Regular length submission

A General Iterative Regularization Framework For Image Denoising

Michael R. Charest Jr.[†], Michael Elad[‡], Peyman Milanfar[§]

Abstract

Many existing techniques for image denoising can be expressed in terms of minimizing a particular cost function. We address the problem of denoising images in a novel way by iteratively refining the cost function. This allows us some control over the tradeoff between the bias and variance of the image estimate. The result is an improvement in the mean-squared error as well as the visual quality of the estimate. We consider three different methods of updating the cost function and compare and contrast them. The framework presented here is extendable to a very large class of image denoising and reconstruction methods. The effectiveness of the proposed methods is amply illustrated on a variety of examples.

Keywords: regularization, image denoising, iterative, bias, variance

I. Introduction

Consider the noisy image $\mathbf{y}$ given by

$$\mathbf{y} = \mathbf{x} + \mathbf{v} \tag{1}$$

where $\mathbf{x}$ is the true image that we would like to recover and $\mathbf{v}$ is zero-mean additive white noise with no assumptions made on its distribution. Note that for ease of notation we carry out all of our analysis with vectors representing 1-D signals, though the treatment is valid in multiple dimensions.

A very general technique for estimating $\mathbf{x}$ from the noisy image $\mathbf{y}$ is to minimize a cost function of the form

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} C(\mathbf{x}, \mathbf{y}). \tag{2}$$

Some specific examples of this are when $C(\mathbf{x}, \mathbf{y}) = H(\mathbf{x}, \mathbf{y}) + J(\mathbf{x})$ and $H(\mathbf{x}, \mathbf{y}) = \frac{1}{2}\|\mathbf{y} - \mathbf{x}\|^2$. The estimate then becomes

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} \left\{ \frac{1}{2}\|\mathbf{y} - \mathbf{x}\|^2 + J(\mathbf{x}) \right\} \tag{3}$$

where $J(\mathbf{x})$ is a convex regularization functional such as those in Table I. The parameter λ controls the amount of regularization.

For the regularization term corresponding to the bilateral filter, $\mathbf{S}^n$ is a matrix shift operator and $\mathbf{W}_n$ is a weight matrix where the weights are a function of both the radiometric and spatial distances between pixels in a local neighborhood.

Figure 1 is an example of the denoising ability of the bilateral filter on an image with added white Gaussian noise. By looking at the estimate residual we notice that we have removed some of the high frequency content of the image along with the noise. This is true of any denoising technique. We now turn our attention to recovering this lost detail.

II. Iterative Regularization Methods

The general framework that we present here seeks to improve our image estimate by iteratively updating the cost function of our choosing. We can express this as

$$\hat{\mathbf{x}}_k = \arg\min_{\mathbf{x}} C_k(\mathbf{x}, \mathbf{y}). \tag{4}$$

Various manifestations of this iterative regularization procedure exist. We present three different algorithms for performing the cost function update, and briefly describe their interrelation. Each algorithm seeks to extract lost detail from the the residual $\mathbf{y} - \hat{\mathbf{x}}_k$ in a unique way.

Using the operator $\mathcal{B}(\cdot)$ to denote the net effect of the minimization in (4) we formulate the iterative regularization methods as

1) $\hat{\mathbf{x}}_{k+1} = \mathcal{B}\left(\mathbf{y} + \sum_{i=1}^{k}(\mathbf{y} - \hat{\mathbf{x}}_i)\right)$,
2) $\hat{\mathbf{x}}_{k+1} = \mathcal{B}(\mathbf{y}) + \left(\sum_{i=1}^{k}\mathcal{B}(\mathbf{y} - \hat{\mathbf{x}}_i)\right)$, and
3) $\hat{\mathbf{x}}_{k+1} = \mathcal{B}(\mathbf{y}) + \sum_{i=1}^{k}\left(\mathcal{B}(\mathbf{y}) - \mathcal{B}(\hat{\mathbf{x}}_i)\right) = (k+1)\mathcal{B}(\mathbf{y}) - \sum_{i=1}^{k}\mathcal{B}(\hat{\mathbf{x}}_i)$.

The first method was recently presented in [1] by Osher et al. The second method is a generalization of Tukey's "twicing" method [5], and the third is a novel method that we introduce and study here.

† Electrical Engineering Dept., Univ. of California, Santa Cruz, 95064 (email: charest@soe.ucsc.edu)
‡ Computer Science Dept., The Technion-Israel Institute of Technology, Haifa, Israel (email: elad@cs.technion.ac.il)
§ Electrical Engineering Dept., Univ. of California, Santa Cruz, 95064 (email: milanfar@ee.ucsc.edu)

Denosing Technique	$J(\mathbf{x})$
Tikhonov	$\frac{\lambda}{2}\|\mathbf{x}\|^2$
Total Variation [1], [2]	$\lambda\|\nabla\mathbf{x}\|_1$
Bilateral [3], [4]	$\frac{\lambda}{2}\sum_{n=-N}^{N}\left[\mathbf{x}-\mathbf{S}^n\mathbf{x}\right]^T\mathbf{W}_n\left[\mathbf{x}-\mathbf{S}^n\mathbf{x}\right]$

TABLE I

VARIOUS DENOISING TECHNIQUES AND THEIR ASSOCIATED REGULARIZATION TERM.

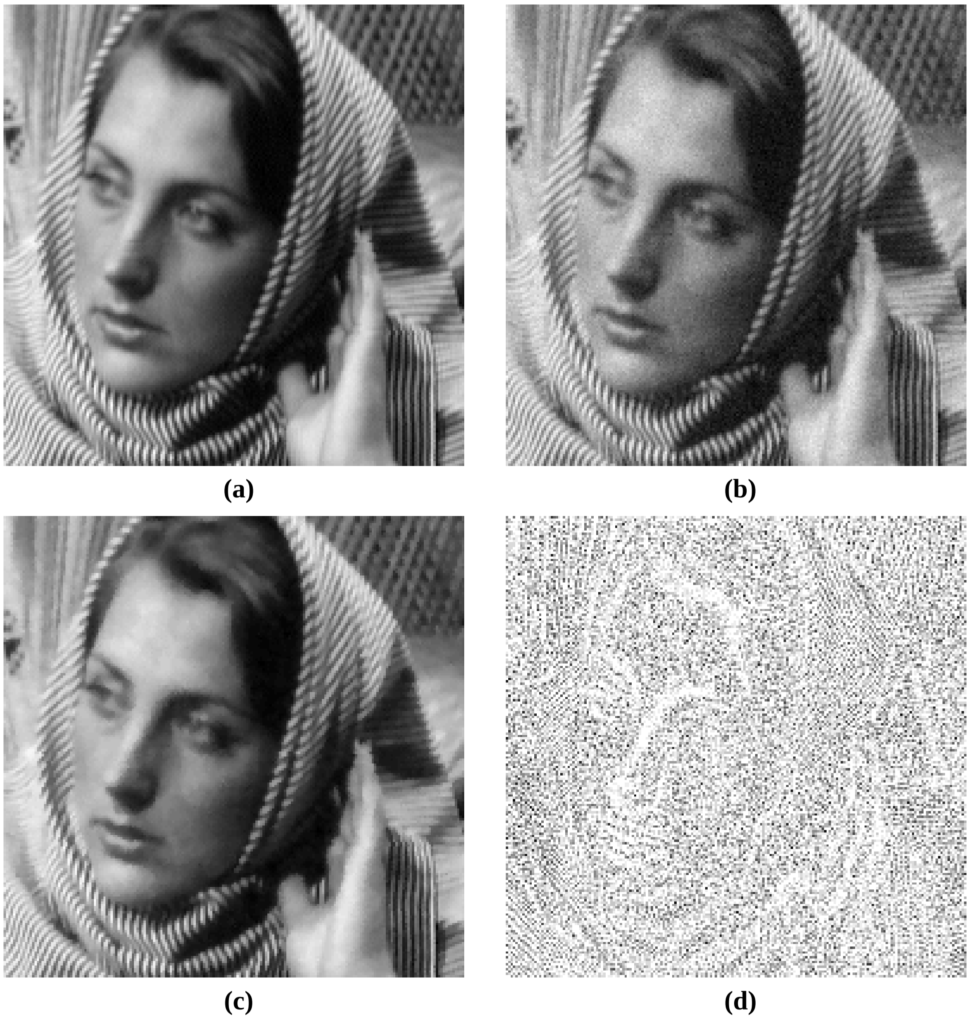


Fig. 1. **(a)** Detail of the original 'Barbara' image **(b)** 'Barbara' with added while Gaussian noise of variance 29.5 (SNR$=20$dB) **(c)** The result of minimizing the Bilateral cost function for the noisy image (b) (MSE$=19.30$) **(d)** The residual (b)-(c)

A. Osher's Iterative Regularization Method

The work of Osher et al. [1] describes an algorithm for iterative regularization using the total variation cost function ($\frac{1}{2}\|\mathbf{y}-\mathbf{x}\|^2+\lambda\|\nabla\mathbf{x}\|_1$) and the more general form $H(\mathbf{x},\mathbf{y})+J(\mathbf{x})$, though practical denoising results are limited to the total variation case. Limiting ourselves to the case when $H(\mathbf{x},\mathbf{y})=\frac{1}{2}\|\mathbf{y}-\mathbf{x}\|^2$ for simplicity, [1] improves the estimate that results from (3) via the following algorithm, which we here call "Osher's Iterative Regularization Method"

$$\hat{\mathbf{x}}_{k+1}=\mathcal{B}\left(\mathbf{y}+\sum_{i=1}^{k}(\mathbf{y}-\hat{\mathbf{x}}_i)\right), \tag{5}$$

with $\hat{\mathbf{x}}_0=0$.

This can also be written as

$$\hat{\mathbf{x}}_{k+1}=\mathcal{B}\left(\mathbf{y}+\mathbf{v}_k\right), \tag{6}$$

with $\hat{\mathbf{x}}_0 = 0$, $\mathbf{v}_0 = 0$, and

$$\mathbf{v}_{k+1} = \mathbf{v}_k + (\mathbf{y} - \hat{\mathbf{x}}_{k+1}). \tag{7}$$

The quantity $\mathbf{v}_k$ can be interpreted as a cumulative sum of the image residuals.

We note that the sum of the residuals have been added back to the noisy image and processed again using a cost function minimization. The intuition here is that if at each iteration, the residual contains more signal than noise, our estimate will improve. A block diagram illustrating this method is shown in Figure 2.

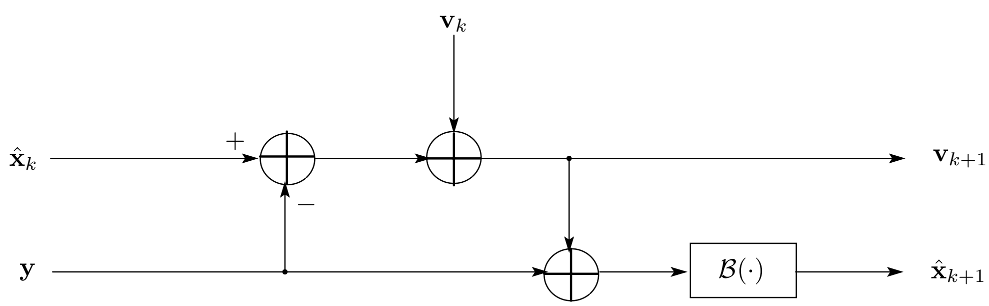


Fig. 2. Osher Block Diagram

B. *Iterative "Twicing" Regularization*

In his book [5] in the mid 1970's Tukey presented a method he called "twicing" where a filtered version of the data residual was added back to the inital estimate $\hat{\mathbf{x}}_0$ as

$$\hat{\mathbf{x}} = \hat{\mathbf{x}}_0 + \mathcal{B}(\mathbf{y} - \hat{\mathbf{x}}_0). \tag{8}$$

Tukey's original motivation for this was to provide an improved method for data fitting that would go beyond a direct fit and incorporate additional "roughness" into the estimate in a controlled way. The same year, motivated by this idea, this concept was used by Kaiser and Hamming [6] as a way of sharpening the response of symmetric FIR linear filters. Both references also mentioned the possibility of iterating this process. Thus we here call the iterated version of Tukey's Twicing, "Iterative Twicing Regularization" (ITR). We can express this as:

$$\begin{aligned}\hat{\mathbf{x}}_{k+1} &= \hat{\mathbf{x}}_k + \mathcal{B}(\mathbf{y} - \hat{\mathbf{x}}_k)\\ &= \hat{\mathbf{x}}_{k-1} + \mathcal{B}\left(\mathbf{y} - \hat{\mathbf{x}}_{k-1}\right) + \mathcal{B}\left(\mathbf{y} - \hat{\mathbf{x}}_k\right)\\ &\quad\vdots\\ &= \hat{\mathbf{x}}_1 + \sum_{i=1}^{k} \mathcal{B}\left(\mathbf{y} - \hat{\mathbf{x}}_i\right) = \mathcal{B}(\mathbf{y}) + \sum_{i=1}^{k} \mathcal{B}\left(\mathbf{y} - \hat{\mathbf{x}}_i\right)\end{aligned} \tag{9}$$

where $\hat{\mathbf{x}}_0 = 0$.

A block diagram illustrating this algorithm is shown in Figure 3.

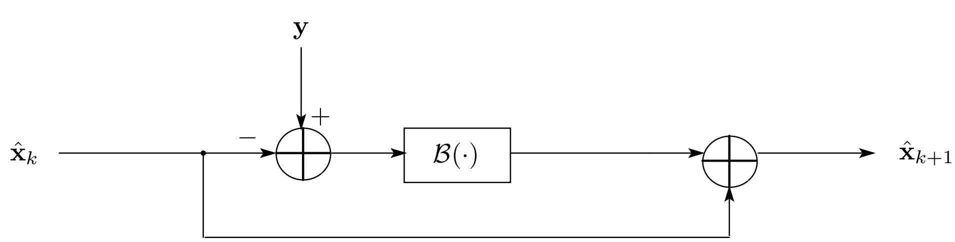


Fig. 3. Iterative Twicing Regularization Block Diagram

C. *Iterative Unsharp Regularization*

The process of unsharp masking has been used for many years to sharpen images [7]. The process consists of subtracting a blurred version of an image from the image itself. The third algorithm for iterative regularization that we present is very similar in spirit to unsharp masking. We call this method "Iterative Unsharp Regularization" (IUR) and formulate it as:

$$\begin{aligned}
\hat{\mathbf{x}}_{k+1} &= \hat{\mathbf{x}}_k + \mathcal{B}(\mathbf{y}) - \mathcal{B}(\hat{\mathbf{x}}_k) \\
&= \hat{\mathbf{x}}_{k-1} + (\mathcal{B}(\mathbf{y}) - \mathcal{B}(\hat{\mathbf{x}}_{k-1})) + (\mathcal{B}(\mathbf{y}) - \mathcal{B}(\hat{\mathbf{x}}_k)) \\
&\vdots \\
&= \hat{\mathbf{x}}_1 + \sum_{i=1}^{k} (\mathcal{B}(\mathbf{y}) - \mathcal{B}(\hat{\mathbf{x}}_i)) = \mathcal{B}(\mathbf{y}) + \sum_{i=1}^{k} (\mathcal{B}(\mathbf{y}) - \mathcal{B}(\hat{\mathbf{x}}_i))
\end{aligned} \tag{10}$$

where $\hat{\mathbf{x}}_0 = 0$.

We illustrate this method in Figure 4.

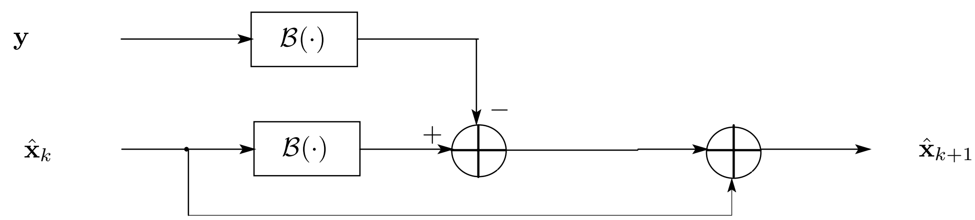


Fig. 4. Iterative Unsharp Regularization Block Diagram

D. Relationships Between the Algorithms

The three iterative regularization methods that we present here appear to differ from one another by various linear distributions of the $\mathcal{B}(\cdot)$ operator.

In general, denoising techniques can be spatially adaptive in order to preserve the high frequency detail in the estimate, as is the case with the Total Variation Filter and the Bilateral Filter. This makes the operator $\mathcal{B}(\cdot)$ non-linear. However, if we naively ignore this fact and manipulate the operator $\mathcal{B}(\cdot)$, we can relate all the aforementioned methods for iterative regularization. We will show that these resulting techniques are useful and illustrate their relative advantages for performing iterative regularization. Thus we argue that the iterative regularization method presented in [1] is actually a subset of a more general framework for iterative regularized denoising.

We note that for all three iterative regularization methods, if we continue to iterate the estimate, $\hat{\mathbf{x}}_k$ converges to the noisy data $\mathbf{y}$. However we do not provide proof of convergence in this summary.

III. Bias-Variance Tradeoff

To measure the effectiveness of the algorithms, the mean-squared error (MSE) is a natural choice. The MSE is defined as

$$mse(\hat{\theta}) = E\left[(\hat{\theta} - \theta)^2\right] \tag{11}$$

where $\hat{\theta}$ is the estimate and θ is the underlying signal. We can rewrite the MSE as

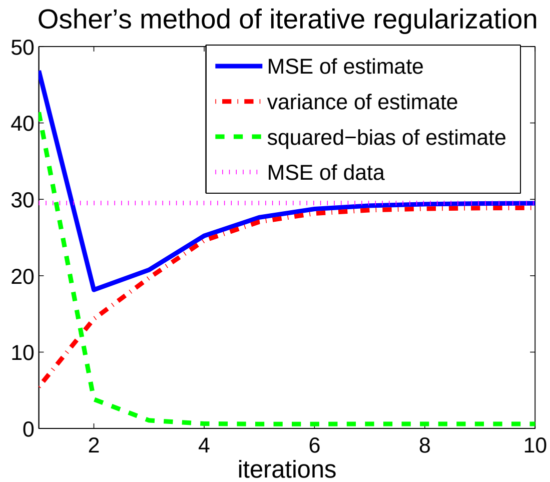
$$\begin{aligned}
mse(\hat{\theta}) &= E\left(\left[\left(\hat{\theta} - E(\hat{\theta})\right) + \left(E(\hat{\theta}) - \theta\right)\right]^2\right) \\
&= E\left[\left(\hat{\theta} - E(\hat{\theta})\right)^2\right] + 2\,E\left[\left(\hat{\theta} - E(\hat{\theta})\right)\left(E(\hat{\theta}) - \theta\right)\right] + E\left[\left(E(\hat{\theta}) - \theta\right)^2\right] \\
&= var(\hat{\theta}) + 0 + \left(E(\hat{\theta}) - \theta\right)^2 = var(\hat{\theta}) + bias^2(\theta).
\end{aligned} \tag{12}$$

Thus, as is well-known, MSE is the sum of the estimate variance and squared-bias [8].

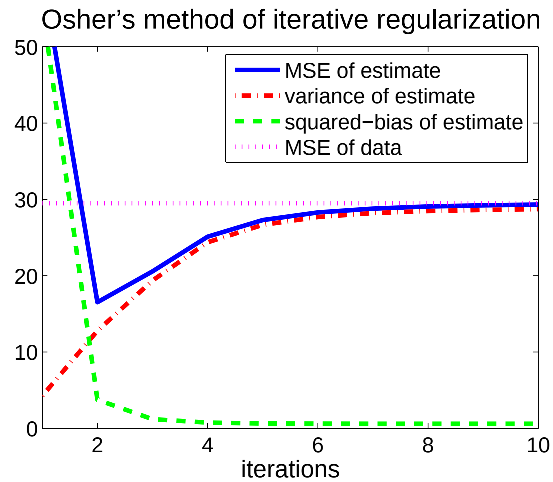
Ref. [9] presents a bias-variance tradeoff analysis for a method called L_2 boosting, which is related to the ITR method. However, some of the key assumptions made in that analysis do not apply to the iterative regularization methods that we present here (namely, in our general analysis $\mathcal{B}(\cdot)$ is a non-linear operator).

In Figure 5 we can see how the MSE, variance, and squared-bias of Osher's iterative regularization method (with Bilateral functional $J(\mathbf{x})$) are affected by iteration number. The image used for this simulation is shown in Figure 6. We use the Bilateral Filter ([10], [3]) to carry out the cost function minimization with parameters $N = 2$, $\sigma_d = 1.1$, and $\sigma_r = 35$. Notice that the squared-bias decreases as we iterate but the variance increases. The mean-square error optimal estimate occurs where the sum of these two values is at a minimum.

For all of the iterative regularization methods presented here, the bias of the first iteration of the estimate is expected to be largest. This is especially true in the case of the Total Variation and Bilateral functionals because we are using cost functions



(a) Bilateral



(b) Total Variation

Fig. 5. MSE, variance, and squared-bias of the estimates $\hat{\mathbf{x}}_k$ of the noisy image shown in Figure 6 (b) using Osher's iterative regularization method (with (a) Bilateral and (b) Total Variation regularization functionals).

(a)

(b) MSE=29.54

Fig. 6. **(a)** The original 'Barbara' image **(b)** 'Barbara' with added while Gaussian noise of variance 29.5 (PSNR= 33.43dB)

that assume that the underlying image is piece-wise constant [2], [10], [3]. Thus the first estimate, $\hat{\mathbf{x}}_1 = \mathcal{B}(\mathbf{y})$ is a piece-wise constant version of the image. That is to say that much of the high-frequency detail in the image has been removed along with most of the noise causing the image estimate to appear piece-wise flat. See Figure 7 for an example.

As we continue to iteratively refine our general cost function (2) we begin to add back some of that lost texture, thus the squared-bias begins to decrease. However, since no method is perfect, we do get a bit of noise added back as well; this causes the variance to increase. At some point in the iterative process we get the best tradeoff of restored texture and suppressed noise; this is our optimal MSE estimate.

We achieve similar results for the ITR and IUR methods, as can be seen in Figures 8, 9 and 10. The MSE, variance,and squared-bias for each of the methods have been calculated via Monte-Carlo simulation (using 50 noise realizations for each method). The Bilateral Filter parameters were chosen to yield the best MSE for each of the methods.

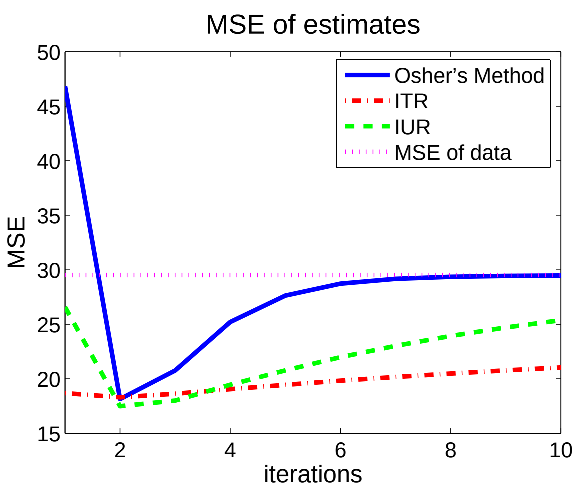
IV. Experiments

Table II provides a brief summary of the combinations of iterative regularization methods and functionals that we illustrate in this section.

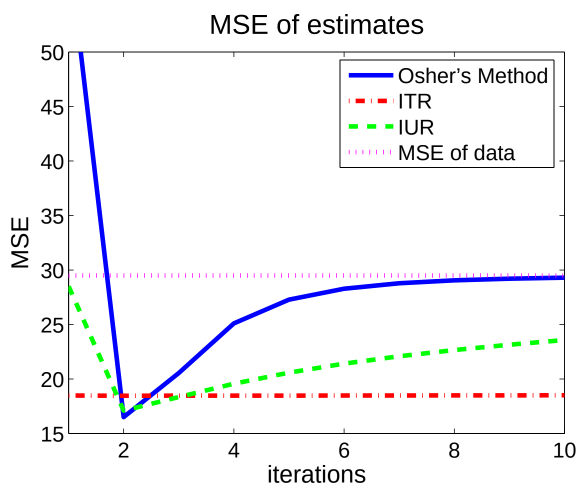
(a) An image with texture

(b) The piece-wise constant version of the image

Fig. 7. Both the Bilateral Filter and the Total Variation filter make an assumption of an underlying piece-wise constant image. Thus the estimates that result from these processes have a piece-wise flat appearance.



(a) Bilateral



(b) Total Variation

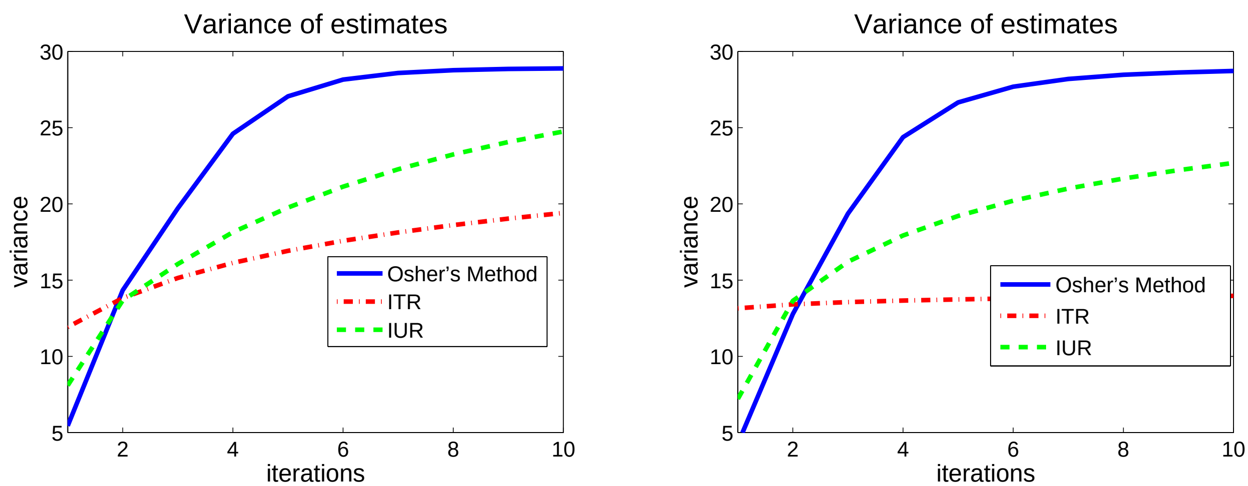
Fig. 8. MSE of the estimates $\hat{\mathbf{x}}_k$ of the noisy image shown in Figure 6 (b) using Osher's iterative regularization method, IUR, and ITR (with (a) Bilateral and (b) Total Variation regularization functionals).

A. Iterative Regularization Using Total Variation Functional

For this experiment we add white Gaussian noise of variance $\sigma^2 = 29.5$ to the image 'Barbara' shown in Figure 6 (a). The resulting noisy image has a PSNR of 33.43dB and is shown in Figure 6 (b). For this experiment we have selected the Total Variation Filter [2] which has some control parameters that determine the filter weights. These parameters, as well as our regularization parameters, are tuned by hand until the we obtain the estimate with the lowest mean-squared error for each of the iterative regularization methods. In all three iterative regularization methods, 50 steepest-descent steps were used to minimize the cost function as each iteration. The values of λ used for Osher's method, ITR, and IUR respectively were: $\lambda = .18$, $\lambda = .7$, and $\lambda = .32$. The best MSE estimates produced by Osher's regularization method, ITR, and IUR are shown in Figure 11 (a), (c), and (e) respectively. The best way to see the subtle differences between the methods is to look at the residual $|\mathbf{y} - \hat{\mathbf{x}}_k|$, thus these are shown as well. A residual that contains less structure and looks more like pure noise is an indication of a better denoising algorithm. The MSE, variance, and squared-bias of these examples correspond to the plots in Figures 5 (a), 8 (a), 9 (a), and 10 (a).

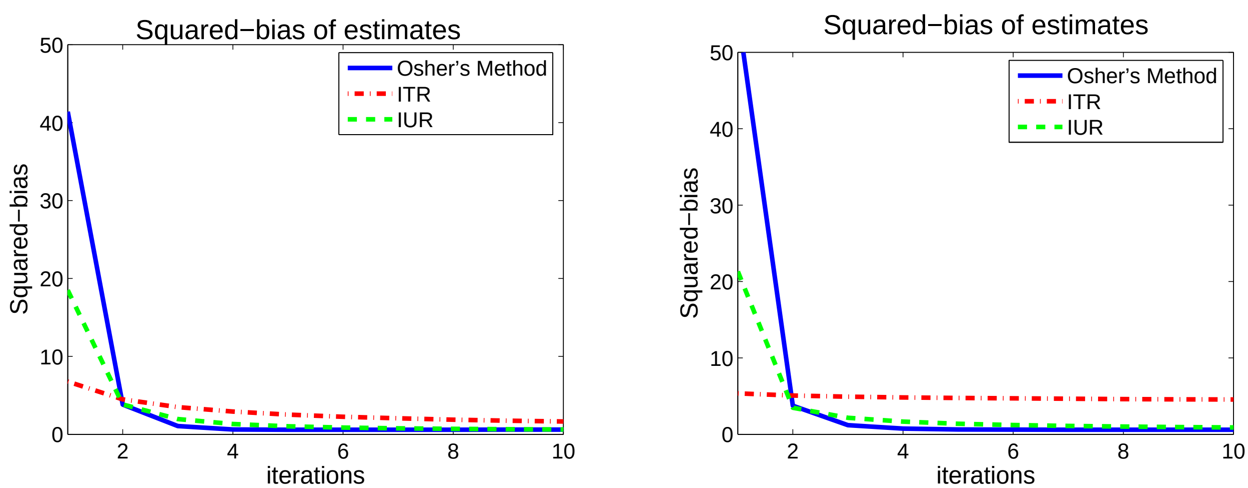
B. Iterative Regularization Using Bilateral Functional

We repeat the same procedure as the previous experiment using the Bilateral Filter instead of the Total Variation Filter.



(a) Bilateral **(b)** Total Variation

Fig. 9. Variance of the estimates $\hat{\mathbf{x}}_k$ of the noisy image shown in Figure 6 (b) using Osher's iterative regularization method, IUR, and ITR (with (a) Bilateral and (b) Total Variation regularization functionals).



(a) Bilateral **(b)** Total Variation

Fig. 10. Squared-bias of the estimates $\hat{\mathbf{x}}_k$ of the noisy image shown in Figure 6 (b) using Osher's iterative regularization method, IUR and ITR (with (a) Bilateral and (b) Total Variation regularization functionals).

The Bilateral Filter has three user defined parameters: kernel size N, geometric spread σ_d, and photometric spread σ_r. We list the values for these parameters that we used in this experiment for completeness and refer the reader to [10] and [3] for more information on the Bilateral Filter. In all three iterative regularization methods we used $N = 2$ and $\sigma_d = 1.1$. The values of σ_r used for Osher's method, ITR, and IUR respectively were: $\sigma_r = 35$, $\sigma_r = 14$, and $\sigma_r = 23$. The best MSE estimates of the three iterative regularization methods and their residuals are shown in Figure 12. The MSE, variance, and squared-bias of these examples correspond to the plots in Figures 5 (b), 8 (b), 9 (b), and 10 (b).

V. Conclusions

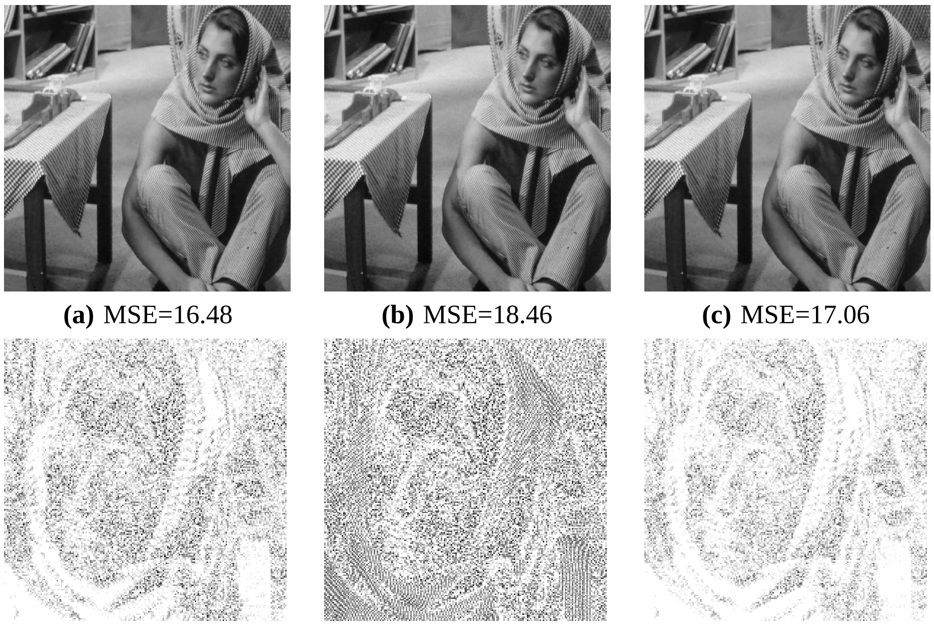
Denoising algorithms that can be formulated as (2) such as the Bilateral Filter and Total Variation Filter are frequently used due to their ease of implementation and effectiveness. We have shown that the iterative regularization methods which we present here can improve on the results of these algorithms.

Osher's method appears to give the best results in the experiment where the Total Variation functional is used. However, IUR gives better results in the Bilateral experiment. Clearly the different iterative regularization methods are useful and the "best" method to use can vary depending on the regularization functional and possibly even the particular image. This leaves much room for further investigation.

	Total Variation	**Bilateral**
Osher's	*Figure 11 (a)*	Figure 12 (a)
ITR	Figure 11 (b)	Figure 12 (b)
IUR	Figure 11 (c)	*Figure 12 (c)*

TABLE II

THE DIFFERENT EXAMPLES WE PRESENT HERE. THE LOWEST MSE RESULT IN EACH COLUMN IS ITALICIZED.



(a) MSE=16.48 **(b)** MSE=18.46 **(c)** MSE=17.06

(d) **(e)** **(f)**

Fig. 11. **(a)** The result of applying Osher's iterative regularization method using the Total Variation Filter **(b)** The result of applying IUR using the Total Variation Filter **(c)** The result of applying ITR the Total Variation Filter **(d)** Detail of the residual of (a) **(e)** Detail of the residual of (b) **(f)** Detail of the residual of (c)

REFERENCES

[1] S. Osher, M. Burger, D. Goldfarb, J. Xu, and W. Yin, "An iterative regularization method for total variation-based image restoration," *SIAM Multiscale Model. and Simu.*, vol. 4, pp. 460–489, 2005.

[2] T. Chan, S. Osher, and J. Shen, "The digital TV filter and nonlinear denoising," *IEEE Trans. on Image Processing*, vol. 10, no. 2, pp. 231–241, 2001.

[3] M. Elad, "On the origin of the bilateral filter and ways to improve it," *IEEE Trans. on Image Processing*, vol. 11, no. 10, pp. 1141–1151, 2002.

[4] S. Farsiu, M. Robinson, M. Elad, and P. Milanfar, "Fast and robust mulitframe super resolution," *IEEE Trans. on Image Processing*, vol. 13, no. 10, pp. 1327–1344, 2004.

[5] J. Tukey, *Exploratory Data Analysis*. Addison-Wesley, 1977.

[6] J. Kaiser and R. Hamming, "Sharpening the response of a symmetric nonrecursive filter by multiple use of the same filter," *IEEE Trans. on Acoustic Speech and Signal Processing*, vol. ASSP-25, no. 5, pp. 415–422, Oct. 1977.

[7] R. Gonzalez and R. Woods, *Digital Image Processing, second edition*. Pearson Education, 2002.

[8] S. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*, A. Oppenheim, Ed. Prentice Hall PTR, 1993.

[9] P. Buhlmann and B. Yu, "Boosting with the L_2 loss: Regression and classification," *Journal of the American Statistical Association*, vol. 98, no. 462, pp. 324–339, 2003.

[10] C. Tomasi and R. Manduchi, "Bilateral filtering for gray and color images," *Proceedings of the IEEE International Conference on Computer Vision*, pp. 836–846, 1998.

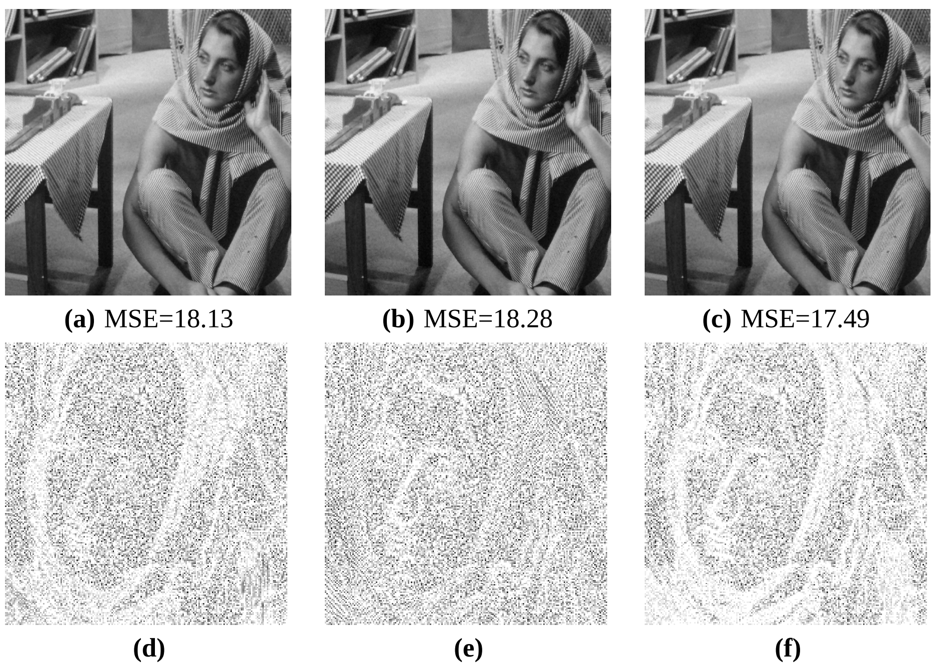


Fig. 12. **(a)** The result of applying Osher's iterative regularization method using the Bilateral Filter **(b)** The result of applying IUR using the Bilateral Filter **(c)** The result of applying ITR the Bilateral Filter **(d)** Detail of the residual of (a) **(e)** Detail of the residual of (b) **(f)** Detail of the residual of (c)