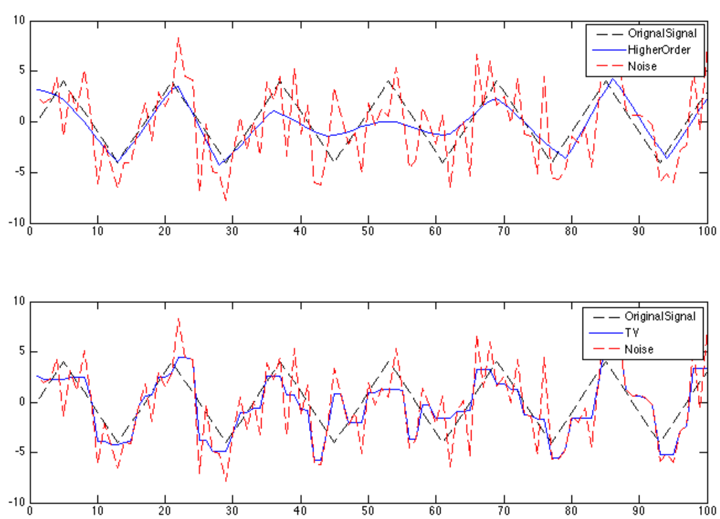
Higher Order Total Variation for Image Denoising

Ben Larson

Dec 12, 2016

1 Introduction

Acquisition of digital images almost always is accompanied by some undesired noise. Many developments have been made to address this issue. In a very important work by Rudin, Osher, and Fatemi used Total Variational methods to denoise images in a way that preserved edges while removing noise[1]. The problem that occurs with total variation methods is staircasing of the image, smooth signals such as ramps are transformed into piecewise sections (consider figure 1). Recent attempts to address this issue include using higher order methods in conjunction with total variation to remove noise, preserve edges, and avoid staircasing. In this paper I investigate several state of the art image denoising algorithms that use higher order total variation.



1.1 Images

In order to denoise images we need create a mathematical model that can represent images. Following the work of Laurent Younes[2]we setup our definitions of an image.
First we define a image as a function: $u : \mathcal{R}^2 \rightarrow \mathcal{R}$. Images are discrete as represented as matrices with dimension MxN. The open subset $\Omega \in \mathcal{R}^2$ is the domain of images. Ω is the space of images we will be solving in our minimization. We model our image with noise[1]:

$$u_0(x, y) = u(x, y) + \eta(x, y)$$

Where u_0 is the noisy image and η is some noise distribution added to the image. To minimize the noise and hence find the optimal solution a minimization over the space of images is given as the minimization of a functional[2]:

$$E(I; I_0) = Q(I) + \lambda C(I, I_0)$$

Where Q(I) is the quality of the image, small for clean images, large for noisy ones. C(I,I_0) is the constraint that the new image shouldn't deviate too far from the original. Defining our constraints as in [1]: 1) Denoised image should have the same mean as the noisy image $\int_\Omega u = \int_\Omega u_0$ and 2) the noisy removed should be: $\sigma^2 = \int_\Omega (u - u_0)^2$.

Optimization of the functionals in this paper requires us to take the Gateaux derivative $f_h(\epsilon) = E(I + \epsilon h; I_0)$ of the energy function and set it equal to zero $f'(0) = 0$. Then follow either gradient descent procedure or for dual methods a fixed point iteration setup as formulated in Chambolle and Chan[3, 4].

2 Higher Order Noise removal

In this section we implement the higher order PDE as developed by Lyskaer in [5]. The authors use a combination of two energies, total variation and higher order PDE defined as:

$$E1 = u = \int_\Omega |\nabla u| - \lambda(u - u_0)^2 dxdy \tag{1}$$

$$E2 = v = \int_\Omega |D^2 u| - \lambda_2 (u - u_0)^2 dxdy \tag{2}$$

They solve both energies per iteration of gradient descent, and update the image:

$$u^{n+1} = \theta^{n+1} u^n + (1 - \theta^{n+1}) v^n$$

$$v^{n+1} = u^{n+1}$$

The higher order terms are defined as:

$$D^2 u = Dxx(u) + Dxy(u) + Dyx(u) + Dyy(u)$$

Where finding Dxx would amount to finding the first derivative with respect to x, then the second derivative with respect to x. The TV term was originally solved in [1] as well as in class:

$$E1 = \int_\Omega |\nabla u| dxdy$$

$$E1 = -\nabla \frac{\nabla u}{|\nabla u|} + \lambda_1 (u - u_0) = 0$$

For the higher order term I apply the Gateaux derivative to this energy function and solve:

$$E2 = \int_\Omega |D^2 u| dxdy = \int_\Omega \sqrt{(D^2 u)^2}$$

$$\int_\Omega \sqrt{(D^2 u + D^2 \epsilon h)^2}$$

$$= \int_\Omega \frac{2(D^2 u + D^2 \epsilon h) D^2 h}{\sqrt{(D^2 u + D^2 \epsilon h)^2}}$$

Find when limit of $\epsilon \to 0$:

$$= \int_\Omega \frac{2(D^2 u)}{\sqrt{(D^2 u)^2}} D^2 h$$

Use integration by parts twice:

$$= \int_\Omega D^2 \left(\frac{(D^2 u)}{|D^2 u|}\right) h$$

$$E2 = \int_\Omega D^2(\frac{(D^2u)}{|D^2u|}) + \lambda_2(u-u_0) = 0$$
$$= D^2(\frac{Dxx(u)}{|D^2(u)|} + \frac{Dxx(u)}{|D^2(u)|} + \frac{Dxx(u)}{|D^2(u)|} + \frac{Dxx(u)}{|D^2(u)|})$$

This notation in [5] then expands this by taking the partial derivative as:

$$= Dxx\left(\frac{Dxx(u)}{|D^2(u)|}\right) + Dxy\left(\frac{Dxx(u)}{|D^2(u)|}\right) + Dyx\left(\frac{Dxx(u)}{|D^2(u)|}\right) + Dyy\left(\frac{Dxx(u)}{|D^2(u)|}\right) \tag{3}$$

Then we have our gradient descent algorithm, with α as the step size, as:

$$v^{n+1} = v^n - \alpha * \left(Dxx\left(\frac{Dxx(u)}{|D^2(u)|}\right) + Dxy\left(\frac{Dxx(u)}{|D^2(u)|}\right) + Dyx\left(\frac{Dxx(u)}{|D^2(u)|}\right) + Dyy\left(\frac{Dxx(u)}{|D^2(u)|}\right) - \lambda_2(v+u_0)\right) = 0$$

While optimizing they define θ to be a matrix applied point wise that weights the results of the two above energies. Where there are edges we want to weight it to follow TV, where there is smooth/ramp type areas we want to increase weight the higher order terms.

$$w = \theta * u + (1-\theta)v$$
$$\theta^{n+1} = \begin{cases} 1 & if|\nabla w| \geq c \\ \frac{1}{2}\cos(\frac{2\pi|\nabla w|}{c}) + \frac{1}{2} & otherwise \end{cases}$$
$$u^{n+1} = \theta^{n+1}u^n + (1-\theta^{n+1})v^n$$
$$v^{n+1} = u^{n+1}$$

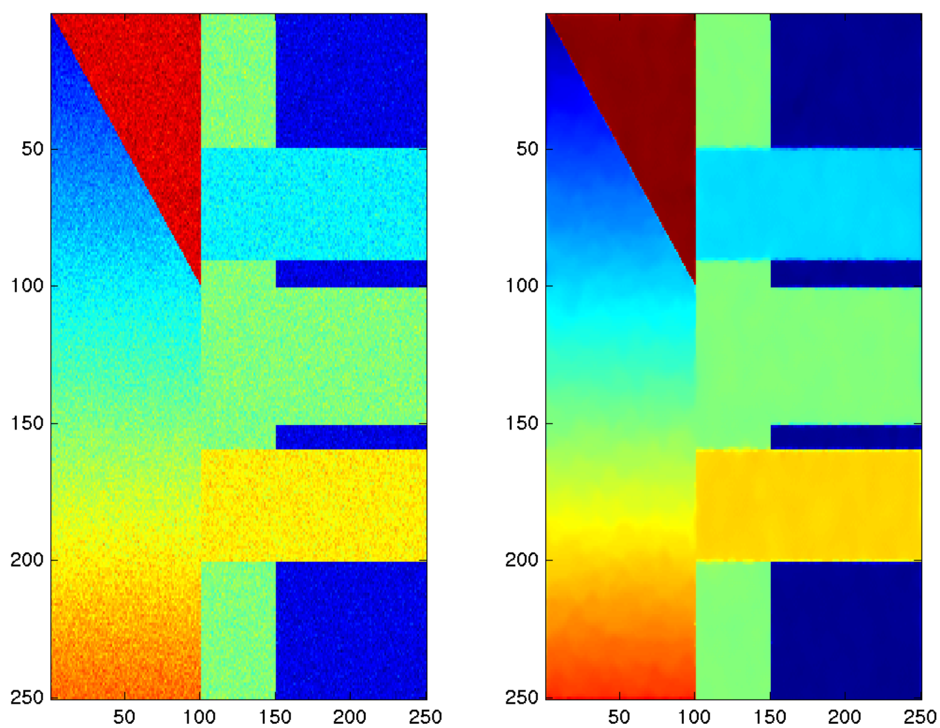
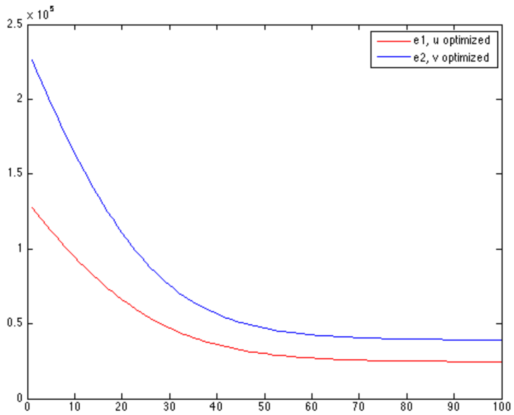
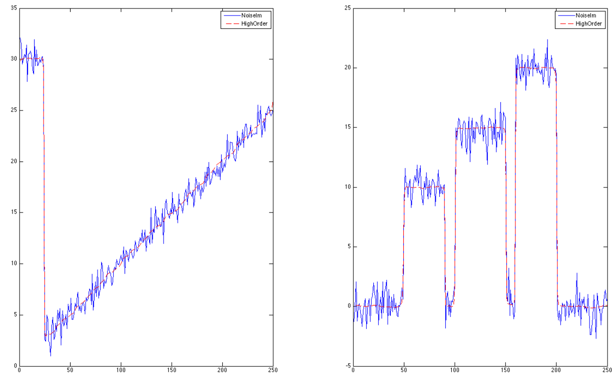


Figure 1: Right original noisy image, left is the resulting image of the convex combination algorithm as described in [5]. The resulting image looks very good with edges preserved, but some blurring. This can maybe be fixed with some parameter tuning. Staircasing seems to be totally eliminated on the ramp.



(a) Two energies are displayed. E1 relates to (1) and E2 to (2).



(b) Red line is the denoised signal. Blue is the orignal noisy data. A 1D slice along the y axis.

3 Higher Order without the partial terms

During the implementation of the above algorithm I found that Lysaker had published an earlier paper [6] where he explored using higher order terms without the mixed partial derivatives. Strictly Dxx(Dxx) and Dyy(Dyy). Surprisingly they performed very similarly for the parameters I chose.

$$E1(u) = \int_\Omega |D^2(u)| + \lambda(u-u_0)^2 = \int_\Omega |(Dxx(u_x)| + |Dyy(u_y)| + \lambda(u-u_0)^2 dxdy \tag{4}$$

$$= \int_\Omega (Dxx\left(\frac{Dxx}{|D(u)^2|}\right) + Dyy\left(\frac{Dyy}{|D(u)^2|}\right))dxdy$$

$$0 = (Dxx\left(\frac{Dxx(u)}{|D(u)^2|}\right) + Dyy\left(\frac{Dyy(u)}{|D(u)^2|}\right)) + \lambda(u-u_0)$$

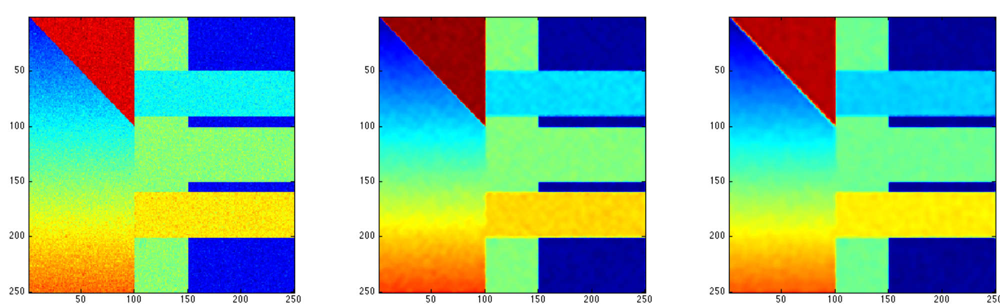


Figure 2: Right is the original image. Middle is without Dxy Dyx partial derivatives. Right is using (3) to minimize the image (without the convex combination).

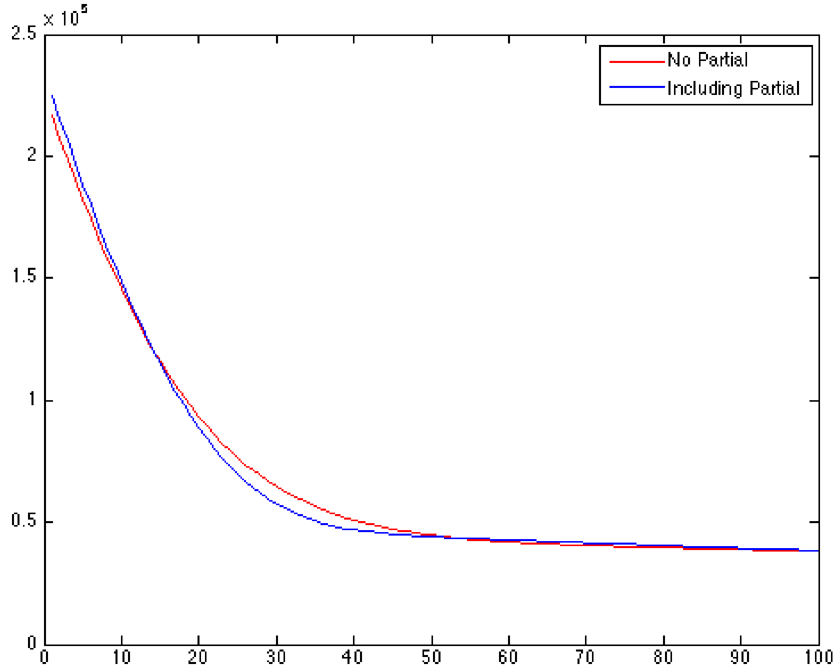
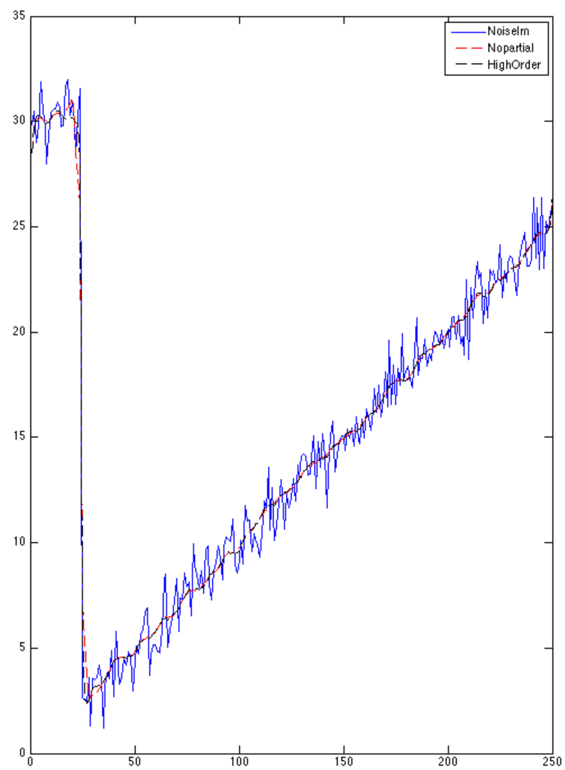
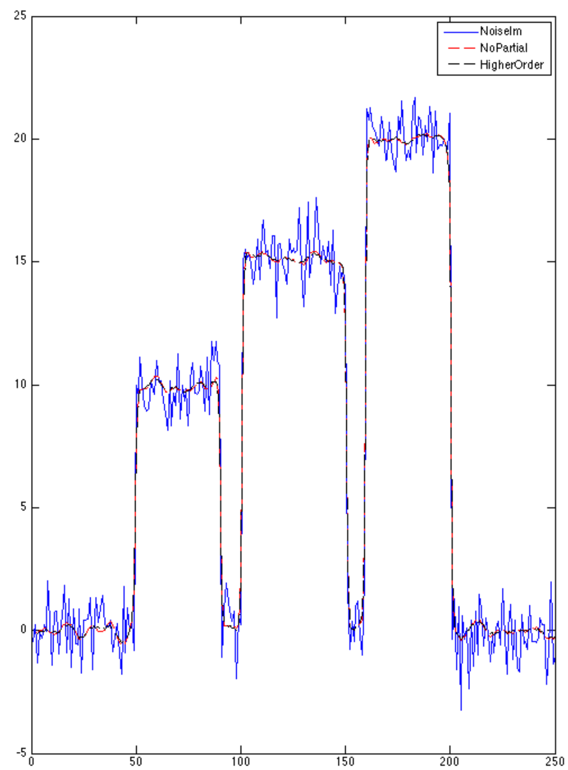


Figure 3: Energy for functional 3 and 4.





4 Higher Order Dual Methods

A problem arises with the above methods when the gradient vanishes. Many of the terms require the abs value of the gradient in the denominator, a small constant is added. In the higher order term $|D^2(u) + \beta|$ or with TV $|\nabla I + \beta|$ where $\beta \approx 1x10^{-6}$. This is undesirable because as it can change the results of the denoising algorithm.
A dual formulation for higher order methods was investigated by authors in [4]. Their method finds the Total Variation as a projection. Following the dual method by Chambolle [3] they proposed the following:

$$J(u) = \int_\Omega |\nabla I|$$

Using Legendre Fenchel Transform

$$J^*(u) = sup_u < u, v >_X - J(u)$$

$$J^*(v) = X_K(v) = \begin{cases} 0 & if v \in B_G \\ +\infty & otherwise \end{cases}$$

$$B_G = \{div\xi : \xi \in C_c^1(\Omega, \mathcal{R}^2), |\xi(x)| \leq 1 \forall x \in \Omega\}$$

A non-linear projection method was used to solve:

$$\inf_u \left(J(u) + \frac{1}{2\lambda}(f - u)^2 \right)$$

Who's solution is given by $u = f - P_{\lambda B_G}(f)$ where P is the orthogonal projector on λB_G. This projection can be found by solving for:

$$\min\{||\lambda div(p) - f||^2, |p| \leq 1 \forall i, j = 1, ..N\}$$

This p is found by taking the derivative of the above and using KKT conditions to get the Lagrangian Multipliers:

$$-(\nabla(\lambda div(p) - g)) + \alpha p = 0$$

With the conditions that $\alpha > 0$ and $|p| = 1$ or $\alpha = 0$, $\nabla(\lambda div(p) - g) = 0$ and $|p| < 1$.

$$p^{n+1} = p^n + \tau((\nabla(\lambda div(p^n) - \frac{g}{\lambda})) - |\nabla(\lambda div(p^n) - \frac{g}{\lambda})|p^{n+1})$$

4.1 Chan, Esedoglu, Park: Fourth Order Dual

Using the above formulation we want to minimize the following energy using the L2 norm on the terms on the right.

$$\inf_{u_1, u_2} \left\{ \int_\Omega |\nabla u_1| + \alpha \int_\Omega |\triangle u_2| + \frac{1}{2\lambda} \int_\Omega (f - u_1 - u_2)^2 dxdy \right\}$$

To solve these energies we break them up into two separate minimization problems. The first minimization holds u1 constant and solves for u2. The primal energy is defined as:

$$E1 = \inf_{u_2} \alpha \int_\Omega |\triangle u_2| + \frac{1}{2\lambda} \int_\Omega (f - u_1 - u_2)^2 dxdy$$

Then the dual:

$$E1_{Dual} = \min_{p, |p| \leq 1} \int \frac{\alpha\lambda}{2}(\triangle p)^2 - (f - u1)\triangle(p) dxdy$$

Solving for the dual again requires the Gateaux derivatives as mentioned previously:

$$\int \frac{\alpha\lambda}{2}(\triangle p + \epsilon\triangle h)^2 - (f - u_1)\triangle(p + \epsilon h) dxdy$$

$$\int \alpha\lambda(\triangle p + \epsilon\triangle h)\triangle h - (f - u1)(\triangle h) + (\triangle(p) + \epsilon\triangle h) * 0 dxdy$$

$$\int \alpha\lambda(\triangle p)\triangle h - (f - u1)\triangle h dxdy$$

Integration by parts:

$$\alpha\lambda\triangle^2(p)h - \triangle(f - u_1)h$$

$$\alpha\lambda\triangle^2(p) - \triangle(f - u_1)$$

$$p^0 = 0, \quad p^{n+1} = \frac{p^n - \tau A^n}{1 + |A^n|}$$

$$A^n = \triangle^2 p^n - \triangle\left(\frac{f - u_1}{\alpha\lambda}\right)$$

$$u_2^n = f - u_1 - \alpha\lambda\triangle p^n \to u_2$$

This is the solution for u2 as our iterations approach ∞. We now solve for the TV dual holding u2 constants. If we set u2 to be zero we get the dual of the TV without higher order methods. This is plotted below.

$$E2 = \inf_{u1} \int_\Omega |\nabla u_1| + \frac{1}{2\lambda}\int_\Omega (f - u_1 - u_2)^2 dxdy$$

$$E2_{Dual} = \min_{p,|p|\leq 1} \int_\Omega \left(\lambda div(p) - (f - u_2)\right)^2 dxdy$$

$$= \int_\Omega \left(\lambda div(p + \epsilon h) - (f - u_2)\right)^2 dxdy$$

$$= 2 * \nabla\left(\lambda div(p) - (f - u_2)\right)$$

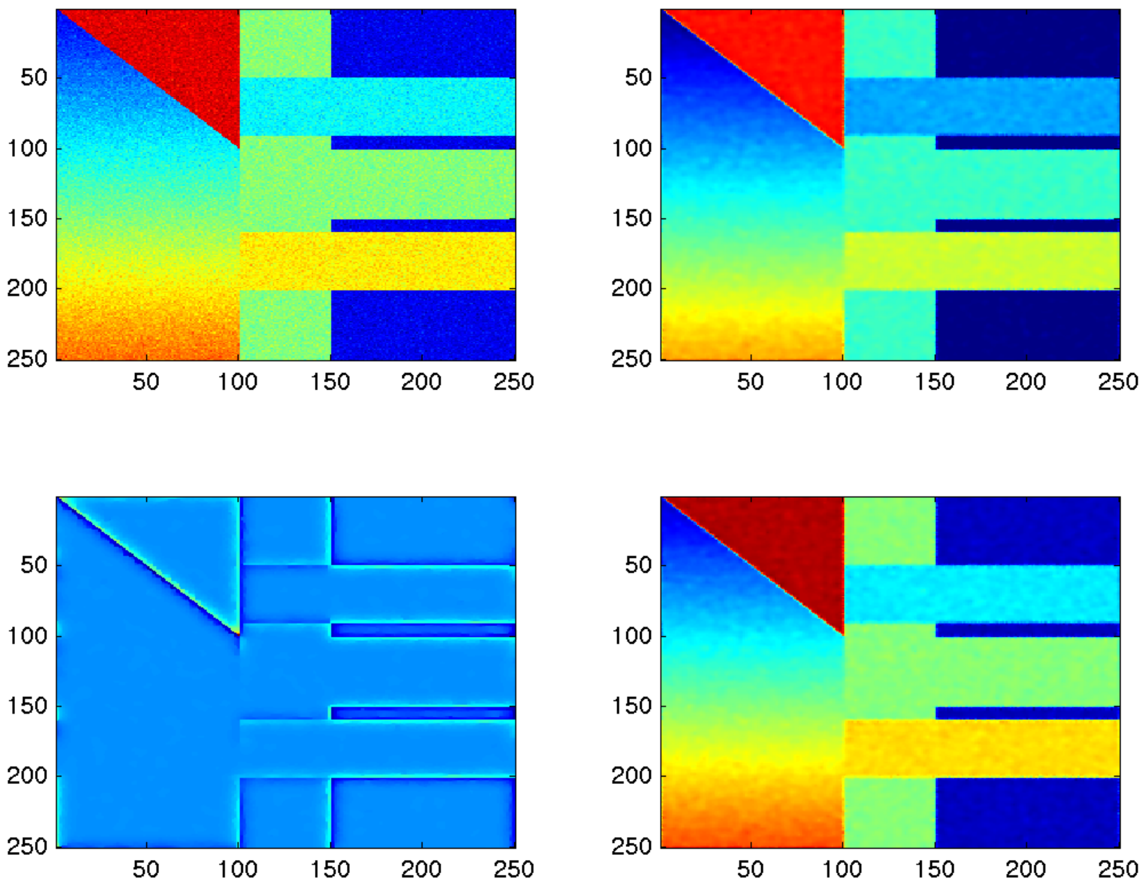
$$p^0 = 0, \quad p^{n+1} = \frac{p^n + \tau A^n}{1 + |A^n|}$$

$$A^n = \nabla\left(div(p^n) - \frac{f - u_2}{\lambda}\right)$$

$$u_1^n = f - u_2 - \lambda div(p^n) \to u_1$$

This is our solution for u1 as our iterations approach ∞. We get our final image by the simple addition:

$$u = u_1 + u_2$$



(a) Caption

Figure 4: Higher Order Primal Dual Results. Top left is the original noisy image. Top right is the final image u=u1+u2. Bottom left is the u1 term. Bottom right is the u2 term (this term was optimized first).

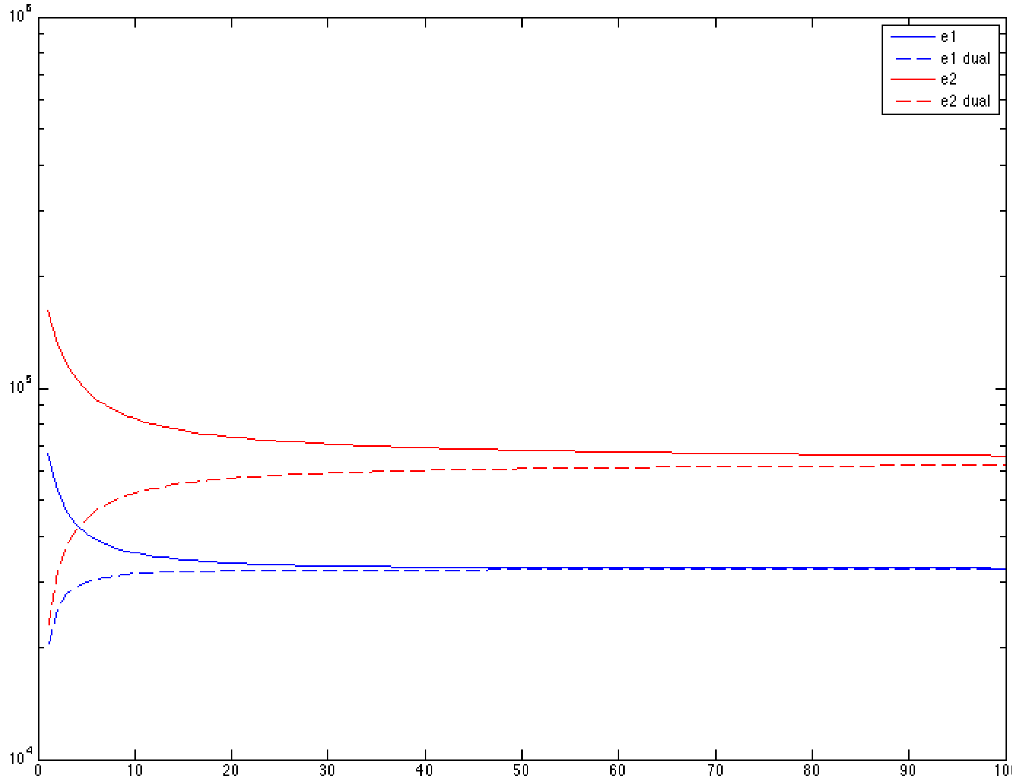


Figure 5: Minimization of the energies of the primal and dual. For both E1 and E2. We notice the duality gap, changing our parameters: α or λ changes this gap.

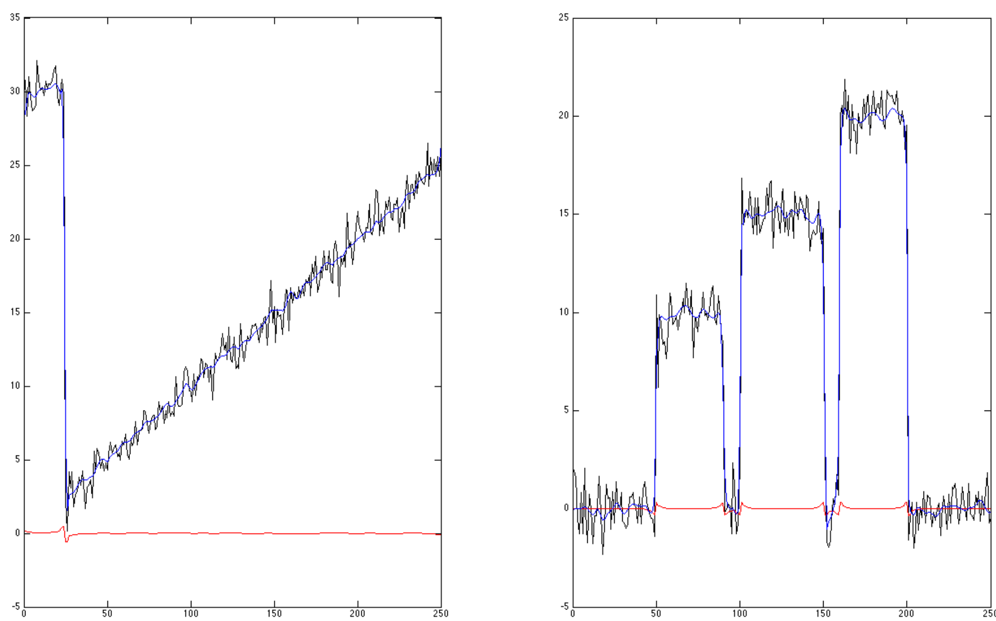


Figure 6: 1D cross section of the final higher order primal dual.

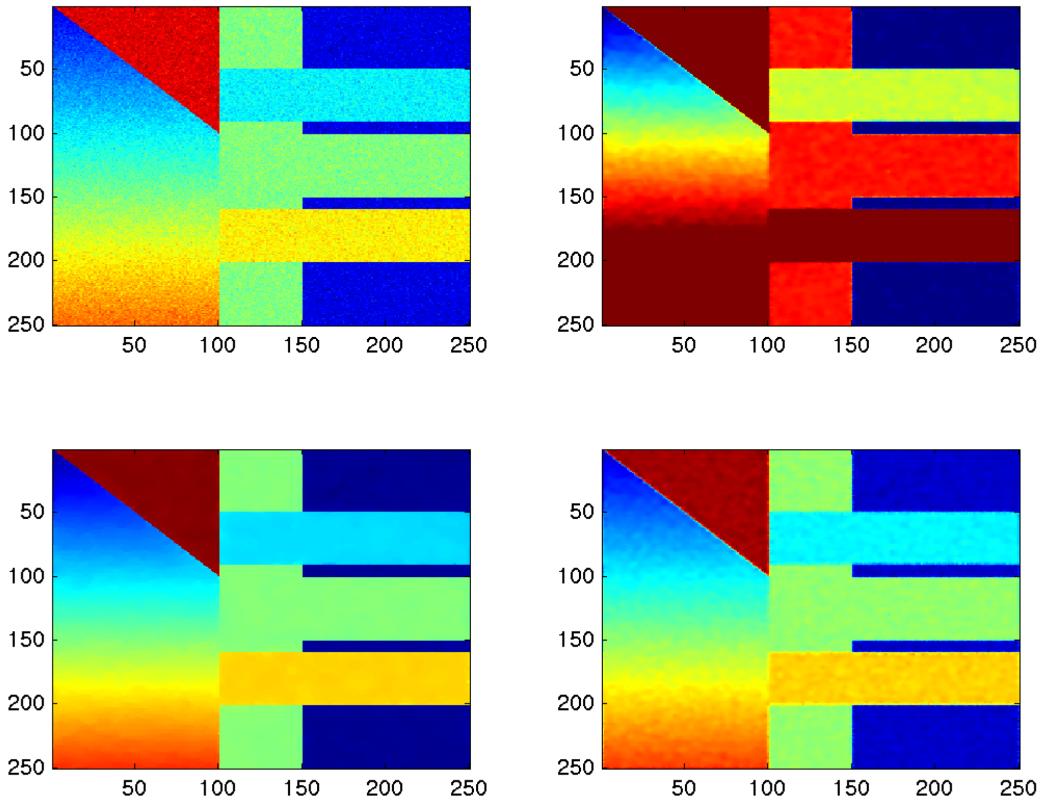


Figure 7: Setting u2 = 0 in the algorithm when solving for u1 solves for the TV dual(Rudin,Osher, Fatemi) model. The higher order dual is also found independently. Top left is the noisy image, top right can be ignored as it is the addition of TV and higher order which is meaningless. Bottom left is the TV dual. Bottom right is the higher order dual.

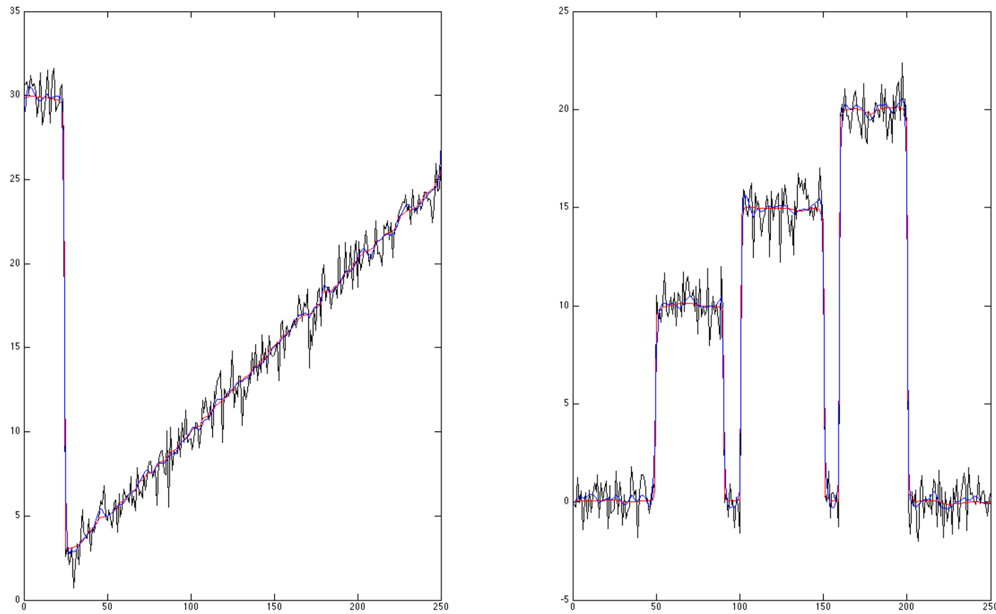


Figure 8: 1D cross seciton/comparison of the dual solutions. Red is the higher order term. Blue is the TV term.

References

[1] Leonid I Rudin, Stanley Osher, and Emad Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D: Nonlinear Phenomena*, 60(1):259–268, 1992.

[2] Laurent Younes. Mathmatical image analysis ams 493.

[3] Antonin Chambolle. An algorithm for total variation minimization and applications. *Journal of Mathematical imaging and vision*, 20(1-2):89–97, 2004.

[4] Tony F Chan, Selim Esedoglu, and Frederick Park. A fourth order dual method for staircase reduction in texture extraction and image restoration problems. In *2010 IEEE International Conference on Image Processing*, pages 4137–4140. IEEE, 2010.

[5] Marius Lysaker and Xue-Cheng Tai. Iterative image restoration combining total variation minimization and a second-order functional. *International journal of computer vision*, 66(1):5–18, 2006.

[6] Marius Lysaker, Arvid Lundervold, and Xue-Cheng Tai. Noise removal using fourth-order partial differential equation with applications to medical magnetic resonance images in space and time. *IEEE Transactions on image processing*, 12(12):1579–1590, 2003.