

NONCONVEX REGULARIZATION FOR SHAPE PRESERVATION

Rick Chartrand

Los Alamos National Laboratory,
rickc@lanl.gov

ABSTRACT

We show that using a nonconvex penalty term to regularize image reconstruction can substantially improve the preservation of object shapes. The commonly-used total-variation regularization, $\int |\nabla u|$, penalizes the length of object edges. We show that $\int |\nabla u|^p$, $0 < p < 1$, only penalizes edges of dimension at least $2 - p$, and thus finite-length edges not at all. We give numerical examples showing the resulting improvement in shape preservation.

Index Terms— Image reconstruction, image edge analysis, image shape analysis.

1. INTRODUCTION

An important task in image processing is to reconstruct an image from a noisy one. Unfortunately, many processes that remove noise will alter the characteristics of objects in the image. A simple example is Gaussian smoothing. Convoluting an image with a Gaussian kernel will remove noise effectively. However, discontinuities will be removed as well, resulting in the blurring of object edges.

A noise removal method that remedies this problem is total variation regularization, also known as the Rudin-Osher-Fatemi (ROF) model [1]. Given a noisy image f , a grayscale-intensity function on $\Omega \subset \mathbb{R}^2$, the image reconstruction u is the minimizer of the following functional:

$$F(u) = \int_{\Omega} |\nabla u| + \lambda \int_{\Omega} |u - f|^2. \quad (1)$$

The first term is the regularization term; using total variation suppresses noise, but allows discontinuities. The second term is the data fidelity term, which keeps the reconstruction close to the data. The squared L^2 norm is best suited for additive, Gaussian noise [2]. Choosing the value of the parameter λ allows one to adjust the relative effect of the two terms.

A drawback of the ROF model for accurate image reconstruction is the following. Let $E \subset \mathbb{R}^2$ be a bounded set with Lipschitz boundary, and let $u = \chi_E$ be the characteristic function of E . Then $\int |\nabla u|$ is the perimeter of E . It follows that total variation regularization penalizes the length of object edges in an image. Shortening edges will decrease the regularization term, while increasing the data fidelity term by

an amount that will depend on the area that is altered. The result is that sharp corners will be rounded, and long, thin features will tend to be removed, as these changes affect edge length more than shape area. The extent to which this occurs depends on λ , the choice of which is usually dictated by the noise level in the image. One is sometimes forced to choose between noise removal and shape preservation.

The remedy we propose is to alter the geometric property to be penalized. We use for our regularization functional

$$F_p(u) = \int_{\Omega} |\nabla u|^p + \lambda \int_{\Omega} |u - f|^2, \quad (2)$$

where $0 < p < 1$. We will show in Section 2 that the “ p -Dirichlet integral” regularization term only penalizes boundaries of dimension at least $2 - p$. Since most edges in images have finite length, penalizing this higher-dimensional property of edges will not penalize most shapes. Yet, as we will see in Section 4, our discrete implementation will still remove noise.

The cost of this approach is that F_p is nonconvex. Neither existence nor uniqueness of minimizers is guaranteed. Nevertheless, the algorithm we will demonstrate in Section 4, in countless examples of all manners of images, has never been observed to fail to converge, or to converge to a minimum that, if not unique, was not a suitable reconstruction. Moreover, the convergence of the algorithm, a generalization of Vogel and Oman’s fixed-point method [3], is rapid.

2. SHAPE BOUNDARY DIMENSION

Let $E \subset \mathbb{R}^2$, and let $u = \chi_E$ be the characteristic function of E , having value 1 on E and 0 elsewhere. We will show that $\int |\nabla u|^p = 0$ unless the boundary ∂E of E has dimension at least $2 - p$. In particular, if E has finite perimeter and $p < 1$, the regularization term of (2) does not penalize E at all. For this purpose, we define $\int |\nabla u|^p$ as the limit of its discrete approximations, under increasingly fine resolution. To be specific, we consider a uniform, square grid G with grid spacing Δx . Our discretization of the partial derivatives of u uses simple forward differences, and Neumann boundary conditions. We have

$$\int |\nabla u|^p \approx \sum_{x \in G} |\nabla u(x)|^p (\Delta x)^2. \quad (3)$$

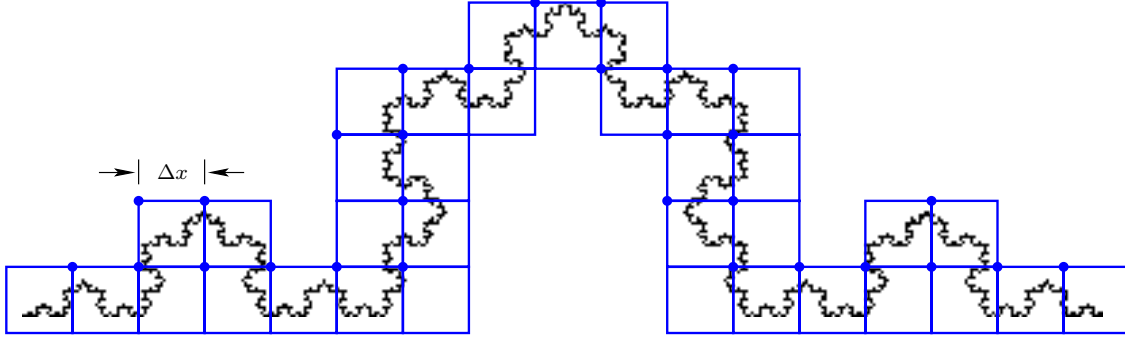


Fig. 1. A portion of the boundary ∂E of the Koch snowflake E of Fig. 4. The squares are those having vertices at grid points, and that intersect ∂E . The dots are those grid points where the discrete gradient of $u = \chi_E$ is nonzero, those where the segment to the right or below crosses ∂E . The dots are a subset of the upper-left corners of the squares.

Here $\nabla u(x)$ is the discrete gradient at x , with each partial derivative being of the form $(u(x_+) - u(x))/\Delta x$, with x_+ being the next grid point to the right or down. The notion of dimension we will use is the upper box-counting dimension. Using the same square grid, let $N_{\Delta x}(\partial E)$ be the number of grid squares that intersect ∂E (see Fig. 1). Then the upper box-counting dimension of ∂E is

$$\dim \partial E = \limsup_{\Delta x \rightarrow 0} \frac{\log N_{\Delta x}(\partial E)}{\log(1/\Delta x)}. \quad (4)$$

Theorem 2.1. *Let $E \subset \mathbb{R}^2$, $u = \chi_E$. Let $0 < p \leq 1$. Then the limit of (3) as $\Delta x \rightarrow 0$ is zero unless the upper box-counting dimension of ∂E is at least $2 - p$.*

Proof. Suppose $\dim \partial E < 2 - p$. Then we can find $r < 2 - p$ such that for all sufficiently small Δx , $\log N_{\Delta x}(\partial E) < r \log(1/\Delta x)$, or $N_{\Delta x}(\partial E) < (\Delta x)^{-r}$.

The key observation is that if the discrete gradient is nonzero at a given grid point, then the grid square having its upper-left corner at that point must intersect ∂E (see Fig. 1). Thus, the number of nonzero terms of (3) is at most $N_{\Delta x}(\partial E)$. Since the values of u are 0 or 1, we have the upper bound

$$\sum_{x \in G} |\nabla u(x)|^p (\Delta x)^2 \leq N_{\Delta x}(\partial E) (2/(\Delta x)^2)^{p/2} (\Delta x)^2. \quad (5)$$

For sufficiently small Δx , this is at most $(\Delta x)^{-r+2-p} 2^{p/2}$. Since $r < 2 - p$, (3) tends to 0 as $\Delta x \rightarrow 0$. \square

3. NUMERICAL IMPLEMENTATION

To compute a minimizer of (2), we use a straightforward generalization of the fixed-point method of Vogel and Oman [3]. Consider the Euler-Lagrange equation of (2):

$$0 = -\nabla \cdot (|\nabla u|^{p-2} \nabla u) + \lambda(u - f), \quad (6)$$

where for convenience we have ignored factors of $1/p$ and $1/2$. We solve (6) iteratively, by substituting the previous iterate u_{n-1} into $|\nabla u|$, then letting u_n be the solution of the

resulting linear equation:

$$(-\nabla \cdot (|\nabla u_{n-1}|^{p-2} \nabla) + \lambda I) u_n = \lambda f. \quad (7)$$

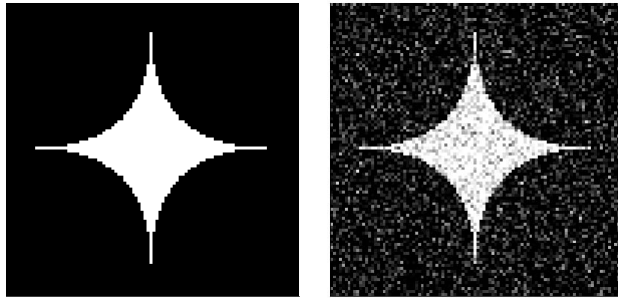
Also, to avoid division by zero, we approximate $|\nabla u_{n-1}|$ by $\sqrt{|\nabla u_{n-1}|^2 + \epsilon}$ for a small ϵ (namely 10^{-6} in the examples below). We begin the iteration with $u_0 = f$, the noisy image. We discretize gradients with forward differences (as in Section 2) and divergences with backward differences, and use Neumann boundary conditions.

We do not prove convergence. Moreover, this procedure can only be expected to produce a local minimum, owing to the nonconvexity of (2). However, in hundreds of examples we have always observed convergence, unless ϵ is too small, in which case the system (7) can become numerically singular. We have no way of knowing whether the computed minimizer is local or global, but in our experience it is always a sensible reconstruction. Furthermore, the convergence is rapid. In the examples below, we find 15 iterations to be ample, with each iteration taking under a second for a 100×100 image, and about 5 seconds for a 512×512 image, in Matlab on a 1.8 GHz laptop.

4. EXAMPLES

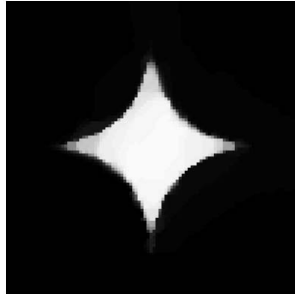
We present three examples of regularization of noisy binary images. In each case, we started with a $\{0, 1\}$ -valued image, the characteristic function of a shape E . We then added Gaussian noise of mean zero and standard deviation 0.25. The algorithm of Section 3 was then run with $p = 1, 3/4, 1/2$, and $1/4$. In each case, the regularization parameter λ was chosen to have the largest value (i.e., the weakest regularization) that still removed noise. (This is done by trial and error, separately for each p).

Our first example is a 100×100 image of an astroid shape with sharp cusps (Fig. 2). We see that with $p = 1$, the cusps are cut off, and the contrast near them slightly decreased.

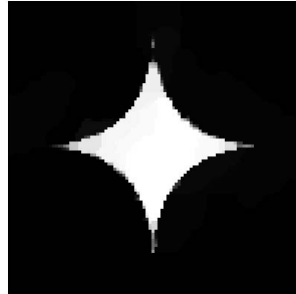


(a) Astroid shape

(b) Noise added



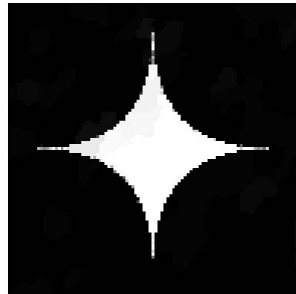
(c) $p = 1$



(d) $p = 3/4$



(e) $p = 1/2$



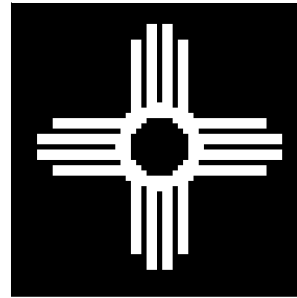
(f) $p = 1/4$

Fig. 2. The 100×100 astroid image has 25% noise added, then is regularized using (2) with four values of p . The larger values of p cut off the cusps, and decrease intensity near them. The smaller values preserve the shape.

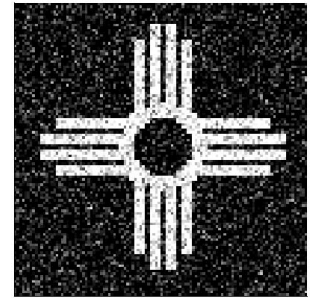
With $p = 3/4$, the cusps are still cut off, but to a lesser degree. With $p = 1/2$ and $p = 1/4$, the cusps and intensity are well-preserved.

The second example is a 112×112 image of a Zia symbol, with thin rays and thin gaps between (Fig. 3). With $p = 1$ and $p = 3/4$, the rays are smeared together, and contrast is lost. The shape and contrast are well-preserved with $p = 1/2$, slightly better with $p = 1/4$.

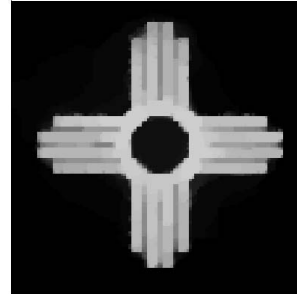
The final example is a 512×512 discrete Koch snowflake (Fig. 4; a zoomed-in version is in Fig. 5). The true Koch snowflake has a boundary of dimension $\log 4 / \log 3 \approx 1.26$. Edge details and contrast near the edges are better preserved with $p \leq 3/4$ than with $p = 1$.



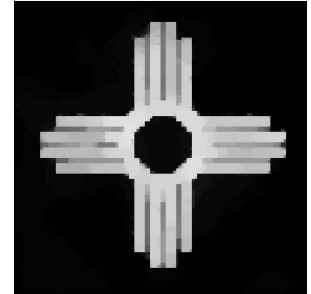
(a) Zia shape



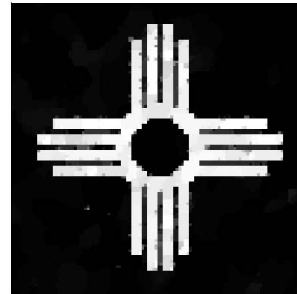
(b) Noise added



(c) $p = 1$



(d) $p = 3/4$



(e) $p = 1/2$

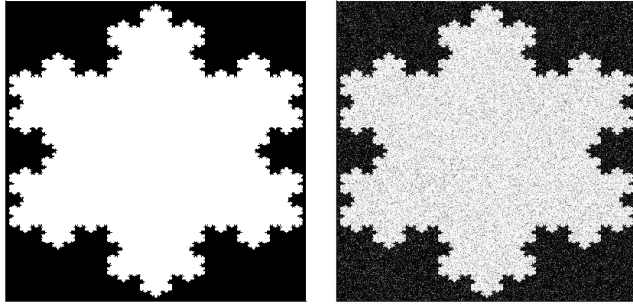


(f) $p = 1/4$

Fig. 3. The 112×112 Zia image has 25% noise added, then is regularized using (2) with four values of p . The larger values of p smear together the rays and gaps between them, and decrease contrast. The smaller values of p preserve the shape.

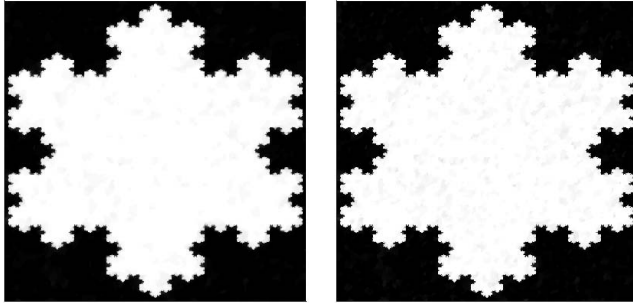
5. REFERENCES

- [1] L. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D*, vol. 60, pp. 259–268, 1992.
- [2] M. Green, "Statistics of images, the TV algorithm of Rudin-Osher-Fatemi for image denoising and an improved denoising algorithm," CAM Report 02-55, UCLA, 2002.
- [3] C. R. Vogel and M. E. Oman, "Iterative methods for total variation denoising," *SIAM J. Sci. Comput.*, vol. 17, no. 1, pp. 227–238, 1996.



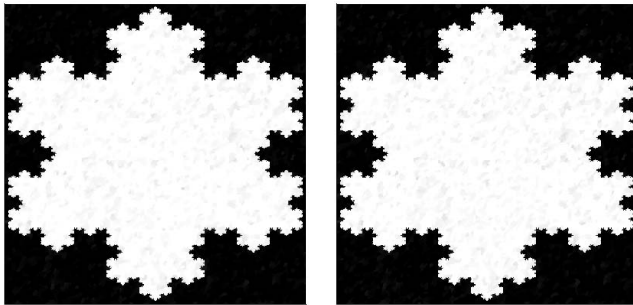
(a) Koch snowflake

(b) Noise added



(c) $p = 1$

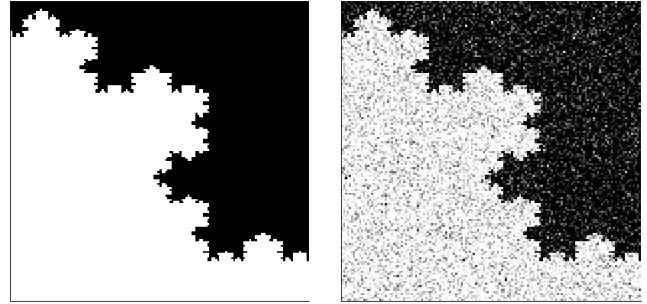
(d) $p = 3/4$



(e) $p = 1/2$

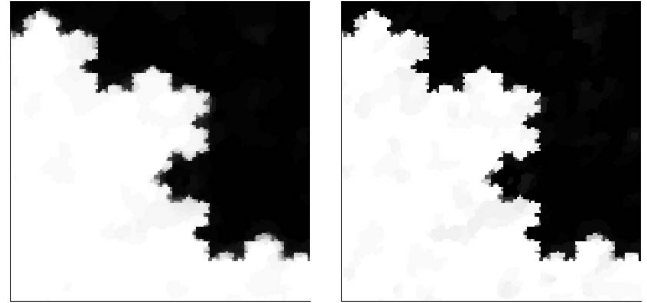
(f) $p = 1/4$

Fig. 4. The 512×512 Koch snowflake image has 25% noise added, then is regularized using (2) with four values of p .



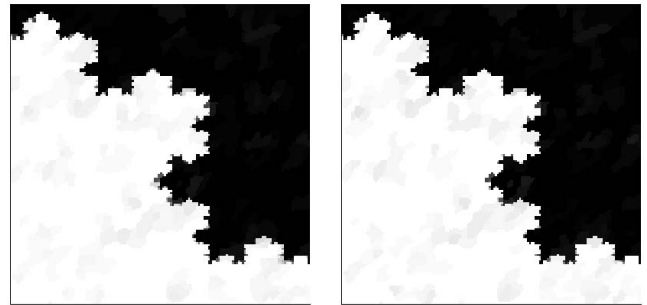
(a) Zoomed Koch snowflake

(b) Noise added



(c) $p = 1$

(d) $p = 3/4$



(e) $p = 1/2$

(f) $p = 1/4$

Fig. 5. A 150×150 portion of Fig. 4, right of center at the top. Edge detail and contrast is better preserved with $p \leq 3/4$.