Structural Similarity-Based Optimization Problems with L^1-Regularization: Smoothing Using Mollifiers

Daniel Otero[1(✉)], Davide La Torre[2,3], and Edward R. Vrscay[1]

[1] Department of Applied Mathematics, Faculty of Mathematics, University of Waterloo, Waterloo, ON N2L 3G1, Canada
{dotero,ervrscay}@uwaterloo.ca
[2] Department of Economics, Management, and Quantitative Methods, University of Milan, Milan, Italy
[3] Department of Applied Mathematics and Sciences, Khalifa University, Abu Dhabi, UAE
davide.latorre@unimi.it, davide.latorre@kustar.ac.ae

Abstract. In this paper we propose a new method of solving optimization problems involving the structural similarity image quality measure with L^1-regularization. The regularization term $\|x\|_1$ is approximated by a sequence of smooth functions $\|x\|_1^\varepsilon$ by means of C_0^∞ functions known as mollifiers. Because the functions $\|x\|_1^\varepsilon$ epi-converge to $\|x\|_1$, the sequence of minimizers of the smooth objective functions converges to a minimizer of the non-smooth problem. This approach permits the use of gradient-based methods to solve the minimization problems as opposed to methods based on subdifferentials.

1 Introduction

Many problems in image processing may be cast into the following form: Given a $y \in \mathbb{R}^m$ and a compact subset $D \subset \mathbb{R}^n$, find

$$\min_{x \in D} \frac{1}{2}\|Ax - y\|_2^2 + \gamma\|x\|_1, \tag{1}$$

where A is an $m \times n$ transformation matrix (e.g., wavelet, Fourier, random matrix, etc.). Such a functional is known as the *Lasso problem*, whose L^1 regularizing term induces sparseness in its solution [1,16,23]. The quadratic term, which is usually called the "fidelity term", keeps the solution close to the observation y.

A great variety of algorithms have been proposed to solve the Lasso problem, e.g., Fast Iterative Soft-Thresholding Algorithm (FISTA) [1] and the Least Angle Regression [16]. These specialized methods usually rely on techniques from subdifferential calculus to overcome the non-differentiability of the regularizing term of (1). Nevertheless, classical methods can be employed by either casting (1) as a Quadratic Program (QP) [16] or by approximating the L^1 norm with a family of smooth functions $\varphi_\varepsilon \in C_c^\infty(\mathbb{R}^n)$ known as *mollifiers* [12].

M. Kamel and A. Campilho (Eds.): ICIAR 2015, LNCS 9164, pp. 33–42, 2015.
DOI: 10.1007/978-3-319-20801-5_4

In particular, in [24], the Gaussian distribution is used as an approximate mollifier to solve a smooth version of (1), which is obtained by convolving each component of the L^1 norm with a standard one-dimensional Gaussian density function of variance ε^2. This technique allows the usage of gradient-based methods for approximating the optimal solution $x^\star$ of (1). In this case, the smooth approximation of the original problem is given by

$$\min_{x\in D} \frac{1}{2}\|Ax-y\|_2^2 + \gamma\|x\|_1^\varepsilon, \tag{2}$$

where $\|x\|_1^\varepsilon$ is equal to

$$\|x\|_1^\varepsilon = \sum_{i=1}^{n} \int_{\mathbb{R}} |x_i - z_i| \hat{\phi}_\varepsilon \left(\frac{z_i}{\varepsilon}\right) dz. \tag{3}$$

Here, $\hat{\phi}(x)$ is the standard normal distribution in one dimension.

In an imaging context, the main drawback of these approaches is that they employ the square of the Euclidian distance as a fitting term, which is not the best choice when it comes to measure visual closeness [28,29]. To overcome this difficulty, many authors have incorporated the Structural Similarity Index Measure (SSIM) as a fidelity term in different types of optimization problems [6,7,17–20]. The SSIM is one of the most popular measures of visual quality, which was introduced in [28], and it has been shown to outperform the square of the Euclidian distance as a measure of visual quality.

Nevertheless, mathematical treatment of the SSIM is difficult, thus simpler versions of the SSIM are desirable. In particular, the definition of this measure as a normalized metric has been employed in [6,17,18]. This simplified version of the SSIM has nice properties such as quasi-convexity [4,18], which allows the use of quasi-convex techniques to solve optimization problems that employ the SSIM as a fidelity term [18].

In [18], several imaging tasks are carried out by solving different types of quasi-convex optimization problems in which the SSIM is minimized subject to a set of convex constraints. One of the problems that is addressed is

$$\begin{aligned} \min_{x} \quad & T(Ax,y) \\ \text{subject to} \quad & \|x\|_1 \leq \lambda. \end{aligned} \tag{4}$$

Here, the fidelity term is given by $T(Ax,y) = 1 - S(Ax,y)$, where $S(\cdot,\cdot)$ is the simplified version of the SSIM as a normalized metric [6,17,18]. The unconstrained counterpart of (4) was studied in [17]. In this case, an algorithm that uses a generalization of the soft-thresholding operator [11,23] is employed for solving the following optimization problem:

$$\min_{x\in D} \; T(Ax,y) + \gamma\|x\|_1 \,. \tag{5}$$

The advantage of these formulations is that the concepts of similarity and sparsity are combined into a single optimization problem; therefore, solutions of this problem are similar to the observation y in the SSIM sense and also sparse.

In this paper, we extend the work of [17] by solving (5) via mollifiers. This approach allows us to use gradient-based methods for solving the non-smooth problem (5). In this case, the following smooth version of (4) is solved:

$$\min_{x \in D} \; T(Ax, y) + \gamma \|x\|_1^{\varepsilon}, \tag{6}$$

where $\|x\|_1^{\varepsilon}$ is obtained by convolving the L^1 norm with a multivariate Gaussian distribution of variance ε^2. As expected, the sequence of minimizers x_{ε}^* of (6) converges to an optimal solution x^* of (5) when $\varepsilon \to 0$. Numerical results that show the performance of the gradient-based method presented in this paper are also included.

2 Smoothing via Mollifiers

In this section we recall some basic notions and properties of mollifiers and introduce a smoothing approach. For each $\varepsilon > 0$, let us consider a family of functions $\varphi_{\varepsilon} \in C_0^{\infty}(\mathbb{R}^n)$ that satisfies the following properties:

1. $\varphi_{\varepsilon}(x) \geq 0$, for all $x \in \mathbb{R}^n$,
2. $support(\varphi_{\varepsilon}) \subseteq \{x \in \mathbb{R}^n : \|x\| \leq \varepsilon\}$,
3. $\int_{\mathbb{R}^n} \varphi_{\varepsilon}(x)dx = 1$.

Such functions are called *mollifiers* [12].

We now provide a way to construct a family of smooth functions approximating any function f in L_{loc}^1 (locally integrable functions). Given a family of mollifiers $\{\varphi_{\varepsilon} : \mathbb{R}^n \to \mathbb{R}_+ | \varepsilon \in \mathbb{R}_+\}$, we can define a smooth function approximation f^{ε} of f through the convolution

$$(f * \varphi_{\varepsilon})(x) := \int_{\mathbb{R}^n} f(x - z)\varphi_{\varepsilon}(z)dz = \int_{\mathbb{R}^n} f(z)\varphi_{\varepsilon}(x - z)dz\,.$$

The sequence $f * \varphi_{\varepsilon}$ is said to be a sequence of mollified functions. Some properties of mollified functions can be considered classical. From a computational perspective let us notice that if $Y_{\varepsilon}(x, \cdot)$ is a random vector with density defined by $z \to \varphi_{\varepsilon}(x - z)$, the above definition can be written as

$$(f * \varphi_{\varepsilon})(x) := \mathbb{E}(f(Y_{\varepsilon}(x, \cdot)))\,,$$

where $\mathbb{E}$ is the expected value of the random variable $f(Y_{\varepsilon}(x, \cdot))$. This stochastic interpretation allows us to avoid the calculation of the above integral by estimating the expected valued of $f(Y_{\varepsilon}(x, \cdot))$ instead.

Theorem 1. *[2] Let $f \in C(\mathbb{R}^n)$. Then $f * \varphi_{\varepsilon}$ converges continuosly to f, i.e. $f * \varphi_{\varepsilon_m}(x_m) \to f(x)$ for all $x_m \to x$. In fact, $f * \varphi_{\varepsilon}$ converges uniformly to f on every compact subset of $\mathbb{R}^n$ as $\varepsilon_m \to 0$.*

The previous convergence property can be generalized.

Definition 1. *[2] A sequence of functions $f_m : \mathbb{R}^n \to \mathbb{R}$ epi-converges to $f : \mathbb{R}^n \to \mathbb{R}$ at x if:*

1. $\liminf_{m\to+\infty} f_m(x_m) \geq f(x)$ *for all* $x_m \to x$;
2. $\lim_{m\to+\infty} f_m(x_m) = f(x)$ *for some sequence* $x_m \to x$.

The sequence f_m epi-converges to f if this holds for all $x \in \mathbb{R}^n$. In this case we say that f is the epi-limit of f_m.

It can be easily checked that when f is the epi-limit of some sequence f_m, then f is lower semicontinuous. Moreover if f_m converges continuously, then it also epi-converges. The notion of epi-convergence ensures the convergence of mimimizers of f_m to the minimizers of f (see [22]).

Definition 2. *[12] A function $f : \mathbb{R}^n \to \mathbb{R}$ is strongly lower semicontinuous (s.l.s.c.) at x if it is lower semicontinuous at x and there exists a sequence $x_m \to x$ with f continuous at x_m (for all m) such that $f(x_m) \to f(x)$. The function f is strongly lower semicontinuous if this holds at all x.*

Theorem 2. *[12] Let $\varepsilon_m \to 0$ if $m \to +\infty$. For any s.l.s.c. function $f : \mathbb{R}^n \to \mathbb{R}$, and any associated sequence f_{ε_m} of mollified functions we have that f is the epi-limit of f_{ε_m}.*

Lemma 1. *The mollified norm $\|x\|_1^\varepsilon$ is greater or equal than its non-smooth counterpart $\|x\|_1$ for any $x \in \mathbb{R}^n$.*

Proof. Let $f(z) = \|x - z\|_1$. Then, by convexity of f and using Jensen's inequality, we have that

$$\|x - \mathbb{E}(z)\|_1 \leq \int_{\mathbb{R}^n} \|x - z\|_1 \varphi_\varepsilon(z) dz. \tag{7}$$

Given that $\mathbb{E}(z) = 0$, we immediately obtain that $\|x\|_1 \leq \|x\|_1^\varepsilon$ for all $x \in \mathbb{R}^n$.

Theorem 3. *Let $g : \mathbb{R}^n \to \mathbb{R}$ and (ε_m) be a sequence of positive real numbers such that $\varepsilon_m \to 0$. The function $g(x) + \gamma\|x\|_1$ is the epi-limit of the sequence of functions $h_m : \mathbb{R}^n \to \mathbb{R}$ defined as*

$$h_m(x) := g(x) + \gamma\|x\|_1^{\varepsilon_m}. \tag{8}$$

Proof. Let (x_m) be a sequence in $\mathbb{R}^n$ such that $x_m \to x$. Since $\|x\|_1^{\varepsilon_m}$ converges to $\|x\|_1$ as m tends to infinity, we have that

$$\lim_{m\to\infty} h_m(x_m) = g(x) + \gamma\|x\|_1. \tag{9}$$

Also, by lemma 1, it follows that for any $x_m \in \mathbb{R}^n$ and any $\varepsilon_m \in \mathbb{R}_+$

$$g(x_m) + \gamma\|x_m\|_1^{\varepsilon_m} \geq g(x_m) + \gamma\|x_m\|_1. \tag{10}$$

Taking $\liminf$ *at both sides over all sequences $x_m \to x$ we obtain that*

$$\liminf_{m\to\infty} g(x_m) + \gamma\|x_m\|_1^{\varepsilon_m} \geq g(x) + \gamma\|x\|_1. \tag{11}$$

This completes the proof.

By means of mollified functions it is possible to define generalized directional derivatives for a non-smooth function f, which, under suitable regularity of f, coincide with Clarke's subdifferential. In [12] (see also [8,9,15] for alternative definitions of generalized derivatives through mollified functions), a generalized gradient w.r.t. the mollifier sequence f_{ε_m} has also been defined in the following way:

$$\partial_\varepsilon f(x) := \left\{ \limsup_{m \to +\infty} \nabla f_{\varepsilon_m}(x_m), x_m \to x \right\}. \tag{12}$$

Theorem 4. *[12] Let $f : \mathbb{R}^n \to \mathbb{R}$ be locally Lipschitz at x; then $\partial_\varepsilon f(x)$ coincides with Clarke's subdifferential at x.*

Theorem 5. *Let $g : \mathbb{R}^n \to \mathbb{R}$ be differentiable and locally Lipschitz at x. Also, let $h(x) = g(x) + \gamma \|x\|_1$. Then, $\partial_\varepsilon h(x)$ coincides with Clarke's subdifferential at x.*

Proof. *Clearly, $\|x\|_1$ is Lipschitz continuous with Lipschitz constant one since*

$$|\|x\|_1 - \|y\|_1| \leq \|x - y\|_1. \tag{13}$$

Moreover, since g is locally Lipschitz at x, it follows that h is locally Lipschitz at x as well. Also, by definition of $\partial_\varepsilon(\cdot)$ one has that

$$\partial_\varepsilon h(x) := \left\{ \limsup_{m \to +\infty} \nabla (g(x_m) + \gamma \|x_m\|_1^{\varepsilon_m}), x_m \to x \right\} \tag{14}$$

$$:= \left\{ \limsup_{m \to +\infty} \nabla g(x_m) + \gamma \nabla (\|x_m\|_1^{\varepsilon_m}), x_m \to x \right\} \tag{15}$$

$$:= \nabla g(x) + \gamma \left\{ \limsup_{n \to +\infty} \nabla (\|x_m\|_1^{\varepsilon_m}), x_m \to x \right\}, \tag{16}$$

where the last equation is indeed the set of Clarke's subgradients of h at x.

In the sequel we will use the following family of smoothing Gaussian functions:

$$\hat{\phi}_\varepsilon(x) = \frac{1}{\varepsilon^n} \hat{\phi}\left(\frac{x}{\varepsilon}\right), \tag{17}$$

where

$$\hat{\phi}(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{\|x\|_2^2}{2}}. \tag{18}$$

It is well known that ϕ_ε is a density function, so its integral over $\mathbb{R}^n$ is equal to one, it is smooth, and ϕ_ε goes to zero when $\|x\| \to +\infty$. However, this sequence is not a proper family of mollifiers as each element $\hat{\phi}_\varepsilon$ does not have a compact support. Nevertheless, it can be proved that, given a function f, the family of smooth functions

$$\hat{f}_\varepsilon(x) = (f * \hat{\phi}_\varepsilon)(x) = \frac{1}{\varepsilon^n} \int_{\mathbb{R}^n} f(x - z) \hat{\phi}\left(\frac{z}{\varepsilon}\right) dz \tag{19}$$

epi-converges to f when $\varepsilon \to 0$. This easily follows by taking a sequence of mollifiers φ_δ with compact support converging to $\hat{\phi}$ when $\delta \to 0$, and then use the convergence properties of mollifiers.

Furthermore, Theorems 3 and 5 guarantee that the sequence of minimizers x_ε^* of (6) converges to a minimizer x^* of (5) when ε tends to zero. In other words, $x_\varepsilon^* \to x^*$ as $\varepsilon \to 0$.

3 SSIM-Based Optimization with Sparsity

The Structural Similarity Index Measure (SSIM) between x and y, where $x, y \in \mathbb{R}^n$, is defined as [28]

$$\text{SSIM}(x,y) = \left(\frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1}\right)\left(\frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2}\right)\left(\frac{\sigma_{xy} + C_3}{\sigma_x\sigma_y + C_3}\right). \tag{20}$$

Here, μ_x and μ_y denote the mean values of x and y, respectively, and σ_{xy} denotes the cross correlation between x and y, from which all other definitions follow. The small positive constants, C_1, C_2 and C_3 provide numerical stability and can be adjusted to accommodate the Human Visual System (HVS) [28,29].

Under the assumption that the vectors x and y have zero mean, the latter expression can be simplified:

$$S(x,y) = \frac{2x^Ty + C}{\|x\|_2^2 + \|y\|_2^2 + C}, \tag{21}$$

where $C = (n-1)C_2$. Reformulation of the SSIM as a normalized metric comes out from the definition of the following distance-dissimilarity function $T(x,y)$ [6,17,18]:

$$T(x,y) = 1 - S(x,y) = \frac{\|x-y\|_2^2}{\|x\|_2^2 + \|y\|_2^2 + C}. \tag{22}$$

Note that $0 \leq T(x,y) \leq 2$. Furthermore, $T(x,y) = 0$ if and only if $x = y$.

Algorithms for solving (6) can be developed by first computing its gradient. To do this, we define the following non-linear functional:

$$f(x) = T(x,y) + \gamma\|x\|_1^\varepsilon. \tag{23}$$

Its gradient is given by

$$\nabla f_\varepsilon(x) = \frac{2S(x,y)A^TAx - 2A^Ty}{\|Ax\|_2^2 + \|y\|_2^2 + C} + \gamma\int_{\mathbb{R}^n} \|z\|_1 \nabla\hat{\phi}_\varepsilon\left(\frac{x-z}{\varepsilon}\right)dz, \tag{24}$$

where $\nabla\hat{\phi}_\varepsilon(x)$ is equal to

$$\nabla\hat{\phi}_\varepsilon\left(\frac{x}{\varepsilon}\right) = \frac{-x}{\sqrt{(2\pi)^n}\varepsilon^{n+2}}e^{-\frac{\|x\|_2^2}{2\varepsilon^2}}. \tag{25}$$

By using (24), and defining $\mathbf{1} = [1,\cdots,1]^T \in \mathbb{R}^m$, we propose the following algorithm for solving (6):

Algorithm 1. Gradient descent for unconstrained SSIM-L^1 optimization via mollifiers

initialize Choose $x = x_0$, λ;
data preprocessing $\bar{y} = \frac{1}{n}\mathbf{1}^T y$, $y = y - \bar{y}\mathbf{1}$;
repeat
$x = x - \lambda \nabla f_\varepsilon(x)$;
until stopping criterion is met (e.g., $\|x^{(new)} - x^{(old)}\|_\infty < \delta$);
return x, $y = y + \bar{y}\mathbf{1}$.

Notice that this algorithm will return an optimal x^* such that the mean of Ax^* is zero. Nevertheless, it is possible to obtain the non-zero mean optimal $x^\star$ by means of the following equation:

$$x^\star = x^* + \bar{y}(A^T A)^{-1} A^T \mathbf{1}, \tag{26}$$

provided that the inverse of $A^T A$ exists (see [17] and [18] for more details).

4 Experiments

In these experiments we solve the approximate sparse reconstruction problem (6) with the proposed gradient-descent algorithm. Its performance is measured by comparing its recovered solutions with the solutions obtained by the algorithm introduced in [17] for solving (5) and the solutions of problem (1). In all computations a set of Discrete Cosine Transform (DCT) coefficients is to be recovered; therefore, problem (1) was solved by means of the soft-thresholding (ST) operator [16,23].

In all the experiments images were divided into non-overlapping 8×8 pixel blocks. As expected, the means of each block are subtracted prior to processing, which are added after the non-overlapping blocks have been processed. This is also done when problem (1) is solved at each pixel block for the sake of a fair comparison between the different methods.

It is worthwhile to mention that for computing the integral of the gradient of (23) we performed a Monte Carlo integration (see Eq. (24)). This can be done by noticing the fact that calculating

$$\int_{\mathbb{R}^n} \|z\|_1 \nabla \hat{\phi}_\varepsilon \left(\frac{x - z}{\varepsilon} \right) dz \tag{27}$$

is equivalent to compute the expected value $\mathbb{E}(\|z\|_1 (x - z))$, where z follows a Gaussian distribution of variance ε^2 and mean equal to x.

In Fig. 1, in the left plot it can be observed an example of the optimal DCT coefficients that are obtained by the different methods that are being compared. Plots in red and green correspond to the solutions obtained by the algorithm introduced in [17] and ST respectively. The blue plot is the optimal solution

that was obtained by the proposed method when $\varepsilon = 0.001$. True sparsity in the solution is not achieved since this occurs in the limit when $\varepsilon \to 0$; nevertheless, it can be seen that the proposed method gives a good non-sparse approximation of the solution of the non-smooth problem (5). This in fact can be useful for providing a good initial guess of a thresholding method that solves (5) [24]. In the plot on the right it can be seen how a sequence of optimal solutions of (6) gets closer to a solution of (5) as ε tends to zero. In this case, the plot in magenta corresponds to the set of optimal DCT coefficients that is obtained by solving problem (5).

As for visual results, these are shown in Fig. 2. In the presented example, a sub-image of the test image *Lena* was employed. In the bottom row the original sub-image and its recovered counterparts can be observed. Regularization was carried out at all non-overlapping pixel blocks in such a way that the number of non-zero DCT coefficients obtained by the algorithm introduced in [17] and the ST operator is always 19. As for the regularization of the proposed algorithm, the values of the regularization parameter that were used were the same that were employed for the algorithm that solves the non-smooth problem (5).

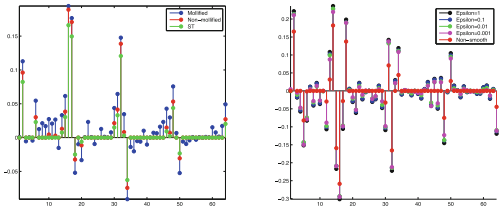

Fig. 1. The plot on the left shows an example of the different solutions that were obtained by the three methods that were compared. The plot on the right shows how a sequence of minimizers x_ε^* of the mollified SSIM-based optimization problem (6) converges to a minimizer x^* of the non-smooth problem (5).

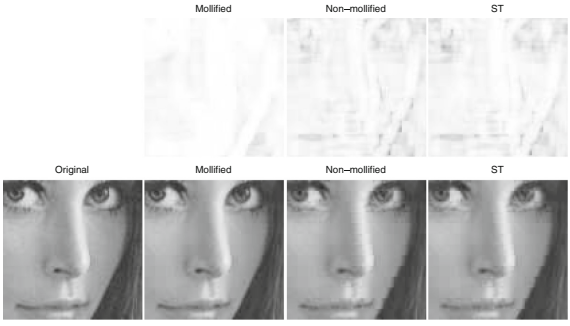

Fig. 2. Visual results for a sub-image from the test image *Lena*. In all cases, regularization is carried out to induce the same degree of sparsity for all methods at each non-overlapping pixel block. In the bottom row the original image and its reconstructions are shown. The corresponding SSIM maps can be seen in the top row.

This was done in this way since in the limit when $\varepsilon \to 0$, both problems (5) and (6) are equivalent. In other words, the strength of the regularization tends to be the same for these two methods. Along with the images of the bottom row, the SSIM maps that depict the similarity between the original sub-image and its reconstructions are shown in the top row. The higher the brightness of these maps at a given location, the higher the SSIM at that particular point [28]. As mentioned in [17], performance of the ST approach and their algorithm is very similar, however, the average $T(Ax, y)$ of the non-mollified SSIM-based optimization problem (0.9156) is higher than the average $T(Ax, y)$ of the L^2 counterpart (0.9117). As for the proposed approach, the recovered image is visually more appealing than the other two methods, and as expected, the average $T(Ax, y)$ is the highest of the three approaches that are being compared (0.9629). This should not be surprising since several recovered DCT coefficients are not set to zero by the proposed algorithm, which is not always the case for the other two methods.

Acknowledgements. We gratefully acknowledge that this research has been supported in part by the Natural Sciences and Engineering Research Council of Canada (NSERC) in the form of a Discovery Grant (ERV).

References

1. Beck, A.A., Teboulle, M.: A fast iterative shrinkage-thresholding algorithm for linear inverse problems. SIAM J. Imaging Sci. Arch. **2**(1), 183–202 (2009)
2. Brezis, H.: Analyse Fonctionelle Theorie et Applications. Masson editeur, Paris (1963)
3. Bruckstein, A., Donoho, D., Elad, M.: From sparse solutions of systems of equations to sparse modeling of signals and images. SIAM Rev. **51**(1), 34–81 (2009)
4. Brunet, D.: A Study of the Structural Similarity Image Quality Measure with Applications to Image Processing, Ph.D. thesis, University of Waterloo (2010)
5. Brunet, D., Vrscay, E.R., Wang, Z.: On the mathematical properties of the structural similarity index. IEEE Trans. Image Proc. **21**(4), 1488–1499 (2012)
6. Brunet, D., Vrscay, E.R., Wang, Z.: Structural similarity-based approximation of signals and images using orthogonal bases. In: Campilho, A., Kamel, M. (eds.) ICIAR 2010. LNCS, vol. 6111, pp. 11–22. Springer, Heidelberg (2010)
7. Channappayya, S.S., Bovik, A.C., Caramanis, C., Heath Jr, R.W.: Design of linear equalizers optimized for the structural similarity index. IEEE Trans. Image Process. **17**(6), 857–872 (2008)
8. Crespi, G.P., La Torre, D., Rocca, M.: Second-order mollified derivatives and optimization. Rendiconti del Circolo Matematico di Palermo **52**(2), 251–262 (2003)
9. Crespi, G.P., La Torre, D., Rocca, M.: Mollified derivatives and second-order optimality conditions. J. Nonlinear Convex Anal. **4**(3), 437–454 (2003)
10. Donoho, D., Elad, M.: Optimality sparse representation in general (non-orthogonal) dictionaries via L^1-minimization. Proc. Nat. Acad. Sci. **100**, 2197–2202 (2003)
11. Donoho, D.: Denoising by soft-thresholding. IEEE Trans. Inf. Theory **41**(3), 613–627 (1995)

12. Ermoliev, Y.M., Norkin, V.I., Wets, R.J.B.: The minimization of semicontinuous functions: mollifier subgradients. SIAM J. Control Optim. **33**, 149–167 (1995)
13. Jongen, H.T., Stein, O.: Smoothing by mollifiers. Part I: semi-infinite optimization. J. Global Optim. **41**(3), 319–334 (2008)
14. Jongen, H.T., Stein, O.: Smoothing by mollifiers. Part II: nonlinear optimization. J. Global Optim. **41**(3), 335–350 (2008)
15. La Torre, D., Rocca, M.: Remarks on second order generalized derivatives for differentiable functions with Lipschitzian Jacobian. Appl. Math. E - Notes **3**, 130–137 (2003)
16. Mairal, J., Bach, F., Jenatton, R., Obozinski, G.: Convex Optimization with Sparsity-Inducing Norms. Optimization for Machine Learning. MIT Press, Cambridge (2011)
17. Otero, D., Vrscay, E.R.: Unconstrained structural similarity-based optimization. In: Campilho, A., Kamel, M. (eds.) ICIAR 2014, Part I. LNCS, vol. 8814, pp. 167–176. Springer, Heidelberg (2014)
18. Otero, D., Vrscay, E.R.: Solving problems that employ structural similarity as the fidelity measure. In: Proceedings of the Internartional Conference on Image Processing, Computer Vision and Patter Recognition IPCV 2014, pp. 474–479. Springer, Heidelberg (2014)
19. Rehman, A., Rostami, M., Wang, Z., Brunet, D., Vrscay, E.R.: SSIM-inspired image restoration using sparse representation. EURASIP J. Adv. Sig. Proc. (2012)
20. Rehman, A., Gao, Y., Wang, J., Wang, Z.: Image classification based on complex wavelet structural similarity. Sig. Proc. Image Comm. **28**(8), 984–992 (2013)
21. Richter, T., Kim. K.J.: A MS-SSIM optimal JPEG 2000 encoder. In: Data Compression Conference, pp. 401–410, Snowbird, Utah, March 2009
22. Rockafellar, R.T., Wets, R.J.-B.: Variational Analysis. Springer Verlag, Berlin (1998)
23. Turlach, B.A.: On algorithms for solving least squares problems under an L^1 penalty or an L^1 constraint. In: Proceedings of the American Statistical Association, Statistical Computing Section, pp. 2572–2577 (2005)
24. Voronin, S., Yoshida, D.: Gradient Based Methods for Non-Smooth Regularization Via Convolution Smoothing http://arxiv.org/pdf/1408.6795.pdf
25. Wainwright, M.J., Schwartz, O., Simoncelli, E.P.: Natural image statistics and divisive normalization: modeling nonlinearity and adaptation in cortical neurons. In: Rao, R., Olshausen, B., Lewicki, M. (eds.) Probabilistic Models of the Brain: Perception and Neural Function, pp. 203–222. MIT Press, Cambridge (2002)
26. Wandell, B.A.: Foundations of Vision. Sinauer Publishers, Sunderland (1995)
27. Wang, Z., Bovik, A.C.: Mean squared error: love it or leave it? a new look at signal fidelity measures. IEEE Signal Proc. Mag. **26**(1), 98–117 (2009)
28. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. IEEE Trans. Image Proc. **13**(4), 600–612 (2004)
29. Wang, Z., Bovik, A.C.: A universal image quality index. IEEE Signal Process. Lett. **9**(3), 81–84 (2002)