On Over-Parameterized Model Based TV-Denoising

Tal Nir and Alfred M. Bruckstein
Computer Science Department
Technion - Israel Institute of Technology
Haifa 32000 Israel
Email: taln,freddy@cs.technion.ac.il

***Abstract*— An over-parametrization model based total variation signal and image denoising method is proposed and analyzed in this paper. In cases where some structural information is provided on the signals or images of interest, this method may lead to substantial improvements in denoising performance.**

I. Introduction

The Total Variation (TV) approach for image denoising was first proposed by Osher-Rudin-Fatemi [1]. In the TV approach, we seek for a minimizer of a functional by a variational approach, where, the regularization term in the functional measures the smoothness of the solution via an L_1 norm. The TV regularization approach is relevant not only to images but also to general (one or higher dimensional) function regularization, or in other problems, such as optical flow computation. For example, Papenberg-Bruhn-Brox-Didas-Weickert [2] proposed to estimate the optical flow using L_1 regularization in both the data and smoothness terms in their variational formulation. The advantages of an L_1 regularizer resulted in many state of the art variational methods. Such a regularizer is defined on the border between convex and non-convex functions, it is able to preserve sharp discontinuities, and it is less sensitive to outliers than its L_2 alternative. In [3] we demonstrate the advantages of representing the optical flow field using over-parameterized models. We here discuss a novel approach to regularization which relies on over-parameterizing the space of functions over which we seek a variational solution.

II. Problem Formulation

Let us start, for pedagogical reasons, with functions in one dimension. Suppose we have an ideal continuous function $f_{Ideal}(x)$ that we wish to estimate from its noisy samples: $f_{Noisy}(x) = f_{Ideal}(x) + n(x)$ where, $n(x)$ is zero mean white Gaussian noise. Let us define the following functional

$$E(f) = \int (f(x) - f_{Noisy}(x))^2 dx + \alpha \int (f'(x))^2 \, dx. \qquad (1)$$

The first term (data/fidelity term) requires that the solution would be close to the measurement and the second term (regularization or smoothness term) is a data independent regularization term which enforces a smooth solution. The resulting Euler-Lagrange equation is

$$f(x) - f_{Noisy}(x) - \alpha f''(x) = 0. \qquad (2)$$

The constant α is the relative weight of the regularization. For $\alpha \to 0$ we obtain the trivial solution of $f(x) = f_{Noisy}$. For $\alpha \to \infty$ we obtain the solution $f''(x) = 0$ from the Euler-Lagrange equation, which results in a linear solution for $f(x)$. However, Equation (1) shows that only the constant solution produces a finite penalty of the functional. Therefore, in this case the resulting solution is constant and by observing the data term, the value of this constant is the average of f_{Noisy} over the integration interval. The solution of the EL equations can be numerically sought after via a gradient descent algorithm which initializes $f(x) = f_{Noisy}(x)$ and then performs gradient descent iterations of the form $f_{k+1}(x) = f_k(x) - \triangle t(f_k(x) - f_{Noisy}(x) - \alpha f_k''(x))$. Let us now replace the L_2 regularizer by a TV L_1 regularizer of the form

$$E(f) = \int (f(x) - f_{Noisy}(x))^2 dx + \alpha \int \sqrt{f'(x)^2 + \varepsilon^2} dx. \qquad (3)$$

The resulting Euler-Lagrange equation

$$2(f(x) - f_{Noisy}(x)) - \alpha \frac{d}{dx}\left(\frac{f'(x)}{\sqrt{f'(x)^2 + \varepsilon^2}}\right) = 0. \qquad (4)$$

After a few algebraic manipulations we obtain

$$2(f(x) - f_{Noisy}(x)) - \alpha \frac{\varepsilon^2 f''(x)}{(f'(x)^2 + \varepsilon^2)^{1.5}} = 0. \qquad (5)$$

One could also replace the L_2 norm in the data term with an L_1 norm of equation (3). In this case, we obtain the functional

$$E(f) = \int \sqrt{(f(x) - f_{Noisy}(x))^2 + \varepsilon^2} dx + \alpha \int \sqrt{f'(x)^2 + \varepsilon^2} dx, \qquad (6)$$

that yields the Euler-Lagrange equation,

$$\frac{f(x) - f_{Noisy}(x)}{\sqrt{(f(x) - f_{Noisy}(x))^2 + \varepsilon^2}} - \alpha \frac{d}{dx}\left(\frac{f'(x)}{\sqrt{f'(x)^2 + \varepsilon^2}}\right) = 0$$

Figure 1 shows a piecewise constant function with one discontinuity and the reconstruction result using the L_1 norm TV regularizer of Equation (3), and observe that the discontinuity is well preserved.

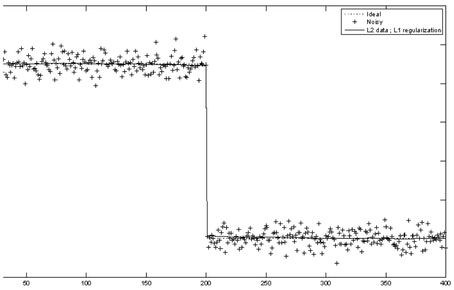

Figure 1. Example 1 with L_2 data term and L_1 regularization.

III. Total Variation of Piecewise Parametric Functions

Let us analyze the case of a clean signal with $f_{Noisy}(x) = f_{Ideal}(x)$, and ask ourselves: what types of functions can the TV denoising functional recover faithfully, i.e. recover without an error? In order for any function to be represented without an error, disregarding the noise, all we have to do is to substitute $f(x) = f_{Ideal}(x)$ in the corresponding Euler-Lagrange equation and check if it is satisfied. In this substitution, the data term vanishes and all we have to do is check the term corresponding to the regularizer. Looking at equations (2) and (5), we observe that for both the L_1 and L_2 regularization methods, the Euler-Lagrange equation is satisfied for $f(x) = f_{Ideal}(x) = f_{Noisy}(x)$ if and only if $f''_{Ideal}(x) = 0$ everywhere. The solution of this simple differential equation leads to a linear function. We conclude that the previously presented denoising functionals can faithfully reproduce linear functions. From a practical point of view, even if the ideal function can not be exactly represented, this does not necessarily mean that the functional is inadequate. We have to keep in mind that due to noise, the restored function would generally be only an approximation to the ideal function, and therefore, practically we would settle for errors in the noise free case that are an order of magnitude less than the errors resulting from the contribution of the noise. In the next section we will see how knowledge of the nature of the ideal function can be incorporated into the TV functional in order to increase the accuracy of the restoration.

IV. An Over-Parameterized Model for the 1D Total Variation Functional

Suppose next that we know that our ideal function is composed of a linear combination of n known "basis functions" $\phi_i(x)$

$$f_{Ideal}(x) = \sum_{i=1}^{n} A_i \phi_i(x). \tag{7}$$

In this case, the problem reduces to parameter estimation, and the use of functional minimization is unnecessary. Suppose however that we know that the ideal function may be represented with the help of basis functions $\phi_i(x)$ with varying coefficients $A_i(x)$, i.e.

$$f_{Ideal}(x) = \sum_{i=1}^{n} A_i(x)\phi_i(x), \tag{8}$$

For most choices of basis functions, we do not impose here any restrictions on the representation possibilities of any $f_{Ideal}(x)$. For more than one basis function, the representation is not unique and there may be infinite ways to represent the same $f_{Ideal}(x)$ function. Suppose for example that $\phi_0(x) = 1$. In this case, $A_0(x) = f_{Ideal}(x)$ and $A_i(x) = 0$ for all $i > 0$ is one possibility to represent any ideal function. If basis function number k, $\phi_k(x)$ is non-zero everywhere, then $A_k(x) = f_{Ideal}(x)/\phi_k(x)$ with all the other coefficients equal to zero, is another possible exact representation of $f_{Ideal}(x)$. If the coefficient functions are piecewise constant, we can still try the approach of parameter estimation, but now one also has to solve a segmentation problem of deciding on the locations of segments that have constant coefficients. In the noise free case simultaneous segmentation and parameter estimation may produce excellent results. In the noisy case on the other hand, the problem is much more difficult. The solution we propose is to incorporate the model of Equation (9) into the functional optimization, so that we now solve for $f(x)$ which is represented by

$$f(x) = \sum_{i=1}^{n} A_i(x)\phi_i(x). \tag{9}$$

Now, the regularization is expressed in terms of the coefficients $A_i(x)$ instead of $f(x)$. Also, the Euler-Lagrange equations are written in terms of the coefficients instead of $f(x)$. For an L_2 functional we would use

$$\begin{aligned} E(A_i) &= \int \left(\sum_{i=1}^{n} A_i(x)\phi_i(x) - f_{Noisy}(x) \right)^2 dx \\ &\quad + \alpha \int \sum A_i'(x)^2 dx. \end{aligned} \tag{10}$$

For an L_1 regularizer and an L_2 data term we get

$$\begin{aligned} E(A_i) &= \int \left(\sum_{i=1}^{n} A_i(x)\phi_i(x) - f_{Noisy}(x) \right)^2 dx \\ &\quad + \alpha \int \sqrt{\sum_{i=1}^{n} A_i'(x)^2 + \varepsilon^2} dx \end{aligned} \tag{11}$$

This is the over-parameterized generalization of the classical total variation functional in $1D$. Note that for piecewise constant coefficients of an ideal function in Equation (8), the solution $f(x)$ can model the ideal solution with no regularization penalty except at the points of discontinuity in the model coefficients. For the sake of simplicity we avoided writing explicit weights for the derivatives of the different coefficients in the above functional since we may choose (equivalently) to scale the basis functions. Since we now minimize the functional with respect to the coefficients, there are now n Euler-Lagrange equations which can be written for the most general choice of basis functions. The Euler-Lagrange equation for coefficient number q is

$$2\left(\sum_{i=1}^{n} A_i(x)\phi_i(x) - f_{Noisy}(x)\right)\phi_q - \alpha A_q''(x) = 0 \quad (12)$$

for the functional of Equation (10), and

$$2\left(\sum_{i=1}^{n} A_i(x)\phi_i(x) - f_{Noisy}(x)\right)\phi_q - \quad (13)$$
$$\alpha\frac{d}{dx}\left(\frac{A_q'(x)}{\sqrt{\sum_{i=1}^{n} A_i'(x)^2+\varepsilon^2}}\right) = 0$$

for the functional of Equation (11). There are n coupled Euler-Lagrange equations for $q = 1...n$. Note that for the choice of one basis function $\phi_1(x) = 1$, the coefficient of this basis function becomes the function itself $f(x) = A_1(x)$ and the Euler-Lagrange Equation (13) reduces to Equation (4). This shows that our method may be viewed as a generalization of the classical total variation framework when the over-parametrization is used with the constant basis function model. Without noise, a perfect reconstruction is achieved for the case of an ideal function with linearly varying coefficients as can be seen by substitution of the ideal function into the Euler-Lagrange Equations (12) and (13). The new regularization strategy obtained by the over-parameterization functional penalizes for changes in the underlying model coefficients instead of penalizing for any change in the function itself. In Example 2, consider the denoising of a piecewise quadratic function with two regions of different quadratic coefficients, shown in Figure 2. Since according to our previous discussion, if the model coefficients are allowed to change linearly and the basis functions are also linear, the overall space of exact representation of noise free ideal functions would be quadratic. The classical and over-parameterized regularization reconstruction errors are shown in Figure 3.

In the examples of this section, we used central first and second derivatives and reflecting boundary conditions. The parameters are set to $\alpha = 80$ and $\varepsilon = 0.1$. We used gradient descent with 8000 iterations. The linear basis functions for the over-parameterized method where: $\phi_1 = 1$ and $\phi_2 = 0.4(x - 200.5)$, where, a scaling factor of 0.4 was empirically chosen, and 200.5 is the midpoint of the span of the x values which range from 1 to 400.

V. Over-Parameterized Image Representations for Total Variation Image Denoising

The derivation of the results from denoising of $1D$ functions to $2D$ functions is straightforward. The classical total variation for $2D$ functions (images) as presented in [1] has an L_2 measure for the data term and an L_1 measure for the regularizer

$$E(f) = \int\int (f(x,y) - f_{Noisy}(x,y))^2 dxdy \quad (14)$$
$$+\alpha\int\int \sqrt{f_x(x,y)^2 + f_y(x,y)^2 + \varepsilon^2 dxdy}$$

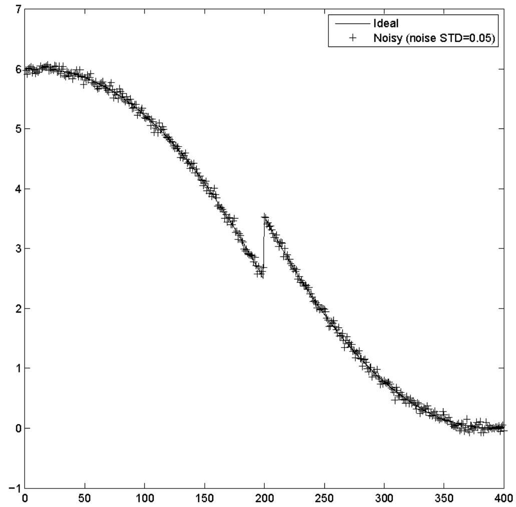

Figure 2. Piecewise quadratic function of example 2.

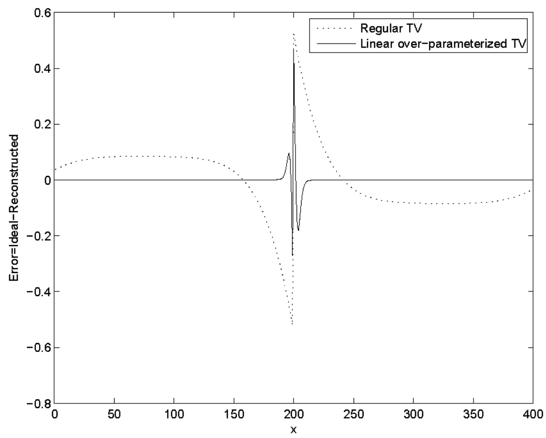

Figure 3. Example 2: reconstruction errors for the regular (dotted) versus linear over-parameterized total variation.

The resulting Euler-Lagrange equation is

$$2(f(x,y) - f_{Noisy}(x,y)) - \quad (15)$$
$$\alpha \cdot \text{div}\left(\frac{\nabla f(x,y)}{\sqrt{\|\nabla f(x,y)\|^2+\varepsilon^2}}\right) = 0$$

The over-parameterized representation of the $2D$ function $f(x,y)$ is

$$f(x,y) = \sum_{i=1}^{n} A_i(x,y)\phi_i(x,y), \quad (16)$$

where now, both the coefficients and basis functions are $2D$ functions. The over-parameterized generalization of the functional of Equation (14) is

$$E(f) = \int\int\left(\sum_{i=1}^{n} A_i(x,y)\phi_i(x,y) - f_{Noisy}(x,y)\right)^2 + \alpha\sqrt{\sum_{i=1}^{n}\|\nabla A_i(x,y)\|^2 + \varepsilon^2}\,dxdy \quad (17)$$

The n Euler-Lagrange equations corresponding to the functional of Equation (17) are obtained for $q = 1...n$

$$E(f) = \left(\sum_{i=1}^{n} A_i(x,y)\phi_i(x,y) - f_{Noisy}(x,y)\right)\phi_q(x,y) - \alpha \cdot \text{div}\left(\frac{\nabla A_q(x,y)}{\sqrt{\sum_{i=1}^{n}\|\nabla A_i(x,y)\|^2 + \varepsilon^2}}\right) \quad (18)$$

Here too, as can be observed from Equation (18), the terms resulting from the regularization are all multiplied by a second derivative (∂_{xx}, ∂_{yy} or ∂_{xy}) of some coefficient (as can be seen after expansion of the div operator), therefore, as in the $1D$ case, the Euler-Lagrange equations are satisfied for any noise free ideal function which can be described by linearly varying coefficients (of the form: $A_i = C_1 + C_2x + C_3y$) multiplying the set of basis functions of our choice.

1) Prisoner's room example: Let us now examine an illustrative synthetic image denoising example: suppose a prisoner in a jail is secretly taking photographs of his cell. If the pictures were taken with a flashlight, they would have looked something like the left image in Figure 4, however since the room was quite dark (right image of Figure 4) and using the flash would be too dangerous, the photo came out dark. When the prisoner finished serving his sentence of imprisonment, he went to see a friend who studied computer vision who wanted to help him get a better image of his cell. First he stretched the gray level values by uniform scaling. Unfortunately, the noise became also more visible (inherent signal to noise problem) as shown in Figure 5 left, so he tried the total variation denoising approach with the results shown on the middle image of Figure 5, where the bars are blurred. He then tried the over-parameterized total variation approach using the information the prisoner (who happened to be a smart engineer) gave him: the light changed linearly on every plane of the room and the light falling on the bars on the window formed a harmonic pattern in the x direction of the image with frequency of $f = 0.25[1/pixel]$. Using this information his friend constructed 3 basis functions: $\phi_1 = 1$, $\phi_2 = sin(2\pi fx)$ and $\phi_3 = cos(2\pi fx)$. The image and error using this method are shown in Figure 6. As can be seen the pattern of the bar is much better reconstructed using the proper basis functions. In terms of signal to noise ratio, the noisy image has a $PSNR = 33.96dB$, the regular total variation has $PSNR = 28.05dB$ (due mainly to errors in the region of the bars), the over-parameterized model has $PSNR = 35.79dB$, all compared to the noise free image.

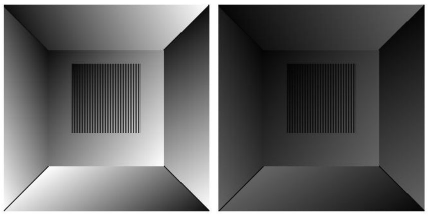

Figure 4. Prisoner's room. Left - Ideal image. Right - Dark photo with noise STD=5.

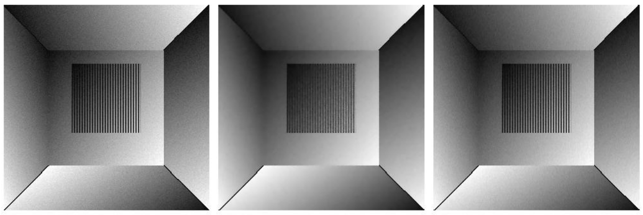

Figure 5. Prisoner's room. Left - Stretched image. Middle - Image after total variation denoising. Right - Image after total variation denoising with the over-parameterized model.

VI. Conclusion

In this paper, we have seen how the total variation approach can be generalized to incorporate an overparameterized model of choice. The model is used by assigning each sample/pixel its own independent set of coefficients. The regularization applied to the derivatives of the coefficients penalizes for deviations from the model parameters. Conditions for having zero representation errors for the noise free case where derived. We believe that this idea can be fruitfully exploited whenever we have some further prior information on the signals to be recovered and a prime example of this is the optic flow estimation problem. Indeed as reported in [3] we obtain the best optic flow estimation results using such an idea.

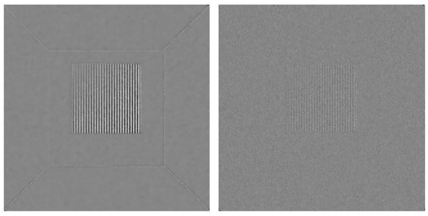

Figure 6. Error image (the same linear scaling from errors to gray-levels was used for both images. Left - Errors for the regular total variation. Right - Errors for the over-parameterized total variation.

References

[1] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," in *Proceedings of the eleventh annual international conference of the Center for Nonlinear Studies on Experimental mathematics : computational issues in nonlinear science*, 1992, pp. 259–268.

[2] N. Papenberg, A. Bruhn, T. Brox, S. Didas, and J. Weickert, "Highly accurate optic flow computation with theoretically justified warping," *Int. J. Comput. Vision*, vol. 67, no. 2, pp. 141–158, 2006.

[3] T. Nir, R. Kimmel, and A. M. Bruckstein, "Over-parameterized variational optical flow," *International Journal of Computer Vision, to appear.*