

# Restauration de données par minimisation de la variation totale, et variantes d'ordre supérieur.

Antonin Chambolle, Pierre-Louis Lions

CEREMADE, CNRS URA 749, Université de Paris-Dauphine, 75775 Paris CEDEX 16, France

## RÉSUMÉ

Nous étudions le problème de restauration de données par minimisation de la variation totale du signal (de l'image, ...), tel que l'ont proposé Rudin et Osher. Cette méthode, très performante pour la reconstruction de signaux constants par morceaux, donne des résultats médiocres sur d'autres types de signaux. Nous proposons une variante, qui s'inscrit dans un cadre très général et donne des résultats meilleurs sur des signaux réguliers, mais non constants, par morceaux.

## ABSTRACT

We study a signal or image restoration method proposed by Rudin and Osher, namely, the constrained total variation (TV) minimization. This very powerful method gives excellent results on nearly piecewise constant signals, but fails on more complicated data. We propose a very general way to combine convex potentials like the TV, and in this setting we introduce a variant that performs better on piecewise regular — but non constant — signals.

## 1 Le modèle de Rudin et Osher.

Rudin et Osher ont proposé le modèle suivant, pour résoudre des problèmes de débruitage et/ou déconvolution d'images (cf par exemple [8]). On suppose que l'image observée,  $u_0(x)$ , est la somme d'un signal oscillant (le "bruit")  $n(x)$ , de moyenne nulle et d'écart quadratique estimé  $\sigma^2$ , et d'une image "correcte"  $u(x)$  dégradée par l'action d'un opérateur  $A$ :

$$u_0 = Au + n \tag{1}$$

$A$  représente en général l'action d'un noyau régularisant, et on suppose que  $A1 = 1$ . On note par  $\Omega \subset \mathbf{R}^N$  l'ensemble sur lequel ces fonctions sont définies, pour simplifier les notations on suppose que  $|\Omega| = 1$ , et enfin que  $A$  est un opérateur linéaire continu de  $L^p(\Omega)$ ,  $p = N/(N - 1)$ . Rudin et Osher proposent de résoudre le problème suivant :

$$\begin{aligned} \text{Minimiser } J(u) &= \int_{\Omega} |\nabla u|, \\ \text{sous les contraintes :} & \\ \int_{\Omega} (Au - u_0) &= 0 \text{ et } \int_{\Omega} |Au - u_0|^2 = \sigma^2. \end{aligned} \tag{2}$$

Ici, la notation  $\int_{\Omega} |\nabla u|$  est impropre : les fonctions considérées  $u$  sont en réalité des fonctions à variation bornée ( $u \in BV(\Omega)$ ) ; leur dérivée n'est pas une fonction, mais une mesure  $Du$  de masse totale finie, et  $J(u) = |Du|(\Omega)$  est la variation totale de  $u$ .

Pour du bruit multiplicatif, le modèle est légèrement différent. On suppose que

$$u_0 = (Au) \times n, \tag{3}$$

$u_0 > 0$  presque partout, et cette fois le signal oscillant  $n(x)$  a pour moyenne 1 (on note encore  $\sigma^2$  son écart quadratique moyen). On cherche alors à résoudre le problème :

$$\begin{aligned} \text{Minimiser } J(u), \\ \text{sous les contraintes :} \\ \int_{\Omega} \frac{u_0}{Au} = 1 \text{ et } \int_{\Omega} \left(\frac{u_0}{Au}\right)^2 = 1 + \sigma^2. \end{aligned} \tag{4}$$

## 2 Résultats mathématiques.

### 2.1 Résultats d'existence.

Les résultats suivants ont été établis dans [3, 4]. Nous n'en donnerons ici aucune démonstration.

**Théorème 1** Soit  $X$  la fermeture dans  $L^2(\Omega)$  de  $L^2(\Omega) \cap A(L^p(\Omega) \cap BV(\Omega))$ , on suppose que  $u_0 \in X$ . Si  $0 < \sigma \leq \|u_0 - \int_{\Omega} u_0\|_2$ , alors (2) a au moins une solution  $u \in BV(\Omega)$  et  $Au \in L^2(\Omega)$  est unique. De plus, la résolution du problème (2) est équivalent à la minimisation du lagrangien

$$J(u) + \frac{\lambda}{2} \int_{\Omega} |Au - u_0|^2 \tag{5}$$

pour une certaine valeur de  $\lambda \geq 0$  (unique dès que  $\sigma < \|u_0 - \int_{\Omega} u_0\|_2$ ).

*Remarque 1.* L'hypothèse  $\sigma \leq \|u_0 - \int_{\Omega} u_0\|_2$  est très naturelle : si on suppose dans le modèle que le "bruit"  $n$  et le signal  $Au$  sont décorrélés, soit  $\int_{\Omega} Au \cdot n = 0$ , alors (1) entraîne  $\int_{\Omega} n^2 \leq \int_{\Omega} |u_0 - \int_{\Omega} u_0|^2$ . (Rappelons que  $|\Omega| = 1$ , et  $\int_{\Omega} u_0$  doit ici être comprise comme la valeur moyenne de  $u_0$ .)

*Remarque 2.* On peut vérifier que  $\sigma$ , vu comme une fonction de  $\lambda$ , décroît strictement (sauf au voisinage de  $\lambda = 0$ ), et même, en étudiant le problème dual, à peu près en  $1/\lambda$ .

Il n'est pas question ici de rentrer dans les détails de la démonstration de ce résultat, précisons simplement qu'une étape intermédiaire importante consiste à vérifier que la résolution de (2) est équivalente à la minimisation de la variation totale  $J(u)$ , sous la seule contrainte inégalité convexe :  $\int_{\Omega} |Au - u_0|^2 \leq \sigma^2$ , et ce dès que  $\sigma \leq \|u_0 - \int_{\Omega} u_0\|_2$ .

Pour (4), on a le résultat suivant, qui est moins complet. En particulier, on est obligé de se limiter au cas où  $A$  est l'opérateur identité.

**Théorème 2** Le problème (4), pour  $A = Id$ , a au moins une solution  $u \in BV(\Omega)$ , dès que

$$1 + \sigma^2 \leq \frac{\int_{\Omega} u_0^2}{\left(\int_{\Omega} u_0\right)^2}. \tag{6}$$



De plus, le lagrangien suivant :

$$J(u) + \lambda \int \left( \frac{u_0}{u} - 1 \right)^2 \quad (7)$$

admet un ou des minima pour tout  $\lambda \geq 0$ , et ces minima sont (à une constante multiplicative près) solution de (4) pour un  $\sigma$  vérifiant (6).

Ici encore, la condition (6) découle naturellement de (3) si on suppose que  $n$  et  $u$  sont décorrélés, au sens où :  $\int_{\Omega} u \cdot n = \int_{\Omega} u \cdot \int_{\Omega} n$  et  $\int_{\Omega} (u \cdot n)^2 = \int_{\Omega} u^2 \cdot \int_{\Omega} n^2$ .

La démonstration de ce résultat passe, comme dans le cas précédent, par l'équivalence entre le problème (4) et un problème de minimisation de la variation totale avec contraintes inégalités. Dans ce cas, les contraintes sont :  $\int_{\Omega} u_0/u \geq 1$  et  $\int_{\Omega} (u_0/u)^2 \leq 1 + \sigma^2$ . Les principaux problèmes viennent ensuite de ce que la première de ces contraintes est concave (ce qui explique la non-convexité du lagrangien).

## 2.2 Méthodes numériques de minimisation de la variation totale.

Les difficultés numériques liées à la résolution du modèle multiplicatif n'ont pas été toutes résolues. Par contre, on dispose pour la minimisation d'un lagrangien tel que (5) de plusieurs méthodes efficaces.

La première, très simple mais pas toujours facile à utiliser dans les variantes plus compliquées de (5) que nous introduiront plus tard, et de plus peu intéressante lorsque  $A$  n'est pas l'identité, consiste à résoudre le problème dual.

Si  $A = Id$ , on a simplement :

$$u = u_0 - \frac{1}{\lambda} \Pi_V(\lambda u_0) = u_0 - \Pi_{\frac{1}{\lambda}V}(u_0) \quad (8)$$

où  $\Pi_X$  désigne la projection orthogonale (dans  $L^2(\Omega)$ ) sur  $X$  et  $V$  est le support de la conjuguée de Legendre-Fenchel  $J^*$  de la variation totale  $J$ . Pratiquement, pour un signal unidimensionnel discret  $u = (u^0, \dots, u^{N-1})$ , on a

$$J(u) = \sum_{i=0}^{N-2} |u^{i+1} - u^i| = \sum_{i=0}^{N-2} |\langle d_i, u \rangle| \quad (9)$$

où  $d_i$  est le vecteur  $(0, \dots, 0, -1, 1, 0, \dots, 0)$  qui vaut  $-1$  au rang  $i$  et  $1$  au rang  $i+1$ . Donc  $J(u) = \sup_{v \in V} \langle v, u \rangle$  avec

$$V = \left\{ v = \sum_{i=0}^{N-2} t_i d_i : t_i \in [-1, 1] \text{ pour tout } i \right\} \quad (10)$$

et  $J^* = \chi_V$ . Projeter sur cette ensemble  $V$  est extrêmement simple (par gradient descendant, en tronquant à chaque fois les valeurs de  $t_i$  qui sortent de l'intervalle  $[-1, 1]$ ), en dimension un on peut ainsi calculer exactement la solution (8) (à la précision de la machine près !) en un temps très court. On peut bien sûr effectuer le même raisonnement en dimension deux.

La deuxième méthode que nous allons décrire est plus générale, et peut s'appliquer en réalité à de nombreuses variantes de (5). Il s'agit d'un algorithme de relaxation inspiré d'articles de D. Geman (cf [7]) et qui apparaît en fait dans de nombreux travaux [1, 6, 9]. Dans le cas particulier qui nous intéresse, on introduit l'énergie suivante :

$$E(u, v) = \int_{\Omega} v |\nabla u|^2 + \frac{1}{v} + |u - u_0|^2. \quad (11)$$

avec  $u \in H^1(\Omega)$  et  $v$  mesurable (par exemple,  $v \in L^\infty(\Omega)$ ). Pour simplifier on a pris  $\lambda = 1$  et  $A = Id$ . Il est très simple

de minimiser  $E$  par rapport à  $u$  ou par rapport à la "variable auxiliaire"  $v \geq 0$ .

**Théorème 3 ([4])** Soit  $v^0 \equiv 1$  et  $\varepsilon > 0$ . On définit deux suites  $u_n, v_n, n \geq 1$ , par :

$$\begin{aligned} u^n &= \arg \min_{u \in H^1(\Omega)} E(u, v^{n-1}) \\ v^n &= \arg \min_{\varepsilon \leq v \leq 1/\varepsilon} E(u^n, v) = \varepsilon \vee \frac{1}{|\nabla u^n|} \wedge \frac{1}{\varepsilon}. \end{aligned} \quad (12)$$

Alors la suite  $u_n$  converge (faiblement dans  $H^1(\Omega)$ , fortement dans  $L^2(\Omega)$ ) vers la fonction  $u$  qui minimise :

$$\int_{\Omega} j_\varepsilon(|\nabla u|) + \frac{1}{2} |u - u_0|^2 \quad (13)$$

où  $j_\varepsilon(x) = x^2/2\varepsilon$  si  $|x| \leq \varepsilon$ ,  $j_\varepsilon(x) = |x| - \varepsilon/2$  si  $\varepsilon \leq |x| \leq 1/\varepsilon$ ,  $j_\varepsilon(x) = \{ \varepsilon x^2 + (1/\varepsilon - \varepsilon) \}/2$  si  $|x| \geq 1/\varepsilon$ .

Notons que si  $\varepsilon$  tend vers zéro, le minimum de (13) tend vers celui de (5).

## 3 Résultats.

Les deux figures 1, 2 montrent des résultats de débruitage pour une image initiale bruitée constante par morceaux, et une qui ne l'est pas.

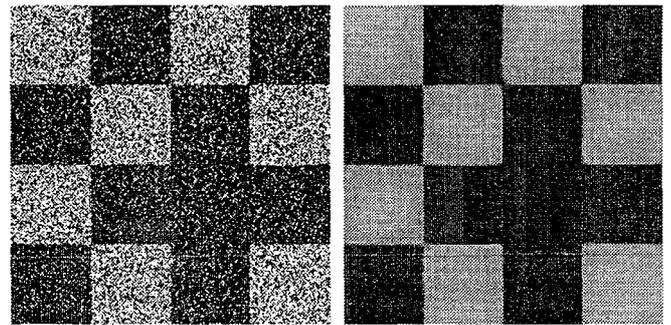


FIG. 1 - Un damier bruité et sa reconstruction ( $\sigma \simeq 60$ , gris clair = 192, gris foncé = 64).



FIG. 2 - "Lenna" avec du bruit, et la reconstruction ( $\sigma \simeq 30$ ).

Dans le premier cas, la reconstruction est quasiment parfaite. Dans le second, si les grandes variations de l'image originale sont assez bien rendues dans l'image reconstruite, on peut observer sur les plages plus régulières l'apparition de petites fractures. En réalité, il est facile de se convaincre que le modèle n'est pas bien adapté à la reconstruction d'images régulières par morceaux, mais non constantes. La figure 3 illustre ce problème en dimension un. Dans ce cas, l'analyse de l'équation d'Euler de (5),

$$-\lambda A^*(Au - u_0) \in \partial J(u) \quad (14)$$

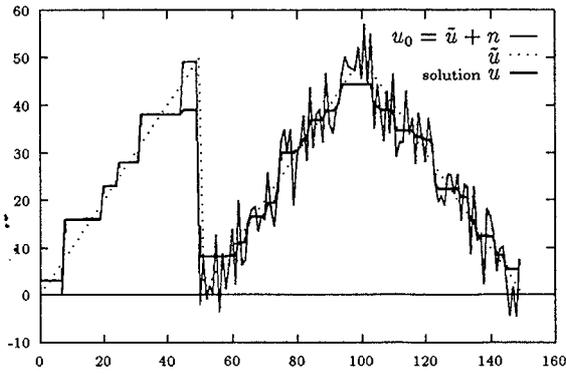


FIG. 3 - Minimisation de la variation totale.

(où  $\partial J(u)$  est le sous-différentiel de  $J$  au point  $u$ ), avec  $A = Id$ , montre que si la solution  $u$  est monotone (ce qui est le cas ici sur  $[0, 50]$ ,  $[50, 100]$  et  $[100, 150]$ ), alors elle alterne entre des plages où elle est constante et des intervalles où elle est égale à la fonction initiale.

## 4 Inf-convolution de deux potentiels convexes.

### 4.1 Cadre général.

Pour pallier les défauts de la méthode, nous proposons de remplacer la variation totale par d'autres potentiels convexes qui tiennent mieux compte des propriétés du signal ou de l'image. Il n'est pas question de remplacer  $|Du|(\Omega)$  par une énergie  $\int_{\Omega} |\nabla u|^p$ , avec  $p > 1$ , puisque les fonctions telles que  $\int_{\Omega} |\nabla u|^p < +\infty$  ne peuvent pas présenter de discontinuités le long de surfaces de dimension  $N - 1$  (et même pas du tout lorsque  $N = 1$ ). De nombreux auteurs proposent de minimiser des potentiels non convexes, avec des croissances sous-linéaires à l'infini. Malheureusement il n'y a généralement pas grand chose de connu sur les flots de gradient descendant de ces potentiels, et ces méthodes ne marchent que parce que la discrétisation limite naturellement les gradients. En contrepartie, les lieux d'apparition des discontinuités sont presque toujours instables.

L'approche que nous proposons consiste à essayer de fabriquer un potentiel convexe qui est minimal (nul) sur certaines classes de fonctions "intéressantes", en combinant deux (ou plusieurs) potentiels qui mesurent chacun une propriété des fonctions que l'on veut caractériser. Soient  $J_1$  et  $J_2$  ces deux potentiels.  $J_1(u)$  peut par exemple être la variation totale de  $u$ , et  $J_2(u)$  la variation totale de  $\nabla u$ . Dans ce cas précis,  $J_1$  s'annule sur les fonctions constantes, alors que  $J_2$  s'annule sur les fonctions affines (alors que  $J_1$  peut être grand); en revanche, une fonction  $u$  peut avoir une petite variation totale  $J_1(u)$  mais présenter des discontinuités, et alors  $J_2(u) = +\infty$ . Il apparaît donc clairement intéressant de considérer l'infimum de ces deux fonctions,  $u \mapsto I(u) = J_1(u) \wedge J_2(u)$ . Mais  $I$  n'est en général pas convexe. On peut s'en sortir de deux manières : en considérant l'enveloppe convexe de  $I$ , i.e.,  $\text{conv}I(u) =$

$$\begin{aligned} & \sup\{j(u) : j \text{ convexe}, j \leq J_1, j \leq J_2\} = \\ & \inf\{\theta J_1(u_1) + (1 - \theta)J_2(u_2) : \theta u_1 + (1 - \theta)u_2 = u, \theta \in [0, 1]\}, \end{aligned} \quad (15)$$

ou mieux, son enveloppe convexe s.c.i. :

$$\text{sc}^- \text{conv}I = (J_1 \wedge J_2)^{**} = (J_1^* \vee J_2^*)^* ; \quad (16)$$

ou bien, en considérant l'inf-convolution de  $J_1$  et  $J_2$ ,

$$J_1 \Delta J_2(u) = \inf\{J_1(u_1) + J_2(u_2) : u_1 + u_2 = u\}, \quad (17)$$

ou son enveloppe s.c.i.,  $\text{sc}^- J_1 \Delta J_2 = (J_1^* + J_2^*)^*$ .

En fait, on constate facilement que les deux notions coïncident si, par exemple,  $J_1$  est homogène de degré un et  $J_2^* \geq 0$  (par exemple dès que  $J_2(0) = 0$ ). Ceci sera toujours vérifié dans nos exemples.

Lorsqu'on ne sait pas calculer explicitement  $J = J_1 \Delta J_2$ , on peut quand même résoudre le problème de minimisation correspondant à (5) en calculant la solution  $u = u_1 + u_2$  de :

$$\min_{u_1, u_2} J_1(u_1) + J_2(u_2) + \frac{\lambda}{2} \int_{\Omega} |A(u_1 + u_2) - u_0|^2. \quad (18)$$

### 4.2 Résultats en dimension un.

Comme en dimension deux nous avons appliqué ce formalisme au cas où :  $J_1(u)$  est la variation totale de  $u$ ;  $J_2(u)$  est la variation totale de  $\nabla u = u'$  (soit formellement " $J_2(u) = \int_{\Omega} |u''|^p$ "). Remarquons qu'en dimension un le problème (18), pour les fonctionnelles  $J_1$  et  $J_2$  proposées, peut s'écrire, en posant  $u = u_1 + u_2$  et  $v = u_2'$  :

$$\min_{u, v} \int_{\Omega} |u' - v| + \alpha |v'| + \frac{\lambda}{2} |Au - u_0|^2. \quad (19)$$

(Le paramètre  $\alpha$  introduit ici un contrôle supplémentaire sur l'énergie.) L'interprétation de l'énergie (19) est très simple : la fonction  $v$  représente une sorte d'approximation de la dérivée du signal que l'on veut reconstruire. On essaie de minimiser la variation de cette fonction  $v$  (de sorte à la rendre plus ou moins constante par morceaux), et, simultanément, de minimiser la variation de  $u$  par rapport à  $\int^x v$  : ainsi, si la fonction  $u$  reconstruite peut encore présenter des discontinuités, elle n'a plus tendance à être constante entre ces sauts, mais seulement (plus ou moins) affine de dérivée  $v$ . Notons qu'on tombe ici sur une méthode qui ressemble beaucoup à des approches déjà connues dans la littérature, cf par exemple [2].

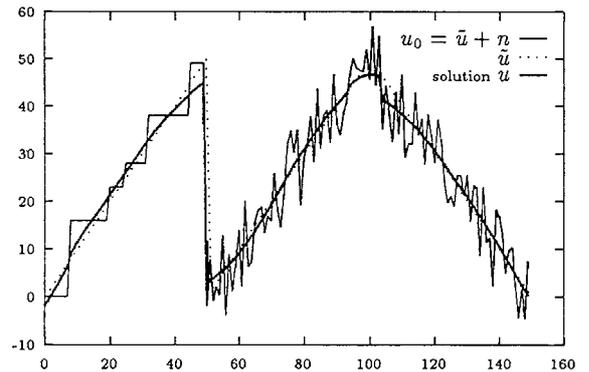


FIG. 4 - Minimisation de (19).

La figure 4 illustre bien ce comportement, et montre une reconstruction bien meilleure que la figure 3. La contrepartie de la méthode est l'introduction du paramètre d'échelle supplémentaire  $\alpha$  à estimer. La même méthode a été testée avec succès en introduisant un troisième potentiel  $J_3(u)$ , la variation totale de  $u''$ .

### 4.3 Résultats en dimension deux.

En dimension deux on se heurte à des problèmes algorithmiques, qui font qu'il n'est pas très intéressant de minimiser



l'équivalent bidimensionnel de (18). En revanche, on peut travailler directement avec l'équivalent de (19) :

$$\min_{u, \mathbf{p}} \int_{\Omega} |Du - \mathbf{p}| + \alpha |D\mathbf{p}| + \frac{\lambda}{2} |Au - u_0|^2 \quad (20)$$

où  $u \in BV(\Omega)$  et  $\mathbf{p} \in BV(\Omega, \mathbb{R}^2)$  est un champ de vecteur qui approxime la dérivée (ou la partie régulière de la dérivée, qui ne contient pas les sauts) de  $u$ . Dans ce cas, puisqu'on ne force pas le vecteur  $\mathbf{p}$  à être un gradient, on minimise une fonctionnelle plus petite que celle qui correspondrait rigoureusement à (18).

Les résultats sont apparemment meilleurs qu'avec la seule variation totale : comme le montre la figure 5, on reconstruit une image qui présente des discontinuités importantes, mais moins de fractures indésirables sur les zones que l'on voudrait régulières. Ce qu'on obtient est visuellement proche des résultats de [5], obtenus par une toute autre méthode.



FIG. 5 - "Lenna" reconstruite par la méthode (20).



FIG. 6 - La "Niçoise" bruitée ( $\sigma \simeq 26$ ).

## Références

- [1] G. Aubert, M. Barlaud, P. Charbonnier, L. Blanc-Féraud, and M. Barlaud. Deterministic edge-preserving regularization in computed imaging. Technical Report TR#94-01, I3S, CNRS URA 1376, Sophia-Antipolis, France, 1994.
- [2] P. N. Belhumeur. A binocular stereo algorithm for reconstructing sloping, creased, and broken surfaces, in the presence of half-occlusion. In *Proc. of the IEEE ICCV*, Berlin, Germany, 1993.
- [3] A. Chambolle. Total variation minimization for image reconstruction: the multiplicative noise case. Preprint, 1995.



FIG. 7 - La "Niçoise" reconstruite par minimisation de la variation totale.



FIG. 8 - La "Niçoise" reconstruite par (20).

- [4] A. Chambolle and P.-L. Lions. Image recovery via total variation minimization and related problems. Technical Report 9509, CEREMADE, CNRS URA 749, Université de Paris-Dauphine, Paris, 1995. (submit to IJCV).
- [5] A. Chambolle and B. J. Lucier. A nonlinear wavelet-based scale space for image smoothing. Preprint, CEREMADE, 1995.
- [6] L. D. Cohen. Auxiliary variables and two-step iterative algorithms in computer vision problems. to appear in *J. Math. Imag. Vision*. See also *Proc. ICCV 95.*, 1995.
- [7] D. Geman and G. Reynolds. Constrained image restoration and the recovery of discontinuities. *PAMI*, 14(3):367-383, 1992.
- [8] P.-L. Lions, S. J. Osher, and L. Rudin. Denoising and deblurring using constrained nonlinear partial differential equations. Technical report, Cognitech Inc., Santa Monica, CA, 1992. submit. to SINUM.
- [9] C. R. Vogel and M. E. Oman. Iterative methods for total variation denoising. Preprint, 1994.