbb{E}|}$ and $\mathbf{u}^{[0]} \in \mathbb{R}^{|\Omega|}$.

  For $k \in \mathbb{N}$

  $$\left\lfloor \begin{array}{l} \mathbf{u}^{[k+1]} = \operatorname{arg\,min}_{\mathbf{u}} \Psi(\mathbf{u},\mathbf{e}^{[k]}) + \frac{c_k}{2}\|\mathbf{u}-\mathbf{u}^{[k]}\|^2 \\ \mathbf{e}^{[k+1]} = \operatorname{arg\,min}_{\mathbf{e}} \Psi(\mathbf{u}^{[k+1]},\mathbf{e}) + \frac{d_k}{2}\|\mathbf{e}-\mathbf{e}^{[k]}\|^2 \end{array} \right.$$

- Under technical assumptions, the sequence $(\mathbf{u}^{[k]},\mathbf{e}^{[k]})_{\ell\in\mathbb{N}}$ converges to a critical point $(\mathbf{u}^*,\mathbf{e}^*)$ of Ψ.

D-MS: PAM iterations

$$\underset{\mathbf{u},\mathbf{e}}{\text{minimize}}\ \Psi(\mathbf{u},\mathbf{e}) := \mathcal{L}(\mathbf{u};\mathbf{z}) + \beta\underbrace{\|(1-\mathbf{e})\odot D\mathbf{u}\|^2}_{\mathcal{S}(D\mathbf{u},\mathbf{e})} + \lambda\mathcal{R}(\mathbf{e})$$

- PAM [Attouch et al. 2010]

  Set $\mathbf{e}^{[0]} \in \mathbb{R}^{|\mathbb{E}|}$ and $\mathbf{u}^{[0]} \in \mathbb{R}^{|\Omega|}$.

  For $k \in \mathbb{N}$

  $$\left\lfloor \begin{array}{l} \mathbf{u}^{[k+1]} = \mathrm{prox}_{\frac{1}{c_k}\Psi(\cdot,\mathbf{e}^{[k]})}(\mathbf{u}^{[k]}) \\ \mathbf{e}^{[k+1]} = \arg\min_{\mathbf{e}} \Psi(\mathbf{u}^{[k+1]},\mathbf{e}) + \frac{d_k}{2}\|\mathbf{e}-\mathbf{e}^{[k]}\|^2 \end{array} \right.$$

- Under technical assumptions, the sequence $(\mathbf{u}^{[k]},\mathbf{e}^{[k]})_{\ell\in\mathbb{N}}$ converges to a critical point $(\mathbf{u}^*,\mathbf{e}^*)$ of Ψ.

D-MS: PAM iterations

$$\underset{\mathbf{u},\mathbf{e}}{\text{minimize}}\ \Psi(\mathbf{u},\mathbf{e}) := \mathcal{L}(\mathbf{u};\mathbf{z}) + \beta\underbrace{\|(1-\mathbf{e})\odot D\mathbf{u}\|^2}_{\mathcal{S}(D\mathbf{u},\mathbf{e})} + \lambda\mathcal{R}(\mathbf{e})$$

- PAM [Attouch et al. 2010]

  Set $\mathbf{e}^{[0]} \in \mathbb{R}^{|\mathbb{E}|}$ and $\mathbf{u}^{[0]} \in \mathbb{R}^{|\Omega|}$.

  For $k \in \mathbb{N}$

  $$\left\lfloor \begin{array}{l} \mathbf{u}^{[k+1]} = \mathrm{prox}_{\frac{1}{c_k}\mathcal{L}(\cdot;\mathbf{z}) + \frac{\beta}{c_k}\mathcal{S}(D\cdot,\mathbf{e})}(\mathbf{u}^{[k]}) \\ \mathbf{e}^{[k+1]} = \arg\min_{\mathbf{e}} \Psi(\mathbf{u}^{[k+1]},\mathbf{e}) + \frac{d_k}{2}\|\mathbf{e}-\mathbf{e}^{[k]}\|^2 \end{array} \right.$$

- Under technical assumptions, the sequence $(\mathbf{u}^{[k]},\mathbf{e}^{[k]})_{\ell\in\mathbb{N}}$ converges to a critical point $(\mathbf{u}^*,\mathbf{e}^*)$ of Ψ.

D-MS: PAM iterations

$$\underset{\mathbf{u},\mathbf{e}}{\text{minimize}}\ \Psi(\mathbf{u},\mathbf{e}) := \mathcal{L}(\mathbf{u};\mathbf{z}) + \beta \underbrace{\|(1-\mathbf{e}) \odot D\mathbf{u}\|^2}_{\mathcal{S}(D\mathbf{u},\mathbf{e})} + \lambda\mathcal{R}(\mathbf{e})$$

- PAM [Attouch et al. 2010]

  Set $\mathbf{e}^{[0]} \in \mathbb{R}^{|\mathbb{E}|}$ and $\mathbf{u}^{[0]} \in \mathbb{R}^{|\Omega|}$.

  For $k \in \mathbb{N}$

  $$\left\lfloor \begin{array}{l} \mathbf{u}^{[k+1]} = \operatorname{prox}_{\frac{1}{c_k}\mathcal{L}(\cdot;\mathbf{z}) + \frac{\beta}{c_k}\mathcal{S}(D\cdot,\mathbf{e})}(\mathbf{u}^{[k]}) \\ \mathbf{e}^{[k+1]} = \operatorname{prox}_{\frac{1}{d_k}\beta\mathcal{S}(D\mathbf{u}^{[k+1]},\cdot) + \frac{\lambda}{d_k}\mathcal{R}}(\mathbf{e}^{[k]}) \end{array} \right.$$

- Under technical assumptions, the sequence $(\mathbf{u}^{[k]},\mathbf{e}^{[k]})_{\ell\in\mathbb{N}}$ converges to a critical point $(\mathbf{u}^*,\mathbf{e}^*)$ of Ψ.

  Difficulty = proximity operator of a sum of two functions.

Proximity operator of a sum of two functions

$$\mathrm{prox}_{\varphi_1+\varphi_2} = \mathrm{prox}_{\varphi_2} \circ \mathrm{prox}_{\varphi_1}?$$

- [Combettes-Pesquet, 2007] $N = 1$, $\varphi_2 = \iota_C$ of a non-empty closed convex subset of C and φ_1 is differentiable at 0 with $h'(0) = 0$.
- [Chaux-Pesquet-Pustelnik,2009] C and φ_2 are separable in the same basis.
- [Yu, 2013][Shi et al., 2017] $\partial\varphi_2(x) \subset \partial\varphi_2(\mathrm{prox}\varphi_1(x))$.
- Other recent results [Pustelnik, Condat, 2017][Yukawa, Kagami, 2017][del Aguila Pla, Jaldén, 2017]

D-MS: PALM iterations

$$\underset{\mathbf{u},\mathbf{e}}{\text{minimize}}\ \Psi(\mathbf{u},\mathbf{e}) := \mathcal{L}(\mathbf{u};\mathbf{z}) + \underbrace{\beta\|(1-\mathbf{e})\odot D\mathbf{u}\|^2}_{\mathcal{S}(D\mathbf{u},\mathbf{e})} + \lambda\mathcal{R}(\mathbf{e})$$

- PALM [Bolte, Sabach, Teboulle, 2013]
  Set $\mathbf{u}^{[0]} \in \mathbb{R}^{|\Omega|}$ and $\mathbf{e}^{[0]} \in \mathbb{R}^{|\mathbb{E}|}$.
  For $k \in \mathbb{N}$
  - Set $\gamma > 1$ and $c_k = \gamma\chi(\mathbf{e}^{[k]})$
  - $\mathbf{u}^{[k+1]} = \mathrm{prox}_{\frac{1}{c_k}\mathcal{L}(\cdot;\mathbf{z})}\Big(\mathbf{u}^{[k]} - \frac{1}{c_k}\nabla_{\mathbf{u}}\mathcal{S}\big(D\mathbf{u}^{[k]},\mathbf{e}^{[k]}\big)\Big)$
  - Set $\delta > 1$ and $d_k = \delta\nu(\mathbf{u}^{[k+1]})$
  - $\mathbf{e}^{[k+1]} = \mathrm{prox}_{\frac{1}{d_k}\lambda\mathcal{R}}\Big(\mathbf{e}^{[k]} - \frac{1}{d_k}\nabla_{\mathbf{e}}\mathcal{S}\big(D\mathbf{u}^{[k+1]},\mathbf{e}^{[k]}\big)\Big)$
- Under technical assumptions, the sequence $(\mathbf{u}^{[k]},\mathbf{e}^{[k]})_{\ell\in\mathbb{N}}$ converges to a critical point $(\mathbf{u}^*,\mathbf{e}^*)$ of Ψ.

Proposed D-MS and algorithmic solution

$$\underset{\mathbf{u},\mathbf{e}}{\text{minimize}}\ \Psi(\mathbf{u},\mathbf{e}) := \mathcal{L}(\mathbf{u};\mathbf{z}) + \underbrace{\beta\|(1-\mathbf{e})\odot D\mathbf{u}\|^2}_{\mathcal{S}(D\mathbf{u},\mathbf{e})} + \lambda\mathcal{R}(\mathbf{e}),$$

- Proposed Semi Linearized PAM (SL-PAM)

  [Foare, Pustelnik, Condat, 2017]

  Set $\mathbf{u}^{[0]} \in \mathbb{R}^{|\Omega|}$ and $\mathbf{e}^{[0]} \in \mathbb{R}^{|\mathbb{E}|}$.

  For $\ell \in \mathbb{N}$

  Set $\gamma > 1$ and $c_k = \gamma\chi(\mathbf{e}^{[k]})$.

  $$\mathbf{u}^{[k+1]} = \operatorname{prox}_{\frac{1}{c_k}\mathcal{L}(\cdot;\mathbf{z})}\left(\mathbf{u}^{[k]} - \frac{1}{c_k}\nabla_{\mathbf{u}}\mathcal{S}\left(D\mathbf{u}^{[k]},\mathbf{e}^{[k]}\right)\right)$$

  Set $d_k > 0$.

  $$\mathbf{e}^{[k+1]} = \operatorname{prox}_{\frac{1}{d_k}\lambda\mathcal{R}+\mathcal{S}(D\mathbf{u}^{[k+1]},\cdot)}\left(\mathbf{e}^{[k]}\right)$$

Proposed D-MS and algorithmic solution

Proposition [Foare, Pustelnik, Condat, 2017]

The sequence $(\mathbf{u}^{[k]}, \mathbf{e}^{[k]})_{k\in\mathbb{N}}$ generated by SL-PAM converges to a critical point of Ψ if

1. the updating steps of $\mathbf{u}^{[k+1]}$ and $\mathbf{e}^{[k+1]}$ have closed form expressions;
2. the sequence $(\mathbf{u}^{[k]}, \mathbf{e}^{[k]})_{k\in\mathbb{N}}$ generated by SL-PAM is bounded;
3. $\mathcal{L}(A\cdot, \mathbf{z})$, $\mathcal{R}$ and $\Psi(\cdot,\cdot)$ are bounded below;
4. Ψ is a Kurdyka-Łojasiewicz function;
5. $\nabla_{\mathbf{u}}$ and $\nabla_{\mathbf{e}}$ are globally Lipschitz continuous with moduli $\nu(\mathbf{e})$ and $\varepsilon(\mathbf{u})$ respectively, and for all $k \in \mathbb{N}$, $\nu(\mathbf{e}^{[k]})$ and $\varepsilon(\mathbf{u}^{[k]})$ are bounded by positive constants.

Proposed D-MS and algorithmic solution

$$\underset{\mathbf{u},\mathbf{e}}{\text{minimize}}\ \Psi(\mathbf{u},\mathbf{e}) := \mathcal{L}(\mathbf{u};\mathbf{z}) + \underbrace{\beta\|(1-\mathbf{e})\odot D\mathbf{u}\|^2}_{\mathcal{S}(D\mathbf{u},\mathbf{e})} + \lambda\mathcal{R}(\mathbf{e}),$$

- Proposed Semi Linearized PAM (SL-PAM)
  [Foare, Pustelnik, Condat, 2017]

  Set $\mathbf{u}^{[0]} \in \mathbb{R}^{|\Omega|}$ and $\mathbf{e}^{[0]} \in \mathbb{R}^{|\mathbb{E}|}$.

  For $\ell \in \mathbb{N}$

  Set $\gamma > 1$ and $c_k = \gamma\chi(\mathbf{e}^{[k]})$.

  $$\mathbf{u}^{[k+1]} = \operatorname{prox}_{\frac{1}{c_k}\mathcal{L}(\cdot;\mathbf{z})}\left(\mathbf{u}^{[k]} - \frac{1}{c_k}\nabla_{\mathbf{u}}\mathcal{S}\left(D\mathbf{u}^{[k]},\mathbf{e}^{[k]}\right)\right)$$

  Set $d_k > 0$.

  $$\mathbf{e}^{[k+1]} = \operatorname{prox}_{\frac{1}{d_k}\lambda\mathcal{R}+\mathcal{S}(D\mathbf{u}^{[k+1]},\cdot)}\left(\mathbf{e}^{[k]}\right)$$

- The sequence $(\mathbf{u}^{[k]},\mathbf{e}^{[k]})_{\ell\in\mathbb{N}}$ converges to a critical point $(\mathbf{u}^*,\mathbf{e}^*)$ of Ψ

  $\Rightarrow$ Difficulty: Computation $\operatorname{prox}_{\frac{1}{d_k}\lambda\mathcal{R}+\mathcal{S}(D\mathbf{u}^{[k+1]},\cdot)}$

Proposed D-MS and algorithmic solution

Proposition [Foare, Pustelnik, Condat, 2017]
We assume that $\mathcal{S}$ is separable, i.e,

$$(\forall \mathbf{e} = (\mathbf{e}_i)_{1 \leq i \leq |\mathbb{E}|}) \qquad \mathcal{R}(\mathbf{e}) = \sum_{i=1}^{|\mathbb{E}|} \sigma_i(\mathbf{e}_i),$$

where $\sigma_i : \mathbb{R}^{|\mathbb{E}|} \to]-\infty; +\infty]$ with a closed form proximity operator expression.

Let $\mathrm{d}_k > 0$, then

$$\operatorname{prox}_{\frac{1}{d_k}\lambda\mathcal{R}+\mathcal{S}(D\mathbf{u}^{[k+1]},\cdot)}(\mathbf{e}^{[k]}) = \left(\operatorname{prox}_{\frac{\lambda\sigma_i}{2\beta(D\mathbf{u}^{[k]})_i^2+d_k}} \left(\frac{\beta(D\mathbf{u}^{[k+1]})_i^2 + \frac{d_k \mathbf{e}_i^{[k]}}{2}}{\beta(D\mathbf{u}^{[k+1]})_i^2 + \frac{d_k}{2}} \right) \right)_{i \in \mathbb{E}}$$

Proposed D-MS and algorithmic solution

$$\underset{\mathbf{u},\mathbf{e}}{\text{minimize}}\ \mathcal{L}(\mathbf{u};\mathbf{z}) + \beta\|(1-\mathbf{e})\odot D\mathbf{u}\|^2 + \lambda\mathcal{R}(\mathbf{e})$$

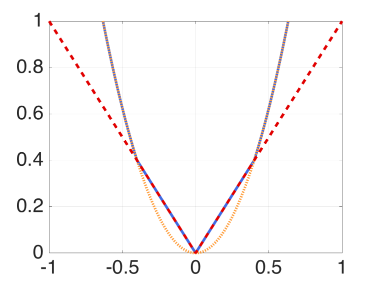
- $\mathcal{R}$: favors sparse solution (i.e. "short $|K|$") and convex.

1. Ambrosio-Tortorelli approximation:

$$\mathcal{R}(\mathbf{e}) = \varepsilon\|D\mathbf{e}\|_2^2 + \frac{1}{4\varepsilon}\|\mathbf{e}\|_2^2 \text{ with } \varepsilon > 0$$

2. ℓ_1-norm: $\mathcal{R}(\mathbf{e}) = \|\mathbf{e}\|_1$

3. Quadratic ℓ_1:
[Foare, Pustelnik, Condat, 2017]
$\mathcal{R}(\mathbf{e}) = \sum_{i=1}^{|\mathbb{E}|} \max\left\{|e_i|, \frac{e_i^2}{4\varepsilon}\right\}.$

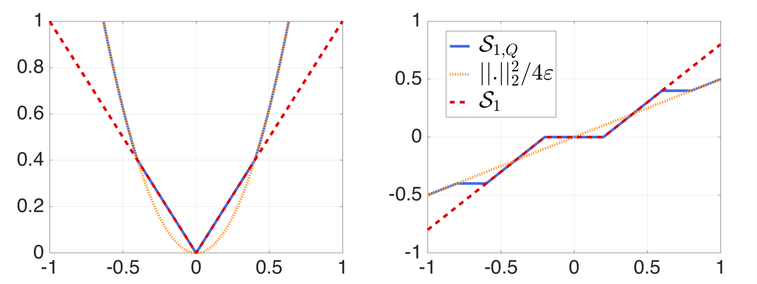


Proposed D-MS and algorithmic solution

Proposition [Foare,Pustelnik,Condat, 2017]

For every $\eta \in \mathbb{R}$ and $\tau, \epsilon > 0$

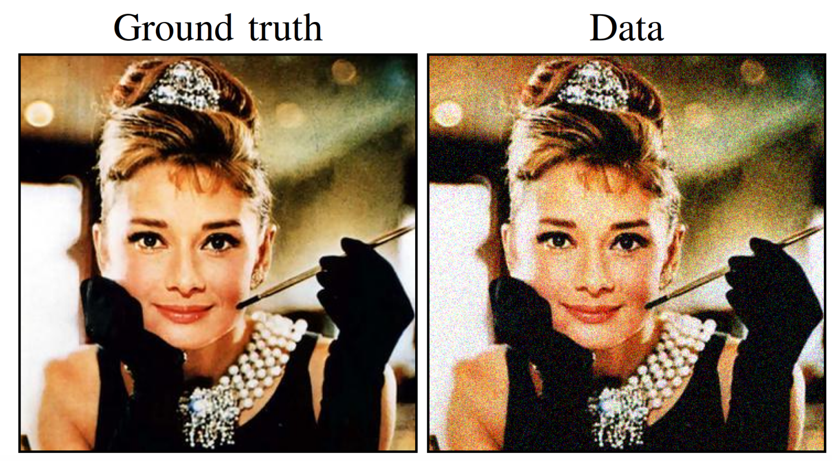
$$\mathrm{prox}_{\tau \max\{|.|, \frac{|.|^2}{4\epsilon}\}}(\eta) = \mathrm{sign}(\eta)\ \max\left\{0, \min\left[|\eta| - \tau, \max\left(4\epsilon, \frac{|\eta|}{\frac{\tau}{2\epsilon} + 1}\right)\right]\right\}$$



Experiments

1. Visual results (proposed versus state-of-the-art) when $\mathcal{L}(\mathbf{u};\mathbf{z}) = \frac{1}{2}\|\mathbf{u}-\mathbf{z}\|_2^2$
2. Quantitative results.
3. Convergence comparisons (proposed versus PALM).
4. Sensitivity to the initialization.
5. Visual results when $\mathcal{L}(\mathbf{u};\mathbf{z}) = \frac{1}{2}\|A\mathbf{u}-\mathbf{z}\|_2^2$.

Experiments

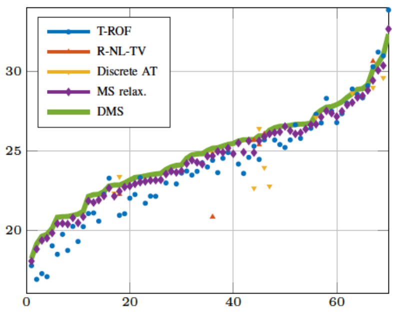


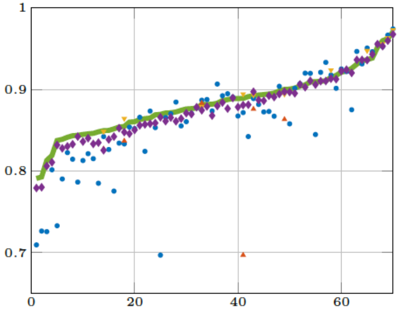
Experiments 1

TV	(Strekalovskiy, Cremers, 2014)	Discret AT (Foare et al., 2016)	Quadratic-ℓ_1

SNR = 24.67 dB	SNR = 22.42 dB	SNR = 23.43 dB	**SNR = 23.75 dB**
SSIM = 0.944	SSIM = 0.855	SSIM = 0.867	**SSIM = 0.877**
Time = 29.25 s	**Time = 1.92s**	Time = 1992s	Time = 57.84s

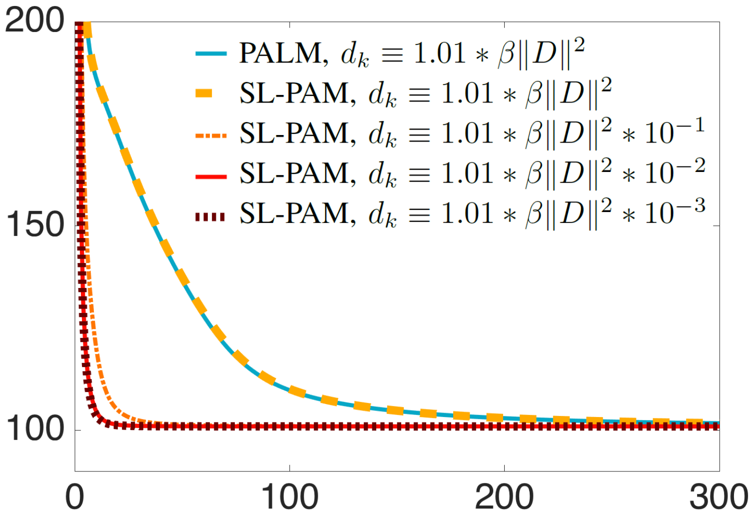
Experiments 2

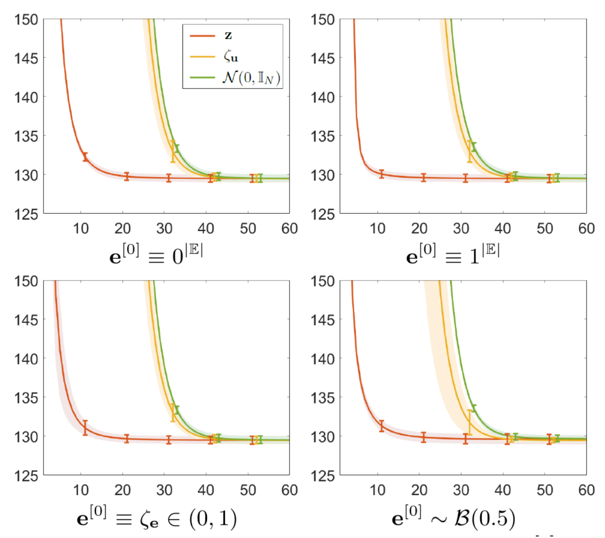




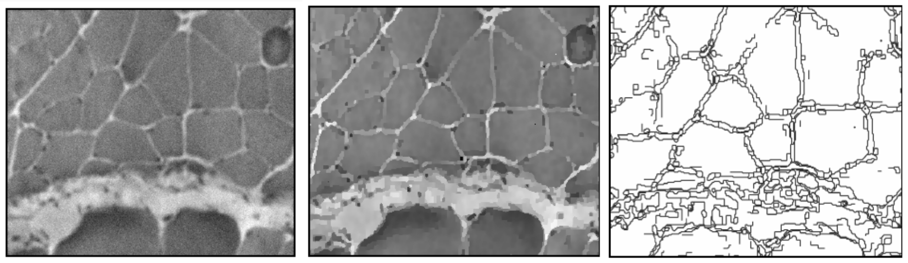
Experiments 3

Convergence PALM versus SL-PALM: $\Psi(\mathbf{u}^{[\ell]}, \mathbf{e}^{[\ell]})$ w.r.t. iterations ℓ




150
145
140
135
130
125
10 20 30 40 50 60
$\mathbf{z}$
$\zeta_{\mathbf{u}}$
$\mathcal{N}(0, \mathbb{I}_N)$
$\mathbf{e}^{[0]} \equiv 0^{|\mathbb{E}|}$
$\mathbf{e}^{[0]} \equiv 1^{|\mathbb{E}|}$
$\mathbf{e}^{[0]} \equiv \zeta_{\mathbf{e}} \in (0, 1)$
$\mathbf{e}^{[0]} \sim \mathcal{B}(0.5)$

Experiments 5



D-MS for images: conclusions

- Efficient (convergence guarantees and time) algorithm for solving a generic D-MS model.
- Flexibility in the data-term: possibility to handle many image degradation to improve interface detection (Poisson noise, blur,...).
- Flexibility in the choice of the discrete difference operator.
- Clear benefit of joint image restoration and interface detection (cf. comparison with Thresholded-TV).

D-MS for images: conclusions

- Efficient (convergence guarantees and time) algorithm for solving a generic D-MS model. $\Rightarrow$ bi-convexity and other properties of the objective function could be considered to have stronger convergence guarantees.
- Flexibility in the data-term: possibility to handle many image degradation to improve interface detection (Poisson noise, blur,...).
- Flexibility in the choice of the discrete difference operator.
- Clear benefit of joint image restoration and interface detection (cf. comparison with Thresholded-TV).

D-MS for images: conclusions

- Efficient (convergence guarantees and time) algorithm for solving a generic D-MS model.
- Flexibility in the data-term: possibility to handle many image degradation to improve interface detection (Poisson noise, blur,...). $\Rightarrow$ Strong assumptions on the degration operator that we would like to relax.
- Flexibility in the choice of the discrete difference operator.
- Clear benefit of joint image restoration and interface detection (cf. comparison with Thresholded-TV).

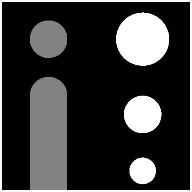
D-MS for images: conclusions

- Efficient (convergence guarantees and time) algorithm for solving a generic D-MS model.
- Flexibility in the data-term: possibility to handle many image degradation to improve interface detection (Poisson noise, blur,...).
- Flexibility in the choice of the discrete difference operator. $\Rightarrow$ Discretization scheme to improve interface detection (choice of D). Convergence to the true MS model (cf. [Belz, Bredies;2019]) .
- Clear benefit of joint image restoration and interface detection (cf. comparison with Thresholded-TV).

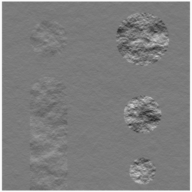
D-MS for images: conclusions

- Efficient (convergence guarantees and time) algorithm for solving a generic D-MS model.
- Flexibility in the data-term: possibility to handle many image degradation to improve interface detection (Poisson noise, blur,...).
- Flexibility in the choice of the discrete difference operator.
- Clear benefit of joint image restoration and interface detection (cf. comparison with Thresholded-TV). $\Rightarrow$ More generally, better to perform one-step procedure rather than two-steps.

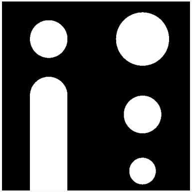
Two-step versus one-step texture segmentation



Mask

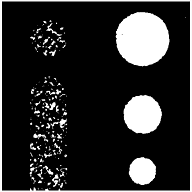


Synthetic texture

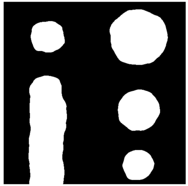


Optimal solution

T-ROF [Cai2013]
[Cai2013]



Matrix factorization
[Yuan2015]



Proposed
[Pascal2019]

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

References

• M. Foare, N. Pustelnik, and L. Condat, Semi-linearized proximal alternating minimization for a discrete Mumford-Shah model, accepted to **IEEE Trans. on Image Processing**, 2019.

• Y. Kaloga, M. Foare, N. Pustelnik, and P. Jensen, Discrete Mumford-Shah on graph for mixing matrix estimation, accepted to **IEEE Signal Processing Letters**, 2019.

• B. Pascal, N. Pustelnik, and P. Abry, Nonsmooth convex joint estimation of local regularity and local variance for fractal texture segmentation, submitted, 2019.

KL

Definition 3 (*Kurdyka–Łojasiewicz property*) Let $\sigma : \mathbb{R}^d \to (-\infty, +\infty]$ be proper and lower semicontinuous.

(i) The function σ is said to have the *Kurdyka–Łojasiewicz (KL) property* at $\overline{u} \in \operatorname{dom} \partial\sigma := \left\{u \in \mathbb{R}^d : \partial\sigma(u) \neq \emptyset\right\}$ if there exist $\eta \in (0, +\infty]$, a neighborhood U of $\overline{u}$ and a function $\varphi \in \Phi_\eta$, such that for all

$$u \in U \cap [\sigma(\overline{u}) < \sigma(u) < \sigma(\overline{u}) + \eta],$$

the following inequality holds

$$\varphi'(\sigma(u) - \sigma(\overline{u}))\, \operatorname{dist}(0, \partial\sigma(u)) \geq 1. \tag{2.4}$$

(ii) If σ satisfy the KL property at each point of $\operatorname{dom} \partial\sigma$ then σ is called a *KL function*.

Convergence

(i) *Sufficient decrease property:* Find a positive constant ρ_1 such that

$$\rho_1 \left\| z^{k+1} - z^k \right\|^2 \leq \Psi(z^k) - \Psi(z^{k+1}), \quad \forall k = 0, 1, \ldots.$$

(ii) *A subgradient lower bound for the iterates gap:* Assume that the sequence generated by the algorithm $\mathcal{A}$ is bounded.[1] Find another positive constant ρ_2, such that

$$\left\| w^{k+1} \right\| \leq \rho_2 \left\| z^{k+1} - z^k \right\|, \quad w^k \in \partial \Psi \left(z^k \right), \quad \forall k = 0, 1, \ldots.$$

(iii) *Using the KL property:* Assume that Ψ is a KL function and show that the generated sequence $\left\{ z^k \right\}_{k \in \mathbb{N}}$ is a *Cauchy sequence*.