A Convex Relaxation of the Ambrosio–Tortorelli Elliptic Functionals for the Mumford–Shah Functional*

Youngwook Kee and Junmo Kim
KAIST, South Korea

Abstract

In this paper, we revisit the phase-field approximation of Ambrosio and Tortorelli for the Mumford–Shah functional. We then propose a convex relaxation for it to attempt to compute globally optimal solutions rather than solving the nonconvex functional directly, which is the main contribution of this paper. Inspired by McCormick's seminal work on factorable nonconvex problems, we split a nonconvex product term that appears in the Ambrosio–Tortorelli elliptic functionals in a way that a typical alternating gradient method guarantees a globally optimal solution without completely removing coupling effects. Furthermore, not only do we provide a fruitful analysis of the proposed relaxation but also demonstrate the capacity of our relaxation in numerous experiments that show convincing results compared to a naive extension of the McCormick relaxation and its quadratic variant. Indeed, we believe the proposed relaxation and the idea behind would open up a possibility for convexifying a new class of functions in the context of energy minimization for computer vision.

1. Introduction

1.1. Observation

Let us consider a minimization problem with a bivariate objective, $f(x,y) = x^2y^2$ on $[0,1] \times [0,2.5] \subset \mathbb{R}^2$, which is illustrated in the top row of Figure 1. Since the function is differentiable with respect to both x and y, one could possibly make use of a naive coordinatewise descent method to attempt to compute a globally optimal solution. Obviously, the function is in itself nonconvex; yet interestingly, one can always find a global minimizer using gradient methods regardless of initialization. Even so, the objective is somewhat problematic in that it prefers a specific set of solutions depending on initial points. For instance, if an initial guess were to be selected on a region such that $y < 2.5x$, minimizers on the other side would not be obtained. The bottom row in Figure 1, on the other hand, shows how a particular solution $(0,0)$ among all possible candidates of the original problem can be surprisingly computed without being prone to getting stuck in local minima. More importantly, unlike the McCormick relaxation (see Section 3 for details), one always has a descent direction everywhere in domain $[0,1] \times [0,2.5] \subset \mathbb{R}^2$ by relaxing $f(x,y)$ with a combination of a quadratic and a linear function.

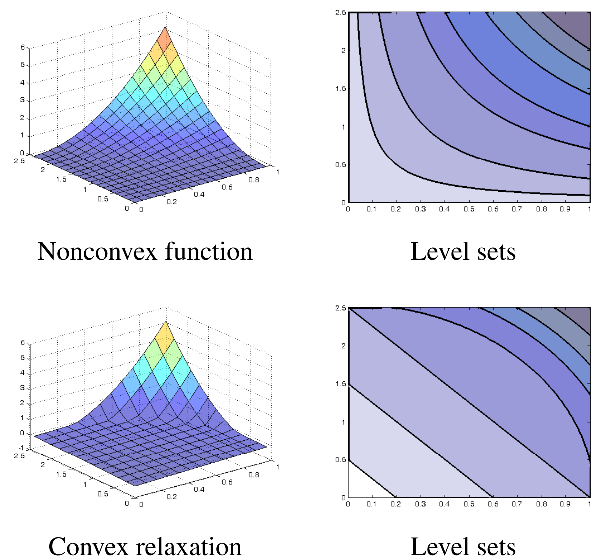


Figure 1. We propose a convex relaxation (**bottom**) for a nonconvex function (**top**) that arises in the Ambrosio–Tortorelli set-up of the Mumford–Shah functional. Compared to McCormick's convex envelope (see Section 3 for details), the proposed relaxation makes it possible to use nonzero descent directions on the entire domain so that it leads to visually better solutions than those derived from the naive relaxation.

At first glance, this observation may not be considered significant in that minimizers[1] of $\min f(x,y)$ on $[0,1] \times [0,2.5]$ are all global minimizers. However, if it appears

*This work was supported in part by an Erasmus Mundus BEAM fellowship (No. L031000107), in part by the National Research Foundation of Korea under Grant 2010-0028680, and in part by the Ministry of Trade, Industry and Energy under Grant 10045252.

[1] *i.e.* $(x,y) \in \{[a,0], 0 \le a \le 1\} \cup \{[0,b], 0 \le b \le 2.5\}$

in a combined form with another (even with a convex objective), in contrast to situations where such a function is the only objective to be minimized, one cannot be sure of the global optimality of a computed solution. Indeed, when it comes to energy minimization problems in computer vision, finding a globally optimal solution becomes a major algorithmic challenge because objectives typically consist of two or three energies—namely the data fidelity term and the regularization term—leading to overall nonconvex objectives.

In light of the observation, we consider throughout the paper how such a nonconvex functional that appears in the Ambrosio–Tortorelli Γ-*convergence* set-up for the Mumford–Shah functional can be convexified assuring near-optimal solutions. It turns out that solutions from the proposed relaxation are energetically and visually better than those computed by minimizing the original nonconvex functional. This is the main contribution of the paper.

1.2. The Mumford–Shah Problem

Let $\Omega \subset \mathbb{R}^n$ be a bounded open set and $g \in L^\infty(\Omega)$ be a given function. The Mumford–Shah problem is given by

$$\min_{u,K} \left\{ \alpha \int_\Omega |u-g|^2 + \int_{\Omega\setminus K} |\nabla u|^2 + \beta \mathcal{H}^{n-1}(K) \right\}, \quad (1)$$
$$u \in W^{1,2}(\Omega\setminus K), K \subset \Omega \text{ closed in } \Omega, \mathcal{H}^{n-1}(K) < \infty,$$

where $\mathcal{H}^{n-1}$ is the $(n-1)$-dimensional Hausdorff measure, and $\alpha, \beta > 0$ are fixed parameters. The functional regularizes large sets of K, and prefers u outside the set K to be close to g in $W^{1,2}$, namely a piecewise smooth approximation of g. The functional was introduced by Mumford and Shah in [21] for image segmentation (*i.e.*, $n = 2$ and Ω may well be a rectangle).

The existence of a minimizer has been proved in [12, 10] where the idea is to use a weak formulation of the problem in a special class of functions of bounded variation $SBV(\Omega)$—readers can find a complete analysis in [2]—as follows:

$$\min_{u} \left\{ \alpha \int_\Omega |u-g|^2 + \int_{\Omega\setminus S_u} |\nabla u|^2 + \beta \mathcal{H}^{n-1}(S_u) \right\}, \quad (2)$$

where $u \in SBV(\Omega)$ and S_u is the discontinuity set of u. Then, for any minimizer u of (2), one can recover a minimizer of (1) by setting $K = \overline{S_u} \cap \Omega$. Since then, the Mumford–Shah functional has been intensively studied, in particular on the regularity of minimizing pairs (u, K), in applied mathematics [20, 11] with free discontinuity problems [2]. Notwithstanding the existence theorem and some results on the regularity, exact computation of solutions for the functional is limited to a significant extent because of its nonconvexity. As a consequence, there has been extensive research on efficient algorithms both in the context of continuous/discrete optimization (we refer readers to [22, 17] and references therein for the continuous and discrete setting, respectively).

The most related work—possibly in line with the *phase-field* approximation of Ambrosio and Tortorelli [3] for which we are going to present a convex relaxation—is those based on the Γ-*convergence* set-up. That is to approximate the functional in (2) by a sequence of regular functionals defined on Sobolev spaces converging to it in the sense of Γ-convergence. We will revisit the Ambrosio–Tortorelli elliptic functionals in Section 2; and refer readers to [5, 6, 8, 15] for different classes of approximating functionals.

In [1], Alberti *et al*. reformulated (2) by means of the flux of a suitable vector field going through the interface of the subgraph of u and provided a sufficient condition for minimality of some pairs of (u, S_u). Unfortunately, it is still open whether there exists a divergence free vector field (calibration) for each minimizer of (2). Later, Pock *et al*. [22] proposed a fast primal-dual algorithm by relaxing the characteristic function of the subgraph of u to $[0, 1]$, and Strekalovskiy *et al*. [24] extended it to a vectorial case. However, not only is it computationally demanding to compute a minimizer of the relaxed problem but also it does not hold the coarea formula [14] so that global minimality is not guaranteed.

Note that we do not compare our results with those from [22] and [24] throughout the paper since our contribution is solely to convexify the Ambrosio–Tortorelli elliptic functionals which converge to the weak formulation (2) in the sense of Γ-convergence. In contrast, the sufficient condition for minimality based on the calibration method [1] is somewhat problematic as mentioned above.

2. The Ambrosio–Tortorelli Functionals

In [3], Ambrosio and Tortorelli proposed a sequence of elliptic functionals to approximate the Mumford–Shah functional in (2) as follows:

$$AT_\epsilon(u,z) = \alpha \int_\Omega |u-g|^2\,dx + \int_\Omega z^2|\nabla u|^2\,dx + \beta \underbrace{\int_\Omega \left(\epsilon|\nabla z|^2 + \frac{(z-1)^2}{4\epsilon} \right) dx}_{L_\epsilon(z)}, \quad (3)$$

where $\epsilon > 0$ is a fixed parameter; then $AT_\epsilon(u,z)$ is well defined on the space $\{(u,z) \in W^{1,2}(\Omega)^2 : 0 \le z \le 1\}$. Here, z is a smooth edge indicator[2] (*i.e.*, $z \to 0$ when $|\nabla u| \to \infty$). Remarkably, the authors have proved that $AT_\epsilon(u,z)$ admits a minimizer and that $AT_\epsilon(u,z)$ converges

[2]Note that z is also called the *phase-field*, and $L_\epsilon(z)$ in (3) is known as the *phase-field energy*.

Figure 2. **A pair of minimizers of the Ambrosio–Tortorelli functional.** In order to compute a pair of minimizers (u^*, z^*) for the Ambrosio–Tortorelli functional, one typically resorts to an alternating optimization technique that does not guarantee a global optimal solution due to the fact that the functional in itself is nonconvex. Yet, in practice it gives quite acceptable solutions for a number of images.

to the Mumford–Shah functional (2) as $\epsilon \to 0$ in the sense of Γ-*convergence*.

Although the Mumford–Shah functional has been nicely approximated by such sequence of elliptic functionals (where one can easily derive its Euler-Lagrange equations and solve these alternatively—see Figure 2), computing a globally optimal solution is indeed a major algorithmic challenge since (3) is still nonconvex—the second term is a function of both u and z. In what follows, therefore, we focus on the second term, $\int_\Omega z^2|\nabla u|^2\,dx$, and propose a convex relaxation for it.

3. Convex Relaxation

For the sake of completeness, we start by recalling McCormick's seminal work on factorable nonconvex problems [18]. The main result that is going to be used for the first relaxation for (3) is summarized as follows [19].

Theorem 1 (McCormick's relaxation of products)**.** *Let $\Omega \subset \mathbb{R}^2$ be a nonempty convex set and $f, f_1, f_2 : \Omega \to \mathbb{R}$ such that $f = f_1 f_2$. Let $f_1^\cup, f_1^\cap : \Omega \to \mathbb{R}$ be a convex and concave relaxation*[3] *of f_1, respectively. Likewise, let $f_2^\cup, f_2^\cap : \Omega \to \mathbb{R}$ be a convex and concave relaxation of f_2. Then for f_1 and f_2 such that $L_1 \le f_1 \le U_1$ and $L_2 \le f_2 \le U_2$, where $L_1, L_2, U_1, U_2 \in \mathbb{R}$,*

$$f^\cup = \max\{g_1 + g_2 - L_1L_2, h_1 + h_2 - U_1U_2\}, \tag{4}$$

where

$$g_1 = \min\{L_2 f_1^\cup, L_2 f_1^\cap\},\ \ g_2 = \min\{L_1 f_2^\cup, L_1 f_2^\cap\},$$
$$h_1 = \min\{U_2 f_1^\cup, U_2 f_1^\cap\},\ \ h_2 = \min\{U_1 f_2^\cup, U_1 f_2^\cap\}, \tag{5}$$

is a convex relaxation of f.

An example Consider $f = x^2y^2$ on $[0,1] \times [0, 2.5] \subset \mathbb{R}^2$, which is illustrated in Figure 3 with its level sets, then $0 \le x^2 \le 1$ and $0 \le y^2 \le 2.5^2$. We substitute f_1 and f_2 with x^2 and y^2, respectively. By Theorem 1, their convex and concave relaxations on $[0,1] \times [0, 2.5]$ are $f_1^\cup = x^2$, $f_1^\cap = x$, $f_2^\cup = y^2$, and $f_2^\cap = 2.5y$. Since $g_1 = 0$, $g_2 = 0$, $h_1 = \min\{2.5^2x^2, 2.5^2x\}$, and $h_2 = \min\{y^2, 2.5y\}$, $f^\cup = \max\{0, 2.5^2x^2 + y^2 - 2.5^2\}$; see the McCormick relaxation in Figure 3.

We now derive a similar relaxation for $\int_\Omega z^2|\nabla u|^2\,dx$ in (3), as follows.

Corollary 1 (McCormick relaxation)**.** *Let $u : \Omega \to \mathbb{R}$ be Lipschitz continuous; then*

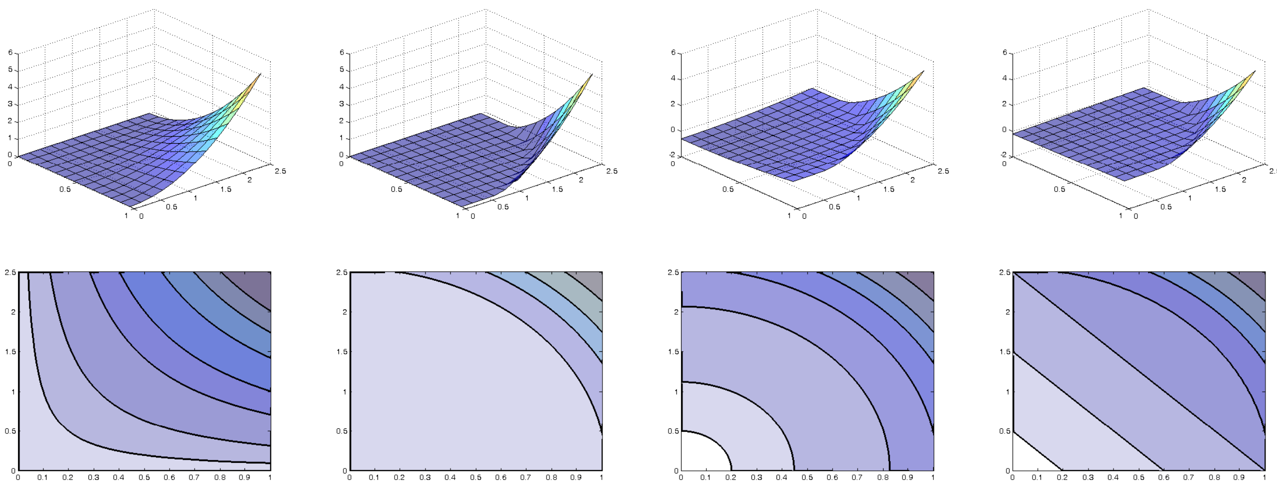
$$F_{\mathrm{Mc}}(u, z) := \int_\Omega \max\{0, L^2z^2 + |\nabla u|^2 - L^2\}\,dx \tag{6}$$
$$\le \int_\Omega z^2|\nabla u|^2\,dx, \tag{7}$$

where L is a Lipschitz constant for u, and $F_{\mathrm{Mc}}(u, z)$ is convex on $\big(W^{1,2}(\Omega) \cap W^{1,\infty}(\Omega)\big) \times W^{1,2}(\Omega)$.

Proof. Substitute f_1 and f_2 with z^2 and $|\nabla u|^2$, then consider its convex and concave relaxations: $f_1^\cup = z^2$, $f_1^\cap = z$, $f_2^\cup = |\nabla u|^2$, and $f_2^\cap = L|\nabla u|$, where $0 \le |\nabla u| \le L$ a.e. x in Ω by Lipschitz continuity (by Rademacher's theorem [13]). □

Let us go back to the example for a while. The convex relaxation $f^\cup$ of f is tight in the sense of [18], yet one can see there is no descent direction when $2.5^2x^2 + y^2 - 2.5^2 < 0$.

[3] A convex and a concave relaxation mean $f^\cup \le f$ convex and $f^\cap \ge f$ concave on Ω, respectively.



$f = x^2y^2$ McCormick relaxation Quadratic relaxation Linear relaxation

Figure 3. **Convex relaxations of a nonconvex factorable function,** $f(x, y) = x^2y^2$ **on** $[0, 1] \times [0, 2.5] \subset \mathbb{R}^2$**.** Although the McCormick relaxation of $f(x, y)$, $f^{\cup} = \max\{0, 2.5^2x^2 + y^2 - 2.5^2\}$, is sufficiently tight, one cannot make use of its gradient information when $2.5^2x^2 + y^2 - 2.5^2 < 0$. For a quadratic relaxation, $f^{\cup}_{\mathrm{Q}} = \max\{0.1(2.5^2x^2 + y^2 - 2.5^2), 2.5^2x^2 + y^2 - 2.5^2\}$, on the other hand, it regularizes solutions on the entire domain quadratically. By replacing $0.1(2.5^2x^2 + y^2 - 2.5^2)$ in the quadratic relaxation with a linear function $0.1(2.5x + y - 2.5)$, we obtain a remarkable property; not only is it tighter than the quadratic relaxation, but it also regularizes "linearly" when $0.1(2.5x + y - 2.5) > 2.5^2x^2 + y^2 - 2.5^2$, which turns out to preserve edges.

Indeed, this is quite problematic in such a situation where one would have to rely on gradient-based algorithms to compute a globally optimal solution for the original problem by means of $f^{\cup}$. Interestingly, by simply replacing the first argument with the second one multiplied by $\eta \in (0, 1]$, we can devise a less tight yet subdifferentiable convex underestimator of $f^{\cup}$, that is, $f^{\cup}_{\mathrm{Q}} = \max\{\eta(2.5^2x^2 + y^2 - 2.5^2), 2.5^2x^2 + y^2 - 2.5^2\}$. We illustrate this "quadratic relaxation" in Figure 3 when $\eta = 0.1$ with its level sets.

Likewise, $F_{\mathrm{Mc}}(u, z)$ can similarly be relaxed further as follows.

Proposition 1 (Quadratic relaxation)**.** *Let u be Lipschitz continuous; then for any $\eta \in (0, 1]$,*

$$F_{\mathrm{Q},\eta}(u, z) := \int_{\Omega} \max\{\eta(L^2z^2 + |\nabla u|^2 - L^2), L^2z^2 + |\nabla u|^2 - L^2\}\, dx \tag{8}$$

$$\leq F_{\mathrm{Mc}}(u, z), \tag{9}$$

where L is a Lipschitz constant for u, and $F_{\mathrm{Q},\eta}(u, z)$ is convex on $\big(W^{1,2}(\Omega) \cap W^{1,\infty}(\Omega)\big) \times W^{1,2}(\Omega)$.

Proof. When $\eta(L^2z^2 + |\nabla u|^2 - L^2) \geq 0$, $F_{\mathrm{Q},\eta} = F_{\mathrm{Mc}}$; when $\eta(L^2z^2 + |\nabla u|^2 - L^2) < 0$, $F_{\mathrm{Q},\eta} \leq 0$ and $F_{\mathrm{Mc}} = 0$. □

Now, let us take a look at $F_{\mathrm{Q},\eta}(u, z)$. In both cases, when $\eta(L^2z^2 + |\nabla u|^2 - L^2) > L^2z^2 + |\nabla u|^2 - L^2$ and $\eta(L^2z^2 + |\nabla u|^2 - L^2) < L^2z^2 + |\nabla u|^2 - L^2$, one can see that the $\int_{\Omega} |\nabla u|^2\, dx$ term is involved, which obviously does not have the edge-preserving property [9]. Although a small value of η somewhat guarantees a small amount of diffusion, the nature of $\int_{\Omega} |\nabla u|^2\, dx$ never vanishes unless $\eta = 0$, which becomes the McCormick relaxation. This is indeed problematic since it becomes a looser lower bound, which would give us a solution far from optimal. As a remedy, we propose the following relaxation of the quadratic relaxation, which in turn preserves edges.

Proposition 2 (Linear approximation[4])**.** *Let u be Lipschitz continuous; and for any $\eta \in (0, 1]$, $F_{\mathrm{L},\eta}(u, z)$ is defined by*

$$F_{\mathrm{L},\eta}(u, z) := \int_{\Omega} \max\{\eta(Lz + |\nabla u| - L), L^2z^2 + |\nabla u|^2 - L^2\}\, dx, \tag{10}$$

where L is a Lipschitz constant for u. Then $F_{\mathrm{L},\eta}(u, z)$ is convex on $\big(W^{1,2}(\Omega) \cap W^{1,\infty}(\Omega)\big) \times W^{1,2}(\Omega)$.

Proof. Pointwise max operation preserves convexity. □

We also illustrate this "linear relaxation" for $f^{\cup}_{\mathrm{Q}}$ in Figure 3 when $\eta = 0.1$ with its level sets. Lastly, we close this section with the following proposition, which shows

[4]Note that the linear approximation is not an underestimator of the second term in (3) on the entire domain. However, we will call this approximation also relaxation throughout the paper; indeed, this is the relaxation that has been referred to as the proposed relaxation.

that when one of the three types of relaxations presented is plugged in the Ambrosio–Tortorelli functional (3), the overall functional becomes a convex relaxation of the functional.

Proposition 3. *Let u be Lipschitz continuous; then for every $\epsilon > 0$ and $\eta \in (0,1]$, a functional $AT_\epsilon^{\mathrm{cvx}} : (W^{1,2}(\Omega) \cap W^{1,\infty}(\Omega)) \times W^{1,2}(\Omega) \to \mathbb{R}$, which maps*

$$(u, z) \longmapsto \alpha \int_\Omega |u - g|^2 \, dx + F(u, z) + \beta L_\epsilon(z) \tag{11}$$

*is convex. Here, $F(u,z)$ is one of the relaxations (*i.e.*, F_{Mc}, $F_{\mathrm{Q},\eta}$, and $F_{\mathrm{L},\eta}$) for $\int_\Omega z^2 |\nabla u|^2 \, dx$.*

Proof. The sum of convex functionals is convex. □

3.1. Euler-Lagrange Equations

So far, we have devised such convex relaxations in a way that they satisfy two properties: 1) differentiability and 2) tightness. We have to admit that there is a heuristic to derive a linear relaxation from the quadratic one; that is, the quadratic relaxation has been lifted by a linear approximation which also maintains differentiability and convexity. Interestingly, it turns out that the linear relaxation has a compelling property that comes into view, once its Euler-Lagrange equations are derived. In the following, therefore, we derive the Euler-Lagrange equations for the relaxations and provide complete insights—how these relaxations work.

McCormick relaxation When $L^2 z^2 + |\nabla u|^2 - L^2 < 0$,

$$2\alpha(u - g) = 0, \tag{12}$$

$$\beta \partial_z L_\epsilon(z) = 0; \tag{13}$$

elsewhere,

$$2\alpha(u - g) - \mathrm{div}(\nabla u) = 0, \tag{14}$$

$$2L^2 z + \beta \partial_z L_\epsilon(z) = 0. \tag{15}$$

As can be observed, when $L^2 z^2 + |\nabla u|^2 - L^2 < 0$, the steady-state solution is g; elsewhere, one can see there is a diffusion term $\mathrm{div}(\nabla u)$ in (14), which blurs images. Indeed, a combination of those equations makes images exhibit lots of speckles (salt and pepper-like noise)—see Figure 5 for details.

Quadratic relaxation When $\eta(L^2 z^2 + |\nabla u|^2 - L^2) < L^2 z^2 + |\nabla u|^2 - L^2$, the corresponding Euler-Lagrange equations of $F_{Q,\eta}(u,z)$ become (14) and (15); elsewhere,

$$2\alpha(u - g) - \eta \, \mathrm{div}(\nabla u) = 0, \tag{16}$$

$$2\eta L^2 z + \beta \partial_z L_\epsilon(z) = 0. \tag{17}$$

On either side, as can be seen, the Euler-Lagrange equations contain $\mathrm{div}(\nabla u)$. Unlike the McCormick relaxation, however, u is regularized with two different diffusion coefficients. Here the amount of diffusion on each pixel is decided by the $\max$ operation, which can be regarded as an inhomogeneous diffusion [25]. That is, pixels around edges ($|\nabla u|$ gets larger and z gets smaller) are more likely to fall into the region of $\eta(|\nabla u|^2 + L^2 z^2 - L^2) < |\nabla u|^2 + L^2 z^2 - L^2$ resultig in a small amount of diffusion $\eta \, \mathrm{div}(\nabla u)$ in (16) compared with $\mathrm{div}(\nabla u)$ in (14).

Linear relaxation Likewise, when $\eta(Lz + |\nabla u| - L) > L^2 z^2 + |\nabla u|^2 - L^2$,

$$2\alpha(u - g) - \eta \, \mathrm{div}\left(\frac{\nabla u}{|\nabla u|}\right) = 0, \tag{18}$$

$$\eta L + \beta \partial_z L_\epsilon(z) = 0. \tag{19}$$

As can be seen in (18), the Euler-Lagrange equation is the same as that of the ROF model [23]. That is, the proposed relaxation regularizes

$$F_{\mathrm{L},\eta}(u, z) = \eta \int_\Omega |\nabla u| \, dx + \eta \int_\Omega L(z - 1) \, dx, \tag{20}$$

when $\eta(|\nabla u| + Lz - L) > |\nabla u|^2 + L^2 z^2 - L^2$. Interestingly, it turns out that $AT_\epsilon^{\mathrm{cvx}}$ in Proposition 3 with (10) selectively regularizes either $\int_\Omega |\nabla u|$ or $\int_\Omega |\nabla u|^2$ depending on the $\max$ operation, which is similar to the Huber norm.

3.2. Vectorial Case

The $(n-1)$-dimensional Hausdorff measure $\mathcal{H}^{n-1}(S_u)$ in (2) not only is itself very difficult to deal with; but becomes more complicated when it is extended to a vectorial case because one has to reasonably take into account color coupling among channels. On the other hand, the benefit of the Ambrosio–Tortorelli approximation is that it can be done straightforwardly once a norm $|\cdot|$ for $\mathbf{u} : \Omega \to \mathbb{R}^n$ is specified because $\mathcal{H}^{n-1}(S_u)$ is well approximated by $L_\epsilon(z)$. Indeed, there are some possible (and well-studied) choices available: 1) the Frobenius norm [7], 2) the Euclidean norm [4], and 3) a generalization of the Jacobian determinant [16].

However, since those norms of a vector-valued function have been developed in the context of vectorial total variation, meaning their dual formulations are available, meticulous care is required to plug them into the Ambrosio–Tortorelli functionals. For example, it has been proved in [16] that the total variation based on the Jacobian determinant is given by

$$\int_\Omega J_1(\mathbf{u}) dx = \int_\Omega \sigma_1(D\mathbf{u}) dx, \tag{21}$$

where J_1 is a generalization of the Jacobian determinant, and $\sigma_1(D\mathbf{u})$ is the largest singular value of the derivative matrix. It is also mentioned that $\sigma_1(\cdot)$ is not differentiable

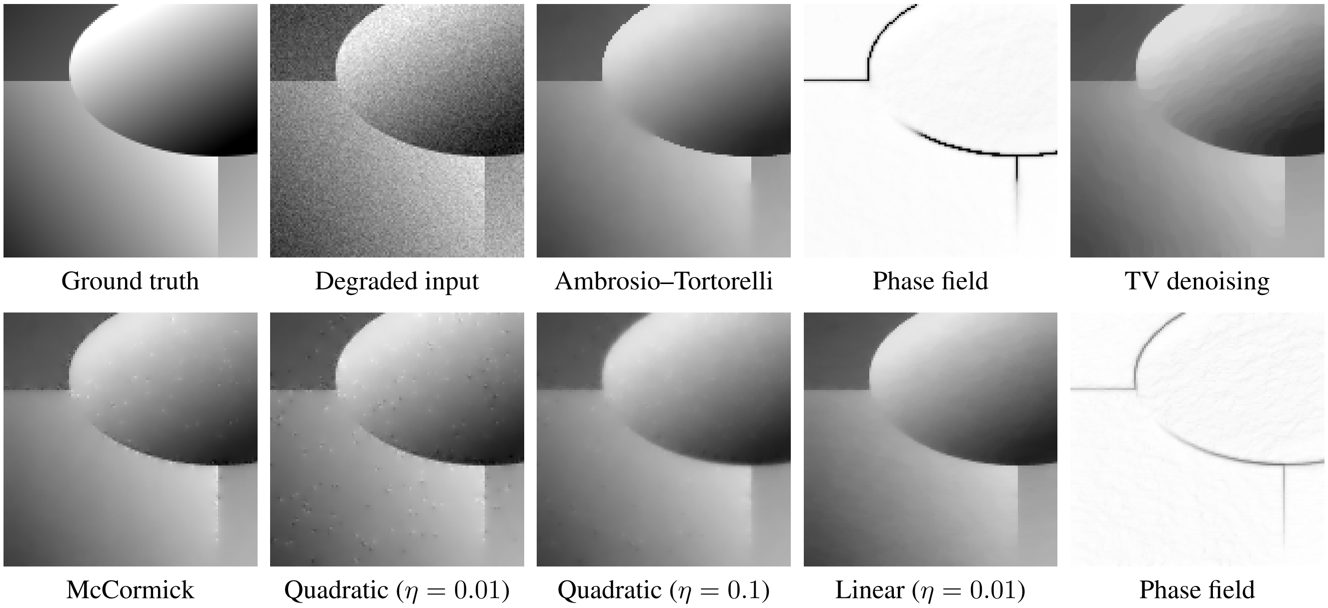


Figure 4. **Piecewise smooth approximations of a synthetic image with different relaxation techniques.** A ground truth image of size 128×128 and one degraded by 5% additive noise are taken from [22]. None of the relaxations exhibit the staircasing effects seen in total variation (TV) denoising. More importantly, compared to the Ambrosio–Tortorelli approximation, a solution of the linear relaxation reconstructs boundaries at the lower right corner. Overall, the linear relaxation shows the best reconstruction result among the possible relaxations where one can see either speckles or blurs.

(they used its dual formulation to minimize it). Obviously, there is no way to differentiate $\sigma_1^2(D\mathbf{u})$.

In [4], Blomgren and Chan observed that the Frobenius norm [7] exhibits color smearing, and proposed the channel-by-channel Euclidean norm. In light of their work, we could extend the second term in (3) as follows

$$\sqrt{\sum_{i=1}^{n} \int_\Omega z^2 |\nabla u_i|^2 \, \mathrm{d}x}, \tag{22}$$

and make use of McCormick's relaxation of composition [19] for further convexification. Nevertheless, we may as well use the Frobenius norm for the vectorial case in this papaer, which can be extended as

$$\int_\Omega z^2 |\nabla \mathbf{u}|_F^2 \, dx := \sum_{i=1}^{n} \int_\Omega z^2 |\nabla u_i|^2 \, dx, \tag{23}$$

since our contribution is solely to decouple the nonconvex factorable term in the Ambrosio–Torotrelli approximation for the Mumford–Shah functional.

4. Experimental Results

In Figure 4, we demonstrate theoretical properties of the (McCormick, quadratic, and linear) relaxations on a synthetic image. Surprisingly, the synthetic image taken from [22] does not make the nonconvexity of the Ambrosio–Tortorelli functionals come into effect. This might explain to some extent why alternating optimization of the functionals works quite well in practice, often giving an energetically better solution ($E = 4.591$) than that of the linear relaxation ($E = 7.4323$, $\eta = 0.01$), where energies are evaluated in the original Ambrosio–Tortorelli functional (3).

However, the Ambrosio–Tortorelli approximation with alternating optimization does not reconstruct the boundaries at the lower right corner whereas the linear relaxation does. Indeed, it turns out that to reconstruct these boundaries is challenging in that neither the McCormick relaxation nor the quadratic variant is capable of carrying out what the linear relaxation does; nor is the functional lifting approach [22]—we refer readers to the reconstruction result of the same image therein. Furthermore, compared to total variation (TV) denoising, none of the relaxations we have presented exhibit a noticeable staircasing effect—see artificial flat regions in the reconstruction result of TV denoising—which is indeed an inherent disaster for the TV regularizer.

Although the McCormick relaxation and its tightness are theoretically interesting, when applied to a task of piecewise smooth approximation it gives rise to lots of speckles with relatively high energy ($E = 27.1154$) since it lacks gradient descent directions. For a solution of quadratic relaxation ($\eta = 0.01$), it turns out that the solution is energetically better ($E = 13.6457$) than that of the McCormick relaxation,

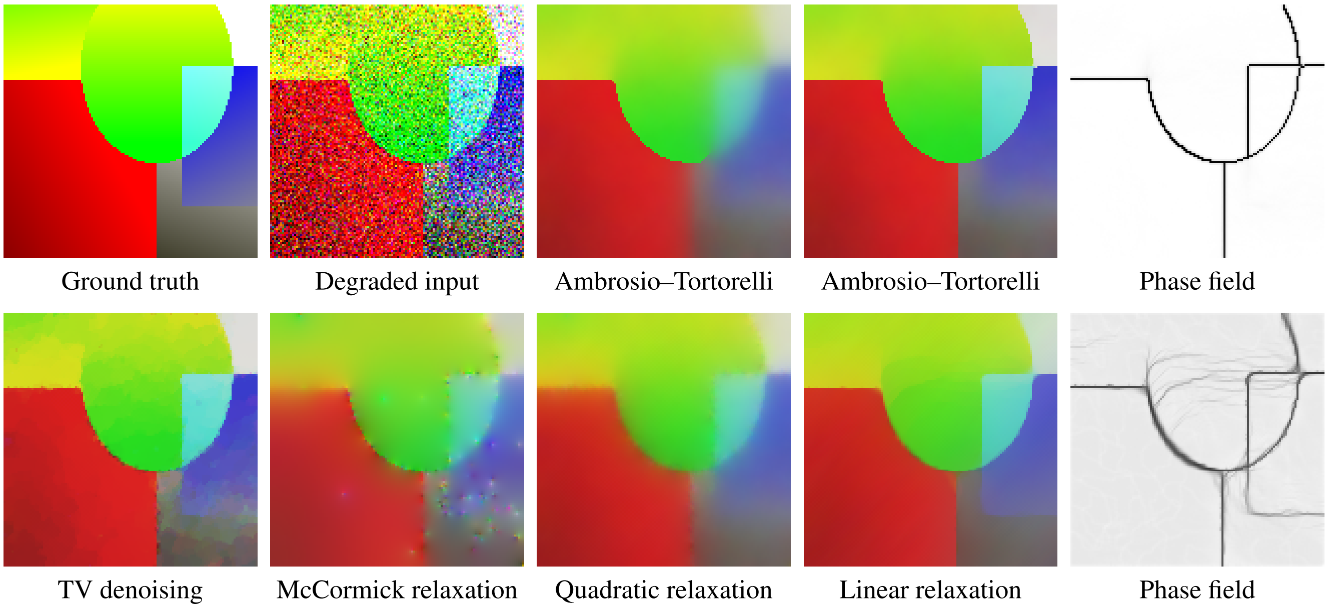


Figure 5. **Piecewise smooth approximations of a vector-valued synthetic image with different relaxation techniques.** A ground truth image of size 128×128 and one degraded by 30% additive noise are taken from [24]. Two piecewise smooth approximations of the Ambrosio–Tortorelli approach (second and third ones from the right on top) are obtained from different initializations (the ground truth and degraded input, respectively). Clearly, the Ambrosio–Tortorelli approximation is prone to getting stuck in local minima. Compared to total variation (TV) denoising, none of the relaxations exhibit the staircasing effects; besides, the linear relaxation reconstructs edges between the blue and black regions whereas none of the other relaxations (including the Ambrosio–Tortorelli approximation) do not.

but gets worse visually; one can clearly see more speckles. A remedy could be to slightly increase the value of η, that is to take into account a looser relaxation, but obviously, its solutions cannot preserve edges since either way it regularizes images by means of the Tikhonov regularizer.

At the beginning of the paper, we considered a nonconvex objective $f(x, y) = x^2y^2$, for which global minimizers are always guaranteed with a naive coordinate-wise gradient method. Also, at the end of Section 1.1, we argued that it becomes problematic when such a function is combined with another—even if it is convex—in that one cannot be sure of its optimality. A convincing example to demonstrate the observation is shown in Figure 5, where none of the solutions of a vectorial extension of the Ambrosio–Tortorelli functionals have the same energy (41.538 and 41.681 each); nor do they look similar. In comparison to a result of total variation (TV) denoising, solutions obtained from the relaxations clearly do not exhibit artificial flat regions. Yet, one can see a significant amount of artifacts (speckles) along the edges in the reconstruction result of the McCormick relaxation (which is basically the same as the scalar case).

For the quadratic relaxation, all the speckles seen in the result of the McCormick relaxation are not observed anymore; yet it overly smoothes edges even with a small value of η (0.08) because of the nature of quadratic regularization. Clearly, the best result in terms of being independent of initialization and significant visual improvements is obtained by the linear relaxation. Moreover, one can see the reconstructed edges between the blue and black regions in the phase field of the linear relaxation (bottom right), which are completely missing in the case of the Ambrosio–Tortorelli functional (top right) even with a ground truth initial. Lastly, a piecewise smooth approximation of a natural image (*Self-Portrait*, September 1889 by Vincent van Gogh) is shown in Figure 6.

5. Conclusion

Inspired by McCormick's seminal work on factorable nonconvex problems, we have shown how a nonconvex functional that appears in the Ambrosio–Tortorelli approximation can be systematically convexified. From a theoretical point of view, however, the relationship of the proposed approach to the original Mumford–Shah functional does not seem quite clear. Indeed, there are still quite a number of theoretical properties to be discovered beyond what we have shown in this paper. Yet, we have demonstrated the capacity of our approach (linear relaxation) in numerous experiments that assure near-optimal solutions of the Ambrosio–Tortorelli functionals. We believe it would provide powerful machinery for optimizing a new class of challenging nonconvex problems that have not been explored so far but appear quite often—namely factorable nonconvex functionals—in the context of variational methods for computer vision.

Input | Ambrosio–Tortorelli | TV denoising | Proposed relaxation

Figure 6. **Piecewise smooth approximation of *Self-Portrait*, September 1889 by Vincent van Gogh.** While the Ambrosio–Tortorelli approximation often gives a fairly acceptable solution, the solution is inconsistent in the sense that one cannot obtain the same solution even with a slightly different initialization. On the other hand, solutions of the proposed method are always guaranteed to be energetically equivalent regardless of any initialization as well as it does not exhibit the staircasing effects.

References

[1] G. Alberti, G. Bouchitté, and G. Dal Maso. The calibration method for the Mumford–Shah functional and free-discontinuity problems. *Calc. Var. Partial Differential Equations*, 16(3):299–333, 2003. 2

[2] L. Ambrosio, N. Fusco, and D. Pallara. *Functions of Bounded Variation and Free Discontinuity Problems*. Oxford: Clarendon Press, 2000. 2

[3] L. Ambrosio and V. M. Tortorelli. Approximation of functional depending on jumps by elliptic functional via Γ-convergence. *Comm. Pure Appl. Math.*, 43(8):999–1036, 1990. 2

[4] P. Blomgren and T. F. Chan. Color TV: total variation methods for restoration of vector-valued images. *IEEE Trans. Image Process.*, 7(3):304–309, 1998. 5, 6

[5] A. Braides. *Approximation of Free-Discontinuity Problems*. Springer, 1998. 2

[6] A. Braides and G. Dal Maso. Non-local approximation of the Mumford–Shah functional. *Calc. Var. Partial Differential Equations*, 5(4):293–322, 1997. 2

[7] X. Bresson and T. F. Chan. Fast dual minimization of the vectorial total variation norm and applications to color image processing. *Inverse Probl. Imaging*, 2(4):455–484, 2008. 5, 6

[8] A. Chambolle. Image segmentation by variational methods: Mumford and Shah functional and the discrete approximations. *SIAM J. Appl. Math.*, 55(3):827–863, 1995. 2

[9] A. Chambolle, V. Caselles, D. Cremers, M. Novaga, and T. Pock. An introduction to total variation for image analysis. In *Theoretical Foundations and Numerical Methods for Sparse Recovery*. De Gruyter, 2010. 4

[10] G. Dal Maso, J.-M. Morel, and S. Solimini. A variational method in image segmentation: existence and approximation results. *Acta Math.*, 168(1):89–151, 1992. 2

[11] G. David. *Singular Sets of Minimizers for the Mumford–Shah Functional*, volume 233. Birkhäuser Basel, 2005. 2

[12] E. De Giorgi, M. Carriero, and A. Leaci. Existence theorem for a minimum problem with free discontinuity set. *Arch. Ration. Mech. Anal.*, 108(4):195–218, 1989. 2

[13] L. Evans. *Partial Differential Equations*. Graduate Studies in Mathematics. American Mathematical Society, 2010. 3

[14] H. Federer. *Geometric Measure Theory*. Springer, 1969. 2

[15] M. Gobbino. Finite difference approximation of the Mumford–Shah functional. *Comm. Pure Appl. Math.*, 51(2):197–228, 1998. 2

[16] B. Goldluecke, E. Strekalovskiy, and D. Cremers. The natural vectorial total variation which arises from geometric measure theory. *SIAM J. Imaging Sci.*, 5(2):537–563, 2012. 5

[17] L. Grady and C. Alvino. The piecewise smooth Mumford–Shah functional on an arbitrary graph. *IEEE Trans. on Image Process.*, 18(11):2547–2561, 2009. 2

[18] G. P. McCormick. Computability of global solutions to factorable nonconvex programs: Part I. convex underestimating problems. *Math. Program.*, 10(1):147–175, 1976. 3

[19] A. Mitsos, B. Chachuat, and P. I. Barton. McCormick-based relaxations of algorithms. *SIAM J. Optim.*, 20(2):573–601, 2009. 3, 6

[20] J.-M. Morel and S. Solimini. *Variational Methods in Image Segmentation*. Birkhäuser Boston, 1995. 2

[21] D. Mumford and J. Shah. Optimal approximations by piecewise smooth functions and associated variational problems. *Comm. Pure Appl. Math.*, 42(5):577–685, 1989. 2

[22] T. Pock, D. Cremers, H. Bischof, and A. Chambolle. An algorithm for minimizing the piecewise smooth Mumford–Shah functional. In *ICCV*, 2009. 2, 6

[23] L. I. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Phys. D*, 60(1):259–268, 1992. 5

[24] E. Strekalovskiy, A. Chambolle, and D. Cremers. A convex representation for the vectorial Mumford–Shah functional. In *CVPR*, 2012. 2, 7

[25] J. Weickert. *Anisotropic Diffusion in Image Processing*, volume 1. Stuttgart: Teubner, 1998. 5