TVL1 Models for Imaging: Global Optimization & Geometric Properties Part I

Tony F. Chan

Math Dept, UCLA

S. Esedoglu

Math Dept, Univ. Michigan

Other Collaborators:

J.F. Aujol & M. Nikolova (ENS Cachan), F. Park, X. Bresson (UCLA)

Research supported by NSF, ONR, and NIH.

Papers: www.math.ucla.edu/applied/cam/index.html

Research group: www.math.ucla.edu/~imagers

Based mainly on the works:

1. Chan, T. F.; Esedoglu, S. *Aspects of total variation regularized L^1 function approximation.* UCLA CAM Report 04-07, **February 2004**. SIAM J. Appl. Math **65**:5 (2005), pp. 1817 - 1837.
2. Chan, T. F.; Esedoglu. S.; Nikolova, M. *Algorithms for finding global minimizers of denoising and segmentation models.* UCLA CAM Report 04-54, **September 2004**. SIAM J. Appl. Math. **66** (2006), pp. 1632 - 1648.
3. Bresson, X.; Esedoglu, S.; Vandergheynst, P.; Thiran, J. P.; Osher, S. *Fast global minimization of the active contours/ snake model.* UCLA CAM Report 05-04, **January 2005**. J. Math. Imaging and Vision. **28**:2 (2007), pp. 151 - 167.

Total Variation & Geometric Regularization

$$TV(u) = \int_{\Omega} |\nabla u| \, dx$$

- *Measures "variation" of u, w/o penalizing discontinuities.*

•*1D: If u is monotonic in [a,b], then TV(u) = |u(b) – u(a)|, regardless of whether u is discontinuous or not.*

- *nD: If u(D) = c* χ*(D), then TV(u) = c* $|\partial D|$.
- *(Coarea formula)* $\int_{R_n} f \, |\nabla u| \, dx = \int_{-\infty}^{+\infty} \Big(\int_{\{u=r\}} f \, ds \Big) dr$

•*Thus TV controls both size of jumps and geometry of boundaries.*

Total Variation Restoration

Regularization: $$TV(u) = \int_{\Omega} |\nabla u| \, dx$$

Variational Model:

$$\min_u f(u) = \alpha TV(u) + \frac{1}{2} \| Ku - z \|^2$$

* First proposed by Rudin-Osher-Fatemi '92.

* Allows for edge capturing (discontinuities along curves).

* TVD schemes popular for shock capturing.

Gradient flow:

$$u_t = -g(u) = \underbrace{\alpha \nabla \cdot \left(\frac{\nabla u}{|\nabla u|} \right)}_{\text{anisotropic diffusion}} \underbrace{- (K^* K u - K^* z)}_{\text{data fidelity}} \qquad \frac{\partial u}{\partial n} = 0$$

TV-L^2 and TV-L^1 Image Models

Rudin-Osher-Fatemi: Minimize for a given image $f(x,y)$:

$$E_2(u,\lambda) := \int_D |\nabla u| + \lambda \int_D (u-f)^2 \, dx$$

Model is convex, with unique global minimizer.

TV-L^1 Model:

$$E_1(u,\lambda) := \int_D |\nabla u| + \lambda \int_D |u-f| \, dx.$$

Model is non-strictly convex; global minimizer not unique.

Discrete versions previously studied by: Alliney'96 in 1-D and Nikolova'02 in higher dimensions, and E. Cheon, A. Paranjpye, and L. Vese'02.

Is this a big deal?

Other successful uses of L^1: Robust statistics; l^1 as convexification of l^0 (Donoho), TV Wavelet Inpainting (C-Shen-Zhou), Compressive Sensing (Candes, Donoho, Romberg, Tao),

Surprising Features of TV+L1 Model

- Contrast preservation
- Data driven scale selection
- Cleaner multiscale decompositions
- Intrinsic geometric properties provide a way to solve **non-convex** shape optimization problems via **convex** optimization methods.

Contrast Loss of ROF Model

- Theorem (Strong-C 96): If f = $1_{B_r(0)}$, $\Delta = B_R(0)$, 0 < r < R;

$$u = (1 - \tfrac{1}{\lambda r})1_{B_r(0)} + \frac{r}{\lambda(R^2 - r^2)}1_{\Delta/B_r(0)}.$$

- Locates edges exactly (robust to small noise).
- Contrast loss proportional to *scale*$^{-1}$:

$$\frac{1}{\lambda r} = \frac{|\partial\Omega|}{2\lambda|\Omega|}.$$

- Theorem (Bellettini, Caselles, Novaga 02): If f = 1_Ω, Ω convex, $\partial\,\Omega$ is $C^{1,1}$, and for every p on $\partial\,\Omega$

$$curv_{\partial\Omega}(p) \leq \frac{|\partial\Omega|}{|\Omega|}.$$

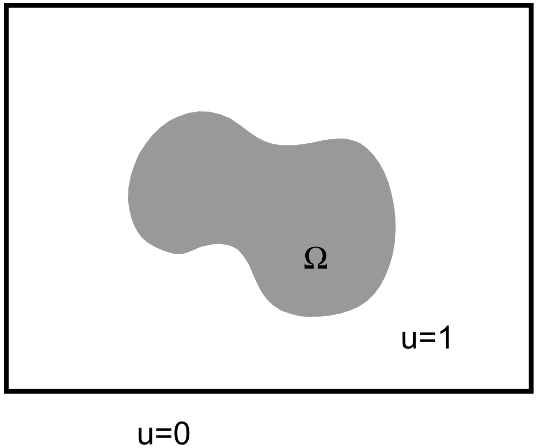
Then $$u = (1 - \frac{|\partial\Omega|}{2\lambda|\Omega|})1_\Omega.$$

TV-L1: Contrast & Geometry Preservation

Contrast invariance: If u(x) is the solution for given image f(x), then cu(x) is the solution for cf(x).

Contrast & Geometry Preservation: Let $f(x) = 1_{\Omega}(x)$, where Ω is a bounded domain with smooth boundary. Then, for large enough λ, the unique minimizer of $E_1(\cdot,\lambda)$ is exactly $f(x)$.

The model recovers such images exactly. Not true for standard ROF.



(Other method to recover contrast loss: Bregman iteration (Osher et al))

Contrast & Geometry Preservation

The solution operator for TV+L^1 acts linearly on some data:

Theorem

Let $D = \mathbb{R}^n$. Assume that $f(x)$ and $g(x)$ are two images with compact, disjoint supports. There exists a minimal separation distance Δ such that if

$$\mathrm{dist}(\mathrm{supp}(f), \mathrm{supp}(g)) > \Delta$$

then the TV+L^1 model acts linearly on

$$\{c_1 f(x) + c_2 g(x) : c_1, c_2 \in \mathbb{R}\},$$

i.e. if u_f minimizes E_1 with f as the given image, and u_g minimizes E_1 with g as the given image, then $c_1 u_f + c_2 u_g$ minimizes E_1 with $c_1 f + c_2 g$ as the given image.

Contrast & Geometry Preservation

Another important fact, shown by Alliney in the discrete case:

Claim

Let $f(x)$ be a given image, and let $u(x)$ be a solution of

$$\min_u \int_D |\nabla u| + \lambda \int_D |f - u| \, dx.$$

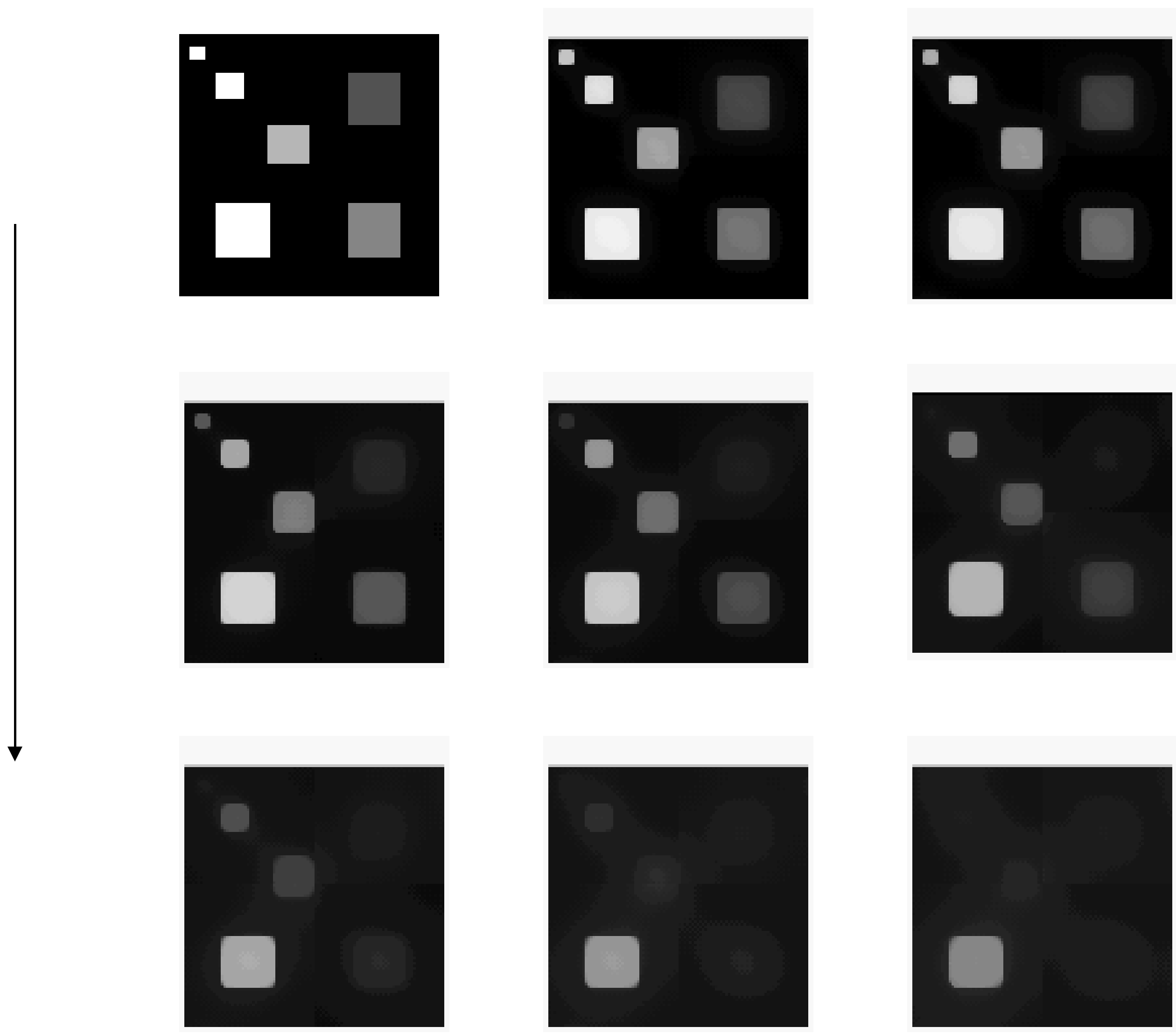
Then, $u(x)$ itself is the solution of

$$\min_v \int_D |\nabla v| + \lambda \int_D |u - v| \, dx.$$

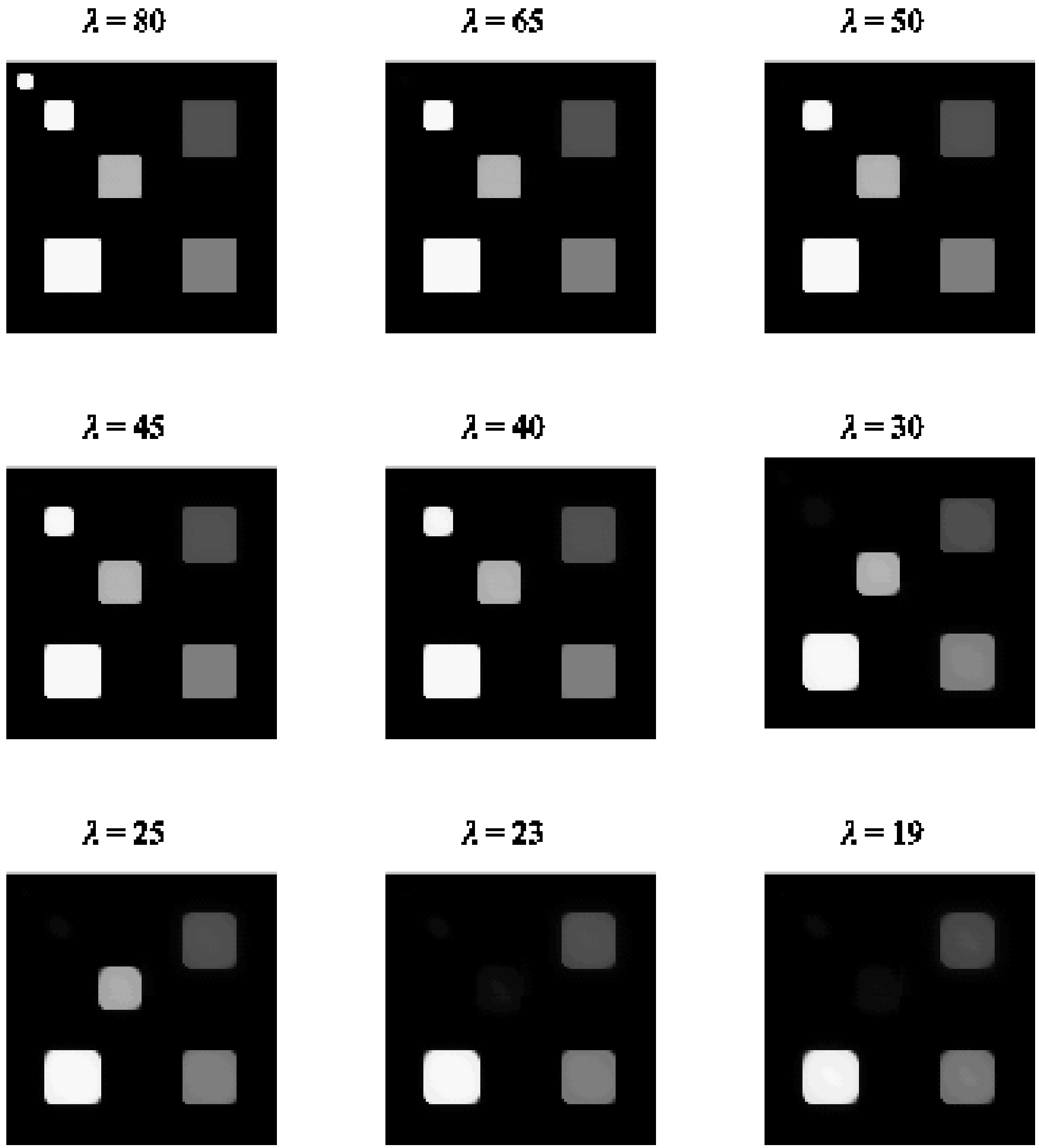
- In other words, denoised images are treated as clean – not true for ROF!

“Scale-space” generated by the original ROF model

Decreasing λ

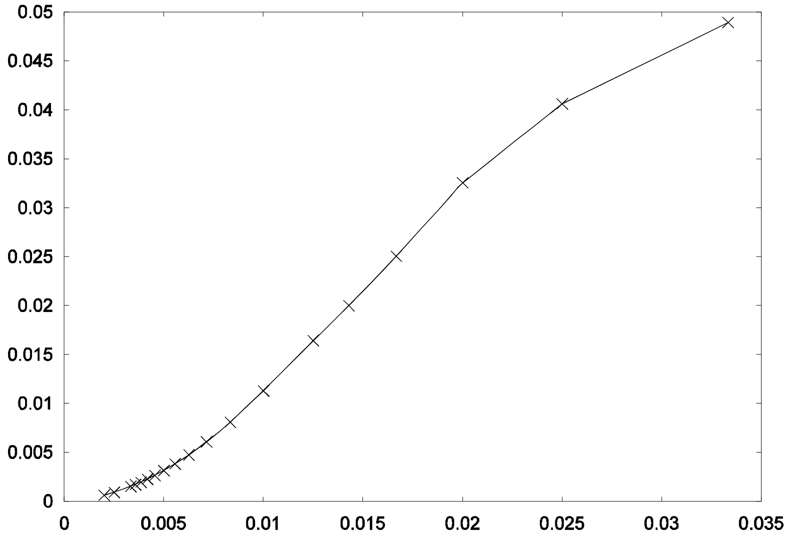


“Scale-space” generated by the $\mathrm{TV} - \mathrm{L}^1$ model

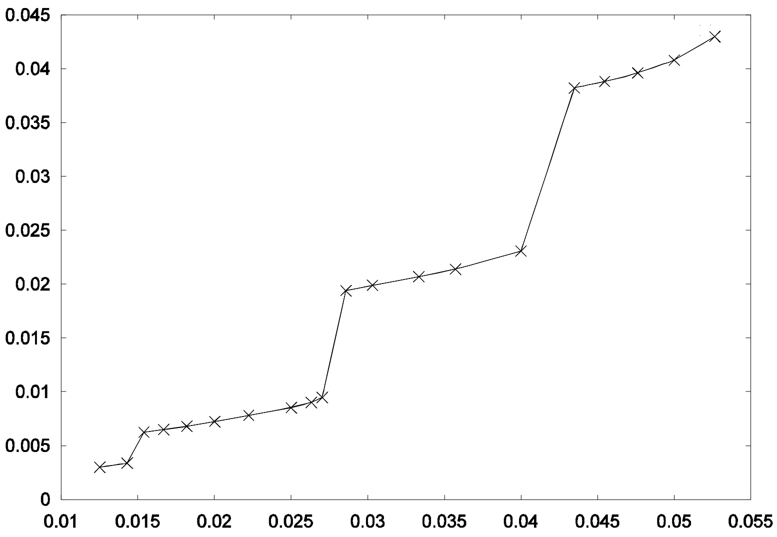


Data Dependent Scale Selection

Plots of $\|u_\lambda(x) - f(x)\|$ vs. λ^{-1}



TVL2

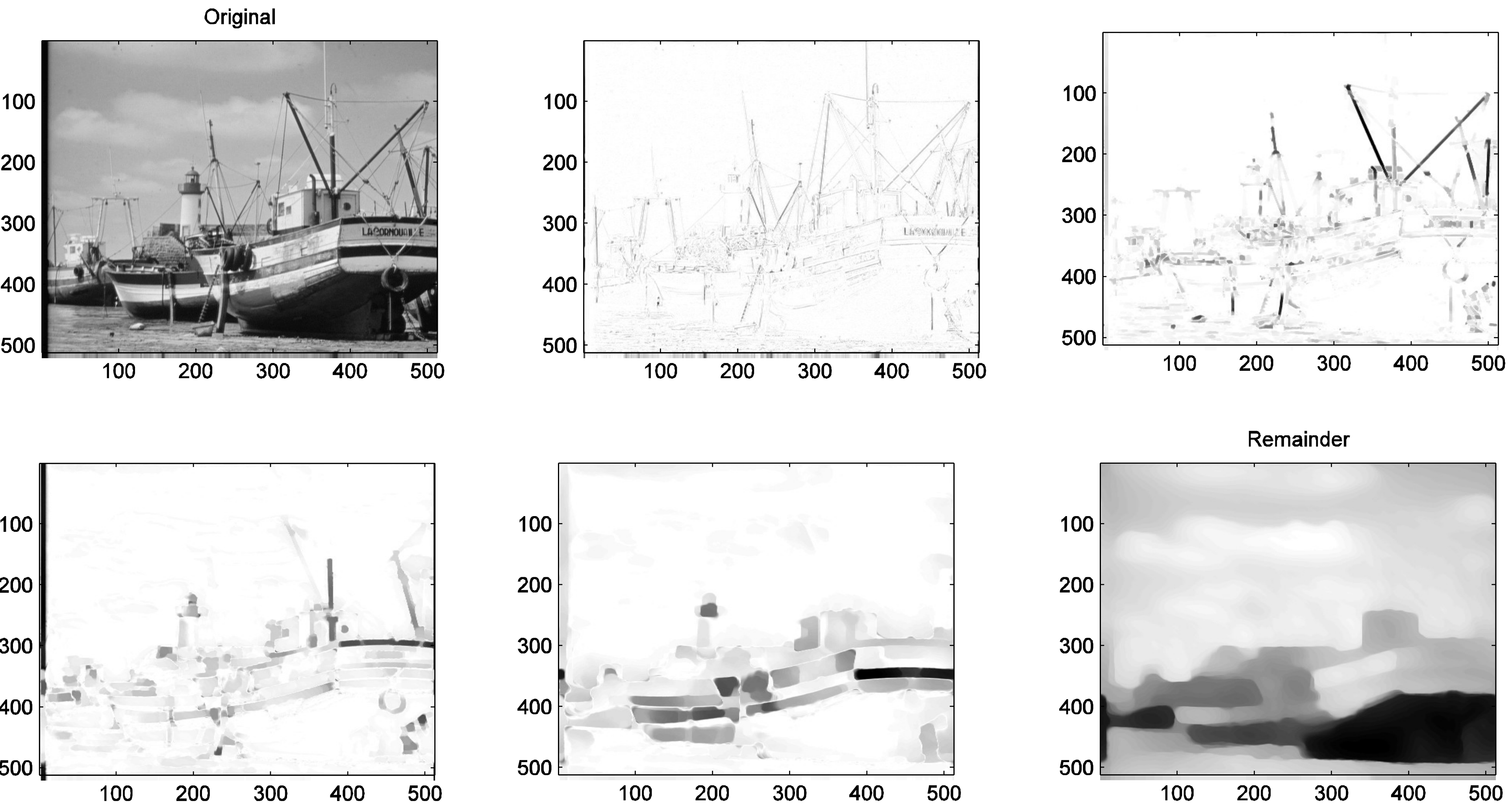


TVL1

Discontinuities of fidelity correspond to removal of a feature (one of the squares).

Multiscale Image Decomposition Example

(related: Tadmor, Nezzar, Vese 03; Kunisch-Scherzer 03)

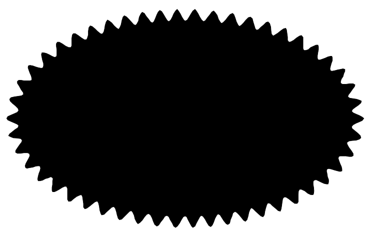


TVL1 decomposition gives well separated & contrast preserving features at different scales. E.g. boat masts, foreground boat appear mostly in only 1 scale

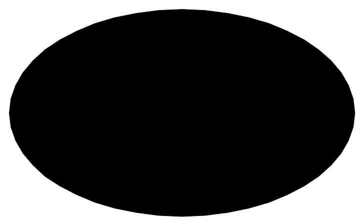
Convexification of Shape Optimization

Motivating Problem: Denoising of Binary Images

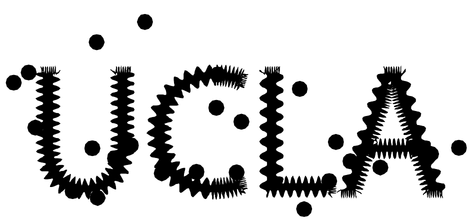
Given a binary observed image $f(x) = 1_{\Omega}(x)$. find a denoised (regularized) version.



Noisy binary image



Denoised binary image



Noisy binary image

Denoised binary image

Applications:

Denoising of fax documents (Osher, Kang).

Understanding many important image models: ROF, Mumford-Shah, Chan-Vese, etc.

Restriction of ROF to Binary Images

Take f(x)=$\mathbf{1}_\Omega$(x) and restrict minimization to set of binary images:

$$\min_{\substack{\Sigma \subset \mathbf{R}^N \\ u(x)=\mathbf{1}_\Sigma(x)}} \int_{\mathbf{R}^\mathbf{N}} |\nabla u| + \lambda \int_{\mathbf{R}^N} \left(u - \mathbf{1}_\Omega\right)^2 dx$$

Considered previously by Osher & Kang, Osher & Vese. Equivalent to the following *non-convex* geometry problem:

$$\boxed{\min_{\Sigma \subset \mathbf{R}^N} \mathrm{Per}(\Sigma) + \lambda |\Sigma \Delta \Omega|}$$

where $S_1 \Delta S_2$ denotes the symmetric difference of the sets S1 and S2.

Existence of solution for any bounded measurable Ω.

Global minimizer not unique in general. Many local minimizers possible.

Example of Local Minima for Geometry

Let the given set be $\Omega = B_R(0)$ with $R \in (\frac{1}{\lambda}, \frac{2}{\lambda})$. Then, it is easy to show that the unique minimizer of the geometry problem is $\Sigma = \emptyset$.

Energy of $\mathbf{1}_{B_r(0)}(x)$ vs. r

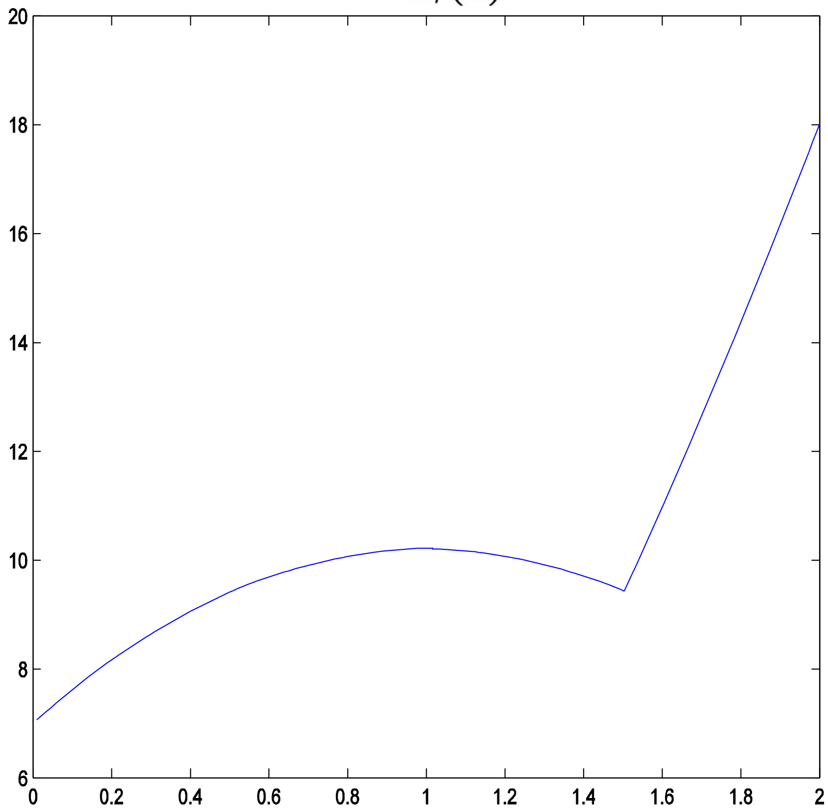
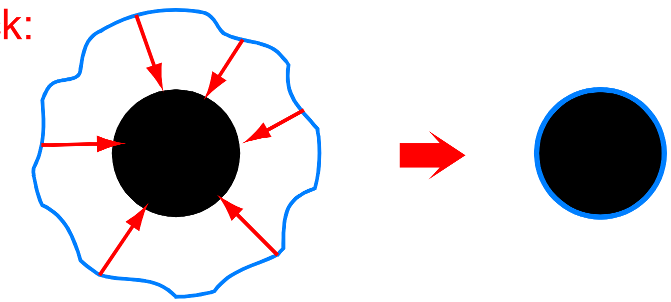
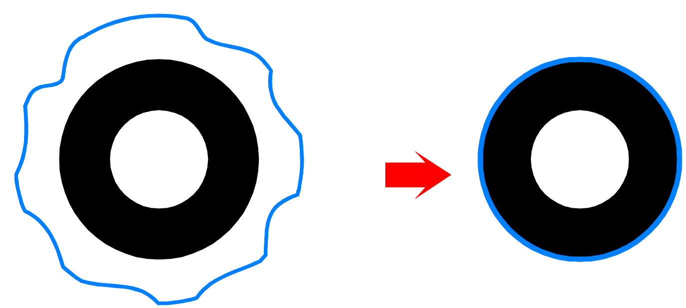


Illustration of how algorithms get stuck:



Other types of local minima:



$\Rightarrow \Sigma = \mathbf{1}_{B_R(0)}(x)$ is a local minimizer w.r.t. L^1-norm.

Global Minimum via TVL1

(C-Esedoglu-Nikolova ’04)

To find a solution (i.e. a global minimizer) u(x) of the non-convex variational problem (same as ROF for binary images):

$$\boxed{\min_{\Sigma \subset \mathbf{R}^N} \mathrm{Per}(\Sigma) + \lambda |\Sigma \Delta \Omega|}$$

it is sufficient to carry out the following steps:

- Find any minimizer of the *convex* TVL^1 energy

$$E_1(u,\lambda) = \int_{\mathbf{R}^N} |\nabla u| + \lambda \int_{\mathbf{R}^N} \Big| u - f \Big| \, dx.$$

  Call the solution found v(x).

- Let $\Sigma = \{ x \in \mathrm{R}^N : v(x) > \mu \}$ for some $\mu \in (0,1)$.

Then Σ is a global minimizer of the original non-convex problem for almost every choice of μ.

Connection of TVL[1] Model to Shape Denoising

- Coarea formula: $$\int_{\mathbf{R}^N} |\nabla u| = \int_{\mathbf{R}} \text{Per}(\{x : u(x) > \mu\})\, d\mu.$$
- "Layer Cake" theorem:
$$\int_{\mathbf{R}^N} |u-f|\, dx = \int_{\mathbf{R}} |\{x : u(x) > \mu\} \Delta \{x : f(x) > \mu\}|\, d\mu.$$

When $f(x) = 1_{\Omega}(x)$

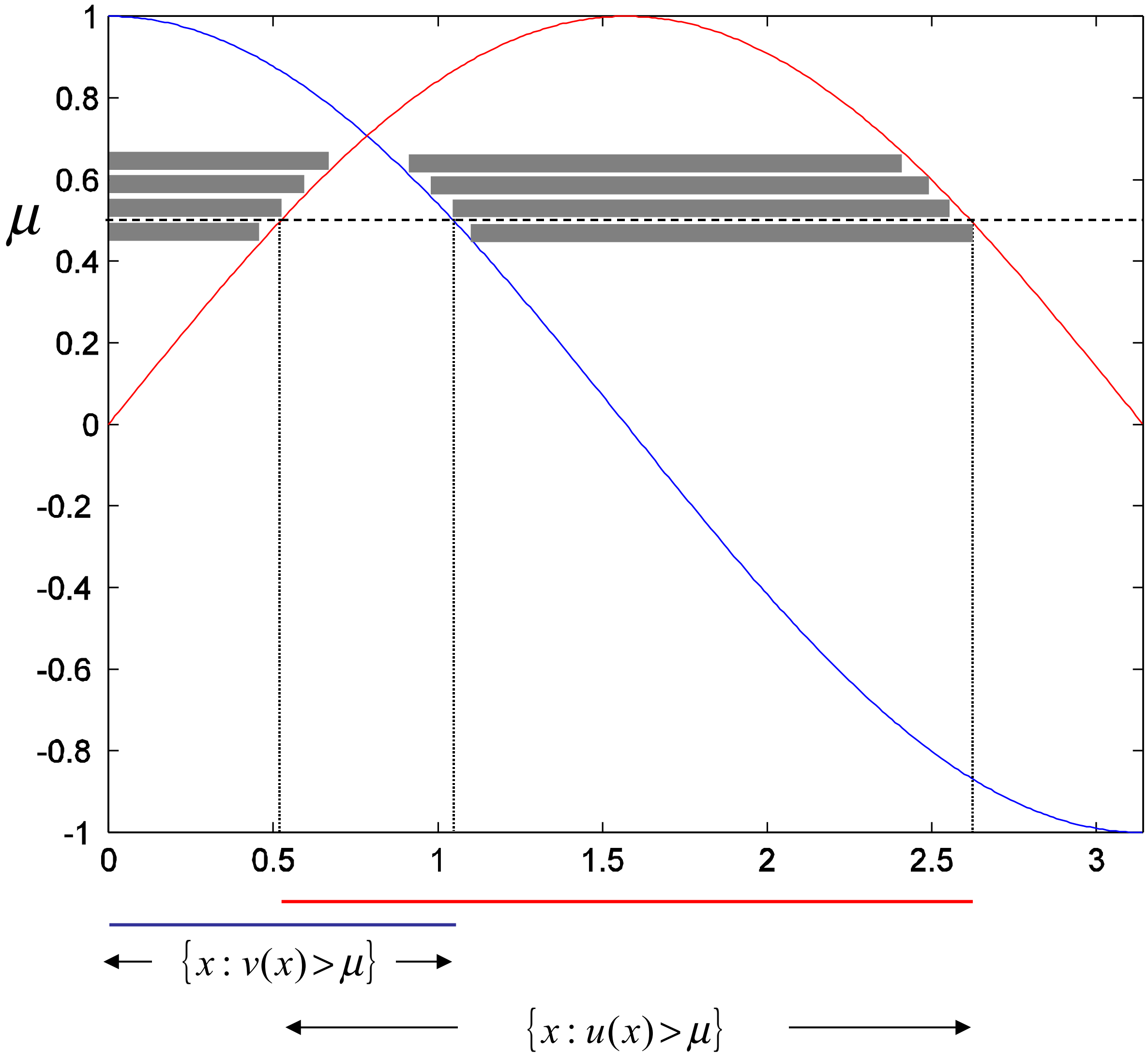
$$E_1(u,\lambda) = \int_0^1 \text{Per}\Big(\underbrace{\{x : u(x) > \mu\}}_{:=\Sigma(\mu)}\Big) + \lambda \Big|\underbrace{\{x : u(x) > \mu\}}_{:=\Sigma(\mu)} \Delta\Omega\Big|\, d\mu.$$

For each upper level set of u(x), we have the *same* geometry problem:

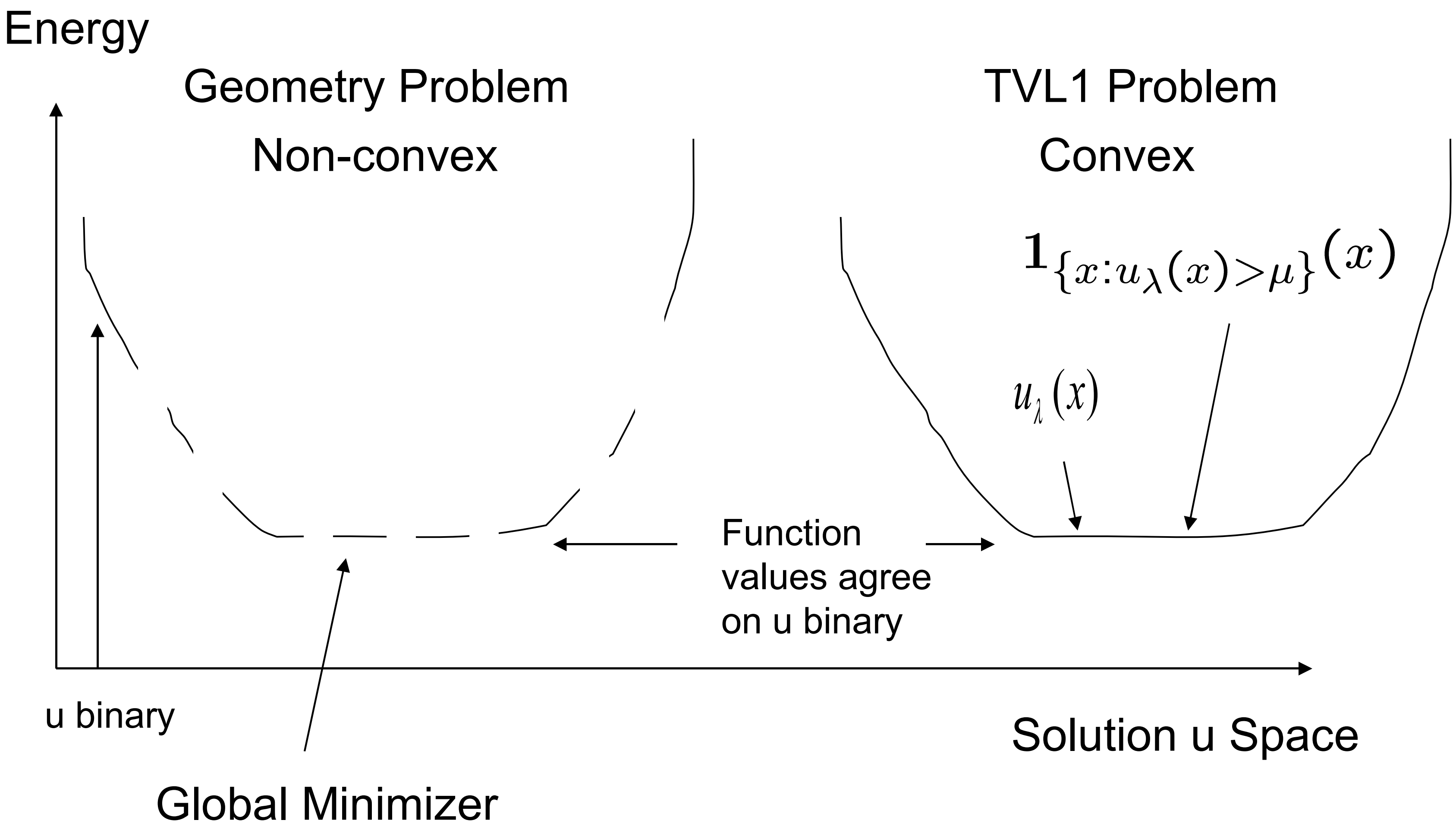
$$\boxed{\min_{\Sigma \subset \mathbf{R}^N} \text{Per}(\Sigma) + \lambda|\Sigma \Delta \Omega|}$$

Illustration of Layer-Cake Formula

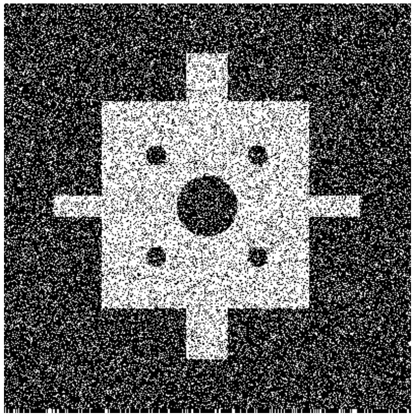
$$\int_D |u-v|\,dx = \int_{-1}^{1} \left|\{x : u(x) > \mu\} \Delta \{x : v(x) > \mu\}\right| d\mu$$



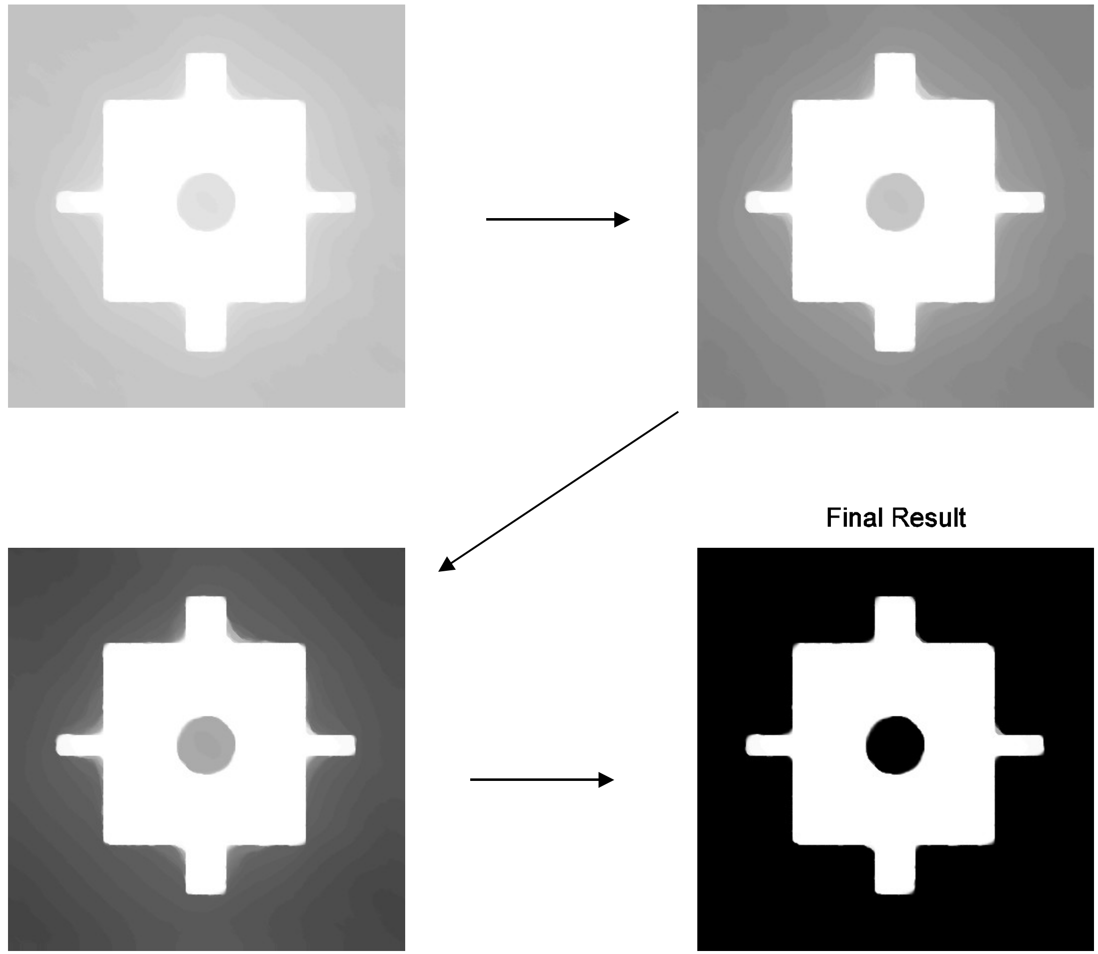
Illustration



Noisy Image



Intermediates, showing the evolution:



Intermediates non-binary! The convex TV-L^1 model opens up new pathways to global minimizer in the energy landscape.

What about other L^p norms?

$$\int_{\mathbb{R}^N} |f|^p \, dx = \int_{\mathbb{R}^N} \int_0^{|f|} p\mu^{p-1} \, d\mu \, dx$$

$$= p \int_0^{\infty} \mu^{p-1} \int_{\mathbb{R}^N} \mathbf{1}_{\{x:|f(x)|>\mu\}}(x) \, dx \, d\mu$$

$$= p \int_{\mathbb{R}} \mu^{p-1} |\{x : |f(x)| > \mu\}| \, d\mu$$

Integrand depends on μ explicitly, and not only on the super level sets of f; so these terms are not purely geometric: Solving different geometric problems at different levels.

Generalization to Image Segmentation

Chan-Vese Model (2001): Simplified Mumford-Shah: Best approximation of f(x) by two-valued functions:

$$u(x) = c_1 \mathbf{1}_{\Sigma}(x) + c_2 \mathbf{1}_{D\setminus\Sigma}(x)$$

Variational CV Segmentation Model:

$$\min_{\substack{c_1,c_2\in\mathbb{R}\\ \Sigma\subset D}} \mathrm{Per}(\Sigma)+\lambda\left\{\int_{\Sigma}(c_1-f)^2\,dx+\int_{D\setminus\Sigma}(c_2-f)^2\,dx\right\}$$

Similar arguments as for shape denoising show CV is equivalent to:

$$\min_{c_1,c_2\in\mathbb{R}} \underbrace{\min_{0\le u(x)\le 1} \int_D |\nabla u| + \lambda\int_D \left\{(c_1-f)^2-(c_2-f)^2\right\} u(x)\,dx}_{\text{CONVEX!}}\,.$$

Theorem: If $(c_1,c_2,u(x))$ is a solution of above formulation, then for a.e. μ in (0,1) the triplet: $\left(c_1, c_2, \mathbf{1}_{\{x:u(x)\geq\mu\}}(x)\right)$

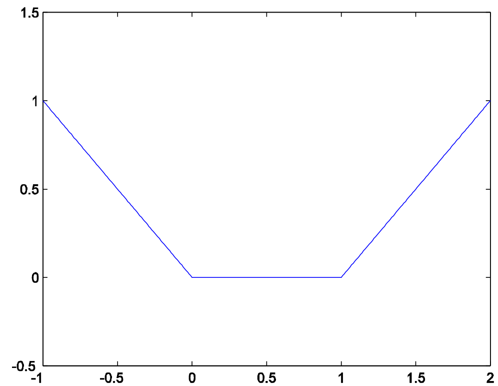
is a global minimizer of the Chan-Vese model.

Incorporating the Constraint

UPSHOT: For fixed c_1, c_2 the inner minimization (i.e. the shape optimization) in our formulation is convex. Also, the constraint on u can be incorporated via exact penalty formulation into an unconstrained optimization problem:

$$\min_u \int_D |\nabla u| + \lambda \int_D \left\{ (c_1 - f)^2 - (c_2 - f)^2 \right\} u(x) + \gamma z(u) \, dx.$$

where z(ξ) looks like:



Turns out: For γ large enough, minimizer u satisfies u(x)⟩\[0,1] for all xin D.

Solve via gradient descent on Euler-Lagrange equation.

Algorithm

1. Gradient descent for the convex formulation:

$$u_t = \nabla \cdot \left(\frac{\nabla u}{|\nabla u|} \right) + \lambda \left\{ (f - c_1)^2 - (f - c_2)^2 \right\} + \frac{1}{\mu} z'(u),$$

2. Update for the constants:

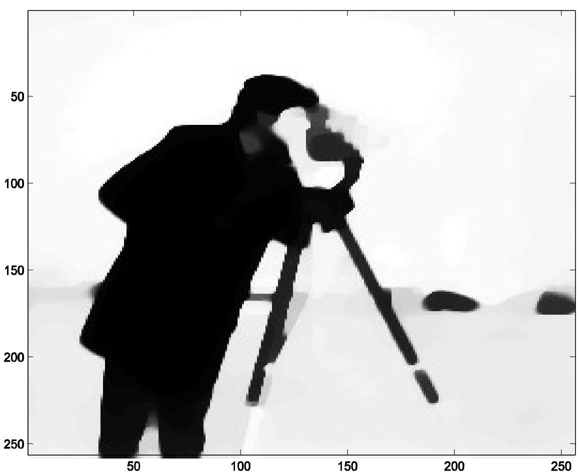
$$c_1 = \frac{\int uf\, dx}{\int u\, dx} \text{ and } c_2 = \frac{\int (1-u) f\, dx}{\int (1-u)\, dx},$$

3. Thresholding at the end:

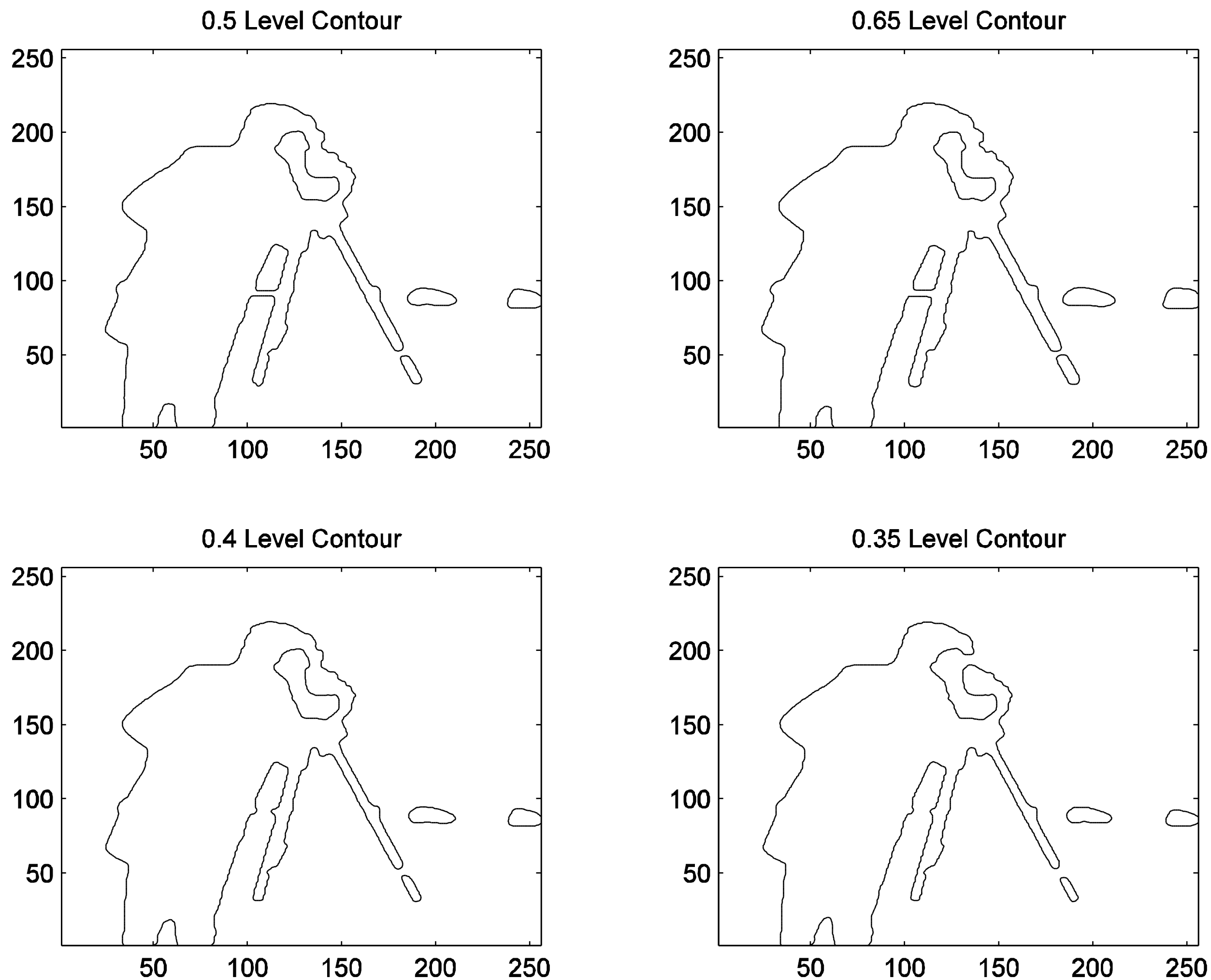
$$\Sigma = \{ x : u(x) > \mu \} \text{ for some } \mu \in (0,1).$$

Sample Computation:

Given image f(x)



u(x) computed



Related and Further Works

Previous Work

The idea of writing total variation based optimization problems in terms of super-level sets goes back (at least) to the works of G. Strang for problems in plasticity:

$$\min_{u \text{ s.t. } \int_D uf\,dx=1} \int_D |\nabla u|$$

It is shown that the minimizer is achieved at a characteristic function for the optimization problem above.

- Strang, G. *L^1 and L^∞ approximation of vector fields in the plane.* Nonlinear PDE in Applied Science (Tokyo, 1982), pp. 273 - 288. North-Holland Math. Stud. 81. Amsterdam, 1983.
- Strang, G. *Maximal flow through a domain.* Mathematical Programming. **26**:2 (1983), pp. 123 - 143.

Contrast & Geometry Perservation

Subsequent work of W. Yin and collaborators:

1. Applications to removing background effects in DNA microarray images.
2. Applications to removing illumination effects (shading) from face images:
   $\Rightarrow$ Better face recognition algorithms.
3. Precise choice of parameter λ in order remove a feature whose dual BV norm is known.

Contrast & Geometry Preservation

- In the fundamental formula

$$E_1(u) = \int_D |\nabla u| + \lambda \int |f - u|\, dx = \int_{\mathbb{R}} \mathrm{Per}(\{x : u > \mu\}) \\ + \lambda |\{x : u > \mu\} \triangle \{x : f > \mu\}|\, d\mu$$

  the integrand is independent of μ.
- $\Rightarrow$ Each level set is processed independently.
- Combined with subsequent results of Chambolle, Darbon & Sigelle, and W. Yin that show that after processing, layers can be stacked back up, we get

Contrast Invariance

Let $\phi : \mathbb{R} \to \mathbb{R}$ be a strictly monotone function. If $u(x)$ minimizes E_1 for the given image $f(x)$, then $\phi(u(x))$ minimizes E_1 for the given image $\phi(f(x))$.

Convex Formulation

Subsequent work by Chambolle:

- Extension to certain multi-phase segmentation models: The non-convex optimization problem:

$$\min_{u\in\{1,2,\ldots N\}} \int_D |\nabla u| + \lambda \sum_{j=1}^{N} \int_{D\cap\{u=j\}} (c_j - f)^2\, dx$$

can be reformulated as the convex optimization problem:

$$\min_{0\leq u_N \leq \ldots u_2 \leq u_1 \leq 1} \sum_{j=1}^{N} \int_D |\nabla u| + \lambda \left(\frac{c_j + c_{j-1}}{2} - f_j \right) u_j .$$

Continuous Max Flow/Min Cut (Bresson-C)

- [Strang 83] defined the continuous analogue to the discrete max flow. He replaced a flow on a discrete network by a vector field p. The continuous max flow (CMF) problem can be formulated as follows:

F,f are the sources and sinks, and w is the capacity constraint

- Application #1: [Strang] F=1, f=0

  If C is a cut then the CMF is given by minimizing the isoparametric ratio:

Isoparametric ratio

Continuous Max Flow/Min Cut

- Application #2: [Appleton-Talbot 06] F=0, f=0 (geodesic active contour)
  The CMF is
  It is a conservation flow (Kirchhoff's law, flow in=flow out).
  [AT] proposed to solve the CMF solving this system of PDEs:

  ⇨ We may notice that these PDEs come from this energy:

  (Weighted TV Norm)

- Application #3: [Bresson-Chan] F=gr, f=0
  Optimizing the CEN model corresponds to solve the CMF problem:

  The previous CMF problem can be solved by the system of PDEs:

  which comes from this energy:

  which is the CEN energy!