Total Variation Minimization and Applications

Antonin Chambolle

CEREMADE - CNRS UMR 7534
Université Paris Dauphine.
and CMAP, Ecole Polytechnique.
Total variation minimization

- An algorithm for minimizing $TV(u) + \frac{1}{2\lambda} ||u - g||^2$

- Applications:
  
  $\rightarrow$ Inverse problems in image processing (denoising, restoration, zooming),

  $\rightarrow$ Evolution of sets by the mean curvature flow.
Main approach

The idea is to minimize numerically $TV + L^2$ norm via the dual problem.

$$J(u) = |Du|(\Omega) =$$

$$\sup \left\{ \int_{\Omega} u \text{div} \, \varphi \, : \, \varphi \in C_c^1(\Omega; \mathbb{R}^N), \|\varphi(x)\| \leq 1 \, \forall x \right\}$$

Problem (primal): given $g \in L^2$,

$$(1) \quad \min_u J(u) + \frac{1}{2\lambda} \|u - g\|_{L^2}^2$$
Dual problem

Several ways to derive the dual problem:

1) Problem is in the form (infconvolution)

\[ F(g) = \min_{u+v=g} J(u) + H(v) \]

\( F = J \triangle H \) is convex l.s.c., so that \( F(g) = F^{**}(g) \)

\( F^{*}(f) = \sup_g \langle f, g \rangle - F(g) \) is the Legendre-Fenchel transform).

Hence one has \( F(g) = \sup_f \langle f, g \rangle - F^{*}(f) \) with

\[ F^{*}(f) = \sup_g \langle f, g \rangle - \min_{u+v=g} (J(u) + H(v)) \]

\[ = \sup_{u,v} g \langle f, u + v \rangle - J(u) - H(v) \]

\[ = J^{*}(f) + H^{*}(f) \]
The dual problem is thus (changing the sign)

$$\min_{f} J^*(f) + H^*(f) - \langle f, g \rangle$$

Here, $H^*(f) = \lambda \|f\|^2 / 2$, hence the problem is

(2)  $$\min_{f} J^*(f) + \frac{\lambda}{2} \| f - (g/\lambda)^2 \|^2 - \frac{1}{2\lambda} \| g \|^2$$
2) A second way to derive the dual problem in this situation (Yosida regularization)

Euler equation: \[
\frac{u - g}{\lambda} + \partial J(u) \ni 0
\]

\[
[p \in \partial J(u) \iff \forall v, J(v) \geq J(u) + \langle p, v - u \rangle] \]

That is, \( \frac{g - u}{\lambda} \in \partial J(u) \).

We have Fenchel’s identity:

\[
p \in \partial J(u) \iff u \in \partial J^*(p) \iff \langle u, p \rangle = J(u) + J^*(p)
\]

We deduce

\[
u \in \partial J^* \left( \frac{g - u}{\lambda} \right)
\]

Letting \( w = g - u \) we get \( \frac{w - g}{\lambda} + \frac{1}{\lambda} \partial J^* \left( \frac{w}{\lambda} \right) \ni 0 \) which is the Euler equation for

\[
\min_w \frac{\|w - g\|^2}{2\lambda} + J^* \left( \frac{w}{\lambda} \right)
\]

It is the same as (2) if we let \( f = w/\lambda \).
What is $J^*$?

If $J$ is the total variation one has

$$J(u) = \sup_{w \in K} \langle u, w \rangle$$

with $K$ given by (the closure in $L^2$ of)

$$\left\{ \text{div} \, \varphi : \varphi \in C_c^1(\Omega; \mathbb{R}^N), \|\varphi(x)\| \leq 1 \ \forall x \right\}.$$ 

Hence $J(u) = \sup_w \langle u, w \rangle - \delta_K(w)$,

$$\delta_K(w) = \begin{cases} 0 & \text{if } w \in K, \\ +\infty & \text{otherwise.} \end{cases}$$

We get $\delta^*_K = J$, yielding $J^* = \delta_K$. Therefore (3) (or (2)) is an orthogonal projection and we find:

(4) $$u = g - \Pi_{\lambda K}(g)$$
To solve the nonlinear projection problem (4) we have to discretize.

A discrete Total Variation is

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

with

$$(\nabla u)_{i,j} = \begin{pmatrix} u_{i+1,j} - u_{i,j} \\ u_{i,j+1} - u_{i,j} \end{pmatrix}$$

(+ B.C.).
One has (as in the continuous setting):

\[ J(u) = \sup_{|\xi_{i,j}| \leq 1} \sum_{i,j} \xi_{i,j} \cdot (\nabla u)_{i,j} \]

\[ = - \sup_{|\xi_{i,j}| \leq 1} \sum_{i,j} (\text{div } \xi)_{i,j} u_{i,j} \]

with \((\text{div } \xi) = \xi^{1}_{i,j} - \xi^{1}_{i-1,j} + \xi^{2}_{i,j} - \xi^{2}_{i,j-1} + \text{B.C.},\)

i.e., \(\text{div} = -\nabla^*\).

We see that, again,

\[ J(u) = \sup_{v \in K} \langle u, v \rangle = \sup_{v} \langle u, v \rangle - \delta_K(v) \]

with \(K = \{\text{div } \xi : |\xi_{i,j}| \leq 1 \; \forall i, j\}\) and

\[ \delta_K(v) = J^*(v) = \begin{cases} 
0 & \text{if } v \in K \\
+\infty & \text{otherwise}
\end{cases} \]
We find that the Dual of (1), for $J$ the discrete Total Variation, is, again,

$$
\min_w \frac{\|w - g\|^2}{2\lambda} + \delta_K \left( \frac{w}{\lambda} \right),
$$

that is

$$
\min_{w \in \lambda K} \|w - g\|^2
$$

Hence $w$ is the projection on $\lambda K$ of $g$ and the solution of (1) is given by

(4) \hspace{1cm} u = g - \Pi_{\lambda K}(g)$
The problem is: $\min_{|\xi_{i,j}| \leq 1} \| \text{div} \xi - g/\lambda \|^2$.

Approach with Lagrange multipliers:

$$\min_{\xi} \| \text{div} \xi - g/\lambda \|^2 + \sum_{i,j} \alpha_{i,j} |\xi_{i,j}|^2.$$ 

The Euler equation is

$$-(\nabla (\text{div} \xi - g/\lambda))_{i,j} + \alpha_{i,j} \xi_{i,j} = 0 \forall i, j$$

with $\alpha_{i,j} \geq 0$ and $\alpha_{i,j} = 0$ whenever $|\xi_{i,j}| < 1$. Computing the norm $| \cdot |$, we find that

$$\alpha_{i,j} = |(\nabla (\text{div} \xi - g/\lambda))_{i,j}|.$$
Gradient Descent

A straightforward descent scheme is the following
\[ \xi_{i,j}^{n+1} = \xi_{i,j}^n + \tau (\nabla (\text{div} \xi^n - g/\lambda))_{i,j} - \tau \alpha_{i,j} \xi_{i,j}^{n+1}, \]
or
\[ \xi_{i,j}^{n+1} = \frac{\xi_{i,j}^n + \tau (\nabla (\text{div} \xi^n - g/\lambda))_{i,j}}{1 + \tau |(\nabla (\text{div} \xi^n - g/\lambda))_{i,j}|}. \]

**Theorem.** The iterations converge as soon as \( \tau \leq 1/\|\text{div}\|_2^2 \) (which is greater or equal to 1/8).

**Proof** (simple). One just shows that
\[ \|\text{div} \xi^{n+1} - g/\lambda\|^2 \leq \|\text{div} \xi^n - g/\lambda\|^2 \]
with < as long as \( \xi^n \) is not a solution of the problem.

**Remark:** Same convergence result for the (more natural) variant
\[ \xi_{i,j}^{n+1} = \Pi_{\{\xi \leq 1\}}(\xi_{i,j}^n + \tau (\nabla (\text{div} \xi^n - g/\lambda))_{i,j}), \]
however (for unknown reasons) it is much slower (even if one can prove the convergence up to \( \tau = 1/4 \), which also works in the previous algorithm).

→ See also [Carter] or [Chan-Golub-Mulet] for primal/dual approaches.
Applications: Image Denoising

- **Classical Model:**
  \[
  g = u + n,
  \]
  
  \[g = (g_{i,j})_{i,j=1}^N\] observed image, 
  
  \[u = (u_{i,j}) \text{ a priori piecewise smooth image},\]
  
  \[n = (n_{i,j}) \text{ Gaussian noise (average 0, variance } \sigma^2 \text{ hence } \frac{1}{N^2} \sum_{i,j} n_{i,j}^2 \sim \sigma^2).\]

  (Or: \(g = Au + n, A = \text{ linear transformation}.\))

- **Problem:** recover \(u\) from \(g\).

- **Tichonov’s Method:**

  (1) \[
  \min_u J(u) + \frac{1}{2\lambda} \|u - g\|^2
  \]

  or

  (1') \[
  \min_u J(u) \text{ subject to } \|u - g\|^2 = N^2 \sigma^2
  \]

- **Choice of \(J\):** \(H^1\) norm \((\sum |\nabla u|^2)\), TV (Rudin-Osher-Fatemi), Mumford-Shah...
(1) with varying $\lambda$
The problem proposed by Rudin-Osher-Fatemi is

\[
(1') \quad \min_u J(u) \quad \text{subject to} \quad \|u - g\|^2 = N^2 \sigma^2
\]

The constraint \(\|u - g\| = N \sigma\) is satisfied if \(\lambda\) in (1) is chosen such that \(\|\Pi_{\lambda K}(g)\| = \lambda \|\text{div} \xi\| = N \sigma\) (where \(\Pi_{\lambda K}(g) = \lambda \text{div} \xi \in \lambda K\)).

We propose the following algorithm for (1'): we fix an arbitrary value \(\lambda_0 > 0\) and compute \(v_0 = \Pi_{\lambda_0 K}(g)\). Then for every \(n \geq 0\), we let \(\lambda_{n+1} = (N \sigma / \|v_n\|) \lambda_n\), and \(v_{n+1} = \Pi_{\lambda_{n+1} K}(g)\). We have the following theorem:

**Theorem.** As \(n \to \infty\), \(g - v_n\) converges to the unique solution of (1').
Resolution of (1’) with $\sigma = 12$. 
Other Applications: Zooming

The setting of the Zooming problem is the following: We have \( u = (u_{i,j})_{i,j=1}^N \in X = \mathbb{R}^{N \times N} \), and \( g \) belongs to a “coarser” space \( Z \subset X \) (for instance, \( Z = \{ u \in X : u_{i,j} = u_{i+1,j} = u_{i,j+1} = u_{i+1,j+1} \text{ for every even } i, j \} \)), \( A \) is the orthogonal projection onto \( Z \), and the problem to solve (as proposed for instance by [Guichard-Malgouyres])

\[
\text{(5)} \quad \min_u J(u) + \frac{1}{2\lambda} \| Au - g \|^2
\]

(for some small value of \( \lambda \)). Since \( Ag = g \), \( Au - g = A(u - g) \) and \( \| Au - g \|^2 = \min_{w \in Z^\perp} \| u - g - w \|^2 \). Hence (5) is equivalent to

\[
\text{(6)} \quad \min_{w \in Z^\perp, u \in X} J(u) + \frac{1}{2\lambda} \| u - (g + w) \|^2
\]
Hence, to solve the zooming problem, one readily sees that the following algorithm will work: letting \( w_0 = 0 \), we compute \( u_{n+1}, w_{n+1} \) as follows

\[
\begin{align*}
    u_{n+1} &= g + w_n - \Pi_{\lambda K}(g + w_n), \\
    w_{n+1} &= \Pi_{Z}(u_{n+1} - g).
\end{align*}
\]

Unfortunately, this method is not very fast. (cf. [Guichard-Malgouyres] for the original introduction of the problem and a different implementation.)

Any linear operator \( A \) can be implemented, with speed of convergence depending of the condition number (and quite slow if \( A \) non invertible, like in this example.)

[Aubert-Bect-Blanc-Féraud-AC]
Meyer introduces the norm $\| \cdot \|_*$ which is dual of the Total Variation:

$$\|v\|_* = \sup_{J(u) \leq 1} \langle u, v \rangle = \min\{\lambda \geq 0, v \in \lambda K\}$$

(it is $+\infty$ if $\sum_{i,j} v_{i,j} \neq 0$).

He proposes to decompose an image $f$ into a sum $u + v$ of a $u$ with low Total Variation and a $v$ containing the oscillations, by solving

$$\min_{f = u + v} J(u) + \mu \|v\|_*$$

The idea:

- $J(u)$ is low when the signal $u$ is very regular (with edges);

- $\|v\|_*$ is low when the signal $v$ is oscillating.
Method

- Osher-Vese: minimize (for \( \lambda \) large)
  \[
  J(u) + \lambda \| f - u - v \|^2 + \mu \| v \|_*
  \]
  that is approximated by
  \[
  J(u) + \lambda \| f - u - \text{div} \xi \|^2 + \mu \| \xi \|_{\ell^p}
  \]
  for \( p >> 1 \).

- We propose the variant (our \( \lambda \) must be small)
  \[
  \min_{u,v} J(u) + \frac{1}{2\lambda} \| f - u - v \|^2 + J^* \left( \frac{v}{\mu} \right)
  \]
  that corresponds to a constraint \( \| v \|_* \leq \mu \).
Algorithm

An advantage of our approach: straightforward algorithm. Let $u_0, v_0 = 0$, then alternate:

- $v_n = \Pi_{\mu K}(f - u_{n-1})$
- $u_n = (f - v_n) - \Pi_{\lambda K}(f - v_n)$
Examples

Original synthetic image and same image with noise ($\sigma = 34$).
Reconstructed with Meyer’s problem and with ROF’s method ($\mu = 55$, $\sigma = 34$).
Difference image with Meyer’s problem and with ROF’s method.
“Texture removal” with Meyer’s problem and with ROF’s method ($\mu = 200$, $\sigma = 40.8$).
Difference image with Meyer’s problem and with ROF’s method.
Another example

A noisy zebra
Reconstructed with Meyer’s problem and with ROF’s method \((\mu = 20, \sigma = 19)\).
Texture removal” with Meyer’s problem and with ROF’s method ($\mu = 200, \sigma = 32.6$).
Differences

Difference image with Meyer’s problem and with ROF’s method.
\[
\min_u J(u) + \frac{1}{2\lambda} \|f - u\|_{H^{-1}}^2
\]

- Dual (cf first derivation of the dual problem)

\[
\min_w J^*(w) + \frac{\lambda}{2} \|\nabla w\|^2 - \langle f, w \rangle
\]

- Algorithm: variant of the TV algorithm (not extremely efficient, \(\tau\) must be quite small)

\[
\xi_{i,j}^{n+1} = \frac{\xi_{i,j}^n - \tau (\nabla(\Delta \text{div} \xi^n - f/\lambda))_{i,j}}{1 + \tau |(\nabla(\Delta \text{div} \xi^n - f/\lambda))_{i,j}|}
\]

\((\Delta = \text{div} \nabla)\). Then \(u = f - \lambda \Delta \text{div} \xi^\infty\).
Denoising with OVS

Reconstructed with OVS’s method and with ROF’s method ($\lambda = 100, \sigma = 33.7$).
Difference image with OVS’s approach and with ROF’s method.
Removing more...

Try to remove the “texture” with OVS’s approach ($\lambda = 1000$).
Denoising of the zebra

Zebra with OVS’s approach ($\lambda = 10$), and difference image.
Let $\Omega \subset \mathbb{R}^N$ and $E \subset\subset \Omega$. Fix $h > 0$ a small time step. Let us solve

\begin{equation}
\min_{w} J(w) + \frac{1}{2h} \int_{\Omega} |w - d_E|^2 dx
\end{equation}

where $d_E(x) = \text{dist}(x, E) - \text{dist}(x, \mathbb{R}^N \setminus E)$. We let $T_h(E) = \{w < 0\}$.

Given $E$, we can define $E_h(t) = T_h[t/h](E)$, a discrete evolution of the set $E$. 
Anisotropic variant

Let $\varphi$ be a convex one–homogeneous function in $\mathbb{R}^N$ (a distance, with $c|x| \leq \varphi(x) \leq c'|x|$ for all $x$).

Let $\varphi^\circ(\xi) = \sup_{\varphi(\eta) \leq 1} \langle \xi, \eta \rangle$ be the polar function. We introduce the anisotropic TV:

$$J_{\varphi}(w) = \int_{\Omega} \varphi^\circ(\nabla w) = \sup \left\{ \int_{\Omega} u \operatorname{div} \psi : \psi \in C^1_c(\Omega; \mathbb{R}^N), \varphi(\psi(x)) \leq 1 \ \forall x \right\}$$

$d^\varphi_E(x) = d^\varphi(x, E) - d^\varphi(x, \mathbb{R}^N \setminus E)$ is the anisotropic signed distance to $E$, with $d^\varphi(x, E) = \inf_{y \in E} \varphi(x - y)$.

We solve

$$(7') \quad \min_w J_{\varphi}(w) + \frac{1}{2h} \int_{\Omega} |w - d^\varphi_E|^2 \, dx$$

and let again $T_h(E) = \{w < 0\}$. 

38
The (formal) Euler Lagrange equation for (7) is

\[-h \text{div} \frac{\nabla w}{|\nabla w|} + w - d_E = 0.\]

At the boundary of $T_hE$, $w = 0$ and we get

\[d_E(x) = -h\kappa_{\{w=0\}}(x)\]

which is an implicit discretization of the Mean Curvature Motion.

→ Is it related to [Almgen-Taylor-Wang] or [Luckhaus-Sturzenecker]? Answer is Yes.

\[(\text{ATW}) \quad \min_{F \subset \mathbb{R}^N} \text{Per}(F) + \frac{1}{h} \int_{F \triangle E} |d_E(x)| \, dx\]

\[\kappa_F(x) + \frac{1}{h}d_E(x) = 0\]

→ same Euler equation.

**Theorem:**

$T_h(E) = \{w < 0\}$ is a solution of (ATW).
Convergence

We deduce (from (ATW)): smoothness of \( \partial T_h E \), Hölder-like continuity in time of \( E_h(t) \), convergence (up to subsequences) of \( E_h(t) \) to some movement \( E(t) \) (in \( L^1 \)). But we also have an important monotonicity property:

**Lemma:**

\[
E \subset E' \implies T_h(E) \subset T_h(E')
\]

[obvious, \( d_E > d_{E'} \implies w > w' \implies T_h E \subset T_h E' \)]

From which we deduce
Theorem: (Convergence to the generalized Mean Curvature Motion) Consider $E$ and $f$ such that $E = \{f < 0\}$, and $u(t)$ the (unique) viscosity solution of the MCM equation
\[
\frac{\partial u}{\partial t} = |\nabla u| \text{div} \frac{\nabla u}{|\nabla u|}
\]
with initial condition $u(t = 0) = f$. Assume at any time, $\Gamma(t) = \partial\{u(t) < 0\} = \partial\{u(t) \leq 0\}$ (no fattening, $\Gamma$ is the \textit{unique} generalized evolution starting from $\partial E$ and is independent of $f$). Then
\[
E_h(t) = T_h^{[t/h]} E \rightarrow E(t)
\]
as $h \to 0$.

(Also for a smooth, elliptic anisotropy $\varphi$.)

41
An advantage of (7) is not only that it yields the monotonicity of (ATW), other properties are also easier to study. Example:

**Theorem** [AC+Vicent Caselles]: Assume $E$ is convex: then $T_hE$ is also convex (any anisotropy). Hence $E_h(t)$ converges to a convex evolution $E(t)$. In the crystalline case, we deduce the existence of an evolution for convex sets (in a quite weak sense, but any dimension), preserving convexity.
Examples of evolutions

An isotropic evolution at different times
Anisotropic evolution (square Wulff shape)
Anisotropic evolution (hexagonal Wulff shape)