# Finite elements discretizations of the total variation

Antonin Chambolle

CMAP, Ecole Polytechnique, CNRS, Palaiseau, France

*joint with T. Pock (T.U. Graz, Austria)*

*Variational methods and optimization in imaging*,
IHP, Paris, 4 Feb., 2019

# Introduction

**Issue:** numerical approximation of the total variation. (Goal is to extend to more general similar functionals...)

Topic Mila was quite interested in: (an)isotropy? Quality of the approximation?

- ▶ Here: focus on Finite elements.
- ▶ Towards a better discretization of the gradients →
- ▶ Non-conforming P1 finite elements?
- ▶ Some good properties and
- ▶ Some issues
- ▶ A self-adaptive approximation in 2D

# Approximations of the total variation

Main difficulties:

- approximate sharp transitions on a discrete grid
- accurately compute the transition energy (perimeter)

# Examples

"Isotropic 2D T.V."

$$TV\,(u)\ =\ \sum_{i,j} \sqrt{(u_{i+1,j} - u_{i,j})^2 + (u_{i,j+1} - u_{i,j})^2}$$

*is* isotropic in the following sense:

# Examples

"Isotropic 2D T.V."

$$TV_h(u) = h \sum_{i,j} \sqrt{(u_{i+1,j} - u_{i,j})^2 + (u_{i,j+1} - u_{i,j})^2}$$

*is* isotropic in the following sense:

**Lemma** as $h \to 0$, $TV_h$ "Γ-converges" to the total variation $\int |Du|$

# Examples

"Isotropic 2D T.V."

$$TV(u) = \sum_{i,j} \sqrt{(u_{i+1,j} - u_{i,j})^2 + (u_{i,j+1} - u_{i,j})^2}$$
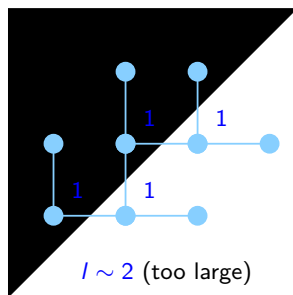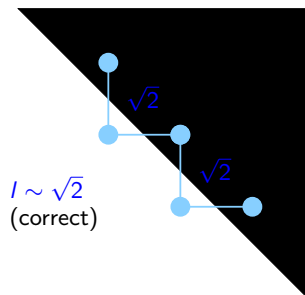
*is* isotropic in the following sense:

**Lemma** as $h \to 0$, $TV_h$ "Γ-converges" to the total variation $\int |Du|$

But

- limit is not *pointwise*;
- is not isotropic at fine scales.

# Examples

Consider a binary image:



On the left, the measure of the length is correct. On the right, it is overestimated (by $\sqrt{2}$ hence more than 40%)

Why is the previous result correct? *Smoothing*

# Examples

In order to approximate well $TV(u)$, one *must* in general smooth $u$. In particular, if $u$ is a characteristic function, its approximation $u_h$ itself cannot be binary but must be the approximation of a smooth approximation of $u$:



This unit circle has "length" $\sim 6.30$ (error is less than $0.3\%$). However, the approximation is smooth especially in some directions.

# Finite differences: improvements

Many other discretizations have been suggested to improve this drawback:

- ▶ AC, Levine, Lucier (SIIMS 2011): "upwind" finite differences;
- ▶ Symmetric treatment of the differences, such as in AC-Pock (SMAI JCM 2015);
- ▶ "Shannon" T.V.: TV of a "Shannon" interpolate (Abergel-Moisan, JMIV 2017);
- ▶ New discrete TV of L. Condat (SIIMS 2017) *very* isotropic.

# Finite elements

Another classical way to approximate total variation / perimeters is to use graph-based pairwise differences

(*cf* [Boykov, Kolmogorov, Cremers, Delong, ECCV 2006] and [Rother, Kolmogorov, Blake 2004] ("GrabCut")).

A third, less conventional in imaging, but natural approach is to discretize the total variations with (low order) finite elements (of course very close to FD on regular meshes).

*A priori*, it is not reasonable to use too high order elements, as solutions of minimization problems involving the total variation are not expected to be smooth (hardly more than Lipschitz, often discontinuous).

On the contrary, one needs to look for representations which accurately represent discontinuities.

# Finite elements

Most existing works are based on P1 conforming finite elements, that is, continuous functions which are piecewise affine. In particular many works by **S. Bartels** and collaborators (SINUM 2012, Math. Comp 2015...) and an interesting TVD "projection operator" on P1 elements which allows to derive easy error estimates (**Bartels, Nochetto, Salgado**, Math. Comp 2015), however for "$l^1$" anisotropic TV.

A recent preprint suggests to go beyond P1 (discontinuous P1 [*cf* also **Repin**, 1989]) and has a dual representation involving Raviart-Thomas vector fields, which also play a role in our approach, *cf* **Herrmann, Herzog, Schmidt, Vidal, Wachsmuth**, "Discrete Total Variation with Finite Elements and Applications to Imaging", arXiv, April 2018 (actually, close to **Condat**'s finite difference approach, 2017)

# P1 Finite elements

But there is one issue with P1 elements.

Consider a mesh $\mathcal{T}$ of simplices (triangles) covering a domain $\Omega \subset \mathbb{R}^d$ ($d = 2$ for simplicity) and the P1 functions $u \in C^0(\Omega)$ such that $\nabla u$ is constant on each $T \in \mathcal{T}$.

Assume one wants to represent a discontinuity: $u = \chi_E \in \{0, 1\}$ at each node of the mesh (the vertices of the simplices). Then clearly on $T \in \mathcal{T}$, $u$ is either constant, or has two vertices with the same value and one with the other value. Hence its gradient is orthogonal to one edge $e \subset \partial T$ of the triangle and

$$|T||\nabla u(T)| = \frac{1}{2}|e|$$

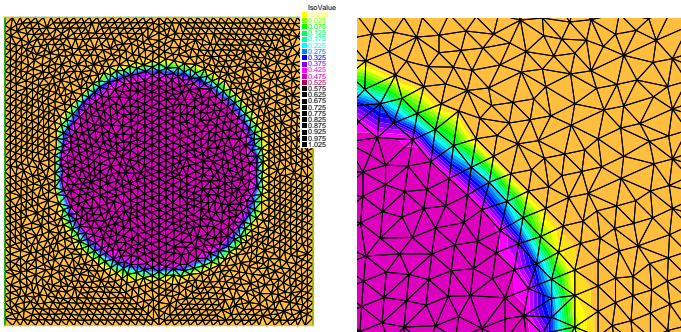which is the length of the segment joining the midpoints of the two other edges.

# P1 Finite elements

Hence

$$TV(u) = \int_\Omega |\nabla u| \approx \text{Per}(E_\mathcal{T})$$

where $E_\mathcal{T} \neq E$ is the set bounded by the line joining all midpoints of edges where $u$ jumps from $0$ to $1$. This line can be quite messy.

# P1 Finite elements: example



Disk length is overestimated by 1.4%.

# How can we do better?

Elementary: to improve the discretization of the gradient, one should...
*discretize the gradient?*
Given $u \in BV(\Omega)$, let

$$g_T := \frac{1}{|T|} \int_T Du.$$

Then, of course, one has

$$\int_\Omega |Du| = \sum_{T \in \mathcal{T}} \int_T |Du| \geq \sum_{T \in \mathcal{T}} |\int_T Du| = \int_\Omega |g| dx$$

where $g = \sum_T g_T \chi_T$.
More precisely, letting $\nu_T = g_T/|g_T|$, $\nu_u = Du/|Du|$,

$$\int_\Omega |Du| = \int_\Omega |g| dx + \frac{1}{2} \int_\Omega |\nu_T - \nu_u|^2 |Du|$$

hence the approximation is good if $Du$ does not rotates too much at the scale of the triangulation.

# How can we do better?

In particular, if $u = \chi_E$ where $E \subset \mathbb{R}^2$ is a set with an interior and exterior ball condition of radius $R$, and $h > 0$ is the size of the triangulation (largest edges), one can deduce that

$$\left(1 - \frac{\pi^2}{18}\left(\frac{h}{R}\right)^2\right) \text{Per}(E) \leq \int |g| dx \leq \text{Per}(E).$$

(Which is quite precise.)

# Crouzeix-Raviart FE

The **Crouzeix-Raviart**, or **nonconforming P1** (CR or ncP1) finite elements are defined as functions which are affine on each simplex and *continuous only through the middle of each facet* of the simplices. A function $u$ is usually projected on the CR space by averaging its value on each facet, and assigning the corresponding result in the middle of the facet. Then, it is easy to show that its affine extension inside a simplex has a gradient which is precisely given by

$$\nabla u_{CR} = \frac{1}{|T|} \int_T Du.$$

It has been observed a few times that this space is interesting to discretize nonlinear energies of the gradient, mostly for nonlinear elasticity or related problems (Di Pietro, Lemaire 2015, Xu, Henao 2011, Henao, Mora-Corral, Xu 2016, Ortner 2011, Ortner, Praetorius 2011, Carstensen-Liu, 2015).

# Crouzeix-Raviart gradients

In 2D, a CR gradient is easily seen to be orthogonal to rotated gradients of P1 functions.

In general, one can show that, given $g$ a "P0"[1] function:

**Lemma:** $g$ is a CR gradient (subject to $\mathcal{T}$) if and only if for every $\phi$ a low order *Raviart-Thomas* vector field (RT0) with $\text{div}\,\phi = 0$, $\phi \cdot \nu = 0$ on $\partial\Omega$,

$$\int_\Omega g \cdot \phi\, dx = 0.$$

(This works in any dimension. See also Carstensen-Liu, 2015 in 2D)

---

[1] P0 is the space of functions which are constant on each $T \in \mathcal{T}$.

# Raviart-Thomas vector fields

RT0 is the space of lower order vector fields (affine on each $T \in \mathcal{T}$) with well defined divergence (conforming [introduced for $H(\text{div})$]). The degrees of freedom are the (constant) fluxes accross each edge/facet of $T$. The basis functions are described as follows

$$\phi(x) = \begin{cases} \frac{l_n}{2|T_+|}(x - \mathbf{p}_+) & x \in T_+ \\ -\frac{l_n}{2|T_-|}(x - \mathbf{p}_-) & x \in T_- \\ 0 & x \notin T_+ \cup T_- \end{cases}$$

(Hence $\phi \cdot \nu = 1$ on the edge.)
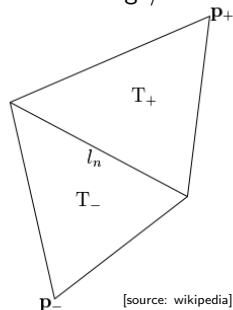


[source: wikipedia]

which has flux exactly $1 \times l_n$ accross the edge $\partial T_+ \cup \partial T_-$ and

$$\text{div}\,\phi = \frac{l_n}{|T_+|}\chi_{T_+} - \frac{l_n}{|T_-|}\chi_{T_-}.$$

# CR Total variation

For $u$ a CR function subject to a mesh $\mathcal{T}$, we simply define

$$TV(u) = \sum_{T \in \mathcal{T}} |T| |\nabla_T u|$$

as $u$ is jumping accross the edges (it is only continuous at the center points), this is *not* the true total variation (it misses the jump part on $\partial T$, so it is below).

For a P0 function $v$ (ie $v = \sum_T v_T \chi_T$), we also introduce

$$TV^0(v) = \min \left\{ TV(u) : u \text{ in CR}, \frac{1}{|T|} \int_T u \, dx = u(c_T) = v_T \; \forall T \in \mathcal{T} \right\}.$$

[$c_T$ is the center of the simplex $T$]

*Consistency:* A $\Gamma$-convergence property still holds ($\mathcal{T} = \mathcal{T}_h$, $h \to 0$).

# CR Total variation

As we have already seen in the beginning of the talk, if $u \in BV(\Omega)$ and $u'$ its projection on CR functions obtained by letting $\nabla_T u' = (1/|T|)Du(T)$ for each $T$ (or averaging $u$ on each edge/facet and putting the value in the middle),

$$\int_\Omega |Du| = TV(u') + \frac{1}{2}\int_\Omega |\nu_{u'} - \nu_u|^2 |Du|.$$

# Two nice properties

The definitions of $TV$ and $TV^0$ enjoy two interesting properties.

The first is the following duality formula: given $u$ in CR and $u^0$ in P0 with $u(c_T) = u_T^0$ for each element $T \in \mathcal{T}$, one has

$$TV(u) \geq$$
$$TV^0(u^0) = \sup \left\{ \int_\Omega u \operatorname{div} \phi \, dx : \phi \text{ RT0 field}, \|\phi(c_T)\| \leq 1 \; \forall T \in \mathcal{T} \right\}$$

[Interesting for optimization/error bounds.]

# Two nice properties

Then, we deduce that, in some sense, $TV$ is "exact" for recovering flat surfaces:

**Proposition:** Let $\nu$ a unit vector, $a \in \mathbb{R}$, $E = \{x \cdot \nu \geq a\}$. Let $u_{\mathcal{T}}$ be the projection of $u = \chi_E$ on CR functions subject to $\mathcal{T}$. Then for any other CR function $v$ with $v = u_{\mathcal{T}}$ on the boundary nodes,

$$TV(v) \geq TV(u_{\mathcal{T}}) = |Du|(\Omega) = \text{Per}(E; \Omega)$$

This is in fact obvious, using the constant field $\nu_E$ as a global "calibration" (the equality follows from the formula

$$\int_{\Omega} |Du| = TV(u') + \frac{1}{2} \int_{\Omega} |\nu_{u'} - \nu_u|^2 |Du|.)$$

# Error bounds

We can also provide error estimates for solutions of standard denoising problems

$$\min_u |Du|(\Omega) + \int_\Omega |u - g|^2 dx \qquad (ROF)$$

**Proposition** If $\bar{u}$ solves $(ROF)$ and $\bar{u}_h$ is the solution of the discrete version [P0, based on $TV^0$], then

- ▶ if $g \in BV$, $\Omega$ is periodic, then

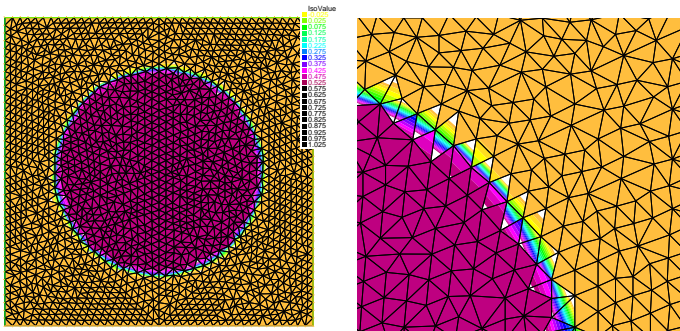$$\|\bar{u}_h - \bar{u}\|_2 \lesssim h^{1/4},$$

- ▶ if the dual problem has a Lipschitz solution, then

$$\|\bar{u}_h - \bar{u}\|_2 \lesssim h^{1/2},$$

[See also Lai-Lucier-Wang'09, Lucier-Wang'11, Bartels, Nochetto, Salgado'15]
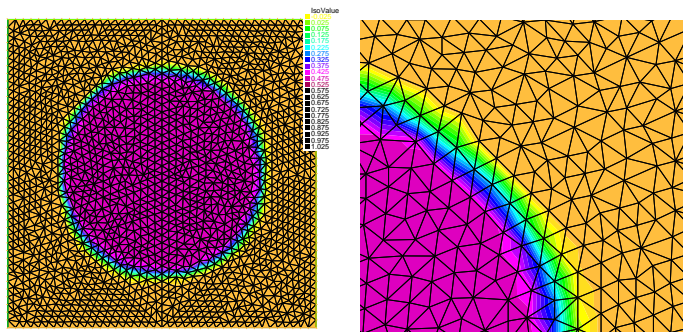
# CR Finite elements: example

The energy and transition region are improved with respect to P1:



Disk length is now overestimated by 0.1% (showing that the optimization is not perfect).

# P1 Finite elements: example

(We recall here the previous P1 example)



Disk length is overestimated by 1.4%.

# But one issue...

Despite the Proposition which shows that our discrete *TV* can be exact regardless of the mesh (which is quite remarkable), there is a price to pay: the solution is not necessarily unique, and can be very diffusive. Consider the following setting:
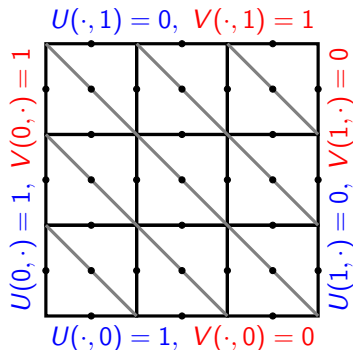


Figure: Problem: minimize $TV(u)$ and $TV(v)$ with $u = U$, $v = V$ on the boundary nodes

# But one issue...

Then, one can show

- ▶ That $u$ is essentially unique and given by the projection of $\chi_{\{x+y<1\}}$ (except on the nodes on the diagonal where the value is arbirary in $[0,1]$);

- ▶ That there are infinitely many solutions $v$, and some which satisfy $0 < v < 1$ on all inner nodes (even if the projection of $\chi_{\{y>x\}}$ is an admissible solution)!
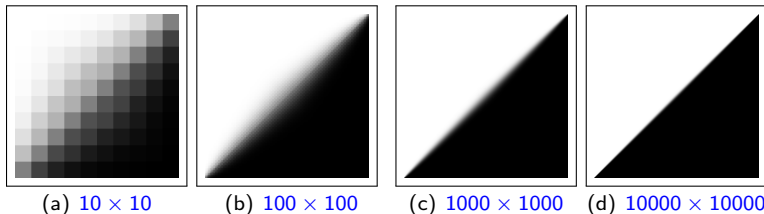


(a) $10 \times 10$  (b) $100 \times 100$  (c) $1000 \times 1000$  (d) $10000 \times 10000$

Figure: Possible solutions $v$ for various discretizations

# Mesh adaption?

▶ The example discussed seems to show that still, a mesh adaption might be desirable to improve the precision of the jumps (even if lower order terms might help improving the quality of a solution).

▶ In 2D, for a quadrangular mesh, there is an interesting point to stress: the nodes are not changed if the squares are cut along the other diagonal. Moreover, flipping the edges in the above example, and keeping the values of $v$ at the nodes, one finds a new $v'$ with $TV(v') \approx 2$ instead of $TV(v) = \sqrt{2}$ (the optimal value).

▶ Hence one should try to minimize the *maximum* of the TV among all possible triangulations with the same nodes. (Different from the P1 conforming case where one should optimize the mesh in order to *minimize* the energy.)
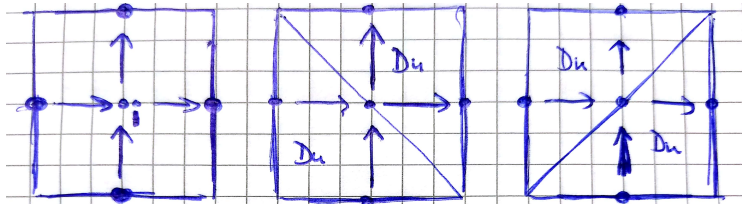
# "Automatic" Mesh adaption in 2D

This suggests an easy strategy to find the "best" CR discretization in 2D. Let $u$ be given at the nodes $\mathbf{i} = (i + 1/2, j + 1/2) \in \mathcal{I}_c$ (centers of the squares) and $(i, j + 1/2)$, $(i + 1/2, j)$ (center of edges) for $i, j$ integers. Define the following finite differences

$$(D_1 u)_{\mathbf{i}} = 2 \begin{pmatrix} u_{\mathbf{i}} - u_{\mathbf{i}-(\frac{1}{2},0)} \\ u_{\mathbf{i}} - u_{\mathbf{i}-(0,\frac{1}{2})} \\ u_{\mathbf{i}+(\frac{1}{2},0)} - u_{\mathbf{i}} \\ u_{\mathbf{i}+(0,\frac{1}{2})} - u_{\mathbf{i}} \end{pmatrix}, \quad (D_2 u)_{\mathbf{i}} = 2 \begin{pmatrix} u_{\mathbf{i}+(\frac{1}{2},0)} - u_{\mathbf{i}} \\ u_{\mathbf{i}} - u_{\mathbf{i}-(\frac{1}{2},0)} \\ u_{\mathbf{i}} - u_{\mathbf{i}-(0,\frac{1}{2})} \\ u_{\mathbf{i}+(0,\frac{1}{2})} - u_{\mathbf{i}} \end{pmatrix}, \quad \forall \mathbf{i} \in \mathcal{I}_c,$$

and then let

$$TV(u) = \sum_{\mathbf{i} \in \mathcal{I}_c} f((D_1 u)_{\mathbf{i}}, (D_2 u)_{\mathbf{i}})$$

# "Automatic" Mesh adaption in 2D



Where

$$f((\xi_{m,n})_{m=1,\ldots,4}^{n=1,2}) =$$
$$\frac{1}{2}\max\left\{\left(\sqrt{\xi_{1,1}^2+\xi_{2,1}^2}+\sqrt{\xi_{3,1}^2+\xi_{4,1}^2}\right), \left(\sqrt{\xi_{1,2}^2+\xi_{2,2}^2}+\sqrt{\xi_{3,2}^2+\xi_{4,2}^2}\right)\right\}.$$

# "Automatic" Mesh adaption in 2D

The derivative operator "$D$" has a lot of redundancy which seems useless at first glance. However, this is the price to pay to have a function $f$ whose *proximity operator* is easy to compute (and explicit). It is given by

$$\arg\min_{\xi \in \mathbb{R}^8} \frac{1}{2\tau}\|\xi - \bar{\xi}\|^2 + f(\xi),$$

and it is not hard to see that it depends only on the length of the vectors which appear in the expression of $f$, so that one merely needs to be able to solve

$$\arg\min_{x \in \mathbb{R}^4, x \geq 0} \frac{1}{2\tau}\|x - \bar{x}\|^2 + \max\{|x_1| + |x_2|, |x_3| + |x_4|\}$$

for $\bar{x} = (\sqrt{\bar{\xi}_{1,1}^2 + \bar{\xi}_{2,1}^2}, \sqrt{\bar{\xi}_{3,1}^2 + \bar{\xi}_{4,1}^2}, \sqrt{\bar{\xi}_{1,2}^2 + \bar{\xi}_{2,2}^2}, \sqrt{\bar{\xi}_{3,2}^2 + \bar{\xi}_{4,2}^2})$, that is, a very low-dimensional $\ell_\infty(\ell_1)$ norm.
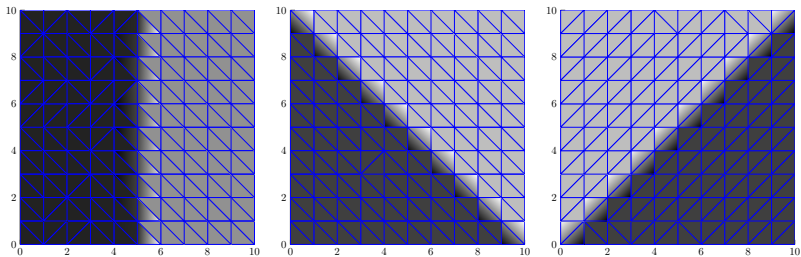
# "Automatic" Mesh adaption in 2D

For this reason, one can solve relatively easily problems of the form

$$\min_{u} TV(u) + G(u) = f(Du) + G(u)$$

for many (usually convex) terms $G$, where the minimum is on the nodal values of $u$. (Proximal first order methods.)

# "Automatic" Mesh adaption in 2D



Automatic adaption of the triangles

# Example: lines



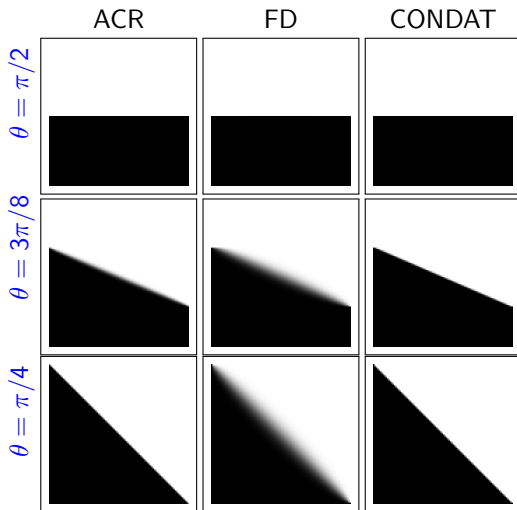Figure: Recovery of lines for several directions (Adaptive CR, Forward Differences, 16NN graph cut, Condat's discretization)

# Example: lines

| $\theta$ | $l$ | ACR | FD | CONDAT |
|---|---|---|---|---|
| $\pi/2$ | 100.00 | **100.00** | **100.00** | **100.00** |
| $3\pi/8$ | 108.24 | **108.59** | 109.61 | 108.79 |
| $\pi/4$ | 141.42 | **141.42** | 142.13 | 140.56 |

Table: Values of the energy for the previous experiments

# Example: circle
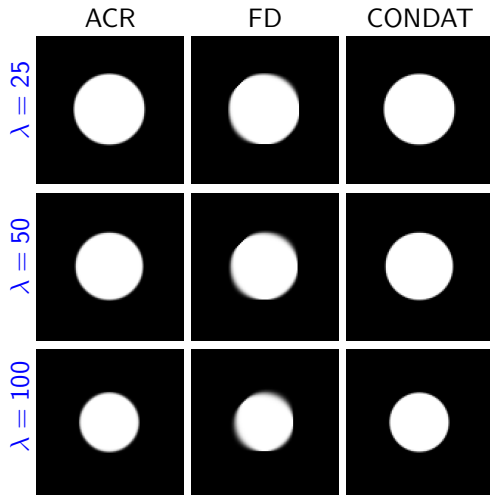


Figure: Minimizing $\lambda \mathrm{Per}(E) + \int_E (|x| - R)dx$ for various values of $\lambda$

# Example: circle

| $\lambda$ | p | ACR | FD | CONDAT |
|-----|--------|--------|--------|------------|
| 25  | 150.52 | 150.71 | 151.80 | **150.38** |
| 50  | 143.31 | **143.37** | 143.65 | 143.40 |
| 100 | 125.66 | **125.64** | 124.74 | 125.83 |

Table: Perimeter of the disc for different values of $\lambda$

A Denoising example

# Perspective

- ▶ 3D? the construction on squares cannot be easily generalized
- ▶ Extensions, other geometric problems (currents?)
- ▶ Extract higher order information, curvature based energies?

**Thank you for your attention**