Journal of Mathematical Imaging and Vision 20: 89–97, 2004
© 2004 Kluwer Academic Publishers. *Manufactured in The Netherlands.*

An Algorithm for Total Variation Minimization and Applications

ANTONIN CHAMBOLLE
CEREMADE–CNRS UMR 7534, Université de Paris-Dauphine, 75775 Paris Cedex 16, France
antonin@ceremade.dauphine.fr

Abstract. We propose an algorithm for minimizing the total variation of an image, and provide a proof of convergence. We show applications to image denoising, zooming, and the computation of the mean curvature motion of interfaces.

Keywords: total variation, image reconstruction, denoising, zooming, mean curvature motion

1. Introduction

The total variation has been introduced in Computer Vision first by Rudin, Osher and Fatemi [17], as a regularizing criterion for solving inverse problems. It has proved to be quite efficient for regularizing images without smoothing the boundaries of the objects.

In this paper we propose a algorithm for minimizing the total variation, that we claim to be quite fast. It is based on a dual formulation, and is related to the works of Chan, Golub, and Mulet [6] or of Carter [3]. However, our presentation is slightly different and we can provide a proof of convergence. We then show how our algorithm can be applied to two standard inverse problems in image processing, that are image denoising and zooming. We refer to [5, 8–10, 15, 19] for other algorithms to solve the same problem (as well as to the other total variation—related papers quoted in this note).

2. Notations and Preliminary Remarks

Let us fix our main notations. To simplify, our images will be 2-dimensional matrices of size $N \times N$ (adaptation to other cases or higher dimension is not difficult). We denote by X the Euclidean space $\mathbb{R}^{N\times N}$. To define the discrete total variation, we introduce a discrete (linear) gradient operator. If $u \in X$, The gradient ∇u is a vector in $Y = X \times X$ given by

$$(\nabla u)_{i,j} = \left((\nabla u)^1_{i,j}, (\nabla u)^2_{i,j}\right)$$

with

$$(\nabla u)^1_{i,j} = \begin{cases} u_{i+1,j} - u_{i,j} & \text{if } i < N, \\ 0 & \text{if } i = N, \end{cases}$$
$$(\nabla u)^2_{i,j} = \begin{cases} u_{i,j+1} - u_{i,j} & \text{if } j < N, \\ 0 & \text{if } j = N, \end{cases}$$

for $i, j = 1, \ldots, N$. Other choices of discretization are of course possible for the gradient, as long as it is a linear operator. Our choice seems to offer a good compromise between isotropy and stability.

Then, the total variation of u is defined by

$$J(u) = \sum_{1\le i,j\le N} |(\nabla u)_{i,j}|, \tag{1}$$

with $|y| := \sqrt{y_1^2 + y_2^2}$ for every $y = (y_1, y_2) \in \mathbb{R}^2$.

Let us observe here that this functional J is a discretization of the standard total variation, defined in the continuous setting for a function $u \in L^1(\Omega)$ (Ω open subset of $\mathbb{R}^2$) by

$$J(u) = \sup\left\{ \int_\Omega u(x)\,\mathrm{div}\,\xi(x)\,dx: \xi \in C^1_c(\Omega;\mathbb{R}^2), |\xi(x)| \le 1\ \forall x \in \Omega \right\} \tag{2}$$

(see for instance [12]). It is well known that J, defined by (2), is finite if and only if the distributional derivative Du of u is a finite Radon measure in Ω, in which case we have $J(u) = |Du|(\Omega)$. If u has a gradient $\nabla u \in L^1(\Omega; \mathbb{R}^2)$, then $J(u) = \int_\Omega |\nabla u(x)|\, dx$. We will work mostly, in this note, in the discrete setting. Let us however make the observation that if some step-size (or pixel size) $h \sim 1/N$ is introduced in the discrete definition of J (defining a new functional J_h equal to h times the expression in (1)), one can show that as $h \to 0$ (and the number of pixels N goes to infinity), J_h "Γ–converges" (see for instance [1]) to the continuous J (defined by (2) on $\Omega = (0,1) \times (0,1)$). This means that the minimizers of the problems we are going to consider approximate correctly, if the pixel size is very small, minimizers of similar problems defined in the continuous setting with the functional (2).

Being J one–homogeneous (that is, $J(\lambda u) = \lambda J(u)$ for every u and $\lambda > 0$), it is a standard fact in convex analysis (we refer to [11] for a quite complete introduction to convex analysis, and to [14] for a monograph on convex optimization problems) that the Legendre–Fenchel transform

$$J^*(v) = \sup_u \langle u, v\rangle_X - J(u)$$

(with $\langle u, v\rangle_X = \sum_{i,j} u_{i,j} v_{i,j}$)[1] is the "characteristic function" of a closed convex set K:

$$J^*(v) = \chi_K(v) = \begin{cases} 0 & \text{if } v \in K \\ +\infty & \text{otherwise.} \end{cases} \tag{3}$$

Since $J^{**} = J$, we recover

$$J(u) = \sup_{v \in K} \langle u, v\rangle_X. \tag{4}$$

In the continuous setting, one readily sees from definition (2) that K is the closure of the set

$$\left\{\operatorname{div} \xi : \xi \in C_c^1(\Omega; \mathbb{R}^2),\, |\xi(x)| \le 1\ \forall x \in \Omega\right\}.$$

Let us now find a similar characterization in the discrete setting. In Y, we use the Euclidean scalar product, defined in the standard way by

$$\langle p, q\rangle_Y = \sum_{1 \le i,j \le N} \left(p^1_{i,j} q^1_{i,j} + p^2_{i,j} q^2_{i,j}\right),$$

for every $p = (p^1, p^2), q = (q^1, q^2) \in Y$. Then, for every u,

$$J(u) = \sup_p \langle p, \nabla u\rangle_Y \tag{5}$$

where the sup is taken on all $p \in Y$ such that $|p_{i,j}| \le 1$ for every i, j. We introduce a discrete divergence $\operatorname{div} : Y \to X$ defined, by analogy with the continuous setting, by $\operatorname{div} = -\nabla^*$ (∇^* is the adjoint of ∇). That is, for every $p \in Y$ and $u \in X$, $\langle -\operatorname{div} p, u\rangle_X = \langle p, \nabla u\rangle_Y$. One checks easily that div is given by

$$(\operatorname{div} p)_{ij} = \begin{cases} p^1_{i,j} - p^1_{i-1,j} & \text{if } 1 < i < N, \\ p^1_{i,j} & \text{if } i = 1, \\ -p^1_{i-1,j} & \text{if } i = N, \end{cases} + \begin{cases} p^2_{i,j} - p^2_{i,j-1} & \text{if } 1 < j < N, \\ p^2_{i,j} & \text{if } j = 1, \\ -p^2_{i,j-1} & \text{if } j = N, \end{cases}$$

for every $p = (p^1, p^2) \in Y$. From (5) and the definition of the operator div, one immediately deduce (4), with K given by

$$\{\operatorname{div} p : p \in Y,\, |p_{i,j}| \le 1\ \forall i, j = 1, \dots, N\}.$$

3. The Algorithm

We propose an algorithm for solving

$$\min_{u \in X} \frac{\|u - g\|^2}{2\lambda} + J(u), \tag{6}$$

given $g \in X$ and $\lambda > 0$. $\|\cdot\|$ is the Euclidean norm in X, given by $\|u\|^2 = \langle u, u\rangle_X$.

The Euler equation for (6) is

$$u - g + \lambda \partial J(u) \ni 0.$$

Here, ∂J is the "sub-differential" of J, defined by $w \in \partial J(u) \Leftrightarrow J(v) \ge J(u) + \langle w, v - u\rangle_X$ for every v (see [11, 14]). The Euler equation may be rewritten $(g - u)/\lambda \in \partial J(u)$, which is equivalent to $u \in \partial J^*((g - u)/\lambda)$ (*cf* [14, Vol. I, Prop. 6.1.2]). Writing this as

$$\frac{g}{\lambda} \in \frac{g - u}{\lambda} + \frac{1}{\lambda} \partial J^*\left(\frac{g - u}{\lambda}\right),$$

we get that $w = (g - u)/\lambda$ is the minimizer of

$$\frac{\|w - (g/\lambda)\|^2}{2} + \frac{1}{\lambda} J^*(w).$$

Since J^* is given by (3), we deduce $w = \pi_K(g/\lambda)$. Hence the solution u of problem (6) is simply given by

$$u = g - \pi_{\lambda K}(g). \tag{7}$$

A possible algorithm for computing u is therefore to try to compute the nonlinear projection $\pi_{\lambda K}$. Notice that this analysis is valid also in the continuous setting (with $X = L^2(\Omega)$).

In dimension 1, the nonlinear projection $\pi_{\lambda K}$ is very easy to solve numerically. The applications are however of limited interest. We now describe our method for computing this projection in dimension 2 (and, theoretically, in any dimension, with little adaptation).

Computing the nonlinear projection $\pi_{\lambda K}(g)$ amounts to solving the following problem:

$$\min\{\|\lambda \operatorname{div} p - g\|^2 : p \in Y, \\ |p_{i,j}|^2 - 1 \le 0 \;\forall\, i, j = 1, \dots, N\}. \tag{8}$$

The Karush-Kuhn-Tucker conditions (*cf* [14, Vol. I, Theorem 2.1.4] or [7, Theorem 9.2-4]) yield the existence of a Lagrange multiplier $\alpha_{i,j} \ge 0$, associated to each constraint in problem (8), such that we have for each i, j

$$-(\nabla(\lambda \operatorname{div} p - g))_{i,j} + \alpha_{i,j} p_{i,j} = 0$$

with either $\alpha_{i,j} > 0$ and $|p_{i,j}| = 1$, or $|p_{i,j}| < 1$ and $\alpha_{i,j} = 0$. In the latter case, also $(\nabla(\lambda \operatorname{div} p - g))_{i,j} = 0$. We see that in any case

$$\alpha_{i,j} = |(\nabla(\lambda \operatorname{div} p - g))_{i,j}|.$$

We thus propose the following semi-implicit gradient descent (or fixed point) algorithm.

We choose $\tau > 0$, let $p^0 = 0$ and for any $n \ge 0$,

$$p^{n+1}_{i,j} = p^n_{i,j} + \tau\big((\nabla(\operatorname{div} p^n - g/\lambda))_{i,j} \\ - |(\nabla(\operatorname{div} p^n - g/\lambda))_{i,j}|\, p^{n+1}_{i,j}\big),$$

so that

$$p^{n+1}_{i,j} = \frac{p^n_{i,j} + \tau(\nabla(\operatorname{div} p^n - g/\lambda))_{i,j}}{1 + \tau|(\nabla(\operatorname{div} p^n - g/\lambda))_{i,j}|}. \tag{9}$$

We now can show the following result.

Theorem 3.1. *Let* $\tau \le 1/8$. *Then*, $\lambda \operatorname{div} p^n$ *converges to* $\pi_{\lambda K}(g)$ *as* $n \to \infty$.

Proof: By induction we easily see that for every $n \ge 0$, $|p^n_{i,j}| \le 1$ for all i, j. Let us fix $n \ge 0$ and let $\eta = (p^{n+1} - p^n)/\tau$. We have

$$\begin{aligned}\|\operatorname{div} p^{n+1} - g/\lambda\|^2 &= \|\operatorname{div} p^n - g/\lambda\|^2 \\ &\quad + 2\tau\langle \operatorname{div}\eta, \operatorname{div} p^n - g/\lambda\rangle + \tau^2\|\operatorname{div}\eta\|^2 \\ &\le \|\operatorname{div} p^n - g/\lambda\|^2 \\ &\quad - \tau\big(2\langle \eta, \nabla(\operatorname{div} p^n - g/\lambda)\rangle - \kappa^2\tau\|\eta\|_Y^2\big).\end{aligned}$$

We denoted $\|\eta\|_Y^2 = \langle \eta, \eta\rangle_Y$, and κ is the norm of the operator $\operatorname{div}: Y \to X$, that we will estimate later on. Now,

$$\begin{aligned}&2\langle \eta, \nabla(\operatorname{div} p^n - g/\lambda)\rangle - \kappa^2\tau\|\eta\|_Y^2 \\ &\quad = \sum_{i,j=1}^{N} 2\eta_{i,j}\cdot(\nabla(\operatorname{div} p^n - g/\lambda))_{i,j} - \kappa^2\tau|\eta_{i,j}|^2,\end{aligned}$$

and since $\eta_{i,j}$ is of the form $\nabla(\operatorname{div} p^n - g/\lambda))_{i,j} - \rho_{i,j}$ (with $\rho_{i,j} = |\nabla(\operatorname{div} p^n - g/\lambda))_{i,j}|p^{n+1}_{i,j}$), we have for every i, j

$$\begin{aligned}&2\eta_{i,j}\cdot(\nabla(\operatorname{div} p^n - g/\lambda))_{i,j} - \kappa^2\tau|\eta_{i,j}|^2 \\ &\quad = (1 - \kappa^2\tau)|\eta_{i,j}|^2 \\ &\qquad + (|(\nabla(\operatorname{div} p^n - g/\lambda))_{i,j}|^2 - |\rho_{i,j}|^2).\end{aligned}$$

Since $|p^{n+1}_{i,j}| \le 1$, $|\rho_{i,j}| \le |(\nabla(\operatorname{div} p^n - g/\lambda))_{i,j}|$. Hence, if $\tau \le 1/\kappa^2$, we see that $\|\operatorname{div} p^n - g/\lambda\|^2$ is decreasing with n, unless $\eta = 0$, that is, $p^{n+1} = p^n$. (This is clear if $\tau < 1/\kappa^2$, and a careful analysis shows that it is also true when $\kappa^2\tau = 1$. Indeed, if $\|\operatorname{div} p^{n+1} - g/\lambda\| = \|\operatorname{div} p^n - g/\lambda\|$ we deduce $|\rho_{i,j}| = |(\nabla(\operatorname{div} p^n - g/\lambda))_{i,j}|$ for each i, j so that either $|\nabla(\operatorname{div} p^n - g/\lambda))_{i,j}| = 0$ or $|p^{n+1}_{i,j}| = 1$. In both cases, (9) yields $p^{n+1}_{i,j} = p^n_{i,j}$.)

Let $m = \lim_{n\to\infty}\|\operatorname{div} p^n - g/\lambda\|$ and $\bar p$ be the limit of a converging subsequence (p^{n_k}) of (p^n). Letting $\bar p'$ be the limit of p^{n_k+1}, we have

$$\bar p'_{i,j} = \frac{\bar p_{i,j} + \tau(\nabla(\operatorname{div}\bar p - g/\lambda))_{i,j}}{1 + \tau|(\nabla(\operatorname{div}\bar p - g/\lambda))_{i,j}|},$$

and repeating the previous calculations we see that since clearly $m = \|\operatorname{div}\bar p - g/\lambda\| = \|\operatorname{div}\bar p' - g/\lambda\|$, it must be that $\bar\eta_{i,j} = (\bar p'_{i,j} - \bar p_{i,j})/\tau = 0$ for every i, j, that is, $\bar p = \bar p'$. Hence

$$-(\nabla(\lambda \operatorname{div}\bar p - g))_{i,j} + |(\nabla(\lambda \operatorname{div}\bar p - g))_{i,j}|\bar p_{i,j} = 0,$$

which is the Euler equation for a solution of (8). One can deduce that $\bar{p}$ solves (8) (see for instance [7, Theorem 9.2-4]) and that $\lambda \operatorname{div} \bar{p}$ is the projection $\pi_K(g)$. Since this projection is unique, we deduce that all the sequence $\lambda \operatorname{div} p^n$ converges to $\pi_K(g)$. The theorem is proved if we can show that $\kappa^2 \leq 8$.

By definition, $\kappa = \sup_{\|p\|_Y \leq 1} \|\operatorname{div} p\|$. Now, (adopting the convention that $p_{0,j} = p_{N,j} = p_{i,0} = p_{i,N} = 0$ for every i, j)

$$\begin{aligned} \|\operatorname{div} p\|^2 &= \sum_{1\leq i,j\leq N} \left(p^1_{i,j} - p^1_{i-1,j} + p^2_{i,j} - p^2_{i,j-1}\right)^2 \\ &\leq 4 \sum_{1\leq i,j\leq N} \left(p^1_{i,j}\right)^2 + \left(p^1_{i-1,j}\right)^2 + \left(p^2_{i,j}\right)^2 \\ &\quad + \left(p^2_{i,j-1}\right)^2 \leq 8\|p\|_Y^2. \end{aligned}$$

Hence $\kappa^2 \leq 8$. □

Remark. Choosing $p^1_{i,j} = p^2_{i,j} = (-1)^{i+j}$ shows that $\kappa^2 \geq 8 - O(1/N)$.

Remark. In practice, it appears that the optimal constant for the stability and convergence of the algorithm is not 1/8 but 1/4. We do not know the reason for this. If $\tau < 1/4$, then it is easy to check that both applications $p^n \mapsto \tilde{p}^n$ and $\tilde{p}^n \mapsto p^{n+1}$ defined respectively by

$$\tilde{p}^n_{i,j} = p^n_{i,j} + \tau(\nabla(\operatorname{div} p^n - g/\lambda))_{i,j} \quad \text{and}$$
$$p^{n+1}_{i,j} = \frac{\tilde{p}^n_{i,j}}{1 + \tau|(\nabla(\operatorname{div} p^n - g/\lambda))_{i,j}|}$$

are contractions, but each in a different norm (the first one for the semi-norm $\|\operatorname{div} p\|$, the second one for the norm $\sup_{i,j} |p_{i,j}|$).

Remark. To our knowledge there exist two other important contributions addressing the same issue, that is the minimization of total variation through a dual approach. One is the paper of Chan, Golub and Mulet [6], the other is the thesis of Carter [3]. In both works, the proposed algorithms are quite different. They share the advantage that they are supposed to work also for "deconvolution" problems, that is, when instead of (6), the problem to solve is

$$\min_{u\in X} \frac{\|Au - g\|^2}{2\lambda} + J(u), \tag{10}$$

with A a linear operator (corresponding in general to a low-pass filtering, that is, a blurring of the image). It is not clear how to adapt our approach to this case, and it is the subject of future studies. We show in Section 5 how to treat the particular case where A is an orthogonal projection (zooming). On the other hand, the advantage of our approach is the existence of the convergence Theorem 3.1, that ensures its efficiency and stability. It also provides a framework for understanding the behavior of the algorithms proposed in [6] and [3], at least in the case $A = Id$.

4. Image Denoising

The idea of minimizing total variation for image denoising, suggested in [17], assumes that the observed image $g = (g_{i,j})_{1\leq i,j\leq N}$ is the addition of an *a priori* piecewise smooth (or with little oscillation) image $u = (u_{i,j})_{1\leq i,j\leq N}$ and a random Gaussian noise, of estimated variance σ^2. It is hence suggested to recover the original image u by trying to solve the problem

$$\min\{J(u) : \|u - g\|^2 = N^2\sigma^2\} \tag{11}$$

(N^2 being the total number of pixels). It can be shown (see for instance [5]) that there exists (both in the continuous and discrete settings, in fact) a Lagrange multiplier $\lambda > 0$ such that, provided $\|g - \langle g\rangle\|^2 \geq N^2\sigma^2$ (with $\langle g\rangle$ the average value of the pixels $g_{i,j}$), this problem has a unique solution that is given by the equivalent problem (6). We have just shown how to numerically solve problem (6), however, since σ is in general less difficult to estimate than λ, we propose another algorithm that tackles directly the resolution of (11). The task is to find $\lambda > 0$ such that $\|\pi_{\lambda K} g\|^2 = N^2\sigma^2$. For $s > 0$, let us set $f(s) = \|\pi_{sK} g\|$. The following lemma states the main properties of f.

Lemma 4.1. *The function* $f(s)$ *maps* $[0, +\infty)$ *onto* $[0, \|g - \langle g\rangle\|]$. *It is non-decreasing, while the function* $s \mapsto f(s)/s$ *is non-increasing. Moreover,* $f \in W^{1,\infty}([0, +\infty))$ *and satisfies, for a.e.* $s \geq 0$,

$$0 \leq f'(s) \leq \frac{f(s)}{s} \leq 2\sqrt{2}N.$$

Proof: Fix $s, s', v = \pi_{sK} g, v' = \pi_{s'K} g$. By definition of the projection, we have

$$\langle g - v, w - v\rangle \leq 0$$

for every $w \in sK$, and, as well,

$$\langle g - v', w - v' \rangle \le 0$$

for every $w \in s'K$. Letting $\theta = s'/s$, and choosing $w = v'/\theta$ in the first inequality and $w = \theta v$ in the second, we find

$$\langle g - v, v' - \theta v \rangle \le 0 \quad \text{and} \quad \langle g - v', \theta v - v' \rangle \le 0.$$

Hence

$$\langle \theta v - v', v - v' \rangle \le 0, \quad \text{that is,}$$
$$\theta f(s)^2 - (1+\theta)\langle v, v' \rangle + f(s')^2 \le 0.$$

Since $\langle v, v' \rangle \le f(s) f(s')$, we find

$$(f(s') - \theta f(s))(f(s') - f(s)) \le 0,$$

that is, $f(s')$ is between $f(s)$ and $\theta f(s)$.

We deduce that $s \mapsto f(s)$ is non-decreasing, while $s \mapsto f(s)/s$ is non-increasing. Notice that for any $s > 0$, $f(s)/s \le \sup_{v \in K} \|v\| \le c = \kappa N$, where $\kappa \le 2\sqrt{2}$ is the norm of the operator div : $Y \to X$, introduced in Section 3. The previous study shows that if $s' \ge s$, we have

$$0 \le f(s') - f(s) \le \theta f(s) - f(s)$$
$$= (s' - s)\frac{f(s)}{s} \le c(s' - s),$$

so that f is c-Lipschitz continuous, and satisfies

$$0 \le f'(s) \le \frac{f(s)}{s} \le c$$

for a.e. $s \ge 0$. Eventually, we can easily show that any $u \in X$ with $\langle u \rangle = 0$ can be written div p for some $p \in Y$, so that there exists $s^* \ge 0$ such that $g - \langle g \rangle \in s^* K$, hence $f(s) = \|g - \langle g \rangle\|$ for every $s \ge s^*$. This ends the proof of the lemma. □

We thus propose the following algorithm, in order to solve (11). We assume $N\sigma$ is between 0 and $\|g - \langle g \rangle\|$. We need to find a value $\bar{\lambda}$ for which $f(\bar{\lambda}) = N\sigma$. We first choose an arbitrary starting value $\lambda_0 > 0$, and compute $v_0 = \pi_{\lambda_0 K}(g)$ with the algorithm described in Section 3, as well as $f_0 = f(\lambda_0) = \|v_0\|$. Then, given λ_n, f_n, we let $\lambda_{n+1} = (N\sigma/f_n)\lambda_n$, and compute $v_{n+1} = \pi_{\lambda_{n+1} K}(g)$ and $f_{n+1} = \|v_{n+1}\|$. We easily deduce from Lemma 4.1 the following theorem.

Theorem 4.2. *As $n \to \infty$, $f_n \to N\sigma$ while $g - v_n$ converges to the unique solution of* (11).

Proof: Assume for instance that $f_0 \le N\sigma$. By induction, we easily show that $\lambda_n \le \lambda_{n+1}$ and that $f_n \le f_{n+1} \le N\sigma$ for any $n \ge 0$. Indeed, if $f_n \le N\sigma$, then $\lambda_{n+1} = (N\sigma/f_n)\lambda_n \ge \lambda_n$, and Lemma 4.1 yields

$$f(\lambda_n) \le f(\lambda_{n+1}) \le (\lambda_{n+1}/\lambda_n) f(\lambda_n),$$

that is, $f_n \le f_{n+1} \le N\sigma$. If $\lambda_n \ge s^*$ (the same s^* introduced in the end of the proof of Lemma 4.1), then $f_n = \|g - \langle g \rangle\| \ge N\sigma$, hence $f_n = N\sigma$ and $\lambda_{n+1} = \lambda_n$. Hence $(\lambda_n)_{n \ge 0}$ and $(f_n)_{n \ge 0}$ are non-decreasing and bounded. Let $\bar{f} = \lim_{n\to\infty} f_n$ and $\bar{\lambda} = \lim_{n\to\infty} \lambda_n$. It is clear that (being f continuous) $\bar{f} = f(\bar{\lambda}) = N\sigma$. Letting $\bar{v} = \pi_{\bar{\lambda} K}(g)$, we deduce that $g - \bar{v}$ is the unique solution of (11). Now, it is straightforward to show that v_n must converge to $\bar{v}$. This proves the theorem. If $f_0 \ge N\sigma$ the proof is identical. □

We show some examples of images processed with this algorithm. In practice, we have observed that we can replace λ with the new value $N\sigma/\|\text{div}\, p^n\|$ after each iteration (9) of the main algorithm of Section 3, and get a very quick convergence to the limit u solving (11).

In the examples of Figs. 2 and 3, the original image is the image of Fig. 1 to which a noise of standard

Figure 1. An image.

Figure 2. The image of Fig. 1 and its reconstruction ($\sigma = 12$).

Figure 3. Same as Fig. 2 with now $\sigma = 25$.

deviation respectively 12 and 25 has been added. The original is a 256×256 square image with values ranging from 0 to 255. The CPU time for computing the reconstructed images is in both case approximately 1.9 seconds, on a 900 MHz Pentium III processor with 2 Mb of cache. The criterion for stopping the iteration just consists in checking that the maximum variation between $p^n_{i,j}$ and $p^{n+1}_{i,j}$ is less than $1/100$. Notice that this algorithm can very easily be parallelized.

5. Zooming

In the case of zooming, the inverse problem that has to be solved is now (in its most simple formulation, as

Figure 4. Left: original 512 × 512 Lena and a 128 × 128 reduction. Middle, the small image expanded by a factor 4. Right, the small image expanded by 4 using the algorithm of Section 5.

proposed by Guichard and Malgouyres, see [13, 16] for a general presentation)

$$\min_{u \in X} \frac{\|Au - g\|^2}{2\lambda} + J(u), \qquad (12)$$

where $g \in X$ is a *coarse* image, that is, belonging to a "coarse" subspace $Z \subset X$, and A is the orthogonal projection onto Z. For instance, Z might be the set of vectors $g_{i,j}$ such that $g_{2k,2l} = g_{2k+1,2l} = g_{2k,2l+1} = g_{2k+1,2l+1}$ for every $k, l \leq N/2$, in which case we expect u to be a zooming of factor 2 of g. We have $Ag = g$, and it is clear that

$$\|Au - g\| = \|A(u - g)\| = \min_{w \in Z^\perp} \|u - g - w\|.$$

Hence (12) may be reformulated as

$$\min_{u \in X, w \in Z^\perp} \frac{\|u - (g + w)\|^2}{2\lambda} + J(u).$$

This provides an obvious algorithm for solving the problem, by alternate minimizations of the energy with respect to w and u. We let $w_0 = 0$ and set for every $n \geq 0$

$$u_n = (g + w_n) - \pi_{\lambda K}(g + w_n)$$

which is computed using the algorithm (9), and

$$w_{n+1} = \pi_{Z^\perp}(u_n - g)$$

which is a straightforward calculation. It is very easy to establish the convergence of this algorithm, as $n \to \infty$, to a minimizer (u, w) of the convex energy $(u, w) \mapsto \|u - (g + w)\|^2/(2\lambda) + J(u)$ (as long as the vectors in $Z^\perp$ have zero average, which is usually the case). We leave it to the reader.

We illustrate the output of this algorithm on Fig. 4. As expected (see [16]), the result is very good. However, we found out that our method is quite slow, and does not seem to be a great improvement with respect to standard methods. Still some work has to been done in order to understand better how the energy is decreased at each iteration, and to try to find faster strategies.

6. Mean Curvature Motion

We mention here quickly another possible application of our algorithm. We do not intend to give to many details in this section (which has *a priori* little applications to imaging and vision). This will be the subject of a forthcoming paper [4]. We present the isotropic case, although the method is very general and also works for anisotropic curvature motion.

Consider a set $E \subset \Omega \subset \mathbb{R}^2$, such that the convex envelope of E is strictly inside Ω. Let d_E be the signed distance to ∂E, such that $d_E \geq 0$ in E and $d_E \leq 0$ in $\Omega \setminus E$. This distance can be computed in a quite efficient way, using a fast-marching algorithm [18]. We choose $h > 0$ and solve then, using our algorithm, a discretization of the problem

$$\min_w \frac{1}{2h} \int_\Omega |w(x) - d_E(x)|^2 \, dx \; + \; J(w) \qquad (13)$$

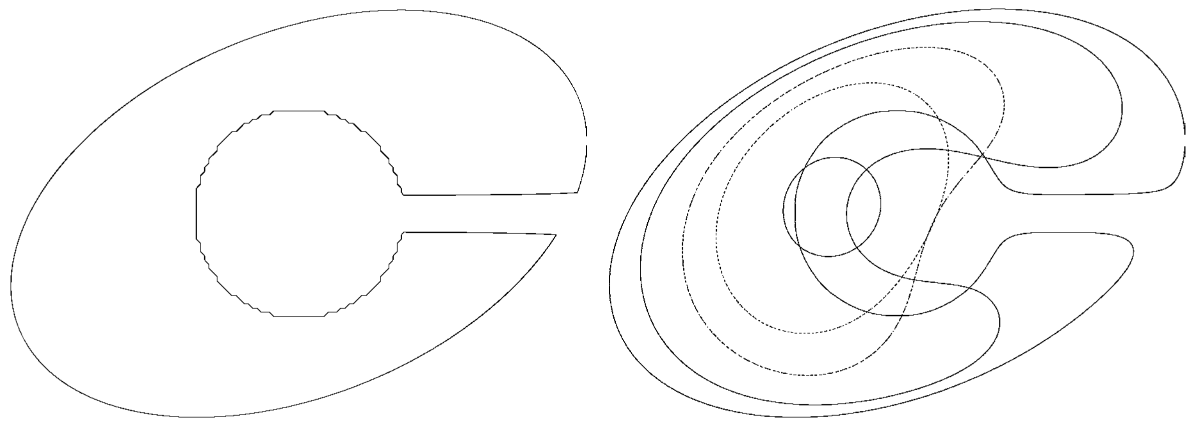

Figure 5. An original curve (left), and its evolution for times $t = 1, 30, 70, 100, 140$ (right).

with J defined by (2). We define the operator T_h by letting $T_h E = \{w > 0\}$, with w the solution of (13).

Given an initial set E_0, we let for $h > 0$ small and every $t > 0$

$$E^h(t) = (T_h)^n E_0$$

with $n = [t/h] =$ the integer part of t/h. Then, if ∂E_0 is smooth, we have the following result.

Theorem 6.1. *There exists $t_0 > 0$ such that, as $h \to 0$, the boundaries $\partial E^h(t)$ converge to $\Gamma(t)$ in the Hausdorff sense for $0 \le t \le t_0$, where $\Gamma(t)$ is the Mean Curvature evolution starting from ∂E_0.*

For the definition the Mean Curvature Motion, we refer to [2] and the huge literature that has followed. This result holds in fact in any dimension. The proof will be given in [4]. Figure 5 shows the evolution of a curve computed with this algorithm.

Note

1. We will sometimes drop the subscript "X", when not ambiguous.

References

1. A. Braides, *Gamma—Convergence for Beginners*, No. 22 in Oxford Lecture Series in Mathematics and Its Applications. Oxford University Press, 2002.
2. K.A. Brakke, *The Motion of a Surface by its Mean Curvature*, Vol. 20 of *Mathematical Notes*. Princeton University Press: Princeton, NJ, 1978.
3. J.L. Carter, "Dual methods for total variation—Based image restoration," Ph.D. thesis, U.C.L.A. (Advisor: T. F. Chan), 2001.
4. A. Chambolle, "An algorithm for mean curvature motion," to appear in *Interfaces Free Bound.*
5. A. Chambolle and P.-L. Lions, "Image recovery via total variation minimization and related problems," *Numer. Math.*, Vol. 76, No. 2, pp. 167–188, 1997.
6. T.F. Chan, G.H. Golub, and P. Mulet, "A nonlinear primal-dual method for total variation-based image restoration," *SIAM J. Sci. Comput.*, Vol. 20, No. 6, pp. 1964–1977, 1999 (electronic).
7. P.G. Ciarlet, *Introduction à l'analyse numérique matricielle et à l'optimisation*, Collection Mathématiques Appliquées pour la Maîtrise. [Collection of Applied Mathematics for the Master's Degree]. Masson: Paris, 1982.
8. P.L. Combettes and J. Luo, "An adaptive level set method for nondifferentiable constrained image recovery," *IEEE Trans. Image Process.*, Vol. 11, 2002.
9. F. Dibos and G. Koepfler, "Global total variation minimization," *SIAM J. Numer. Anal.*, Vol. 37, No. 2, pp. 646–664, 2000 (electronic).
10. D.C. Dobson and C.R. Vogel, "Convergence of an iterative method for total variation denoising," *SIAM J. Numer. Anal.*, Vol. 34, No. 5, pp. 1779–1791, 1997.
11. I. Ekeland and R. Temam, *Convex Analysis and Variational Problems*. Amsterdam: North Holland, 1976.
12. E. Giusti, *Minimal Surfaces and Functions of Bounded Variation*. Birkhäuser Verlag: Basel, 1984.
13. F. Guichard and F. Malgouyres, "Total variation based interpolation," in *Proceedings of the European Signal Processing Conference*, Vol. 3, pp. 1741–1744, 1998.
14. J.-B. Hiriart-Urruty and C. Lemaréchal, *Convex Analysis and Minimization Algorithms. I, II*, Vol. 305–306 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag: Berlin, 1993 (two volumes).
15. Y. Li and F. Santosa, "A computational algorithm for minimizing total variation in image restoration," *IEEE Trans. Image Processing*, Vol. 5, pp. 987–995, 1996.

16. F. Malgouyres and F. Guichard, "Edge direction preserving image zooming: A mathematical and numerical analysis," *SIAM J. Numer. Anal.*, Vol. 39, No. 1, pp. 1–37, 2001 (electronic).
17. L.I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D*, Vol. 60, pp. 259–268, 1992.
18. J.A. Sethian, "Fast marching methods," *SIAM Rev.*, Vol. 41, No. 2, pp. 199–235, 1999 (electronic).
19. C.R. Vogel and M.E. Oman, "Iterative methods for total variation denoising," *SIAM J. Sci. Comput.*, Vol. 17, No. 1, pp. 227–238, 1996. Special issue on iterative methods in numerical linear algebra (Breckenridge, CO, 1994).

Antonin Chambolle studied mathematics and physics at the Ecole Normale Supérieure in Paris and received the Ph.D. degree in applied mathematics from the Université de Paris-Dauphine in 1993. Since then he has been a CNRS researcher at the CERE-MADE, Université de Paris-Dauphine, and, for a short period, a researcher at the SISSA, Trieste, Italy. His research interest include calculus of variations, with applications to shape optimization, mechanics and image processing.