On Globally Optimal Local Modeling: From Moving Least Squares To Over-Parametrization

Shachar Shem-Tov, Guy Rosman,
Gilad Adiv, Ron Kimmel and Alfred M. Bruckstein *

Abstract - This paper discusses a variational methodology, which involves locally modeling of data from noisy samples, combined with global model parameter regularization. We show that this methodology encompasses many previously proposed algorithms, from the celebrated moving least squares methods to the globally optimal over-parametrization methods recently published for smoothing and optic flow estimation. However, the unified look at the range of problems and methods previously considered also suggests a wealth of novel global functionals and local modeling possibilities. Specifically, we show that a new non-local variational functional provided by this methodology greatly improves robustness and accuracy in local model recovery compared to previous methods. The proposed methodology may be viewed as a basis for a general framework for addressing a variety of common problem domains in signal and image processing and analysis, such as denoising, adaptive smoothing, reconstruction and segmentation.

1 Introduction

A fundamental problem in both image and signal processing is that of recovering a function, a curve or a surface (i.e., a signal or an image) from its noisy and distorted samples. Significant research effort was invested in this problem and the results obtained so far are quite remarkable. The most important ingredient in the success of any method that extracts signals from noise is, of course, the set of assumptions

Shachar Shem-Tov, Guy Rosman, Ron Kimmel, Alfred M. Bruckstein
Technion - Israel Institute of Technology, 2000 Haifa, Israel,
e-mail: shemtov/rosman/ron/freddy@cs.technion.ac.il

Gilad Adiv
Rafael, Israel, e-mail: gilad3a@gmail.com

* This research was supported by the Israel Science foundation (ISF) grant no. 1551/09.

that summarizes our prior knowledge about the properties of the signal that effectively differentiates it from the noise. These assumptions range from some vague general requirements of smoothness on the signals, to quite detailed information on the structure or functional form of the signals that might be available due to prior knowledge on their sources.

The prior information on signal/image is often expressed in the form of a parameterized model. For instance, in speech recognition [1] slowly varying coefficients of the short-time Fourier transform (STFT) are used to locally describe and model highly fluctuating spectral characteristics over time. In object recognition the choice of the correct spatial support for objects i.e., the segmentation, is a fundamental issue [2], hence, in general scene understanding, support maps are used to represent segmentation of images into homogeneous chunks, enabling the representation of objects as disjoint regions with different local modeling parameters [3]. In geometric modeling, B-splines (which are essentially a continuous set of piecewise polynomials), are used for local curve and surface approximation, interpolation and fitting from noisy samples [4]. In model-based texture segmentation [5], the selection of an appropriate support set for the local model, is important to obtain a good local texture representation, which then serves as a basis for the segmentation process. In sparse coding [6, 7], the main goal is to model data vectors (signals) as a linear combination of a few elements (support set) from a known dictionary. Sparse coding has proven to be very effective for many signal or image processing tasks, as well as advances in computer vision tasks such as object recognition.

One of the widespread and successful methods for local signal modeling, is the celebrated *moving least squares* local fitting method (MLS), which in recent years has evolved to become an important tool in both image and signal processing and in computational graphics. In [8], Levin explored the moving least-squares method and applied it to scattered-data interpolation, smoothing and gradient approximation. In [9, 10, 11] the moving least squares technique was employed for modeling surfaces from point-sampled data, and proved to be a powerful approach. This was followed by the work of Fleishman et al. [10, 12], incorporating robust statistics mechanisms for outlier removal. Common to these works is the locality of the fitting procedure and the lack of global assumptions expressing prior knowledge on the variations of local parameters.

The aim of this paper is to show how one can design variational functionals that exploit local fitting of models and global smoothness assumptions on the variations of model parameters, that are natural for various types of signals. A first attempt, at such a variational methodology, was made by Nir et al. [13, 14]. This work was subsequently broadened and generalized by Bruckstein in [15], by the realization that over-parametrization methods naturally follow from combining moving least squares, or other local fitting methods, with global priors on parameter variations. The local modeling relates to a wealth of classical methods, such as Haralick's and Watson's *facet model* for images [16] and extends them in many ways. More importantly, the discussion and experimental results reported in this paper point at a rather general methodology for designing functionals for variational model estima-

tion and signal reconstruction and focusing on the denoising problem is merely an illustrative test case.

The importance of the proposed variational framework, lies in the fact that it allows for directly incorporating knowledge of the problem domain, hence it is easily extendable to address numerous problem areas, such as denoising, deconvolution and optical flow in image and signal processing and various other fields of research. Moreover, due to the structure of the proposed functionals, our variational framework is able, unlike many common methods, to accurately recover the underlying model of a signal while addressing its main aim.

2 The local modeling of data

For the sake of simplicity, we shall here limit our discussion to one dimensional signals and address the generic problem of denoising. Let $f(x)$ be a one dimensional signal and $f_{noisy}(x) = f(x) + n(x)$ be it's noisy counterpart. Also, denote by $\{f_j = f(x_j) + n(x_j)\}$ the set of samples of the noisy signal. Suppose that f can be (locally) described by a parameterized model of the form

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

where $A = \{A_i\}_{i=1}^n$ is a set of parameters, and $\phi = \{\phi_i\}_{i=1}^n$ is a set of *'basis'* signals. *Local modeling* is the process of estimating $A = \{A_i\}$ from the noisy signal f_{noisy} or it's samples $\{f_j\}$, in the neighborhood of a point x. As a simple example, we can consider the Taylor approximation as a parameterized model with polynomial basis functions $\phi_i = x^{i-1}$.

Suppose we would like to use this local modeling for denoising our noisy data $f_{noisy}(x)$ or $\{f_j\}$. Then, around $x = x_0$, we want to estimate $A_i(x_0)$, i.e., the parameters of the model (1), by solving:

$$\arg \min_{[A_1, A_2, \ldots, A_n]} \left\| f_{noisy}(x_0) - \sum_{i=1}^{n} A_i \phi_i(x_0) \right\|, \tag{2}$$

in some local neighborhood of x_0 and a distance norm $\|\cdot\|$. This minimization gives us the best local estimate of $\hat{f}(x_0)$

$$\hat{f}(x_0) = \sum_{i=1}^{n} A_i(x_0) \phi_i(x_0). \tag{3}$$

Repeating this process for every location x give us the "moving" best estimate of f.

The choice of the distance or measure of error is of a great importance. One common choice is the weighted least squares distance, as considered, for example, by Farnebäck [17], as a generalization to the facet model:

$$\left\| f_{noisy}(x) - \sum_{i=1}^{n} A_i(x)\phi_i(x) \right\|_w = \int \left(f_{noisy}(y) - \sum_{i=1}^{n} A_i(x)\,\phi_i(y) \right)^2 w(y-x)\,dy. \quad (4)$$

This is essentially a weighted L_2 norm, where $w(\cdot)$ is a weight function which localizes the estimation of the parameters $A_i(x_0)$ (This error measure can also be applied to a sampled signal). Both the continuous and the discrete cases, described above, yield a process to compute the local model parameters $A_i(x)$ and therefore for estimating the signal $\hat{f}(x)$, not only at x_0 but at all x. This process is the well-known moving least squares estimation process.

We wish to emphasize that although our proposed methodology, discussed in the next section, focuses on the L_2 norm, this is by no means the only choice of error measure. The choice of the error measure may be adapted to each particular problem of interest. Other measures, such as the *least sum of absolute values* distance [18] or L_1 norm, can readily be substituted into the cost functionals.

3 Global priors on local model parameter variations

In the previous discussion we did not impose any conditions on the model parameters A_i, and in the definition of the neighborhood around x_0, via the weight functions, we did not use any idea of adaptation to the data.

Suppose now that we are given some structural knowledge of the signal $f(x)$. We would like to use this knowledge to improve the estimation process. For example, suppose we a priori know that $f(x)$ is a piecewise polynomial signal over a set of intervals, i.e., we know that:

$$f(x) = \sum_{i=1}^{n} A_i^r x^{i-1} \quad for \quad x \in [x_r, x_{r+1}] \quad (5)$$

but we do not know the sequence of breakpoints $\{x_r\}$. Using polynomial basis functions $\left[1, x, x^2, ..., x^{n-1}\right]$, we know a priori that a good estimate for $f(x)$ may be provided by piecewise constant sets of parameters $\{A\}$, over $[x_r, x_{r+1}]$ segments, and changes in the parameters occur only at the breakpoints $\{x_r\}$. Such knowledge provides us the incentive to impose some global prior on the parameters, such that the minimum achieved by the optimization process, will indeed favor them to be piecewise constant. This may be achieved by supplementing the moving least squares local fitting process with constraints on the variations of the parameters in the minimization process. Thus, we shall force the estimated parameters not only to provide the best weighted local fit to the data, but also to be consistent with the local fitting over adjacent neighborhoods. This is where we diverge and extend the facet model, which assigns the basis function's parameters at each point, using solely a local least square fit process.

In this paper, the assumed structural knowledge of the signal implies that good estimates can be achieved by a piecewise constant model i.e., a model whose pa-

rameters are piecewise constant. Therefore the focus of this work will be to design functionals which also impose global priors on the model parameters. We shall demonstrate that one can design a functional that does indeed fulfil this requirement.

The mix of local fitting and global regularity is the main idea and power behind *over-parameterized* variational methods and is what makes them a versatile problem solving tool. By adapting the local fitting process and incorporating the global prior (in a way that will be described in the following sections), this methodology can be readily applied to address problems in various domains.

4 The over-parameterized functional

In [14] Nir and Bruckstein presented a first attempt at noise removal based on an over-parameterized functional. This functional was, similarly to many well known functionals, a combination of two terms, as follows:

$$E(f,A) = E_D(f,A) + \alpha E_S(A). \tag{6}$$

Here α, is some fixed relative weight parameter, E_D is a data or fidelity term, and E_S is a regularization term. The data term E_D was chosen to:

$$E_D(f,A) = \int \left(f_{noisy}(x) - \sum_{i=1}^{n} A_i(x)\phi_i(x) \right)^2 dx. \tag{7}$$

Note that this functional implies a neighborhood of size 0 in the distance measure (4), which means that this data term only penalizes for *point-wise* deviations of the estimated signal via $\sum_{i=1}^{n} A_i(x)\phi_i(x)$ from f_{noisy}.

The smoothness or regularization term, E_S, was defined to penalize variations of the parameters A_i as follows

$$E_S(A) = \int \Psi \left(\sum_{i=1}^{n} A_i'(x)^2 \right) dx \tag{8}$$

where $\Psi\left(s^2\right) = \sqrt{s^2+\varepsilon^2}$.

The resulting functional yields a channel-coupled *total variation* (TV) regularization process for the estimation of the model parameters. Note that (8) is an approximated L_1 type of regularizer (sometimes referred to as the Charbonnier penalty function). This regularizer causes the functional to be more robust to outliers, and allows for smaller penalties for high data differences (compared to a quadratic regularizer), while maintaining convexity and continuity [19, 20]. The regularization term was designed to impose the global prior on the parameters. It is channel-coupled to "encourage" the parameters to change simultaneously, thus preferring a piecewise constant solution as described in Section 3.

In our experiments (which are discussed in Section 7), as well as in [14] , this functional displayed good performance for noise removal compared with Rudin Osher and Fatemi's [21] classical total variation noise removal functional. A similar functional, with data term modifications, was used by Nir et al. in [13] for optical flow estimation, producing state of the art results.

4.1 The over-parameterized functional weaknesses

Despite the good performance displayed by the over-parameterized functional, it still lacks with regard to the following shortcomings, that were clear in our experiments:

Discontinuities smearing: As mentioned, the regularization term is an approximate L_1 regularizer. A precise L_1 regularizer is indifferent to the way signal discontinuities appear, i.e., the same penalty is given to a smooth gradual signal change, and to sharp discontinuities (as long as the total signal difference is the same). See for example Pock's PhD work [22] for a detailed example. We consider this property as a shortcoming, because we expect the reconstructed parameters to be piecewise constant, where discontinuities appear as relatively few and sharp changes, hence this regularizer does not convey the intended global prior on the parameters, and does not prefer a "truly" piecewise constant solution.
In practice the problem is even more severe: first the selection of ε constant, in the Charbonnier penalty function, proves to be problematic. Choosing a bigger ε value causes the functional to lose the ability to preserve sharp discontinuities and actually prefers to smooth out discontinuities. On the other hand, choosing a smaller ε value degenerates the penalty function. In fact, for any choice of ε, this penalty function will tend to smooth sharp discontinuities. Second, as discussed above the TV-L_1 model suffers from the so called *staircasing* effect, where smooth regions are recovered as piecewise constant staircases in the reconstruction. See the work of Savage et al. [23] and references therein, for a detailed review of such effects.

Origin biasing: The over-parameterized functional's global minimum may depend on the selected origin of the model. In the over-parameterized methodology we consider basis functions which are defined globally across the signal domain. This definition requires us to fix an arbitrary origin for the basis functions. As the true parameters value may vary with a change of the origin choice, thus the value of the regularization term may also vary. For instance, consider the value of the constant term in a linear function, which determines the point at which the line crosses the y-axis. Change of the y-axis location i.e., the origin location, will incur achange in the value the constant term. This dependency on the choice of the basis function origin is termed Origin biasing. A detailed example, regarding the optical flow over-parameterized functional with affine flow basis functions, is given in the work of Trobin et al. [24].

We also note that the data term presented above only penalizes for point-wise deviation form the model, hence it imposes only a point-wise constraint on the functional's minimum, relying only on the regularization term to impose the global constraint. A discussion why this is a problematic issue is given in Section 5.3.

Overall, it is evident, that despite producing good results when applied to various applications, the over-parameterized functional model is fundamentally flawed when attempting to accomplish parameter reconstructions. On one hand the over-parameterized model provides a solution domain wider than the TV model, for the functional to "choose" from, thus often enabling convergence to excellent denoising solutions, on the other hand the constraints applied to the solution domain, through the functional, are not strong enough as to impose convergence to piecewise-constant-parameters solution, as demonstrated in Section 7.

5 The non-local over-parameterized functional

To overcome the shortcomings described in Section 4.1, we shall modify the functional, both in the data term and in the regularization term, as described below

5.1 The modified data term: a non-local functional implementing MLS

In order to overcome the point-wise character of the data term, and to impose a neighborhood constraint in the spirit of (4) in the data term, we extend it to what is commonly referred to as a *non-local functional* [15, 25, 26]. This is done simply by means of defining a weighting function which considers more then point-wise differences.

Making the data term a non-local functional, requires the parameters to model the signal well over a neighborhood of each point. We note that the robustness of the parameters estimate increases with the size of the support set. On the other hand, increasing the size of the support set too much, may reduce the functional's ability to detect discontinuities and to preserve them.

The non-local data term functional is

$$E_D = \int_x \int_y \left(f_{noisy}(y) - \sum_{i=1}^{n} A_i(x)\,\phi_i(y) \right)^2 w(x,y)\,dydx \tag{9}$$

and it conforms to the functional form of the weighted least squares fit distance defined in Section 2. Note that there is yet no restriction on the size or shape of the support set around each point that is induced by the weighting function $w(x,y)$.

For the 1D case thoroughly explored in this work we defined a simple yet powerful sliding window weighting function, as described in Section 5.1.1. The search for optimal data dependent weighting functions is an interesting possibility and will be the subject of future research.

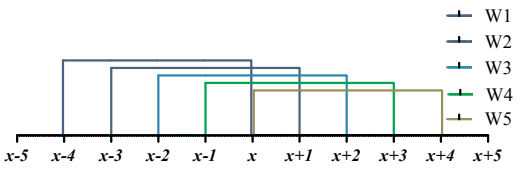
5.1.1 Weighting scheme

The choice of the weighting scheme is of great importance. For each location x_0, the weighting scheme expresses the support set from which the parameters are to be reconstructed from i.e., the set of samples which are used for the implicit MLS fitting process calculated in the minimization process. Thus the weighting scheme should be designed in such a manner which, at points near a discontinuity, will prevent the combination of samples from different sides of the discontinuity, thus enables the preservation of signal discontinuities in the reconstruction process. This implies that a simple box or gaussian window around each location x_0 will not suffice, as it may spread a discontinuity across several samples around the actual location of the discontinuity (which was evident in various experiments we performed).

In order to avoid this behavior, we defined a sliding window weighting function, which is closely related to concepts of ENO schemes used for the approximation of hyperbolic conservation laws first introduced in the fundamental work of Harten et al. [27] (a thorough description was presented by Shu in [28]). For each point x of the signal, we choose a window W of length N such that, $x \in W$ and $r(x) = \sum_{j=1}^{N} r(x, y_j)$ is minimal, where $y_j \in W$ and $r(x, y_j)$ is the non-weighted least squares fit distance at x:

$$r(x, y_j) = \left(f_{noisy}(y_j) - \sum_{i=1}^{n} A_i(x)\, \phi_i(y_j) \right)^2 . \tag{10}$$

Specifically, we chose a successive set of points of size N which include x and also minimizes $r(x)$. For example, for a window of size 5, there are 5 window possibilities as described below:



We mark the chosen window by w^x, and the selected weight function will then be:

$$w(y-x) = \begin{cases} \frac{1}{N} & if \quad y \in w^x \\ 0 & otherwise \end{cases} \tag{11}$$

By no means do we claim that this is the best choice of the weighting function. This is but one possible, adaptive weighting function selection process, that enhances the data term to impose a non point-wise constraint, while allowing for the preservation and sharpening of discontinuities in the signal and that was found to yield good results. Note that the choice of the window is sensitive to noise, therefore for achieving better results we updated the selected window at each location throughout the minimization process as described in section 6.

5.2 The modified regularization term

We extend the over-parameterized regularization term using the *Ambrosio-Tortorelli (AT)* scheme [29] in a similar manner as done in [30] for the TV functional, while retaining the L_1 penalty function as proposed by Shah [31, 32], and applied to the over-parametrization functional by Rosman et al. [33]. This functional transforms the regularization into an implicit segmentation process, each segment with it's own set of parameters. In effect, the AT scheme allows the regularization term to prefer few parameter discontinuities and to prevent discontinuities from smearing into neighboring pixels via the diffusion process, thus allowing piecewise smooth solution. This directly addresses the discontinuities smearing effect described in Section 4.1. The choice of L_1 regularizer, as opposed to the L_2 regularizer in original AT scheme , is due to the fact that a L_1 regularizer better encourages a piecewise constant solution, which is the intended global prior we wish to impose on the solution. Exploration of different sub-$L1$ regularizer function, is beyond the scope of this paper.

The chosen AT regularization term is:

$$E_{S,AT} = \int (1-v_{AT})^2 \Psi \left(\sum_{i=1}^{n} ||A_i'||^2 \right) + \rho_1 (v_{AT})^2 + \rho_2 \left\| v_{AT}' \right\|^2 \tag{12}$$

where v_{AT} is a diffusivity function, ideally serving as an indicator of the parameters discontinuities set in the signal model. Suppose we have a piecewise linear signal, and an ideal solution (A^*, v_{AT}^*) where A^* is piecewise constant, and the diffusivity function v_{AT}^* is 1 at the linear regions boundaries and 0 elsewhere. With such a solution, we expect two neighboring points, belonging to different regions, to have a very small, in fact negligible, diffusivity interaction between them. This is controlled by the value of v_{AT}^* at those points, which effectively cancels the diffusivity interaction between the different sets of linear parameters. Furthermore, the cost associated with this solution is directly due to the discontinuity set measure in the signal i.e., to

$$\int \rho_1(v_{AT})^2 + \rho_2 \left\| v'_{AT} \right\|^2 \tag{13}$$

hence the penalty no longer depends on the size of the signal difference at discontinuities. Moreover, the AT regularization addresses the origin biasing effect, described in Section 4.1, by making the functional much less sensitive to the selected origin. This is due to the fact that ideally we consider only piecewise constant parameters solutions. These solutions nullifies the regularization term energy at every location except for discontinuities where the energy depends solely on energy (13). Therefore the ideal piecewise constant solution becomes a global minimizer of the functional.

5.3 Effects of the proposed functional modifications

An obvious question arises: why do we need to modify both data and regularization terms? To answer this, we first notice that using only the non-local data term improves the local parameter estimates, but cannot prevent the discontinuities smearing effect. A moving least squares process, with small window, will yield excellent parameter estimates, but is unable to prevent the diffusion process from combining data from neighboring segments, thus smoothing and blurring the estimates at boundaries locations. Therefore we need to add the AT scheme to the functional, which has the ability to prohibit the diffusion process at discontinuities.

But then one might expect that the AT scheme without the non-local data term would suffice, by segmenting the regularization into piecewise constant regions, and relying on the data term and the global regularization to recover the correct parameters for each segment. In practice this is not the case. Consider the following illustrative test case: y_s a discontinuous signal, depicted in Figure 1, and suppose we initialize the model parameters to a smoothed version of y_s achieved by calculating the moving least squares fit solution, with a centralized window, on the clean signal. We will now discuss applying different functionals for reconstructing the model parameters from clean signal y_s and the initialized smooth parameters.

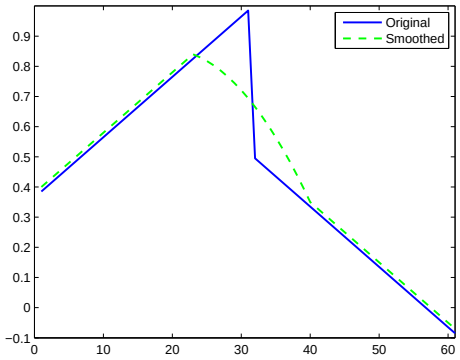


Fig. 1: Solid line - y_s a piecewise linear signal with one discontinuity. Dashed line - a smooth version of y_s with which we initialized the model parameters

Figures 2 and 3, depict snapshots of the reconstructed signal, parameters, and the resulting AT indicator function v_{AT} at different time steps of the minimization process. We used $\alpha = 0.05$ and $\alpha = 10$ respectively. These are snapshots of the minimization process of the point-wise over-parameterized functional, modified with the AT regularization term (MOP). Both minimizations were carried out until convergence. It is evident that the reconstructed parameters are not piecewise constant, as we would like to have from our prior information on the signal. Also, it is important to note that in the first series the signal is perfectly reconstructed, effectively nullifying the data term, and in the second series the signal is smoothed, however the data term energy is still very low.

In contrast, Figure 4 depicts a series of snapshots of the minimization process of the non-local over-parameterized functional (NLOP). Note that here the parameters are perfectly reconstructed, the AT indicator function v_{AT} receives a value close to one only in the vicinity of the discontinuity and also the signal is perfectly reconstructed.

Assuming that the result achieved by the NLOP functional is very close to the global minimum of both functionals, we calculated using interpolation, hypothetical steps of a minimization from the solution achieved by the MOP functional to the global minimum. We then calculated the energy of the MOP functional on each hypothetical step. It becomes obvious that the energy of the functional is raising before dropping to the energy level of the global minimum. Separate calculation of the energy of the data and the regularization terms, indicates that most of the functional energy is concentrated in the regularization term. In the transition between the local and global minimum solutions, the regularization term energy raises and dictates the total energy change, while the data term contribution is negligible.

The solutions to which the MOP functional converges and energy considerations, lead us to the conclusion that the MOP functional is converging into a local minimum solution. This ”trapping” effect is alleviated in the NLOP functional, where the presumed local minimum, achieved by the MOP, is no longer a local minimum.

This is due to the fact that it is much more difficult to drive the energy of the non-local data term close to zero and it contributes significantly to drive the parameters toward their correct values. Thus, the minimization process does not halt and continues toward the global minimum.

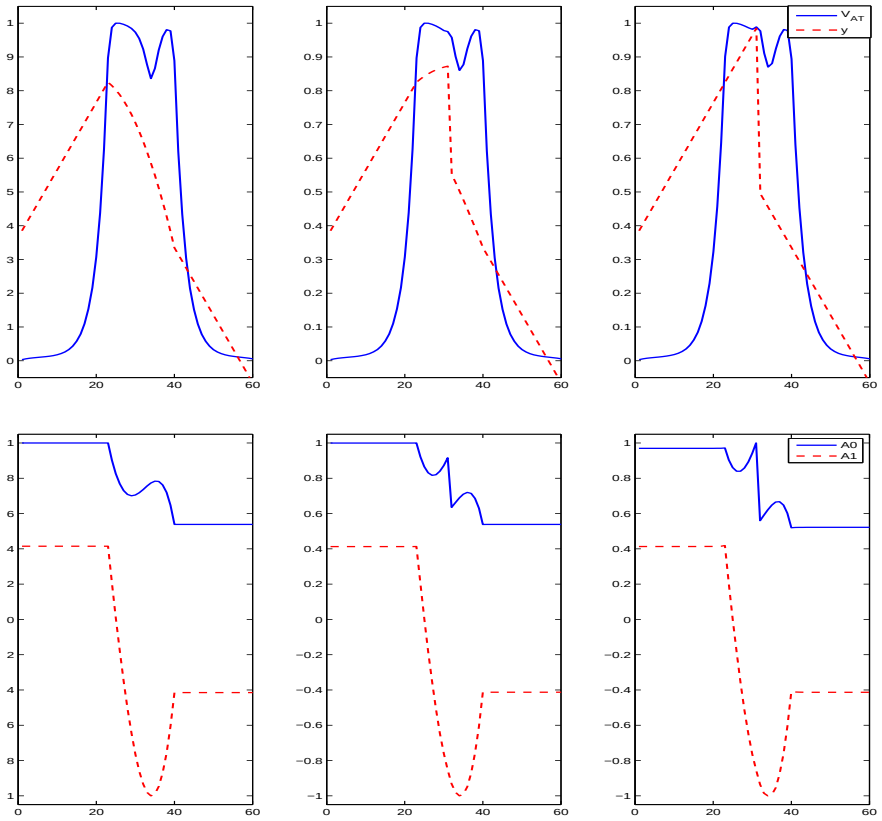


Fig. 2: From left to right, snapshots at various times, of the point-wise over-parameterized functional modified with the AT regularization term, with relative weight of E_s $\alpha = 0.05$, $\rho_1 = 7.5$ and $\rho_2 = 5$. The top image displays the reconstructed signal and the v_{AT} indicator function. The bottom image displays the reconstructed parameters.

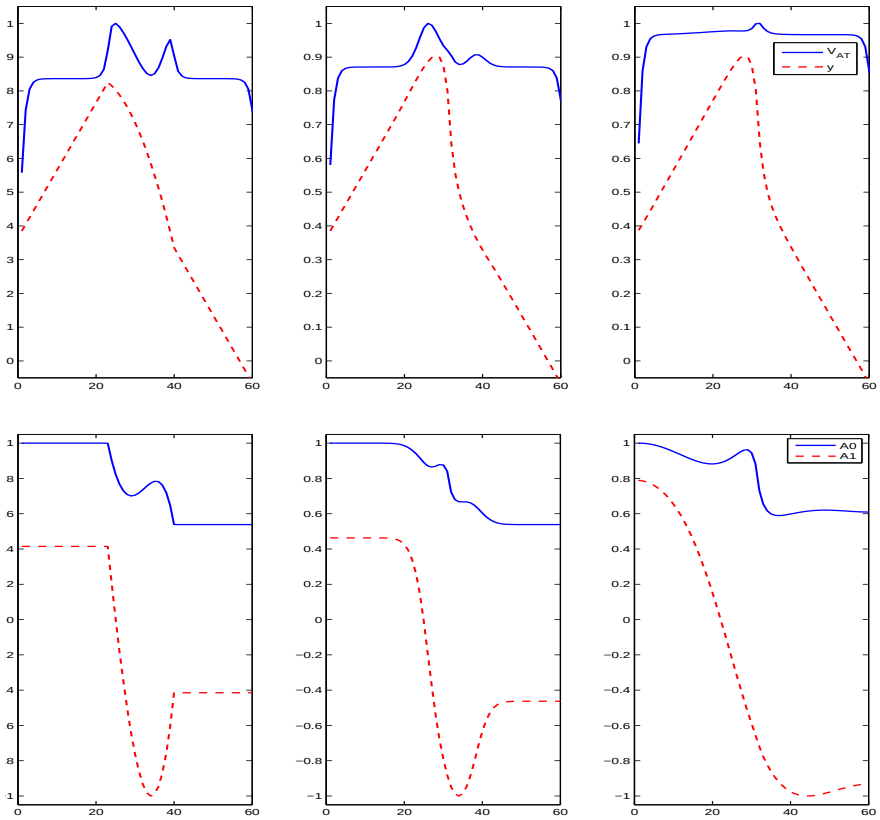


Fig. 3: From left to right, snapshots at various times, of the Mpoint-wise over-parameterized functional modified with the AT regularization term, with relative weight of E_s $\alpha = 10$, $\rho_1 = 7.5$ and $\rho_2 = 5$. The top image displays the reconstructed signal and the v_{AT} indicator function. The bottom image displays the reconstructed parameters.

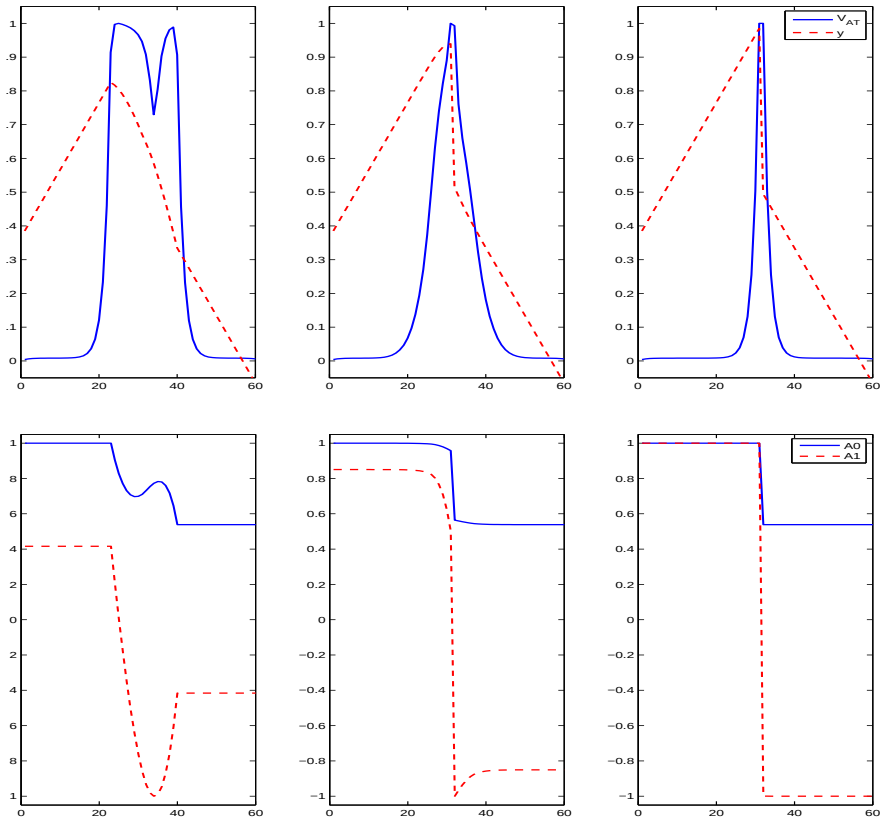


Fig. 4: From left to right, snapshots at various times, of the NLOP functional. The top image displays the reconstructed signal and the v_{AT} indicator function. The bottom image displays the reconstructed parameters.

5.4 Euler-Lagrange equations

Once we designed the functionals to be minimized, by interchangeably fixing the $A_i(x), i = 1...n$ and $v_{AT}(x)$ functions, we readily obtain the Euler-Lagrange equations which characterize the minimizers of the functional.

5.4.1 Minimization with respect to $A_q(x), q = 1...n$ (parameter minimization step).

Fixing $v_{AT}(x)$, we obtain

$$\forall q = 1...n, \quad \nabla_{A_q} E_D - \frac{d}{dx}\left(\nabla_{A'_q} \alpha E_{S,AT}\right) = 0. \tag{14}$$

the variation of the data term with respect to the model parameter functions $A_q(x)$ is given by

$$\nabla_{A_q} E_D = 2 \int_y \left(f_{noisy}(y) - \sum_{i=1}^{n} A_i(x)\, \phi_i(y) \right) \phi_q(y) w(x,y)\, dy. \tag{15}$$

For the smoothness term, the Euler-Lagrange equations are

$$\begin{aligned} \frac{d}{dx}\left(\nabla_{A'_q} \alpha E_{S,AT}\right) &= 4\alpha (1 - v_{AT})\, v'_{AT} \Psi' \left(\sum_{i=1}^{n} ||A'_i||^2 \right) A'_q \\ &+ 2\alpha (1 - v_{AT})^2 \frac{d}{dx} \left(\Psi' \left(\sum_j ||A'_i||^2 \right) A'_q \right) \end{aligned} \tag{16}$$

thus, the energy is minimized by solving the following nonlinear system of equations at each point x, $\forall q = 1...n$

$$\begin{aligned} 2 \int_y \left(f_{noisy}(y) - \sum_{i=1}^{n} A_i(x)\, \phi_i(y) \right) \phi_q(y) w(x,y)\, dy \\ - 4\alpha (1 - v_{AT})\, v'_{AT} \Psi' \left(\sum_{i=1}^{n} ||A'_i||^2 \right) A'_q \\ - 2\alpha (1 - v_{AT})^2 \frac{d}{dx} \left(\Psi' \left(\sum_j ||A'_i||^2 \right) A'_q \right) = 0. \end{aligned} \tag{17}$$

5.4.2 Minimization with respect to $v_{AT}(x)$ (AT minimization step).

Fixing the functions $A_i(x), i = 1...n$, we obtain

$$-2\alpha (1 - v_{AT}) E_s + 2\rho_1 (v_{AT}) - \rho_2 (v''_{AT}) = 0. \tag{18}$$

6 Implementation

We used central first and second derivatives and reflecting boundary conditions. In all the methods, we used various α, ρ_1 and ρ_2 constants, depending on the noise level, as is common in noise reduction methods (this was done for all the considered algorithms, with the appropriate parameters). In all the examples, we assumed a

sampling interval of $dx = 1$. For the minimization process we used gradient descent with 200000 iterations. This minimization method is a notoriously slowly converging method, but it is fast enough for our 1D example, and we intend to pursue a faster implementations in future work.

We performed an AT minimization step every 100 parameter minimization steps, and updated the weighting function every 1000 parameter minimization steps. We used a window size of 10 sample points both for the NLOP functional, and for the weighted least square fit method (used for initialization and reconstruction). We note that using a bigger window size resulted in significantly superior results on all tests, but this may not be the case in other signals. Choosing too big a window may cause an overlap between adjacent discontinuities and prevent the functional from correctly recovering them.

6.1 Initialization

In order to prevent trapping into local minima, we initialize the model parameters from the noisy data by means of a robust MLS fitting i.e., we compute each point's parameters by choosing the best least square fitting approximation via a sliding window least square calculation. This was done in exactly the same manner as the sliding window weighting function described in Section 5.1.1. The parameters chosen are those which generated the minimal reconstruction error. This computation provided a robust initialization, which already preserves, to some extent, discontinuities in the signal. This claim is strengthened by experiments we performed comparing with a regular moving least square fitting initialization. We found that with the robust initialization, the functional converges to a better solution, and nicely exposes the true discontinuities in the signal models.

7 Experiments and results

We conducted various experiments in order to verify the performance of the proposed functional. In this paper we focus mainly on 1D examples leaving 2D extensions to images for future publications, nevertheless we exhibit some initial experiments conducted on 2D synthetic images which yield a good idea of the performance to be expected in the 2D case.

7.1 1D experiments

We begin with the selection of the basis functions. We consider linear basis functions of the form:

$$\begin{cases} \phi_1 = 1 \\ \phi_2 = x \end{cases} \tag{19}$$

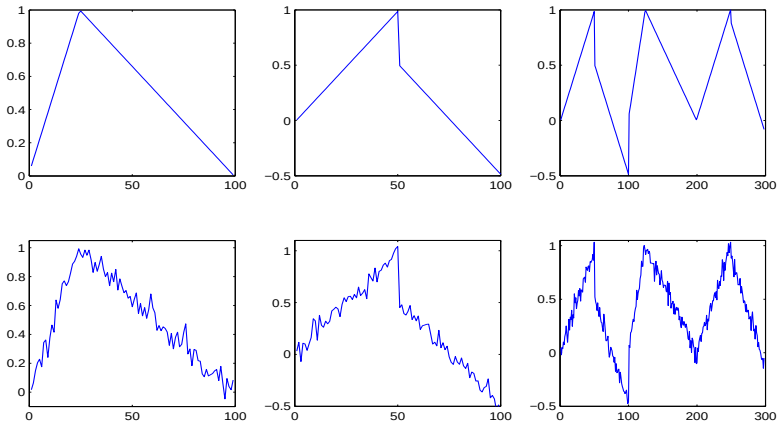
This seemingly simple choice of functions, enables us to make comprehensive tests of the functional performance. Under this choice, the functional is expected to have the best performance on piecewise linear signals. We note that this is an arbitrary choice of basis functions, and one should choose other basis functions appropriate for the signal domain.

To perform the tests we devised a set of synthetic 1D signals, and added white Gaussian noise with standard deviation ranging from $STD = 0.01$ up to $STD = 0.25$. These signals can be separated to two main groups:

- The first group is comprised of noisy piecewise linear signals. This group is interesting because it enables us to test the parameter reconstruction performance as well as noise reduction capabilities, under the optimal basis functions. The piecewise linear signals are depicted in Figure 5.
- The second group is comprised of noisy nonlinear signals, such as higher degree polynomials. This group is interesting only with regard to the noise removal performance, because generating the ground truth parameters for linear basis functions is problematic and we do not expect the linear parameters to be piecewise constant. Naturally, we could choose appropriate basis functions for these signals too, but we wanted to demonstrate that our functional preforms surprisingly well, even when applied with suboptimal basis functions. The nonlinear signals are depicted in Figure 6.

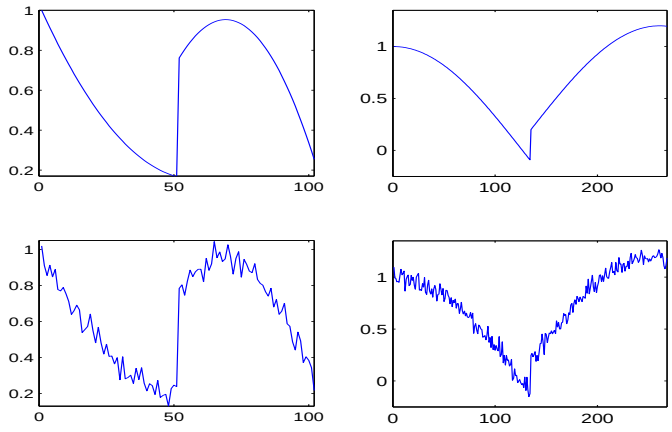
In order to the check the performance of NLOP functionals, and to test them against other denoising algorithms, we also implemented the following noise removal algorithms. The first two are the classic TV functional and the original overparameterized functional (*OP*). We chose these functionals due to their relation to the NLOP functional, enabling us to show the improvement that the NLOP has on predecessor functionals. The third and final algorithm, is the state of the art *K-SVD* noise removal algorithm, firstly proposed by Aharon et al. [7]. We used implementation published by Rubinstein et al. [34].

We compared the various algorithms noise reduction performance and, more importantly, we compared their parameter reconstruction capability. For the latter comparison, we reconstructed parameters from the K-SVD denoised signal, using our robust least square fitting method (see Section 6.1), and compared the results with both our NLOP functional and the original OP functional.



(a) One jump (b) One discontinuity (c) Many jumps & discontinuities

Fig. 5: The piecewise Linear test signals and their STD 0.05 noisy counterparts.



(a) Polynomial (b) Trig

Fig. 6: The nonlinear test signals and their STD 0.05 noisy counterparts. Left: a polynomials signal of degree 2. Right: a signal of combined sine and cosine functions.

7.2 1D Results

Noise removal performance testing, was done by comparing the L_2 norm of the *residual noise*, that is the difference between the cleaned reconstructed signal and the original signal:

$$E_{L_2} = \left\| f(x) - \sum_{i=1}^{n} A_i(x)\phi_i(x) \right\|_2 . \tag{20}$$

Parameters reconstruction performance testing (which with the linear basis function, is only relevant on the piecewise linear signals), was done by calculating the L_2 norm of the difference between the reconstructed parameters and the original parameters, with whom we generated the signal. Figure 7 depicts an example of the recovered parameters and the v_{AT} indicator function obtained for the "Many jumps & discontinuities" signal.

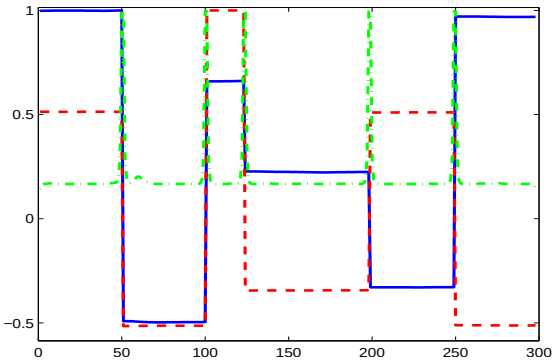
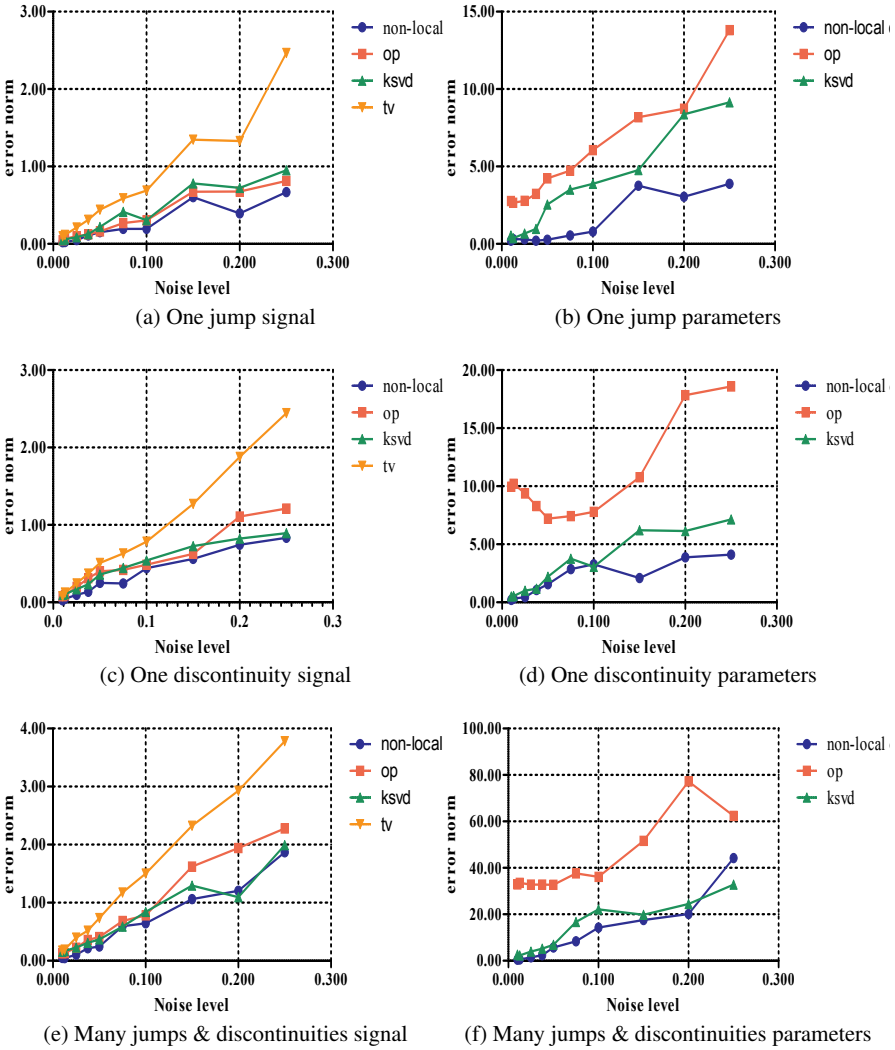


Fig. 7: Example of parameters reconstruction. Note that the indicator function tends to 1 where the signal has parameters discontinuities and tends to 0 in almost any other location.

Figure 8, displays graphs comparing performance of the various algorithms on piecewise linear signals. The left graphs display the noise removal error norms, while the right graphs display parameters reconstruction error norms.

On various noise removal performance graphs we can see the excellent performance of both OP and NLOP functionals, reinforcing the claims of outstanding noise removal performance obtained by the OP functional and maintained by our new functional. Also, we can see that when the signal contains a discontinuity, such as in the "One discontinuity" signal (as apposed to a continuous signal with only parameters discontinuity such as the "One jump" signal), the NLOP functional has greater ability to cope with the discontinuity, thus generating better results than the OP functional. In Figure 9 we display comparison of the residual noise of all the algorithms on the "One discontinuity" and the "Many jumps & discontinuities" signals.



(a) One jump signal (b) One jump parameters

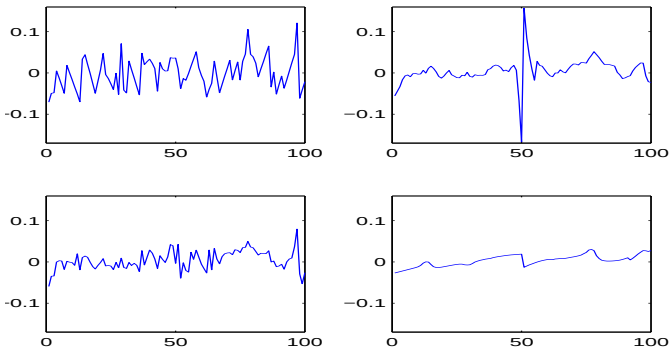
(c) One discontinuity signal (d) One discontinuity parameters

(e) Many jumps & discontinuities signal (f) Many jumps & discontinuities parameters

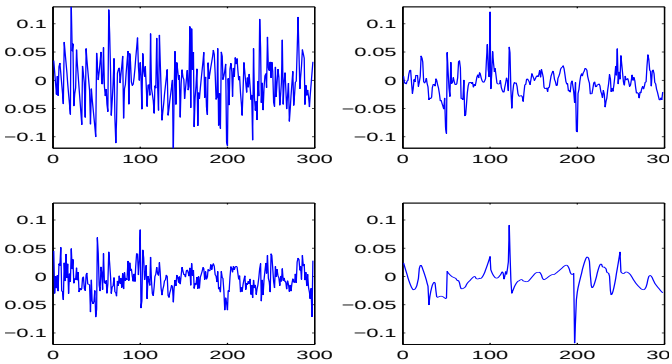
Fig. 8: Signal noise removal and parameters reconstruction comparison for the various piecewise Linear signals. On the left column depicted are graphs compering the L_2 norm of the residual noise of the various algorithms. On the right column depicted are graphs compering the L_2 norm of the various algorithms reconstructed parameters compared to the expected parameters.

When considering parameters reconstruction performance, we see a totally different picture. We see that on most cases, particularly in signals which contains signal discontinuity, the NLOP functional and the K-SVD algorithms both outperform the OP functional. This result demonstrates our claim, in Section 4.1, that

the OP functional lacks the possibility to well enforce a global prior on the reconstructed parameters. Figure 10 compares the reconstruction results of NLOP functional, OP functional and the reconstruction from the denoised K-SVD signal, on the "One discontinuity" signal. Note the reconstruction of NLOP functional is close to a piecewise constant solution, while the OP reconstruction is seemingly a smoothly changing function. In the K-SVD reconstruction, where at each point a set of parameters is chosen regardless of the choice made for it's adjacent neighbors, the lack of influence of enforcement of a global constraint is evident.

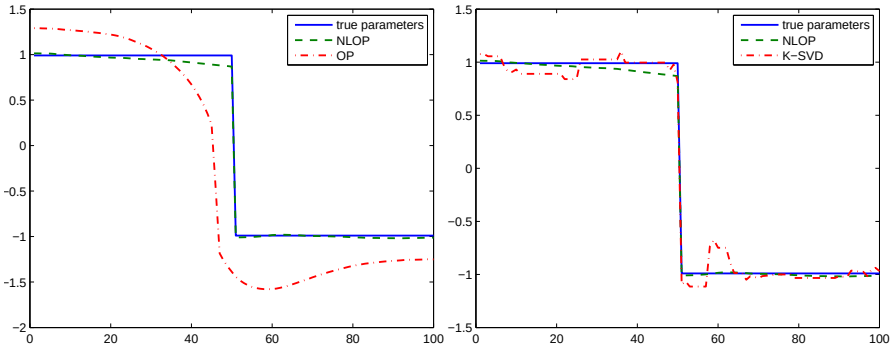


(a) Top left: TV. Top right: OP. Bottom left: K-SVD. Bottom right: NLOP.



(b) Top left: TV. Top right: OP. Bottom left: K-SVD. Bottom right: NLOP.

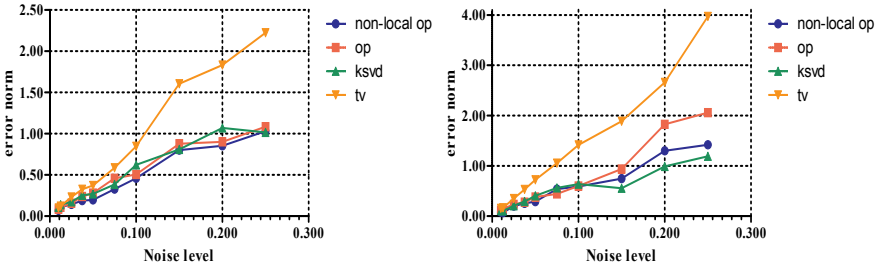
Fig. 9: Residual noise comparison of noise removal. (a): one discontinuity signal with noise $STD = 0.0375$. (b): "many jumps & discontinuities" signal with noise $STD = 0.05$.



(a) NLOP vs OP (b) NLOP vs K-SVD

Fig. 10: This figure compares the reconstructed second parameters on the various algorithms, when denoising the ”One discontinuity” signal. Left image: Comparison of parameter reconstruction between NLOP functional and OP functional on one discontinuity signal. Note how far the OP reconstruction is from a piecewise solution, while generating an excellent denoising result (seen in the relevant graph). Right image: Comparison of parameter reconstruction between NLOP functional and K-SVD algorithm on one discontinuity signal. Note the apparent lack of global constraint on the parameters in the K-SVD reconstruction.

In Figure 11, we compare the noise removal performance on the nonlinear signals. We can see that the NLOP functional still exhibits the best performance but it is not unchallenged by the OP functional. This is due to the strong constraints the NLOP functional has in trying to enforce the linear basis functions, i.e., trying to find a piecewise linear solution suitable for the given signal.



(a) polynomial signal (b) trig signal

Fig. 11: Signal noise removal comparison of the nonlinear signals.

Indeed, in order to see that our functional performance is not restricted by the basis function, and to verify that indeed better performance is achieved if we choose a better set of basis functions to model the signal domain, we performed several tests with higher degree polynomials. We display in Figure 12, results achieved by denoising the polynomial signal displayed in Figure 6, while changing the NLOP functional basis functions to a 2nd degree (TODO: change to 2nd) polynomial. In general we expect a polynomial signal to be best recovered by polynomial basis functions of the same degree. This is clearly depicted in the graph displayed in Figure 12, where we see better performance by NLOP with polynomial basis function compared to NLOP with linear basis functions.

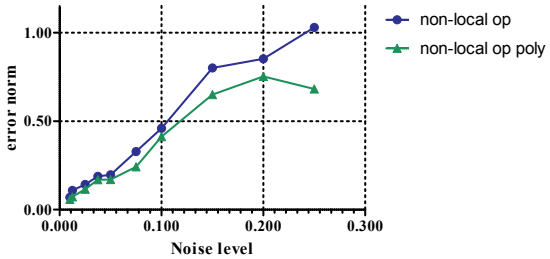


Fig. 12: Comparison of noise removal performance on the polynomial signal. In this figure we compare performance of the NLOP functional with linear basis functions (marked by non-local OP), and NLOP functional with polynomial basis functions (marked by non-local OP poly).

Another test we performed with polynomial basis functions, was on a C_1 continuous 2nd degree polynomial displayed in Figure 13. This is an interesting case, as both this signal and it's first derivative are continuous, and only the second derivative is discontinuous. We found that this signal proved challenging for the MLS initialization method, causing it to misplace the point of discontinuity by several points. This initialization error was not detected by the NLOP functional, which maintained it throughout the minimization process, as displayed for example in Figure **??**. The location of the discontinuity point depends on the random noise. We wish to emphasize that the reconstructed solutions achieved by the minimization of the NLOP functional have piecewise constant reconstructed parameters, which generate a piecewise smooth polynomial solution. This reconstructed signal may as well be the signal from which the noisy signal was generated. Also, the solution achieved by the NLOP functional outperformed the K-SVD method, as displayed in the graphs in Figure 14.

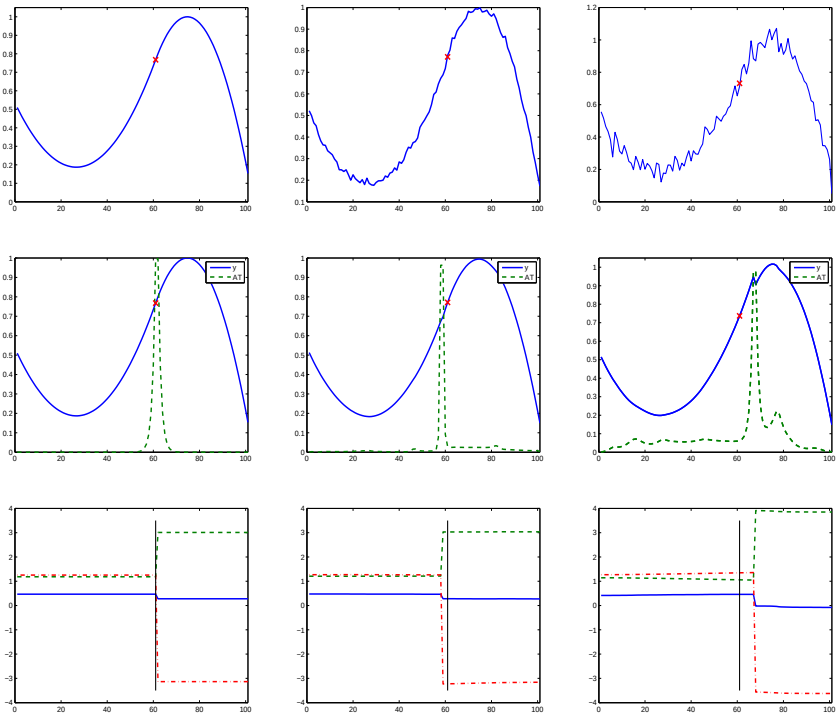
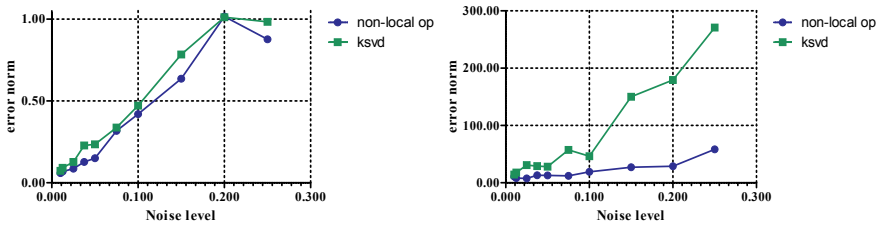


Fig. 13: C_1 continuous 2nd degree polynomial. Point of the second derivative discontinuity is marked by a red cross (or vertical black line). Top row: images of the signals. Middle row: images of the reconstructed signals and the v_{AT} indicator functions. Bottom row: the reconstructed parameters. From left to right: clean signal, 0.01 STD noise, 0.0375 STD noise.



(a) Signal (b) Parameters

Fig. 14: Signal noise removal and parameters reconstruction comparison for the C_1 continuous 2nd degree polynomial.

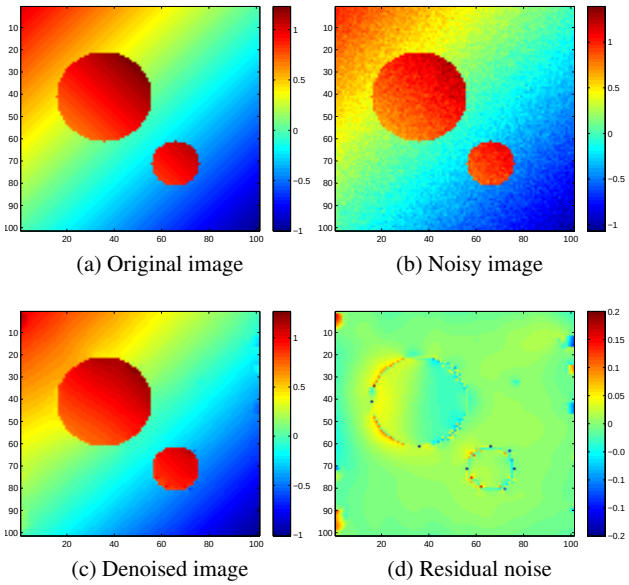
7.3 2D example

We also ran initial tests on a 2D example. The non-local over-parameterized method extends naturally to higher dimensions. We implemented the 2D case in a similar manner to the 1D case, and chose a linear basis functions of the form

$$\begin{cases} \phi_1 = 1 \\ \phi_2 = x \\ \phi_3 = y \end{cases} \tag{21}$$

In Figure 15, we show the 2D example, the NLOP functional noise removal result. In Figure 16, we display the parameters which were reconstructed in the minimization process and the generated v_{AT} indicator function. We can see that the v_{AT} indicator function managed to segment the signal in a good manner, although still delineated some ghost segments especially near the image edges. Note that the recovered parameters are almost piecewise constant as expected.

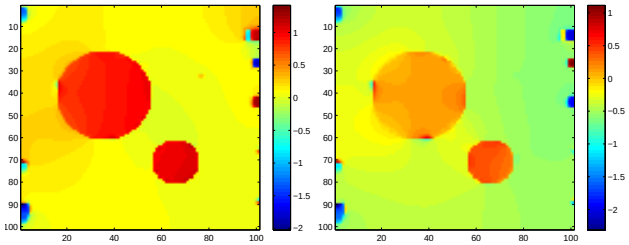
A more thorough discussion of the 2D case is out of the scope of this paper. We intend to explore it extensively in the near future.



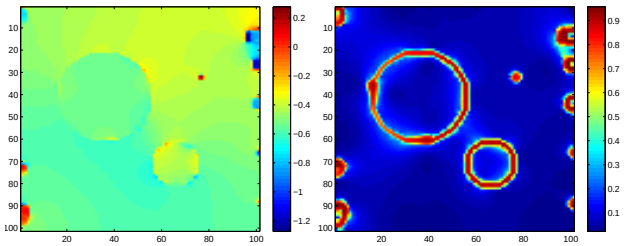
(a) Original image (b) Noisy image

(c) Denoised image (d) Residual noise

Fig. 15: A 2D noise removal example of the 2D NLOP functional.



(a) Parameter 1 (b) Parameter 2



(c) Parameter 3 (d) v_{AT}

Fig. 16: Example of the reconstructed parameters from the noisy image and v_{AT} AT scheme indicator function.

8 Conclusion

A general over-parameterized variational framework for signal and image analysis was presented. This framework can be applied to various image processing and vision problems, such as noise removal, segmentation and optical flow computation. This framework is closely related to the powerful moving least squares method, enhancing it by globally constraining the parameter variations in space based on knowledge available on the problem domain. This knowledge enables a model based reconstruction of the considered signal, by effectively recovering parameters of an a priori, assumed to be known, set of "basis" or "dictionary functions".

The new variational framework relies on the successful over-parameterized functional, and significantly improves it by making it non-local and giving it the power not only to generate excellent results for the problem domain (such as noise removal), but also to reconstruct the underling model parameters that might capture prior information on the problem.

This paper may be viewed as the basis of extensive future research. First of all we wish to thoroughly explore the extension of the non-local over-parametrization functional into $2D$ settings, where the choice of the support set, expressed by the weighting function, becomes a challenging matter. We also intend to return to the optical flow problem, trying to recover a 3D scene representation while calculating

the optical flow. Finally we would like to extend the proposed framework to other problem domains in computer vision.

References

1. T. Quatieri, *Discrete-time speech signal processing: principles and practice*. Upper Saddle River, NJ, USA: Prentice Hall Press, first ed., 2001.
2. T. Malisiewicz and A. A. Efros, "Improving spatial support for objects via multiple segmentations," in *British Machine Vision Conference (BMVC)*, September 2007.
3. T. J. Darrell and A. P. Pentland, "Cooperative robust estimation using layers of support," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 17, pp. 474–487, 1991.
4. E. Cohen, R. F. Riesenfeld, and G. Elber, *Geometric modeling with splines - an introduction*. A K Peters, 2001.
5. M. Haindl and S. Mikes, "Model-based texture segmentation," in *International Conference on Image Analysis and Recognition*, pp. 306–313, 2004.
6. B. A. Olshausen and D. J. Fieldt, "Sparse coding with an overcomplete basis set: a strategy employed by v1," *Vision Research*, vol. 37, pp. 3311–3325, 1997.
7. M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: Design of dictionaries for sparse representation," in *In: Proceedings of SPARS'05*, pp. 9–12, 2005.
8. D. Levin, "The approximation power of moving least-squares," *Mathematics of Computation*, vol. 67, pp. 1517–1531, 1998.
9. D. Levin, "Mesh-independent surface interpolation," *Geometric Modeling for Scientific Visualization*, pp. 37–49, 2003.
10. M. Alexa, J. Behr, D. Cohen-Or, S. Fleishman, D. Levin, and C. T. Silva, "Point set surfaces," in *IEEE Visualization 2001*, pp. 21–28, IEEE Computer Society, October 2001.
11. N. Amenta and Y. J. Kil, "Defining point-set surfaces," *ACM Trans. Graph.*, vol. 23, pp. 264–270, August 2004.
12. S. Fleishman, D. Cohen-Or, and C. T. Silva, "Robust moving least-squares fitting with sharp features," *ACM Trans. Graph.*, vol. 24, pp. 544–552, July 2005.
13. T. Nir, A. M. Bruckstein, and R. Kimmel, "Over-parameterized variational optical flow," *International Journal of Computer Vision*, vol. 76, no. 2, pp. 205–216, 2008.
14. T. Nir and A. Bruckstein, "On over-parameterized model based TV-denoising," in *International Symposium on Signals, Circuits and Systems (ISSCS 2007)*, July 2007.
15. A.M.Bruckstein, "On globally optimal local modeling: From moving least squares to overparametrization," in *Workshop on Sparse Representation of Multiscale Data and Images: Theory and Applications*, Institute of Advanced Study, Nanyang, Technological University, Singapore, December 2009.
16. R. M. Haralick and L. Watson, "A facet model for image data," *Computer Graphics and Image Processing*, vol. 15, no. 2, pp. 113 – 129, 1981.
17. G. Farnebäck, "Spatial domain methods for orientation and velocity estimation," Lic. Thesis LiU-Tek-Lic-1999:13, Dept. EE, Linköping University, SE-581 83 Linköping, Sweden, March 1999. Thesis No. 755, ISBN 91-7219-441-3.
18. A. Tarantola, *Inverse Problem Theory and Methods for Model Parameter Estimation*. Philadelphia, PA, USA: Society for Industrial and Applied Mathematics, 2004.
19. M. J. Black and P. Anandan, "The robust estimation of multiple motions: parametric and piecewise-smooth flow fields," *Comput. Vis. Image Underst.*, vol. 63, pp. 75–104, January 1996.
20. E. Memin and P. Perez, "A multigrid approach for hierarchical motion estimation," in *Computer Vision, 1998. Sixth International Conference on*, pp. 933 –938, jan 1998.
21. L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Phys. D*, vol. 60, pp. 259–268, November 1992.

22. T. Pock, *Fast Total Variation for Computer Vision*. Phd thesis, Graz University of Technology, 2008.
23. J. Savage and K. Chen, "On multigrids for solving a class of improved total variation based staircasing reduction models," *In International Conference on PDE-Based Image Processing and Related Inverse Problems*, 2006.
24. W. Trobin, T. Pock, D. Cremers, and H. Bischof, "An unbiased second-order prior for high-accuracy motion estimation," in *DAGM Symposium*, pp. 396–405, Springer-Verlag, 2008.
25. S. Kindermann, S. Osher, and P. W. Jones, "Deblurring and denoising of images by nonlocal functionals," *Multiscale Modeling and Simulation*, vol. 4, no. 4, pp. 1091–1115, 2005.
26. G. Gilboa and S. Osher, "Nonlocal operators with applications to image processing," *Multiscale Modeling and Simulation*, vol. 7, no. 3, pp. 1005–1028, 2008.
27. A. Harten and S. Osher, "Uniformly high-order accurate nonoscillatory schemes," *SIAM J. Numer. Anal.*, vol. 24, pp. 279–309, April 1987.
28. C. W. Shu, "Essentially non-oscillatory and weighted essentially non-oscillatory schemes for hyperbolic conservation laws," pp. 325–432, Springer, 1997.
29. L. Ambrosio and V. M. Tortorelli, "Approximation of functional depending on jumps by elliptic functional via Γ-convergence," *Communications on Pure and Applied Mathematics*, vol. 43, no. 8, pp. 999–1036, 1990.
30. T. F. Chan and J. Shen, "On the role of the bv image model in image restoration," *AMS Contemporary Mathematics, accepted (by two referees)*, vol. 330, pp. 25–41, 2003.
31. J. Shah, "A common framework for curve evolution, segmentation and anisotropic diffusion," in *Proceedings of the 1996 Conference on Computer Vision and Pattern Recognition (CVPR '96)*, Computer Vision and Pattern Recognition, (Washington, DC, USA), pp. 136–, IEEE Computer Society, 1996.
32. J. Shah, "Curve evolution and segmentation functionals: application to color images," in *In Proceedings IEEE ICIP'96*, pp. 461–464, 1996.
33. G. Rosman, S. Shem-Tov, D. Bitton, T. Nir, G. Adiv, R. Kimmel, A. Feuer, and A. M. Bruckstein, "Over-parameterized optical flow using a stereoscopic constraint," *Scale space and variational methods in computer vision*, 2011.
34. R. Rubinstein, M. Zibulevsky, and M. Elad, "Efficient implementation of the k-svd algorithm using batch orthogonal matching pursuit," technical report, Technion - Israel Institute of Technology, 2008.