On the well-posedness of non-convex total variation

Tuomo Valkonen
University of Liverpool

STRIKE!!!

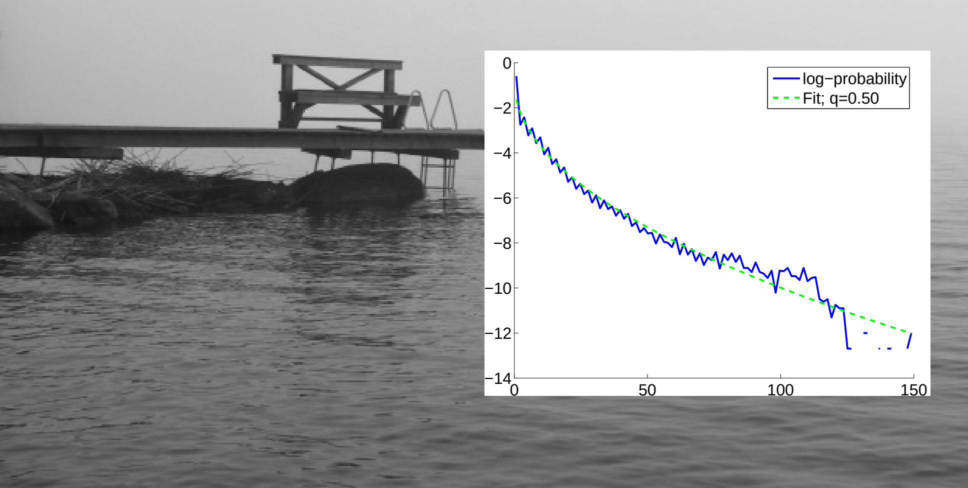
SIAM IS16, May 2016, Albuquerque

Joint work with M. Hintermüller and T. Wu

This work was done while the speaker was in Cambridge and Quito. In the latter he was supported by a Prometeo fellowship of the Senescyt. In Cambridge, his work was supported by the King Abdullah University of Science and Technology (KAUST) Award No. KUK-I1-007-43, and EPSRC first grant Nr. EP/J009539/1 "Sparse & Higher-order Image Restoration", and the Austrian Science Fund (FWF) grant. The other authors have been supported by Austrian FWF SFB F32 "Mathematical Optimization in Biomedical Sciences" and the START-Award Y305.

Gradient histograms

Model $t \mapsto \log P(\{x \mid \|\nabla u(x)\| = t\})$ by $-\alpha t^q$, $(\alpha, q > 0)$.



For recollection: Bayesian interpretation

The denoising problem with prior R and Gaussian noise

$$\min_u \frac{1}{2}\|u - f\|^2_{L^2(\Omega)} + \alpha R(u)$$

corresponds to the MAP estimate

$$\max_u \frac{P(f|u)P(u)}{P(f)}$$

where $P(f|u)$ is the Gaussian noise distribution, and the prior

$$P(u) = C\exp(-\alpha R(u)).$$

The TV^q model

This leads us to the image prior

$$\mathrm{TV}^q(u) := \int_\Omega \|\nabla u(x)\|^q \, dx, \quad (q \in (0,1)),$$

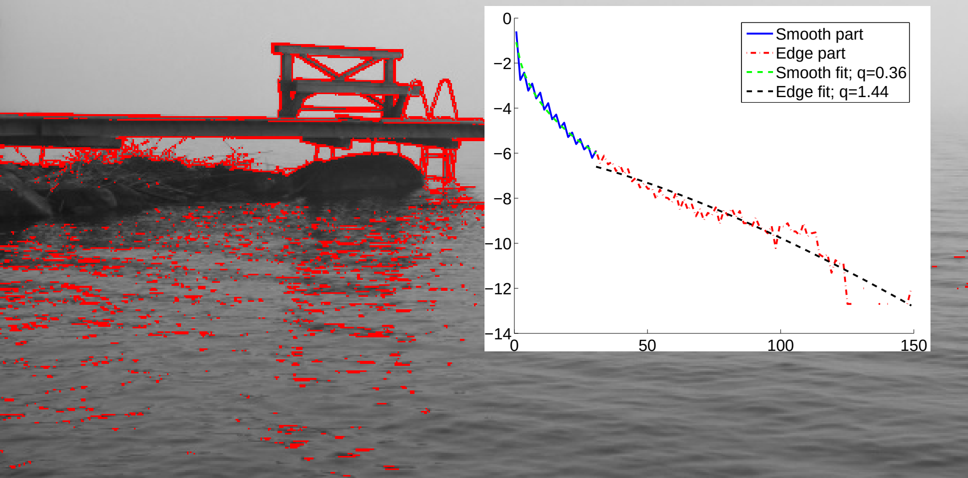
studied in (Huang and Mumford 1999; Hintermüller and Wu 2013; Hintermüller and Wu 2014; Ochs et al. 2013).

Related models for enforcing piecewise constant solutions: (Geman and Geman 1984; Nikolova 2002; Nikolova et al. 2008; Chen and Zhou 2010).

However, are such models theoretically justified?

And is t^q the full story in terms of statistics?

Edge detection, histogram, and t^q fit

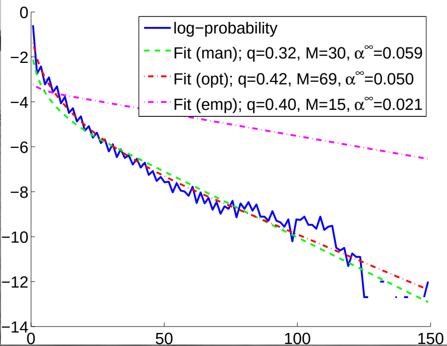


Linearised model fit

Varying M, α, q, we fit to the the function $-\alpha\varphi$ for

$$\varphi_{M,q}(t) := \begin{cases} t^q & 0 \le t \le M, \\ (1-q)M^q + qM^{q-1}t, & t > M. \end{cases}$$

We also define the *asymptotic alpha*, $\alpha^\infty := \alpha\varphi^\infty$.



Is this theoretically justified?

The TV^q model

For an energy functional φ, define

$$\widetilde{\mathrm{TV}}_{\mathrm{c}}^{\varphi}(u) = \int_{\Omega} \varphi(\|\nabla u(x)\|)\, dx, \quad (u \in C^1(\Omega)),$$

and extend this weak* lsc. to $u \in \mathrm{BV}(\Omega)$ by

$$\mathrm{TV}_{\mathrm{c}}^{\varphi}(u) := \liminf_{\substack{u^i \overset{*}{\rightharpoonup} u, \\ u^i \in C^1(\Omega)}} \widetilde{\mathrm{TV}}_{\mathrm{c}}^{\varphi}(u^i).$$

Theorem

If $\varphi(t) = t^q$ for $q \in (0,1)$, then $TV_{\mathrm{c}}^{\varphi}(u) \equiv 0$.

How about the linearised model?

Theorem

Even for the linearised model $\varphi = \varphi_{M,q}$ *with*

$$\varphi_{M,q}(t) := \begin{cases} t^q & 0 \le t \le M, \\ (1-q)M^q + qM^{q-1}t, & t > M, \end{cases}$$

we have $TV_c^{\varphi}(u) \equiv \varphi^{\infty} TV(u) = qM^{q-1} TV(u)$*.*

Difficulties

We need to replace weak* convergence – but with what?

- Weak* convergence is too weak; it demands convex integrands (cf. Bouchitté and Buttazzo 1990; Fonseca and Leoni 2007).
- Strict convergence is also not enough.
- Strong convergence in $\mathrm{BV}(\Omega)$ does not allow approximating piecewise constant functions by smooth functions, so too strong.

Area-strict convergence

Definition

Suppose $\Omega \subset \mathbb{R}^n$ with $n \geq 2$. Then $u^i \to u$ *area-strictly* in $\mathrm{BV}(\Omega)$ if

$$U^i \to U \quad \text{strictly in } \mathrm{BV}(\Omega; \mathbb{R}^{n+1})$$

with the notation $U(x) := (x/\|x\|, u(x))$.

In other words $u^i \to u$ strongly in $L^1(\Omega)$, $Du^i \overset{*}{\rightharpoonup} Du$ weakly* in $\mathcal{M}(\Omega; \mathbb{R}^n)$, and $\mathcal{A}(u^i) \to \mathcal{A}(u)$ for the *area functional*

$$\mathcal{A}(u) := \int_\Omega \sqrt{1 + \|\nabla u(x)\|^2}\, dx + |D^s u|(\Omega).$$

It can be shown that area-strict convergence is stronger than strict convergence, but weaker than norm convergence.

Area-strict continuity

Theorem (Rindler and Shaw 2013)

Let Ω *be a bounded domain with Lipschitz boundary. Let* $f \in C(\mathbb{R}^n)$ *satisfy*

$$|f(A)| \leq C(1 + |A|),$$

and suppose f^∞ *exists. Then the functional*

$$\mathcal{F}(u) := \int_\Omega f(\nabla u(x))\, dx + \int_\Omega f^\infty\left(\frac{dD^s u}{d|D^s u|}(x)\right) d|D^s u|(x)$$

is area-strictly continuous on $BV(\Omega)$.

Application of area-strict continuity

Corollary

Suppose $\varphi \in C(\mathbb{R}^{0,+})$, φ^∞ *exists, and* $\varphi(t) \leq C(1+t)$, $(t \in \mathbb{R}^{0,+})$. *Then the functional*

$$TV_{\mathrm{as}}^{\varphi}(u) := \int_\Omega \varphi(\|\nabla u(x)\|)\,dx + \varphi^\infty |D^s u|(\Omega), \quad (u \in BV(\Omega)),$$

is area-strictly continuous on $BV(\Omega)$.

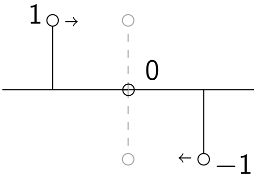
The problem now is: how do we obtain area-strict convergence of a minimising sequence to the denoising problem

$$\min_{u \in \mathrm{BV}(\Omega)} \frac{1}{2}\|f-u\|^2_{L^2(\Omega)} + \mathrm{TV}^{\varphi}_{\mathrm{as}}(u)?$$

Annihilation

Question: What strict convergence lacks that weak* can exhibit?

Answer: Annihilation effects.



How to avoid them?

Multiscale analysis

With $\eta_0 > 0$ and $\{\rho_\epsilon\}_{\epsilon>0}$ a family of mollifiers satisfying the semigroup property $\rho_{\epsilon+\delta} = \rho_\epsilon * \rho_\delta$, we define

$$\eta(\mu) := \eta_0 \sum_{\ell=1}^{\infty} \int_{\mathbb{R}^n} (|\mu| * \rho_{2^{-i}})(x) - |\mu * \rho_{2^{-i}}|(x)\, dx, \quad (\mu \in \mathcal{M}(\Omega; \mathbb{R}^n)).$$

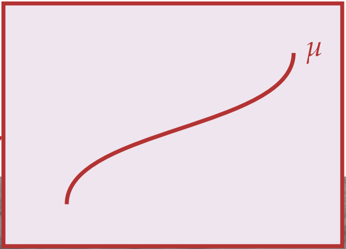
Theorem ($T.V.$ 2011; $T.V.$ 2012)

If $\sup_i \eta(\mu) < \infty$ *and* $\mu^i \overset{*}{\rightharpoonup} \mu$, *then* $|\mu^i|(\Omega) \to |\mu|(\Omega)$.

Multiscale analysis

With $\eta_0 > 0$ and $\{\rho_\epsilon\}_{\epsilon>0}$ a family of mollifiers satisfying the semigroup property $\rho_{\epsilon+\delta} = \rho_\epsilon * \rho_\delta$, we define

$$\eta(\mu) := \eta_0 \sum_{\ell=1}^{\infty} \int_{\mathbb{R}^n} (|\mu| * \rho_{2^{-i}})(x) - |\mu * \rho_{2^{-i}}|(x)\, dx, \quad (\mu \in \mathcal{M}(\Omega; \mathbb{R}^n)).$$



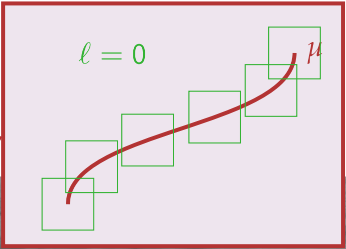
Theorem (*T.V.* 2011; *T.V.* 2012)

If $\sup_i \eta(\mu)<\infty$ *and* $\mu^i \overset{*}{\rightharpoonup} \mu$, *then* $|\mu^i|(\Omega$

Multiscale analysis

With $\eta_0 > 0$ and $\{\rho_\epsilon\}_{\epsilon>0}$ a family of mollifiers satisfying the semigroup property $\rho_{\epsilon+\delta} = \rho_\epsilon * \rho_\delta$, we define

$$\eta(\mu) := \eta_0 \sum_{\ell=1}^{\infty} \int_{\mathbb{R}^n} (|\mu| * \rho_{2-i})(x) - |\mu * \rho_{2-i}|(x)\, dx, \quad (\mu \in \mathcal{M}(\Omega; \mathbb{R}^n)).$$



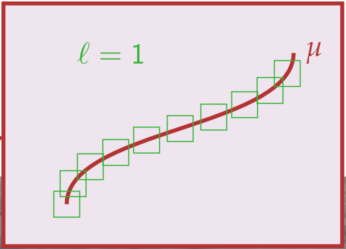
Theorem (*T.V.* 2011; *T.V.* 2012)

If $\sup_i \eta(\mu) < \infty$ *and* $\mu^i \overset{*}{\rightharpoonup} \mu$, *then* $|\mu^i|(\Omega$

Multiscale analysis

With $\eta_0 > 0$ and $\{\rho_\epsilon\}_{\epsilon>0}$ a family of mollifiers satisfying the semigroup property $\rho_{\epsilon+\delta} = \rho_\epsilon * \rho_\delta$, we define

$$\eta(\mu) := \eta_0 \sum_{\ell=1}^{\infty} \int_{\mathbb{R}^n} (|\mu| * \rho_{2^{-i}})(x) - |\mu * \rho_{2^{-i}}|(x)\, dx, \quad (\mu \in \mathcal{M}(\Omega; \mathbb{R}^n)).$$



Theorem (*T.V.* 2011; *T.V.* 2012)

If $\sup_i \eta(\mu) < \infty$ *and* $\mu^i \overset{*}{\rightharpoonup} \mu$*, then* $|\mu^i|(\Omega)$

Multiscale analysis

With $\eta_0 > 0$ and $\{\rho_\epsilon\}_{\epsilon>0}$ a family of mollifiers satisfying the semigroup property $\rho_{\epsilon+\delta} = \rho_\epsilon * \rho_\delta$, we define

$$\eta(\mu) := \eta_0 \sum_{\ell=1}^{\infty} \int_{\mathbb{R}^n} (|\mu| * \rho_{2^{-i}})(x) - |\mu * \rho_{2^{-i}}|(x)\, dx, \quad (\mu \in \mathcal{M}(\Omega; \mathbb{R}^n)).$$

Theorem (*T.V.* 2011; *T.V.* 2012)

If $\sup_i \eta(\mu) < \infty$ *and* $\mu^i \overset{*}{\rightharpoonup} \mu$, *then* $|\mu^i|(\Omega) \to |\mu|(\Omega)$.

The fixed TV^q model

Theorem

Let $\varphi = \varphi_{M,q}$ be the linearised t^q integrand. Suppose $\Omega \subset \mathbb{R}^n$ is bounded with Lipschitz boundary. Then the functional

$$G(u) := \frac{1}{2}\|f - u\|^2_{L^2(\Omega)} + \alpha TV^{\varphi}_{\mathrm{as}}(u) + \eta(DU), \quad (u \in BV(\Omega)),$$

admits a minimiser $u \in BV(\Omega)$.

Remark (convergence of minimising sequences)

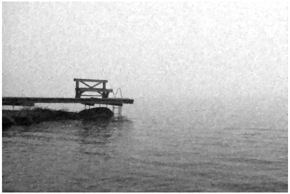
The linearisation of φ is needed for a bound in $\mathrm{BV}(\Omega)$ and weak* convergence. Then a bound on η gives area-strict convergence.

For numerical experiments

- We use a modification of the method of (Hintermüller and Wu 2013; Hintermüller and Wu 2014).
- We vary the cut-off M while keeping the asymptotic α fixed.
  - Defined by $\alpha^\infty := \alpha\varphi^\infty$.
  - Justification: for TV, $\alpha^\infty = \alpha$, so same edge regularisation.
- Empirically optimal q discovered by trial and error.

(a) Original

(b) Noisy image



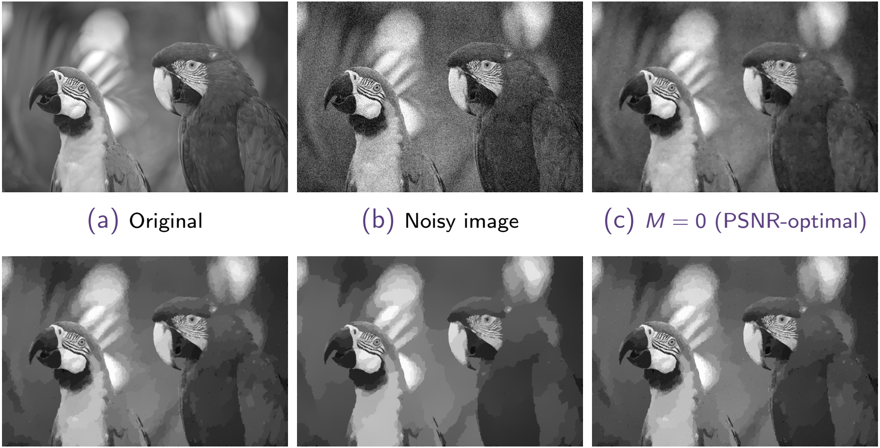
(c) $M = 0$

(d) $M = 10$ (PSNR-optimal)

(e) $M = 40$ (SSIM-optimal)

(f) $M = \infty$

Figure: Pier photo denoising results with noise level $\sigma = 30$ (Gaussian), for varying cut-off M, fixed $q = 0.4$ and fixed $\alpha^\infty = 0.0207$.



(a) Original

(b) Noisy image

(c) $M = 0$ (PSNR-optimal)

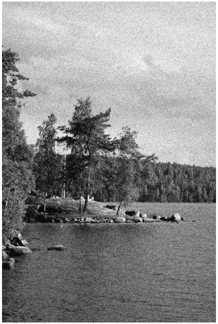
(d) $M = 15$ (SSIM-optimal)

(e) $M = 40$

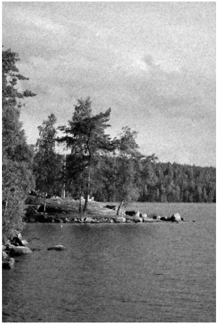
(f) $M = \infty$

Figure: Parrot photo denoising results with noise level $\sigma = 30$ (Gaussian), for varying cut-off M, fixed $q = 0.5$ and fixed $\alpha^{\infty} = 0.0253$.

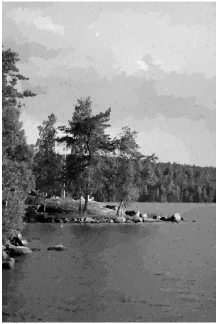
(a) Original



(b) Noisy image



(c) $M = 0$



(d) $M = 20$ (PSNR-optimal)

Figure: Summer photo denoising results with noise level $\sigma = 60$ (Gaussian), for varying cut-off M, fixed $q = 0.3$ and fixed $\alpha^{\infty} = 0.00430$.

Effect of the η term

(a) $\eta_0 = 7\text{e}^{-03}$, $\epsilon_1 = 1$

(b) $\eta_0 = 7\text{e}^{-03}$, $\epsilon_1 = 2$

(c) $\eta_0 = 7\text{e}^{-04}$, $\epsilon_1 = 1$

(d) $\eta_0 = 7\text{e}^{-04}$, $\epsilon_1 = 2$

(e) $\eta_0 = 7\text{e}^{-05}$, $\epsilon_1 = 1$

(f) $\eta_0 = 7\text{e}^{-05}$, $\epsilon_1 = 2$

Conclusion

We may conclude:

- The cut-off M for linearising t^q
  - Is required theoretically
  - Can be seen in image gradient statistics
  - Improves results in practise
- The multiscale regularisation η is a “theoretical artefact” that has yet to be justified in practise.

Thank you for your attention!