Block-coordinate proximal algorithms for scale-free texture segmentation†

B. Pascal[1], N. Pustelnik[1], P. Abry[1], J.-C. Pesquet[2]

April, 28th 2018

[1] Univ Lyon, ENS de Lyon, Univ Claude Bernard Lyon 1, CNRS,
Laboratoire de Physique, F-69342 Lyon, France, firstname.lastname@ens-lyon.fr

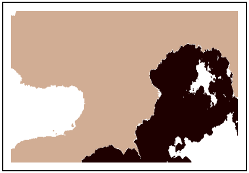
[2] Center for Visual Computing, INRIA, CentraleSupélec, Univ. Paris-Saclay,
9 rue Joliot Curie, 91190 Gif sur Yvette, France, jean-christophe.pesquet@centralesupelec.fr

† Supported by ANR-16-CE33-0020 MultiFracs, France.

Texture segmentation

Segmentation task

k-means

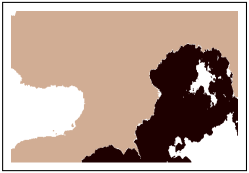


Piecewise constant image

Texture segmentation

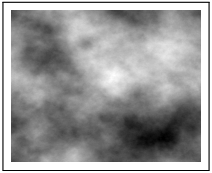
Segmentation task

k-means

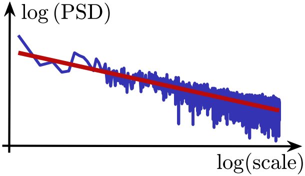


Piecewise constant image

Monofractal scale invariant texture



Slope: fractal parameter *h* [Abry1995]

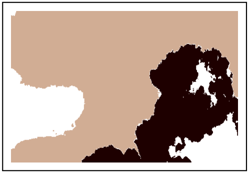


High resolution necessary

Texture segmentation

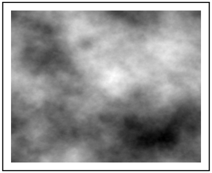
Segmentation task

k-means

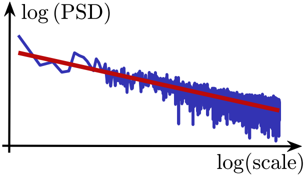


Piecewise constant image

Monofractal scale invariant texture



Slope: fractal parameter h [Abry1995]



High resolution necessary

I) Detect constant h areas

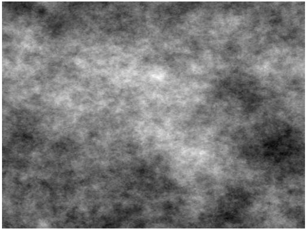
Estimation of local h

II) Effective implementation

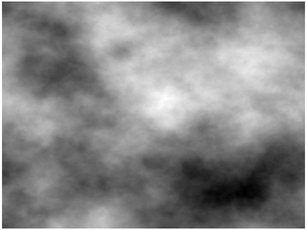
Block-coordinate algorithm

Monofractal textures

Synthetic texture with constant local regularity



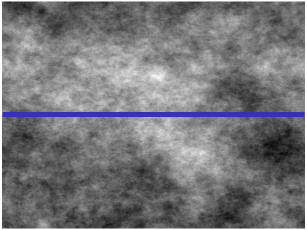
$h = 0.3$



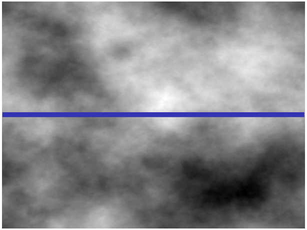
$h = 0.9$

Monofractal textures

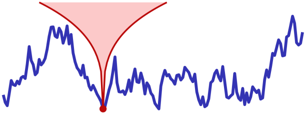
Synthetic texture with constant local regularity

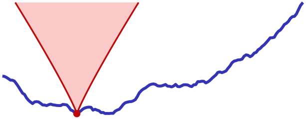


$h = 0.3$



$h = 0.9$





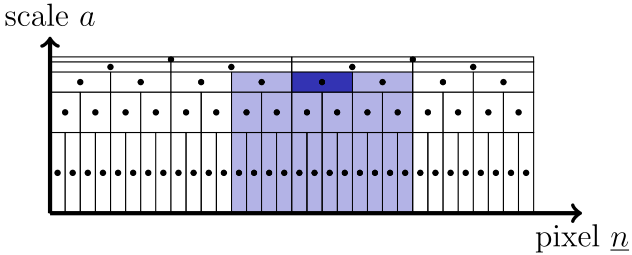
IDEA: fit local behavior with power law functions

$$|f(x) - f(y)| \leq C|x - y|^{h(x)}, \quad h(x) \equiv 0.3\,(\text{left}), \quad 0.9\,(\text{right})$$

Wavelet transform and leader coefficients

(i) **DWT** of image X: $w_{a,\underline{n}}(X)$ at scale a and pixel $\underline{n}$,

(ii) **Local surpremum** of $|w_{a,\underline{n}}(X)|$: $\mathcal{L}_{a,\underline{n}}(X)$ (*leaders*)



Wavelet transform and leader coefficients

(i) **DWT** of image X: $w_{a,\underline{n}}(X)$ at scale a and pixel $\underline{n}$,

(ii) **Local surpremum** of $|w_{a,\underline{n}}(X)|$: $\mathcal{L}_{a,\underline{n}}(X)$ (*leaders*)

Linear regression $\hat{h}(\underline{n})$ [Wendt2009]

$$\boxed{\log\left(\mathcal{L}_{a,\underline{n}}\right) \simeq \log \eta(\underline{n}) + h(\underline{n}) \log(a)}$$

Multiscale analysis

Estimation of local regularity

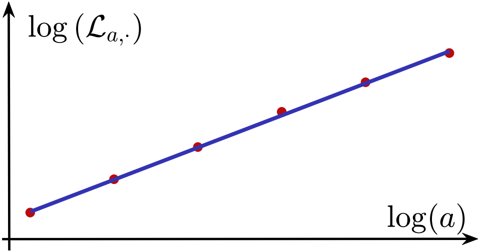
Wavelet transform and leader coefficients

(i) **DWT** of image X: $w_{a,\underline{n}}(X)$ at scale a and pixel $\underline{n}$,

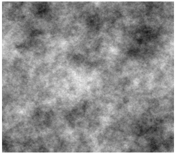
(ii) **Local surpremum** of $|w_{a,\underline{n}}(X)|$: $\mathcal{L}_{a,\underline{n}}(X)$ (*leaders*)

Linear regression $\hat{h}(\underline{n})$ [Wendt2009]

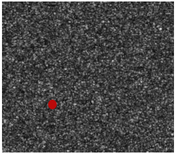
$$\log\left(\mathcal{L}_{a,\underline{n}}\right) \simeq \log \eta(\underline{n}) + h(\underline{n}) \log(a)$$



Texture X

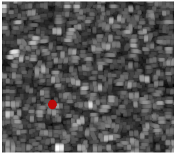


Scale $a = 2^1$



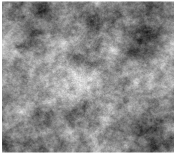
⋮

Scale $a = 2^6$



Multiscale analysis
Estimation of local regularity

Texture X

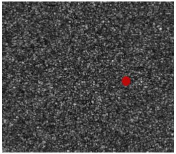


Wavelet transform and leader coefficients

(i) **DWT** of image X: $w_{a,\underline{n}}(X)$ at scale a and pixel $\underline{n}$,

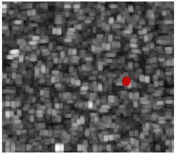
(ii) **Local surpremum** of $|w_{a,\underline{n}}(X)|$: $\mathcal{L}_{a,\underline{n}}(X)$ (*leaders*)

Scale $a = 2^1$



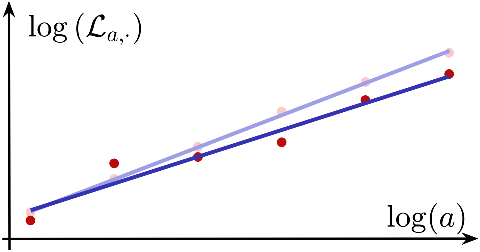
⋮

Scale $a = 2^6$



Linear regression $\hat{h}(\underline{n})$ [Wendt2009]

$$\log\left(\mathcal{L}_{a,\underline{n}}\right) \simeq \log \eta(\underline{n}) + h(\underline{n}) \log(a)$$



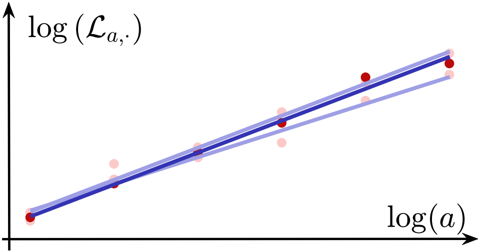
Wavelet transform and leader coefficients

(i) **DWT** of image X: $w_{a,\underline{n}}(X)$ at scale a and pixel $\underline{n}$,

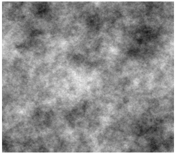
(ii) **Local surpremum** of $|w_{a,\underline{n}}(X)|$: $\mathcal{L}_{a,\underline{n}}(X)$ (*leaders*)

Linear regression $\hat{h}(\underline{n})$ [Wendt2009]

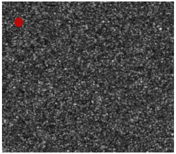
$$\log\left(\mathcal{L}_{a,\underline{n}}\right) \simeq \log \eta(\underline{n}) + h(\underline{n}) \log(a)$$



Texture X

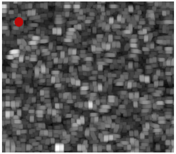


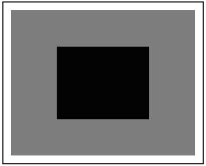
Scale $a = 2^1$



⋮

Scale $a = 2^6$

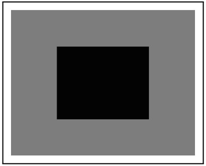




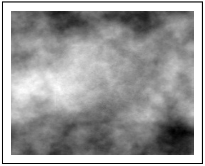
Piecewise constant h

Piecewise monofractal textures

Synthetic data

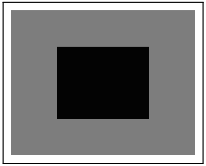


Piecewise constant h

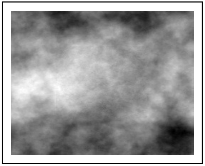


Texture sample X

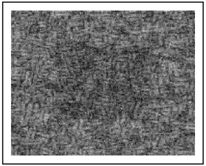
Piecewise monofractal textures
Synthetic data



Piecewise constant h



Texture sample X

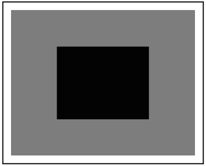


Linear fit $\widehat{h}$

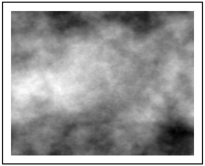
Linear fit is not satisfactory !

Piecewise monofractal textures

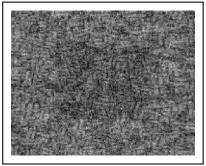
Synthetic data



Piecewise constant h



Texture sample X



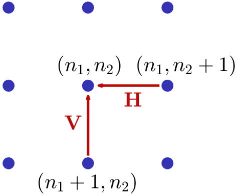
Linear fit $\widehat{h}$

Linear fit is not satisfactory !

Objective function enforcing piecewise constancy [Pustelnik2016]

$$\widehat{\widehat{h}} \in \underset{h}{\text{Argmin}} \quad \underbrace{\mathbf{DF}(h, \mathcal{L}(X))}_{\text{Data Fidelity}} + \underbrace{\lambda \mathbf{TV}(h)}_{\text{Total Variation}}$$

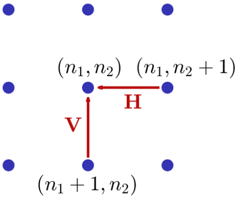
aim: enforce piecewise behavior of estimate



Discrete difference operator

$$(\mathbf{D}h)_{n_1,n_2} = \frac{1}{2}\begin{pmatrix} h_{n_1,n_2+1} - h_{n_1,n_2} \\ h_{n_1+1,n_2} - h_{n_1,n_2} \end{pmatrix}$$

aim: enforce piecewise behavior of estimate

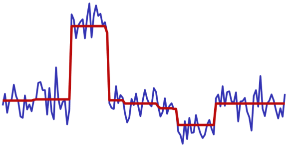


Discrete difference operator

$$(\mathbf{D}h)_{n_1,n_2} = \frac{1}{2}\begin{pmatrix} h_{n_1,n_2+1} - h_{n_1,n_2} \\ h_{n_1+1,n_2} - h_{n_1,n_2} \end{pmatrix}$$

Total variation penalization

$$\|\mathbf{D}h\|_{2,1} = \sum_{n_1=1}^{N-1}\sum_{n_2=1}^{N-1}\sqrt{(\mathbf{H}h)^2_{n_1,n_2} + (\mathbf{V}h)^2_{n_1,n_2}}$$



Total Variation estimation

Convex optimization

$$\widehat{\widehat{h}} \in \operatorname*{Argmin}_{h} \underbrace{\frac{1}{2}\|h - \widehat{h}\|_2^2}_{\mathbf{DF}(h,\mathcal{L}(X))} + \lambda\|\mathbf{D}h\|_{2,1}$$

where $\widehat{h}$ is the linear fit estimate and $\lambda > 0$ the regularization parameter.

Total Variation estimation

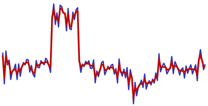
Convex optimization

$$\widehat{\widehat{h}} \in \operatorname*{Argmin}_{h} \underbrace{\frac{1}{2}\|h - \widehat{h}\|_2^2}_{\mathbf{DF}(h,\mathcal{L}(X))} + \lambda \|\mathbf{D}h\|_{2,1}$$

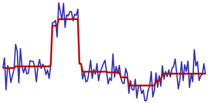
where $\widehat{h}$ is the linear fit estimate and $\lambda > 0$ the regularization parameter.

Difficulties:

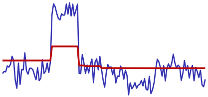
- non-smooth objective function: proximal algorithms
- high computational cost
- regularization parameter: fine tuning of λ



Too small

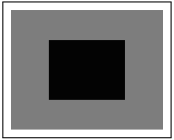


Optimal

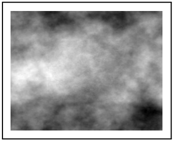


Too large

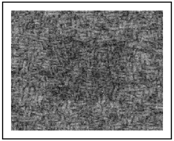
Results on synthetic data
Linear regression then TV regularization



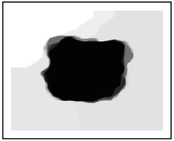
Mask



Texture sample

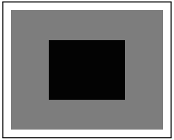


Linear fit $\widehat{h}$ ✗

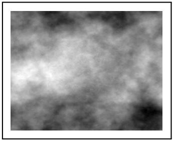


TV denoised $\widehat{\widehat{h}}$ ✓

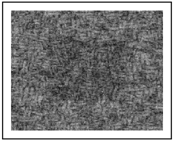
Results on synthetic data
Linear regression then TV regularization



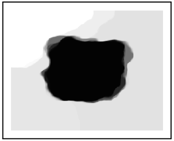
Mask



Texture sample



Linear fit $\widehat{h}$ ✗

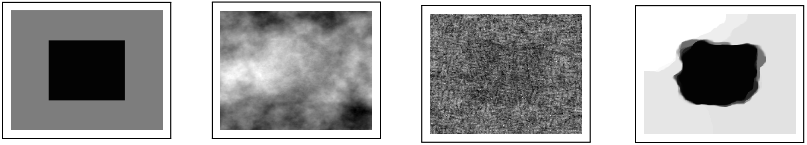


TV denoised $\widehat{\widehat{h}}$ ✓

Issues/Difficulties:

(i) fine tuning of the regularization parameter: $\lambda \in [\lambda_{\min}, \dots, \lambda_{\max}]$

Results on synthetic data
Linear regression then TV regularization



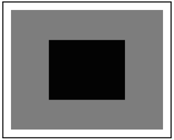
Mask — Texture sample — Linear fit $\widehat{h}$ ✗ — TV denoised $\widehat{\widehat{h}}$ ✓

Issues/Difficulties:

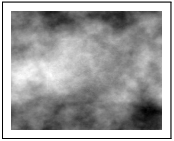
(i) fine tuning of the regularization parameter: $\lambda \in [\lambda_{\text{min}}, \ldots, \lambda_{\text{max}}]$

(ii) λ_{opt} (our problem) $\gg \lambda_{\text{opt}}$ (image denoising): large number of iterations

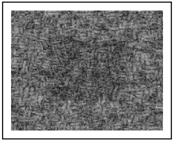
Results on synthetic data
Linear regression then TV regularization



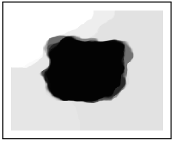
Mask



Texture sample



Linear fit $\widehat{h}$ ✗



TV denoised $\widehat{\widehat{h}}$ ✓

Issues/Difficulties:

(i) fine tuning of the regularization parameter: $\lambda \in [\lambda_{\min}, \dots, \lambda_{\max}]$

(ii) λ_{opt} (our problem) $\gg$ λ_{opt} (image denoising): large number of iterations

(iii) computational cost (time & memory):

Image size	256×256	512×512	1024×1024
Computational time (per λ)	3 min	16 min	86 min

Over-relaxation: FISTA **[Dossal2014]**
Fast **I**terative **S**hrinkage **T**hresholding **A**lgorithm
Acceleration ✓ Memory ✗

Alternating methods: ADMM **[Chambolle2015]**
Proximal **A**lternating **D**escent
Acceleration ✓ Approximation of **TV** ✗

Block-coordinate approaches: (Random) block selection

- Stochastic gradient *(machine learning)* **[Le Roux2012]**
- Primal and/or dual splitting *(image processing)* **[Repetti2015, Feriel2017, Chambolle2017]**

Proximal algorithms

TV denoising

$$\widehat{\widehat{h}} \in \operatorname*{Argmin}_{h} \frac{1}{2}\|h - \widehat{h}\|_2^2 + \lambda \|\mathbf{D}h\|_{2,1}$$

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

$$y^{[k+1]} = \operatorname{prox}_{\gamma \; \lambda\|.\|_{2,1}^*}\left(y^{[k]} + \gamma \; \mathbf{D} \; h^{[k]}\right)$$

$$h^{[k+1]} = h^{[k]} - \mathbf{D}^*\left(y^{\;[k+1]} - y^{[k]}\right)$$

end

Dual forward-backward algorithm [Feriel2017]

TV denoising

$$\widehat{\widehat{h}} \in \operatorname*{Argmin}_{h} \frac{1}{2}\|h - \widehat{h}\|_2^2 + \lambda\|\mathbf{D}h\|_{2,1}$$

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

$$y^{[k+1]} = \text{prox}_{\gamma\ \lambda\|.\|_{2,1}^*}\left(y^{[k]} + \gamma\ \mathbf{D}\ \ h^{[k]}\right)$$

$$h^{[k+1]} = h^{[k]} - \mathbf{D}^*\left(y^{\ [k+1]} - y^{[k]}\right)$$

end

Dual forward-backward algorithm [Feriel2017]

Proximal algorithms

TV denoising

$$\widehat{\widehat{h}} \in \operatorname*{Argmin}_{h} \frac{1}{2}\|h - \widehat{h}\|_2^2 + \lambda\|\mathbf{D}h\|_{2,1}$$

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

$$y^{[k+1]} = \text{prox}_{\gamma \ \lambda\|.\|_{2,1}^*}\left(y^{[k]} + \gamma \ \mathbf{D} \ h^{[k]}\right)$$

$$h^{[k+1]} = h^{[k]} - \mathbf{D}^*\left(y^{\ [k+1]} - y^{[k]}\right)$$

end

Dual forward-backward algorithm [Feriel2017]

Convergence of the iterates $\gamma \ < 2/\,\|\,\mathbf{D}\,\|^2$

(Global operator $\|\mathbf{D}\| = \sqrt{2}$)

Proximal algorithms

TV denoising

$$\widehat{\widehat{h}} \in \operatorname*{Argmin}_{h} \frac{1}{2}\|h - \widehat{h}\|_2^2 + \lambda \sum_{\ell=1}^{L} \|\mathbf{D}_\ell h\|_{2,1}$$

Block strategies

$$\|\mathbf{D}h\|_{2,1} = \sum_{\ell=1}^{L} \|\mathbf{D}_\ell h\|_{2,1}$$

$$\mathbf{D} = [\mathbf{D}_1, \ldots, \mathbf{D}_\ell, \ldots, \mathbf{D}_L]^\top$$

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

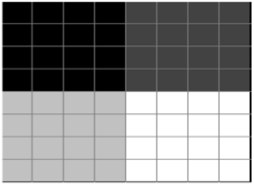
  for $\ell \in \{1, \ldots, L\}$ **do**

$$y_\ell^{[k+1]} = \operatorname{prox}_{\gamma_\ell \lambda \|\cdot\|_{2,1}^*} \left(y_\ell^{[k]} + \gamma_\ell \mathbf{D}_\ell h^{[k]} \right)$$

$$h^{[k+1]} = h^{[k]} - \mathbf{D}_\ell^* \left({y_\ell}^{[k+1]} - y_\ell^{[k]} \right)$$

  end

end



Block dual forward-backward algorithm [Feriel2017]

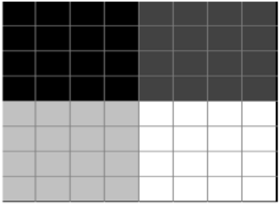
Convergence of the iterates

$$\gamma_\ell < 2/\,\|\,\mathbf{D}_\ell\|^2$$

Choice of the blocks

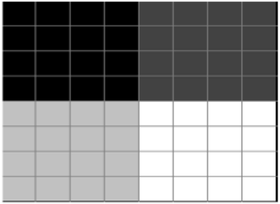
Regions and lattices

Hh or **V**h



Regions

H*h* or **V***h*



Regions

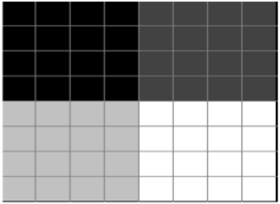
$$\forall \ell \in \{1, \ldots, 4\},$$
$$\|\mathbf{D}_\ell\| = \sqrt{2} \simeq 1.4142$$
$$= \|\mathbf{D}\|$$

no gain !

Choice of the blocks

Regions and lattices

H *h* or **V** *h*



Regions

$$\forall \ell \in \{1, \ldots, 4\},$$
$$\|\mathbf{D}_\ell\| = \sqrt{2} \simeq 1.4142$$
$$= \|\mathbf{D}\|$$

no gain !

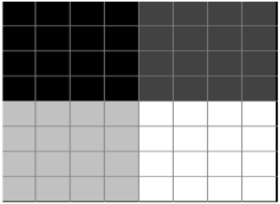
Descent steps

$$\gamma_\ell < 1$$

Choice of the blocks

Regions and lattices

Hh or **V**h



Regions

Hh or **V**h

Lattices

$$\forall \ell \in \{1, \ldots, 4\},$$
$$\|\mathbf{D}_\ell\| = \sqrt{2} \simeq 1.4142$$
$$= \|\mathbf{D}\|$$

no gain !

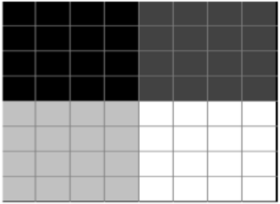
Descent steps

$$\gamma_\ell < 1$$

Choice of the blocks

Regions and lattices

Hh or **V**h



Regions

Hh or **V**h

Lattices

$\forall \ell \in \{1, \ldots, 4\}$,

$\|\mathbf{D}_\ell\| = \sqrt{2} \simeq 1.4142$

$= \|\mathbf{D}\|$

no gain !

$\forall \ell \in \{1, \ldots, 4\}$,

$\|\mathbf{D}_\ell\| = \sqrt{3}/2 \simeq 0.8660$

$< \|\mathbf{D}\|$

gain of factor $\simeq 1.6$

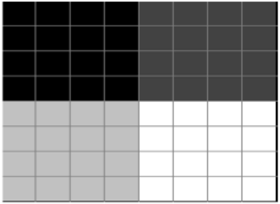
Descent steps

$$\gamma_\ell < 1$$

Choice of the blocks

Regions and lattices

$\mathbf{H}h$ or $\mathbf{V}h$



Regions

$\mathbf{H}h$ or $\mathbf{V}h$

Lattices

Regions:

$$\forall \ell \in \{1, \ldots, 4\},$$
$$\|\mathbf{D}_\ell\| = \sqrt{2} \simeq 1.4142$$
$$= \|\mathbf{D}\|$$

no gain !

Lattices:

$$\forall \ell \in \{1, \ldots, 4\},$$
$$\|\mathbf{D}_\ell\| = \sqrt{3}/2 \simeq 0.8660$$
$$< \|\mathbf{D}\|$$

gain of factor $\simeq 1.6$

Descent steps

$$\gamma_\ell < 1$$

$$\gamma_\ell < 8/3$$

Duality gap

$$\delta(h, y) = \underbrace{\frac{1}{2}\|h - \widehat{h}\|_2^2 + \lambda\|\mathbf{D}h\|_{2,1}}_{\text{primal functional}} + \underbrace{\frac{1}{2}\|-\mathbf{D}^* y + \widehat{h}\|_2^2 - \frac{1}{2}\|\widehat{h}\|_2^2 + \iota_{2,\infty(\lambda)}(y)}_{\text{dual functional}}$$

Duality gap

$$\delta(h,y) = \underbrace{\frac{1}{2}\|h-\widehat{h}\|_2^2 + \lambda\|\mathbf{D}h\|_{2,1}}_{\text{primal functional}} + \underbrace{\frac{1}{2}\|-\mathbf{D}^*y+\widehat{h}\|_2^2 - \frac{1}{2}\|\widehat{h}\|_2^2 + \iota_{2,\infty(\lambda)}(y)}_{\text{dual functional}}$$

Convergence

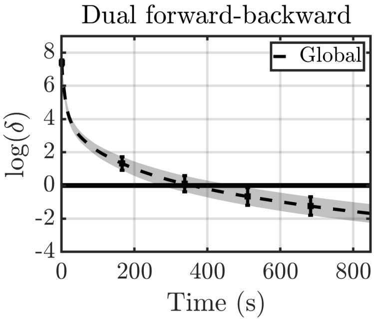
$$\log\delta(h^{[k]},y^{[k]}) \underset{k\to\infty}{\longrightarrow} -\infty$$

Duality gap

$$\delta(h, y) = \underbrace{\frac{1}{2}\|h - \widehat{h}\|_2^2 + \lambda\|\mathbf{D}h\|_{2,1}}_{\text{primal functional}} + \underbrace{\frac{1}{2}\|-\mathbf{D}^* y + \widehat{h}\|_2^2 - \frac{1}{2}\|\widehat{h}\|_2^2 + \iota_{2,\infty(\lambda)}(y)}_{\text{dual functional}}$$

Convergence

$$\log \delta(h^{[k]}, y^{[k]}) \underset{k\to\infty}{\longrightarrow} -\infty$$

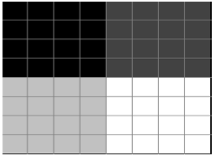


Duality gap

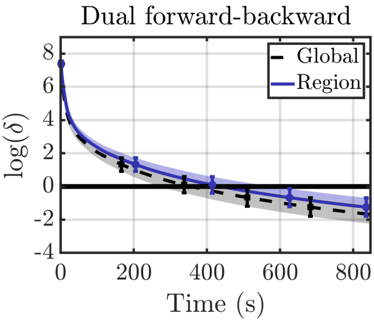
$$\delta(h, y) = \underbrace{\frac{1}{2}\|h - \widehat{h}\|_2^2 + \lambda\|\mathbf{D}h\|_{2,1}}_{\text{primal functional}} + \underbrace{\frac{1}{2}\|-\mathbf{D}^* y + \widehat{h}\|_2^2 - \frac{1}{2}\|\widehat{h}\|_2^2 + \iota_{2,\infty(\lambda)}(y)}_{\text{dual functional}}$$

Convergence

$$\log \delta(h^{[k]}, y^{[k]}) \underset{k\to\infty}{\longrightarrow} -\infty$$



Region



Forward-backward algorithms convergence

Region vs. lattice strategies

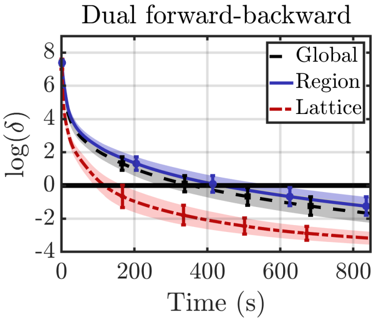
Duality gap

$$\delta(h,y) = \underbrace{\frac{1}{2}\|h-\widehat{h}\|_2^2 + \lambda\|\mathbf{D}h\|_{2,1}}_{\text{primal functional}} + \underbrace{\frac{1}{2}\|-\mathbf{D}^* y + \widehat{h}\|_2^2 - \frac{1}{2}\|\widehat{h}\|_2^2 + \iota_{2,\infty(\lambda)}(y)}_{\text{dual functional}}$$

Convergence

$$\log\delta(h^{[k]}, y^{[k]}) \underset{k\to\infty}{\longrightarrow} -\infty$$

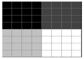
Lattice



Extension to primal-dual algorithms

Block forward-backward

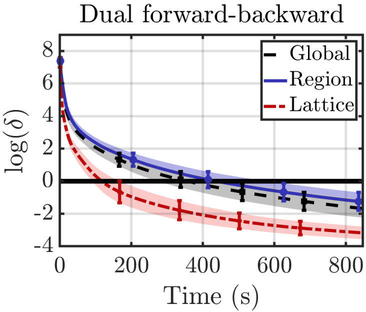
$$\gamma_\ell < \frac{1}{\|\mathbf{D}_\ell\|^2} \quad \text{[Feriel2017]}$$


✗

Region


✓

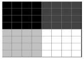
Lattice



Extension to primal-dual algorithms

Block forward-backward

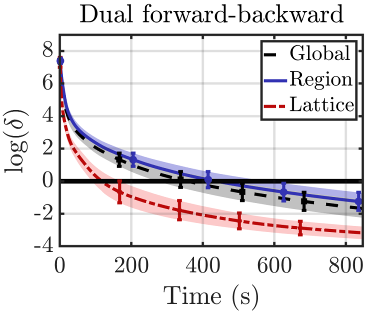
$$\gamma_\ell < \frac{1}{\|\mathbf{D}_\ell\|^2} \quad \text{[Feriel2017]}$$


✗

Region


✓

Lattice



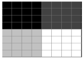
Block primal-dual

$$\sigma_\ell \tau < \frac{1}{L\|\mathbf{D}_\ell\|^2} \quad \text{[Chambolle2017]}$$

Extension to primal-dual algorithms

Block forward-backward

$$\gamma_\ell < \frac{1}{\|\mathbf{D}_\ell\|^2} \quad \text{[Feriel2017]}$$



✗

Region

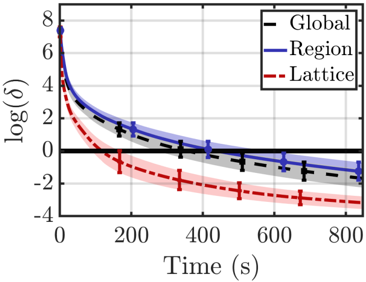
✓

Lattice

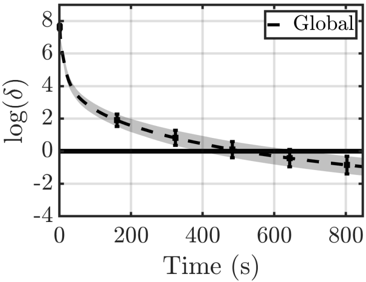
Block primal-dual

$$\sigma_\ell \tau < \frac{1}{L\|\mathbf{D}_\ell\|^2} \quad \text{[Chambolle2017]}$$

Dual forward-backward



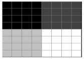
Primal dual



Extension to primal-dual algorithms

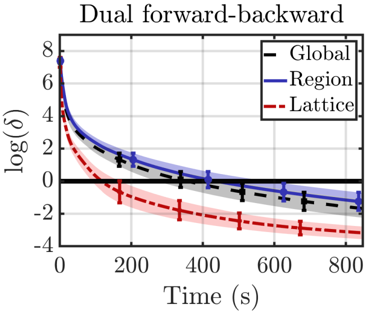
Block forward-backward

$$\gamma_\ell < \frac{1}{\|\mathbf{D}_\ell\|^2} \quad \text{[Feriel2017]}$$



Region ✗

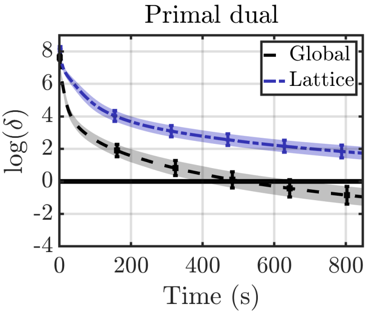
Lattice ✓



Block primal-dual

$$\sigma_\ell \tau < \frac{1}{L\|\mathbf{D}_\ell\|^2} \quad \text{[Chambolle2017]}$$

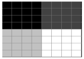
Lattice ✗



Extension to primal-dual algorithms

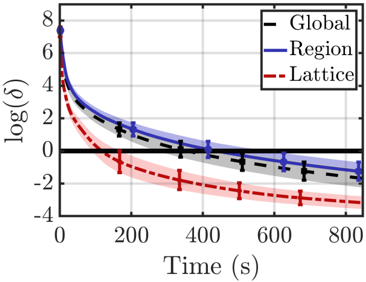
Block forward-backward

$$\gamma_\ell < \frac{1}{\|\mathbf{D}_\ell\|^2} \quad \text{[Feriel2017]}$$


✗
Region


✓
Lattice

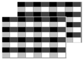
Dual forward-backward



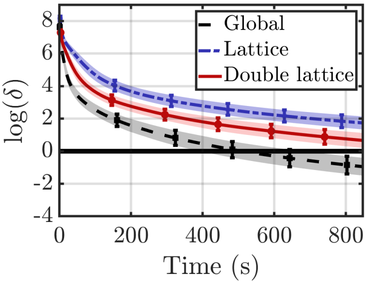
Block primal-dual

$$\sigma_\ell \tau < \frac{p_\ell}{\|\mathbf{D}_\ell\|^2} \quad \text{[Chambolle2017]}$$


✗
Lattice


✗
Double lattice

Primal dual

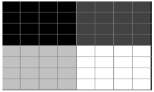


Take home message

Block strategies

Lower memory requirements ✓

[Repetti2015]



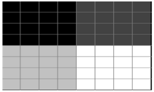
Larger computational time ✗

Take home message

Block strategies

Lower memory requirements ✓

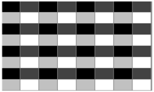
[Repetti2015]



Larger computational time ✗

Lattice splitting

Lower memory requirements ✓

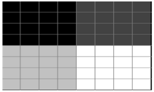


Reduced computational time ✓

Take home message

Block strategies

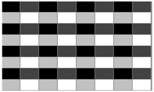
Lower memory requirements ✓
[Repetti2015]



Larger computational time ✗

Lattice splitting

Lower memory requirements ✓



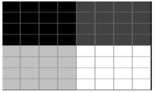
Reduced computational time ✓

Possibilities
- explore large range of regularization parameter λ,
- process high resolution images,
- analyze huge amount of data.

Take home message

Block strategies

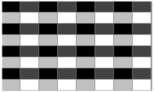
Lower memory requirements ✓ [Repetti2015]



Larger computational time ✗

Lattice splitting

Lower memory requirements ✓



Reduced computational time ✓

Possibilities
- explore large range of regularization parameter λ,
- process high resolution images,
- analyze huge amount of data.

Applications
- medical imaging (cancer detection, ...) [Marin2017],
- meteorology (clouds characterization, ...) [Arrault1997],
- art (painting authentication, ...) [Abry2013].