Mathematical methods for Image Processing

François Malgouyres

Institut de Mathématiques de Toulouse, France

invitation by
Jidesh P., NITK Surathkal

funding
Global Initiative on Academic Network

Oct. 23–27

Plan

Image Restoration

We look for a perfect image $v \in \mathbb{R}^{N^2}$ (or $\mathbb{R}^{N\times N}$) from noise corrupted linear measurements $u \in \mathbb{R}^P$:

$$u = Hv + b$$

where

$$H : \mathbb{R}^{N^2} \longrightarrow \mathbb{R}^P$$

and $b \in \mathbb{R}^P$ is some error (e.g. Gaussian noise of standard deviation σ).

- **Examples:** H is the identity (denoising), a convolution, a sampling operator (in a transformed space or not), ...
- **Applications :** Camera, Remote sensing imaging, Medical imaging (CT, IRM...), Biological imaging (many new microscopy acquisition devices), ...

INSTITUT de MATHEMATIQUES de TOULOUSE

Ill-posed problems by the example

Assume

$$H : \mathbb{R}^{N \times N} \longrightarrow \mathbb{R}^{N \times N}$$

is a convolution with $h \in^{N \times N}$.

$u = h * v + b$ becomes, in Fourier domain,

$$\hat{u}_{k,l} = \hat{h}_{k,l} \hat{v}_{k,l} + \hat{b}_{k,l} \qquad , \forall k, l = 0..N-1.$$

- If $\hat{h}_{k,l} = 0$,

$$\hat{v}_{k,l} \text{ is lost}$$

- If $\hat{h}_{k,l} \neq 0$,

$$\hat{v}_{k,l} = \frac{\hat{u}_{k,l}}{\hat{h}_{k,l}} - \frac{\hat{b}_{k,l}}{\hat{h}_{k,l}}$$

  The noise is amplified by a factor $\frac{1}{\hat{h}_{k,l}}$

Regularize the solution

Choose R that is

- small for the typical data you are restoring
- large for "H^{-1} of the error/noise"

and solve

$$\begin{cases} w^* = \text{Argmin}_w \;\; R(w) \\ \text{under the constraint } \|Hw - u\|_2 \leq \tau \end{cases}$$

or an unconstrained version:

$$w^* = \text{Argmin}_{w \in \mathbb{R}^{N^2}} \;\; R(w) + \lambda \|Hw - u\|_2^2.$$

- These problem can be obtained as MAP estimates (Bayesian framework)
- Can lead to bounds on the error $\|w^* - v\| \leq ...$ ("Compressed sensing" framework, developed Wednesday and Thursday)

Bayesian Modeling

Assume images are distributed according to the prior

$$\mathbb{P}(w) \propto e^{-\mu R(w)}.$$

Assume b is i.i.d. centered Gaussian noise of standard deviation $\sigma > 0$:

$$\mathbb{P}(b) \propto e^{-\frac{\|b\|_2^2}{2\sigma^2}}.$$

We have, for the data u obtained from w

$$\mathbb{P}(u|w) \propto e^{-\frac{\|u-Hw\|_2^2}{2\sigma^2}}.$$

Applying *Bayes* law, we get the posterior distribution

$$\begin{aligned}
\mathbb{P}(w|u) &= \frac{\mathbb{P}(u|w)\mathbb{P}(w)}{\mathbb{P}(u)}, \\
&\propto \mathbb{P}(u|w)\mathbb{P}(w), \\
&\propto e^{-\frac{\|u-Hw\|_2^2}{2\sigma^2}} e^{-\mu R(w)}.
\end{aligned}$$

Bayesian Modeling

The Maximum A Posteriori (MAP) estimate

$$\text{maximize } \mathbb{P}(w|u)$$

equivalently

$$\text{minimize } -\log\left(\mathbb{P}(w|u)\right)$$

This leads to

$$w^{MAP} = \text{Argmin}_{w \in W} \frac{\|u - Hw\|_2^2}{2\sigma^2} + \mu R(w).$$

Identical to

$$w^* = \text{Argmin}_{w \in \mathbb{R}^{N^2}} \; R(w) + \lambda \|Hw - u\|_2^2,$$

when $\lambda = 2\sigma^2\mu$.

Bayesian Modeling: Comments

- Assumption on R is often very wrong
- Differ from the "compressed sensing" approach because R is built independently of the noise and H.
- Powerful for designing models when many variables interact
- Leads to well founded strategies to tune the parameters (λ, in the example): See Expectation-Maximization algorithms.

The total variation

Definition

Let $w \in L^1\left([0,N]^2\right)$, if the following supremum is finite, we say w is of bounded variation, denote $w \in BV\left([0,N]^2\right)$ and define

$$TV(w) = \sup\left\{\int_{[0,N]^2} w \ \operatorname{div}\varphi \ dxdy \quad , \quad \varphi \in C^1\left(]0,N[^2, \mathbb{R}^2\right) \ \ et\ |\varphi| \leq 1\right\} < +\infty$$

Above $C^1\left(]0,N[^2,\mathbb{R}^2\right)$ contains C^1 functions from $]0,N[^2$ into $\mathbb{R}^2$, $|.|$ is the Euclidean norm in $\mathbb{R}^2$ and $|\varphi| \leq 1$ means

$$\begin{array}{lll} & |(\varphi_1(x,y),\varphi_2(x,y))| \leq 1 & ,\forall (x,y) \in]0,N[^2. \\ \text{i.e} & \varphi_1(x,y)^2 + \varphi_2(x,y)^2 \leq 1 & ,\forall (x,y) \in]0,N[^2. \end{array}$$

The total variation, Example 1 : when w is C^1

For any $\varphi = (\varphi_1, \varphi_2) \in C^1\left(]0,N[^2, \mathbb{R}^2\right)$, tel que $|\varphi| \leq 1$

$$\begin{aligned}
\int_{[0,N]^2} w \ \operatorname{div} \varphi \ dxdy &= \int_{[0,N]^2} w \left(\frac{\partial \varphi_1}{\partial x} + \frac{\partial \varphi_2}{\partial y} \right) \ dxdy, \\
&= -\int_{[0,N]^2} \frac{\partial w}{\partial x} \ \varphi_1 + \frac{\partial w}{\partial y} \ \varphi_2 \ dxdy, \\
&= -\int_{[0,N]^2} \nabla w \ . \ \varphi \ dxdy, \\
&\leq \int_{[0,N]^2} |\nabla w| \ dxdy. \qquad (1)
\end{aligned}$$

where $\nabla w \ . \ \varphi$ is the inner product in $\mathbb{R}^2$.

The total variation, Example 1 : when w is C^1

We have, for all $\varphi \in C^1\left(]0,N[^2, \mathbb{R}^2\right)$ such that $|\varphi| \leq 1$

$$\int_{[0,N]^2} w \ \operatorname{div}\varphi \ dxdy \leq \int_{[0,N]^2} |\nabla w| \ dxdy \qquad ,$$

and therefore

$$TV(w) \leq \int_{[0,N]^2} |\nabla w| \ dxdy.$$

In fact, we can prove the converse inequality and state:

Proposition

If $w \in BV\left([0,N]^2\right)$ and w is C^1, we have

$$TV(w) = \int_{[0,N]^2} |\nabla w| \ dxdy.$$

The total variation, Example 2: Characteristic function of a set

Theorem

If $E \subset [0,N]^2$ is an open set with a smooth boundary (For instance Lipschitz) and if $w = \mathbf{1}_{|E}$ then $w \in BV\left([0,N]^2\right)$ and

$$TV(w) = \mathcal{H}^1(\partial E),$$

where $\mathcal{H}^1$ is the Hausdorff[a] measure of dimension 1.

[a] The Haussdorf measure of dimension 1 of a set is the lenght of this set, if the set is "1D"; it is 0, if the dimension of the set is strictily smaller than 1; it is $+\infty$, if it is of dimension strictly larger than 1.

The total variation, Co-area formula

For a function $w \in L^1\left([0,N]^2\right)$ and $t \in \mathbb{R}$, we denote the t level set of w by:

$$\mathcal{L}_w(t) = \{(x,y) \in [0,N]^2, w(x,y) \geq t\}.$$

Theorem (Co-area formula)

If $w \in BV\left([0,N]^2\right)$, then

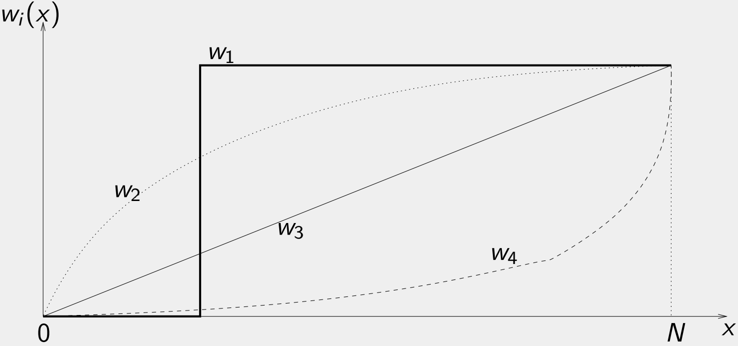
- *for almost every $t \in \mathbb{R}$,*
$$TV\left(\mathbf{1}_{\mathcal{L}_w(t)}\right) < \infty;$$
- *the function*
$$t \longmapsto TV\left(\mathbf{1}_{\mathcal{L}_w(t)}\right) \qquad \text{is measurable;}$$
- *and*
$$TV(w) = \int_{-\infty}^{\infty} TV\left(\mathbf{1}_{\mathcal{L}_w(t)}\right) dt.$$

The total variation, Example 3: if w is a non-decreasing signal

If $w \in BV([0,N])$ is non-decreasing and C^1,

$$\begin{aligned} TV(w) &= \int_0^N |w'| \, dx \\ &= \int_0^N w' \, dx \\ &= w(N) - w(0). \end{aligned}$$

The total variation, Example 3: if w is a non-decreasing signal



$$TV(w_1) = TV(w_2) = TV(w_3) = TV(w_4)$$

The total variation, on the grid

For any $w \in \mathbb{R}^{N^2}$, we set

$$TV(w) = \sum_{m,n=1}^{N} |\nabla w_{m,n}|,$$

where $|.|$ is the Euclidean norm in $\mathbb{R}^2$,

$$\nabla w_{m,n} = \begin{pmatrix} \partial_m w_{m,n} \\ \partial_n w_{m,n} \end{pmatrix} = \begin{pmatrix} w_{m+1,n} - w_{m,n} \\ w_{m,n+1} - w_{m,n}, \end{pmatrix}$$

(We assume w periodic.)

The total variation, on the grid

$$TV(w) = \sum_{m,n=1}^{N} |\nabla w_{m,n}|,$$

- **Good properties :** continuous, convexe, a semi-norm over $\mathbb{R}^{N^2}$, a norm over $\{w \in \mathbb{R}^{N^2}, \sum_{m,n=1}^{N} w_{m,n} = 0\}$
- **Bad properties :**
  - **Not coercive** over $\mathbb{R}^{N^2}$ $(TV(w+c) = TV(w))$: Not a problem as long as the data fidelity term does not make the image mean diverge.
  - **Not differentiable**, as soon as there is $(m,n) \in \{1,\ldots,N\}^2$ such that

    $$|\nabla w_{m,n}| = 0.$$

    Many ways to avoid this problem. In this lecture, we smooth TV

    $$TV_\varepsilon(w) = \sum_{m,n=1}^{N} \varphi_\varepsilon(|\nabla w_{m,n}|^2),$$

    where $\varepsilon > 0$ is small and $\varphi_\varepsilon(t) = \sqrt{t+\varepsilon}$, for $t > 0$.

INSTITUT de MATHEMATIQUES de TOULOUSE

The total variation, on the grid: Computing $\nabla TV_\varepsilon(w)$

For

$$\varphi_\varepsilon(t) = \sqrt{t+\varepsilon} \quad ,\forall t > 0,$$

we have

$$\varphi'_\varepsilon(t) = \frac{1}{2\sqrt{t+\varepsilon}} \quad ,\forall t > 0. \tag{1}$$

We denote the finite difference operators :

$$\begin{array}{rcl} \partial_m : \mathbb{R}^{N^2} & \longrightarrow & \mathbb{R}^{N^2}, \\ (w_{m,n})_{1\leq m,n\leq N} & \longmapsto & (w_{m+1,n} - w_{m,n})_{1\leq m,n\leq N}, \end{array}$$

and

$$\begin{array}{rcl} \partial_n : \mathbb{R}^{N^2} & \longrightarrow & \mathbb{R}^{N^2}, \\ (w_{m,n})_{1\leq m,n\leq N} & \longmapsto & (w_{m,n+1} - w_{m,n})_{1\leq m,n\leq N}. \end{array}$$

The total variation, on the grid: Computing $\nabla TV_\varepsilon(w)$

For any w, $w' \in \mathbb{R}^{N^2}$ we have

$$\begin{aligned}
\langle w, \partial_m w' \rangle &= \sum_{m,n=1}^{N} w_{m,n} \left(w'_{m+1,n} - w'_{m,n} \right), \\
&= \sum_{m,n=1}^{N} w_{m,n} w'_{m+1,n} - \sum_{m,n=1}^{N} w_{m,n} w'_{m,n}, \\
&= \sum_{m,n=1}^{N} w_{m-1,n} w'_{m,n} - \sum_{m,n=1}^{N} w_{m,n} w'_{m,n}, \\
&= \sum_{m,n=1}^{N} \left(w_{m-1,n} - w_{m,n} \right) w'_{m,n}, \\
&= \langle \partial_m^* w, w' \rangle.
\end{aligned}$$

Therefore

$$\partial_m^* w_{m,n} = w_{m-1,n} - w_{m,n} \quad , \forall (m,n) \in \{1, \ldots, N\}^2.$$

The total variation, on the grid: Computing $\nabla TV_\varepsilon(w)$

$$
\begin{aligned}
& TV_\varepsilon(w+w') - TV_\varepsilon(w) \\
= & \sum_{m,n=1}^{N} \varphi_\varepsilon\Big(|\nabla(w+w')_{m,n}|^2\Big) - \varphi_\varepsilon\Big(|\nabla w_{m,n}|^2\Big), \\
= & \sum_{m,n=1}^{N} \varphi_\varepsilon\Big(|\nabla w_{m,n}|^2 + 2(\partial_m w_{m,n}\partial_m w'_{m,n} + \partial_n w_{m,n}\partial_n w'_{m,n}) + o(|\nabla w'_{m,n}|)\Big) \\
& \qquad -\varphi_\varepsilon\Big(|\nabla w_{m,n}|^2\Big), \\
= & \sum_{m,n=1}^{N} 2\ \varphi'_\varepsilon\Big(|\nabla w_{m,n}|^2\Big)(\partial_m w_{m,n}\partial_m w'_{m,n} + \partial_n w_{m,n}\partial_n w'_{m,n}) + o(|\nabla w'_{m,n}|).
\end{aligned}
$$

We denote $X \in \mathbb{R}^{N^2}$ and $Y \in \mathbb{R}^{N^2}$ such that

$$
X_{m,n} = 2\ \varphi'_\varepsilon\Big(|\nabla w_{m,n}|^2\Big)\ \partial_m w_{m,n} \qquad \text{and} \qquad Y_{m,n} = 2\ \varphi'_\varepsilon\Big(|\nabla w_{m,n}|^2\Big)\ \partial_n w_{m,n},
$$

The total variation, on the grid: Computing $\nabla TV_\varepsilon(w)$

We get

$$\begin{aligned} TV_\varepsilon(w+w') - TV_\varepsilon(w) &= \sum_{m,n=1}^{N} X_{m,n}\partial_m w'_{m,n} + \sum_{m,n=1}^{N} Y_{m,n}\partial_n w'_{m,n} \\ &\quad + \sum_{m,n=1}^{N} o(|\nabla w'_{m,n}|) \\ &= \langle X, \partial_m w'\rangle + \langle Y, \partial_n w'\rangle + o(\|\nabla w'\|_1), \\ &= \langle \partial_m^* X + \partial_n^* Y, w'\rangle + o(\|w'\|_2). \end{aligned}$$

since

$$\sum_{m,n=1}^{N} |\nabla w'_{m,n}| \leq \sum_{m,n=1}^{N} |\nabla w'_{m,n}|_1 \leq 4 \sum_{m,n=1}^{N} |w'_{m,n}| \leq 4N\|w'\|_2.$$

The total variation, on the grid: Computing $\nabla TV_\varepsilon(w)$

Finally,

$$\nabla TV_\varepsilon(w) = \partial_m^* X + \partial_n^* Y,$$

with

$$X_{m,n} = 2\,\varphi'_\varepsilon\left(|\nabla w_{m,n}|^2\right)\,\partial_m w_{m,n} \qquad \text{and} \qquad Y_{m,n} = 2\,\varphi'_\varepsilon\left(|\nabla w_{m,n}|^2\right)\,\partial_n w_{m,n},$$

We also have (admitted), for all w and w'

$$\|\nabla TV_\varepsilon(w') - \nabla TV_\varepsilon(w)\| \leq \frac{8}{\sqrt{\varepsilon}}\|w' - w\|$$

The total variation : To go further

- Non-local Total Variation: see Li-Malgouyres-Zeng
- Numerical methods: Based on Graph cuts (see Chambolle and Darbon), primal-dual approach (see Chambolle, Chambolle-Pock)
- Staircase effect: see Nikolova
- Theoretical Justification: Compressed sensing with Co-sparse/Analysis prior (see Gribonval)