Alternate Direction Method of Multipliers for Unconstrained Structural Similarity-Based Optimization

Daniel Otero[1], Davide La Torre[2,3], Oleg V. Michailovich[4], and Edward R. Vrscay[5(✉)]

[1] Department of Electronic Engineering,
Faculty of Engineering, Pontificia Universidad Javeriana, Bogotá, Colombia
oterod@javeriana.edu.co
[2] Dubai Business School, University of Dubai, Dubai, UAE
dlatorre@ud.ac.ae
[3] Department of Economics, Management, and Quantitative Methods,
University of Milan, Milan, Italy
davide.latorre@unimi.it
[4] Department of Electrical and Computer Engineering, Faculty of Engineering,
University of Waterloo, Waterloo, ON N2L 3G1, Canada
olegm@uwaterloo.ca
[5] Department of Applied Mathematics, Faculty of Engineering,
University of Waterloo, Waterloo, ON N2L 3G1, Canada
ervrscay@uwaterloo.ca

Abstract. Recent studies have demonstrated that the Structural Similarity Index Measure (SSIM) is the top choice for quantifying both visual quality and image similarity. Although the SSIM is not convex, it has been successfully employed in a wide range of imaging tasks over the last years. In this paper, the authors propose a new method based on the Alternate Direction Method of Multipliers (ADMM) for solving an unconstrained SSIM-based optimization problem. We focus our analysis on the case in which the regularizing term is convex. The paper also includes numerical examples and experiments that showcase the effectiveness of the proposed method.

1 Introduction

As is well known, it is customary to employ Euclidean-based metrics in a wide variety of image processing tasks. These metrics are quite convenient to use since they are mathematically tractable and easily computed. For instance, the squared Euclidean distance is convex and differentiable, which makes it well suited as, say, the objective function of an optimization problem. As such, it is usually included as the fidelity term in imaging tasks that can be carried out by optimizing a cost function. In general, such functions have the following form:

$$\min_x \frac{1}{2}\|A(x) - y\|_2^2 + \lambda h(x), \tag{1}$$

where $A(\cdot)$ is generally a linear operator (e.g., blurring kernel, subsampling operator, etc.), y is a given observation, $h(x)$ is a regularizing term, and the constant λ is a regularization parameter.

The role of the fidelity term $\|A(x) - y\|_2^2$ is to maintain the solution to (1) close to the observed data y. As for the regularization term $h(x)$, this has two main purposes: (i) It prevents over-fitting and (ii) ensures the solution will have certain expected features which are based on prior information or assumptions. For example, if the optimal solution is assumed to have bounded variation, a typical regularization term is $h(x) = \|x\|_{TV}$, where $\|\cdot\|_{TV}$ is the Total Variation (TV) seminorm [6–8].

Despite the advantages that Euclidean-based metrics offer, it has been shown that they are not appropriate for measuring similarity between images [19, 20]. Given this, many measures of visual quality have been proposed in an attempt to model the Human Visual System (HVS). In particular, the Structural Similarity Index Measure (SSIM), originally proposed by Wang *et al.* [20], has become the top choice for quantifying both visual quality and image similarity.

Although the SSIM is not convex and not as mathematically tractable as the Euclidean-based metrics, it has been successfully employed in a wide range of imaging tasks over the last years. For instance, in [3] the authors find the best approximation coefficients in the SSIM sense when an orthogonal transformation is used (e.g., Discrete Cosine Transform (DCT), Fourier, etc.). Very briefly, a contrast-enhanced version of the best ℓ_2-based approximation is obtained. Based on this result, Rehman *et al.* [15] address the SSIM version of the image restoration problem proposed by Elad *et al.* in [10], where the denoising of images is performed using sparse and redundant representations over learned dictionaries. Furthermore, in [15], the authors also introduce a super-resolution algorithm—also based on the SSIM—to recover from a given low resolution image its high resolution version.

Another interesting application for reconstruction and denoising was proposed in [9]. Here, the authors define the statistical SSIM index (statSSIM), an extension of the SSIM for wide-sense stationary random processes. By optimizing the statSSIM, an optimal filter in the SSIM sense is found. The non-convex nature of the statSSIM is overcome by reformulating its maximization as a quasiconvex optimization problem, which is solved using the bisection method [1,9]. Nevertheless, it is not mentioned that the SSIM—under certain conditions—is a quasi-convex function (see [4]). As a result, it can be minimized using quasiconvex programming techniques, which permits the consideration of a much broader spectrum of SSIM-based optimization problems. Such techniques have already been introduced in [11,13].

More recently, Brunet *et al.* proposed a systematic framework for the design of SSIM-based restoration algorithms [5]. Applications such as optimal SSIM image denoising and soft-thresholding are introduced in this work. Other imaging techniques based on the SSIM can also be found in [16,18]. In these works, optimization of rate distortion, video coding and image classification are explored using the SSIM as a measure of performance.

Also, in [12], the authors introduced a general framework for carrying out *unconstrained SSIM-based optimization*. In particular, two algorithms are proposed for solving optimization problems of the form

$$\min_{x} T(\Phi(x), y) + \lambda h(x), \tag{2}$$

where Φ is usually a linear transformation, λ is a regularization parameter, $h(x)$ is a convex regularizing term, and $T(\cdot, \cdot)$ is a dissimilarity measure, which is given by

$$T(x, y) = 1 - \text{SSIM}(x, y). \tag{3}$$

In this paper, we propose a new method based on the Alternate Direction Method of Multipliers (ADMM) for solving problem (2). In particular, we focus our attention on the case in which the regularizing term $h(x)$ is convex. Experiments that showcase the effectiveness of the proposed method are also included.

2   The Structural Similarity Index Measure (SSIM)

Structural similarity (SSIM) [20] provides a measure of visual closeness of two images (or local image patches) by quantifying similarities in three fundamental characteristics: luminance, contrast and structure. Luminances are compared in terms of a relative change in means. Contrasts are compared in terms of relative variance. Finally, structures are compared in terms of the correlation coefficient between the two images. The SSIM value is computed by simply taking the product of these changes.

In what follows, we let $x, y \in \mathbb{R}^n$ denote two n-dimensional signal/image blocks. The SSIM between x and y is defined as [20],

$$\text{SSIM}(x, y) = \left(\frac{2\mu_x \mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1} \right) \left(\frac{2\sigma_x \sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2} \right) \left(\frac{\sigma_{xy} + C_3}{\sigma_x \sigma_y + C_3} \right). \tag{4}$$

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, C_3 provide numerical stability and can be adjusted to accommodate the HVS. Note that $-1 \leq \text{SSIM}(x, y) \leq 1$. Furthermore, $\text{SSIM}(x, y) = 1$ if and only if $x = y$. As such, x and y are considered to be more similar the closer $\text{SSIM}(x, y)$ is to 1.

Setting $C_3 = C_2/2$ leads to the following definition of the SSIM index found in [20] and used in [3] and elsewhere,

$$\text{SSIM}(x, y) = \left(\frac{2\mu_x \mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1} \right) \left(\frac{2\sigma_{xy} + C_2}{\sigma_x^2 + \sigma_y^2 + C_2} \right). \tag{5}$$

Since the statistics of images vary greatly spatially, the $\text{SSIM}(x, y)$ is computed using a sliding window of 8×8 pixels. The final result, i.e., the so-called *SSIM index*, is basically an average of the individual SSIM measures.

A further simplification results when x and y have zero mean, i.e., $\mu_x = \mu_y = 0$. In this special case, we obtain the following expression

$$\text{SSIM}(x,y) = \frac{2x^T y + C}{\|x\|_2^2 + \|y\|_2^2 + C},\tag{6}$$

where $C = (n-1)C_2$ (see [11,12] for more details). For the remainder of this paper, unless otherwise stated, we shall be working with zero mean vectors, so that Eq. (6) will be employed in all computations of the SSIM.

The corresponding distance/dissimilarity function $T(x,y)$ in Eq. (3) becomes

$$T(x,y) = 1 - \text{SSIM}(x,y) = \frac{\|x-y\|_2^2}{\|x\|_2^2 + \|y\|_2^2 + C}.\tag{7}$$

Note that $0 \leq T(x,y) \leq 2$. Furthermore, $T(x,y) = 0$ if and only if $x = y$. As mentioned earlier, since $\text{SSIM}(x,y)$ is a measure of similarity, $T(x,y)$ can be considered as a measure of dissimilarity between x and y.

3   Unconstrained SSIM-Based Optimization

We shall focus on unconstrained SSIM-based optimization problems of the form,

$$\min_x T(\Phi x, y) + \lambda h(x),\tag{8}$$

where $\lambda > 0$ is a regularization parameter and $h : \mathbb{R}^n \to \mathbb{R}$ is a regularization functional, which is often defined to be convex. Notice that the first term in (8) is not convex, thus the entire cost function is not convex either. This implies that the existence of a unique global minimizer of (8) cannot be generally guaranteed. Despite this, it is still possible to devise efficient numerical methods capable of converging to either a locally or a globally optimal solution, as will be shown in the following Section of the paper.

3.1   ADMM-Based Approach

In order to solve problem in (8) we follow an approach based on the Augmented Lagrangian Method of Multipliers (ADMM). This methodology is convenient since it allows us to solve a wide variety of unconstrained SSIM-based optimization problems by splitting the cost function to be minimized into simpler optimization problems that are easier to solve.

The problem in (8) can be solved efficiently by taking advantage of the fact that the objective function is separable. Let us write Problem (8) in its equivalent constrained form:

$$\min_{x,z} T(\Phi x, y) + \lambda h(z),$$
$$\text{subject to } x - z = 0,\tag{9}$$

where $z \in \mathbb{R}^n$ [2]. Clearly, (9) is equivalent to problem (8), thus by solving it, we automatically obtain a minimizer of the original optimization problem in (8).

As is customary in the ADMM methodology, let us first form the corresponding augmented Lagrangian of (9),

$$L_\rho(x, z, u) = T(\Phi x, y) + \lambda h(z) + \frac{\rho}{2}\|x - z + u\|_2^2, \tag{10}$$

where $u = v/\rho$ is a scaled dual variable [2]. As expected, the iterations of the proposed algorithm for solving (9) will be the minimization of Eq. (10) with respect to variables x and z in an alternate fashion, and the update of the dual variable u, which accounts for the maximization of the dual function $g(u)$:

$$g(u) := \inf_{x,y} L_\rho(x, z, u). \tag{11}$$

Thus, we define the following iteration for minimizing the cost function of the equivalent counterpart of problem (8):

$$x^{k+1} := \underset{x}{\operatorname{argmin}} \left(T(\Phi x, y) + \frac{\rho}{2}\|x - z^k + u^k\|_2^2 \right), \tag{12}$$

$$z^{k+1} := \underset{z}{\operatorname{argmin}} \left(h(z) + \frac{\rho}{2\lambda}\|x^{k+1} - z + u^k\|_2^2 \right), \tag{13}$$

$$u^{k+1} := u^k + x^{k+1} - z^{k+1}. \tag{14}$$

Observe that the x-update can be computed using the algorithm introduced in [12] for differentiable regularizing terms. Furthermore, when h is convex, the z-update is equal to the *proximal operator* of $(\lambda/\rho)h$ [14]. Recall that for a convex function $f : \mathbb{R}^n \to \mathbb{R}$ its proximal operator $\mathbf{prox}_f : \mathbb{R}^n \to \mathbb{R}^n$ is defined as

$$\mathbf{prox}_f(v) := \underset{x}{\operatorname{argmin}} \left(f(x) + \frac{1}{2}\|x - v\|_2^2 \right). \tag{15}$$

It then follows that

$$z^{k+1} := \mathbf{prox}_{\frac{\lambda}{\rho}h}(x^{k+1} + u^k). \tag{16}$$

Given the latter, we introduce the following algorithm for solving Problem (8).

Algorithm I: ADMM-BASED METHOD FOR UNCONSTRAINED
SSIM-BASED OPTIMIZATION

initialize $x = z = x_0$, $u = \mathbf{0}$;
data preprocessing $y = y - \frac{1}{n}\mathbf{1}^T y$;
repeat
    $x := \operatorname{argmin}_x \left(T(\Phi x, y) + \frac{\rho}{2}\|x - z + u\|_2^2 \right)$;
    $z := \operatorname{argmin}_z \left(h(z) + \frac{\rho}{2\lambda}\|x - z + u\|_2^2 \right)$;
    $u := u + x - z$;
until stopping criterion is satisfied.
return x.

---

4 Applications

As anticipated, by choosing different types regularization terms and linear operators, a wide variety of SSIM-based imaging tasks can be performed. In this section, due to space limitations, we review just one application that has been barely studied, namely, SSIM-TV denoising. For the interested reader, more applications and experimental results can be found in [11–13].

4.1 SSIM-TV Denoising

Until now we have worked with vectors, nevertheless, the proposed algorithm can be easily adapted for dealing with matrices, which are the digital counterparts of images. In the particular case of denoising, images can be denoised in the following fashion. Let $Y \in \mathbb{R}^{m \times n}$ be a noisy image. Also, let $V : \mathbb{R}^{m \times n} \to \mathbb{R}^{mn \times 1}$ be a linear transformation that converts matrices into column vectors, that is,

$$V(A) = \text{vec}(A) = [a_{11}, a_{21}, \dots, a_{(m-1)n}, a_{mn}]^T, \tag{17}$$

where $A \in \mathbb{R}^{m \times n}$.

As mentioned before, it is more convenient to employ an average of local SSIMs as a fidelity term. Let $\{Y_i\}_{i=1}^N$ be a partition of the given image Y such that $\cup_{i=1}^N Y_i = Y$. Further, let $\{X_i, Z_i\}_{i=1}^N$ also be partitions of the variables X and Z such that $\cup_{i=1}^N X_i = X$ and $\cup_{i=1}^N Z_i = Z$. Also, let $MT : \mathbb{R}^{m \times n} \times \mathbb{R}^{m \times n} \to \mathbb{R}$ be given by

$$MT(X, Y) = \frac{1}{N} \sum_{i=1}^N T(V(X_i), V(Y_i)). \tag{18}$$

Then, the optimization problem that is to be solved is

$$\min_X MT(X, Y) + \lambda \|X\|_{TV}, \tag{19}$$

where the regularizing term is a discretization of the isotropic TV seminorm for real-valued images [6].

If $\{Y_i, X_i \, Z_i\}_{i=1}^N$ are partitions of non-overlapping blocks, the problem in (19) can be solved by carrying out the following iterations,

$$X_i^{k+1} := \underset{X_i}{\text{argmin}} \left(T(V(X_i), V(Y_i)) + \frac{N\rho}{2} \|X_i - Z_i^k + U_i^k\|_F^2 \right), \tag{20}$$

$$Z^{k+1} := \underset{Z}{\text{argmin}} \left(\|Z\|_{TV} + \frac{\rho}{2\lambda} \|Z - X^{k+1} - U^k\|_F^2 \right), \tag{21}$$

$$U^{k+1} := U^k + X^{k+1} - Z^{k+1}, \tag{22}$$

where $\|\cdot\|_F$ is the Frobenius norm and U_i is an element of the partition of the dual variable U. As expected, $\cup_{i=1}^N U_i = U$, and $U_i \cap U_j = \varnothing$ for all $i \neq j$. Notice that the Z-update may be computed efficiently by using the algorithm introduced by Chambolle in [6]. The extension of this algorithm when a weighted average of local SSIMs is used as a measure of similarity between images is straightforward.

We close this section by mentioning that to the best of our knowledge, the contributions reported in [11,13,17] along with the applications presented above are the only approaches in the literature that combine TV and the SSIM.

5 Experiments

In the following experiments, the denoising of some images corrupted with Additive White Gaussian Noise (AWGN) was performed. Although from a maximum a posteriori (MAP) perspective the ADMM-SSIM approach is not optimal, it is worthwhile to see how denoising is carried out when the SSIM-based metric is employed as a fidelity term.

It is important to mention that in order to reduce blockiness in the reconstructions the mean of each non-overlapping pixel block is not subtracted prior to processing. This implies that the fidelity term defined in (18) is not equivalent, but is based on the dissimilarity measure introduced in Sect. 2. Despite this, the experiments presented below suggest that this fidelity measure may be used as a substitute of the SSIM.

In all experiments, we employed non-overlapping pixel blocks. Performance of the ℓ_2- and SSIM-based approaches is assessed by computing the MSSIM of the original images and their corresponding reconstructions. Here, the MSSIM is simply the average of the SSIM values of all non-overlapping blocks.

As expected, the noiseless approximation is obtained by solving Problem (19). To evaluate the performance of the proposed ADMM-SSIM method, we compare it with its ℓ_2 counterpart, namely,

$$\min_X \ \|X - Y\|_2^2 + \lambda \|X\|_{TV}. \tag{23}$$

Naturally, Chambolle's algorithm can be employed for solving this optimization problem [6]. In order to compare the effectiveness of the proposed approach and Chambolle's method (TV), regularization was carried out in such a way that the TV seminorms of the reconstructions yielded by both methods are the same.

In Fig. 1, some visual results are shown. We employed the test image *Lena*. The noisy image, as well as the SSIM map, can be observed in the first row. The reconstructed and original images are presented in the second row. The TV seminorm of the reconstruction is 2500 for *Lena*. The Peak Signal-to-Noise Ratio (PSNR) prior to denoising was 18.067 dB in all experiments.

It is evident that the proposed method performs significantly better than its ℓ_2 counterpart. Notice that some features of the original *Lena* are better reconstructed (e.g., the eyes in *Lena*), whereas in the ℓ_2 reconstruction these features are considerably blurred. This is mainly due to the fact that the noise does not completely hide some of the more important attributes of the original image. Since the fidelity term enforces the minimizer of problem (19) to be visually as similar as possible as the given noisy observation, while denoising is still accomplished, the reconstruction yielded by the ADMM-SSIM approach is visually more similar to the noiseless image. As for MSSIM values, these are 0.4386 and 0.6468 for the ℓ_2 and ADMM-SSIM reconstructions, respectively.

Fig. 1. Some visual results for the denoising of the test image *Lena*. The TV semi-norm of both reconstructed images is 2500. **Top row:** Noisy image along with SSIM maps between each reconstructed image and the original. **Bottom row:** Original and ADMM-SSIM- and ℓ^2-based denoised images, with MSSIM values 0.6468 and 0.4386, respectively.
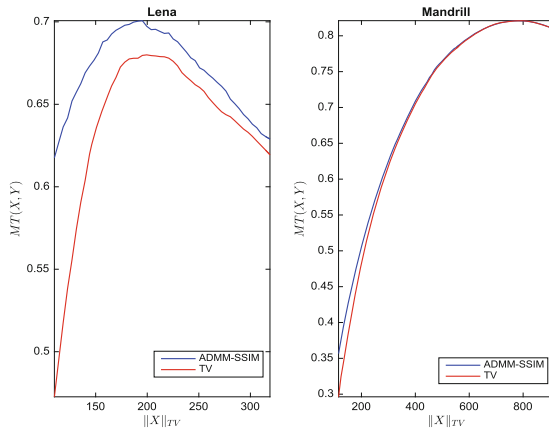

Fig. 2. The behaviour of the average SSIM of reconstructed images obtained from the proposed SSIM-based method and the classical ℓ_2 method as a function of the TV seminorm of the reconstruction. **Left:** The *Lena* image. **Right:** The *Mandrill* image. In the case of the *Lena* image, the SSIM-based approach clearly outperforms the classical ℓ_2 method. For the *Mandrill* image, however, the performance of both methods is, in general, very similar.

In order to have a general idea of the effectiveness of the SSIM-based methodology when regularization varies, in Fig. 2, we show the behaviour of the MSSIM as a function of the TV seminorm of the reconstructions obtained by both the ADMM-SSIM and the ℓ_2 approaches. The plot on the left shows the behaviour

of the MSSIM for a noisy image patch of *Lena* whereas the plot on the right shows the results for a corrupted image patch of *Mandrill*. As expected, the plot on the right hand side shows that for images with low regularity—such as *Mandrill*—the ADMM-SSIM and ℓ_2 methods exhibit similar effectiveness over a wide range of regularization values. On the other hand, for the image *Lena*, one observes a significant difference between the performances of the two methods. This suggests that when strong regularization is required, it is more advantageous to employ SSIM-based techniques over ℓ_2 methods if certain visual features need to be recovered, provided that the reconstruction possesses some degree of regularity.

References

1. Boyd, S., Vandenberghe, L.: Convex Optimization. Cambridge University Press, Cambridge (2004)
2. Boyd, S., Parikh, N., Chu, E., Peleato, B., Eckstein, J.: Distributed optimization and statistical learning via the alternating direction method of multipliers. Found. Trends Mach. Learn. **3**, 1–122 (2010)
3. 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). https://doi.org/10.1007/978-3-642-13772-3_2
4. Brunet, D., Vrscay, E.R., Wang, Z.: On the mathematical properties of the structural similarity index. IEEE Trans. Image Process. **21**, 1488–1499 (2012)
5. Brunet, D., Channappayya, S., Wang, Z., Vrscay, E.R., Bovik, A.: Optimizing image quality. In: Handbook of Convex Optimization Methods in Imaging Science, pp. 15–41 (2017)
6. Chambolle, A.: An algorithm for total variation minimization and applications. J. Math. Imaging Vis. **20**, 89–97 (2004)
7. Chambolle, A., Caselles, V., Cremers, D., Novaga, M., Pock, T.: An introduction to total variation for image analysis. Theor. Found. Numer. Methods Sparse Recover. **9**, 263–340 (2010)
8. Chambolle, A., Pock, T.: A first-order primal-dual algorithm for convex problems with applications to imaging. J. Math. Imaging Vis. **40**, 120–145 (2011)
9. Channappayya, 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**, 857–872 (2008)
10. Elad, M., Aharon, M.: Image denoising via sparse and redundant representations over learned dictionaries. IEEE Trans. Image Process. **15**, 3736–3745 (2006)
11. Otero, D.: Function-valued mappings and SSIM-based optimization in imaging, Ph.D. thesis, University of Waterloo, Waterloo, ON, Canada (2015)
12. Otero, D., Vrscay, E.R.: Unconstrained structural similarity-based optimization. In: Campilho, A., Kamel, M. (eds.) ICIAR 2014. LNCS, vol. 8814, pp. 167–176. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-11758-4_19

13. Otero, D., Vrscay, E.R.: Solving optimization problems that employ structural similarity as the fidelity measure. In: Proceedings of the International on Image Processing, Computer Vision and Pattern Recognition, pp. 474–479. CSREA Press (2014)
14. Parikh, N., Boyd, S.: Proximal algorithms. Found. Trends Optim. **1**, 123–231 (2013)
15. Rehman, A., Rostami, M., Wang, Z., Brunet, D., Vrscay, E.R.: SSIM-inspired image restoration using sparse representation. EURASIP J. Adv. Signal Process. **2012**, 1–12 (2012)
16. Rehman, A., Gao, Y., Wang, J., Wang, Z.: Image classification based on complex wavelet structural similarity. Signal Process. Image Commun. **28**, 984–992 (2013)
17. Shao, Y., Sun, F., Li, H., Liu, Y.: Structural similarity-optimal total variation algorithm for image denoising. In: Sun, F., Hu, D., Liu, H. (eds.) Foundations and Practical Applications of Cognitive Systems and Information Processing. AISC, vol. 215, pp. 833–843. Springer, Heidelberg (2014). https://doi.org/10.1007/978-3-642-37835-5_72
18. Wang, S., Rehman, A., Wang, Z., Ma, S., Gao, W.: SSIM-motivated rate-distortion optimization for video coding. IEEE Trans. Circuits Syst. Video Techn. **22**, 516–529 (2012)
19. Wang, Z., Bovik, A.C.: A universal image quality index. IEEE Signal Process. Lett. **9**, 81–84 (2002)
20. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. IEEE Trans. Image Process. **13**, 600–612 (2004)