Bilevel learning of variational models

Carola-Bibiane Schönlieb
Includes joint work with M. Benning, L. Calatroni, C. Chung, J. C. De Los Reyes, T. Valkonen, and V. Vlacic

Department for Applied Mathematics and Theoretical Physics
Cantab Capital Institute for the Mathematics of Information
University of Cambridge, UK

Warwick
22 February 2016

UNIVERSITY OF CAMBRIDGE

CAMBRIDGE IMAGE ANALYSIS
DAMTP - UNIVERSITY OF CAMBRIDGE

Outline

Outline

1. Bilevel learning of variational models

2. Conclusions and Outlook

Outline

Generic image analysis model

For given data f we seek a regularised image u by minimising

$$\mathcal{J}(u) = \underbrace{R(u)}_{\text{Prior}} + \lambda \underbrace{\int \phi(T(u), f)}_{\text{Data model}} \to \min_u,$$

where

- $R(u)$ is the prior (regularising) term: modelling a-priori information about the minimiser u in terms of regularity, e.g. $R(u) = \int u^2 \, dx$ which results in $u \in L^2$.
- T linear/nonlinear forward operator, $\phi(T(u), f)$ generic distance function, the data fidelity term of the functional which forces the minimiser u to obey (to a certain extent) the forward model.
- The parameter $\lambda > 0$ balances data model and prior.

Modelling ingredients

The result heavily depends on the correct modelling. There are two main degrees of freedom

- Image model: R, prior, regularity of the image, basis function representation, sparsity, . . .
- Data model: $T, \ \phi, \ \lambda$, physical understanding, statistics, heuristics, . . .

. . . and in both cases, we can try to extract this information directly from the data (experiments).

Modelling ingredients

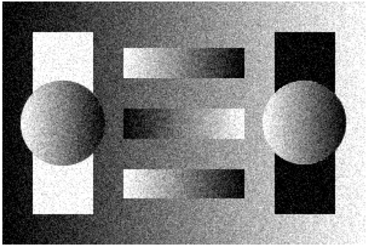
The result heavily depends on the correct modelling. There are two main degrees of freedom

- Image model: R, prior, regularity of the image, basis function representation, sparsity, ...
- Data model: $T, \ \phi, \ \lambda$, physical understanding, statistics, heuristics, ...

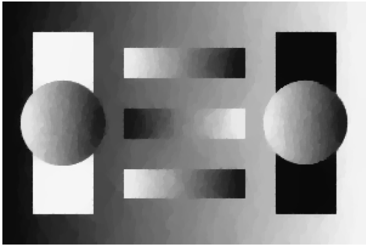
... and in both cases, we can try to extract this information directly from the data (experiments).

TV versus 2nd order TV regularisation

$$\min_u \left\{\alpha \int_\Omega d|Du|(x) + \|u-f\|_2^2\right\}$$



Noisy image



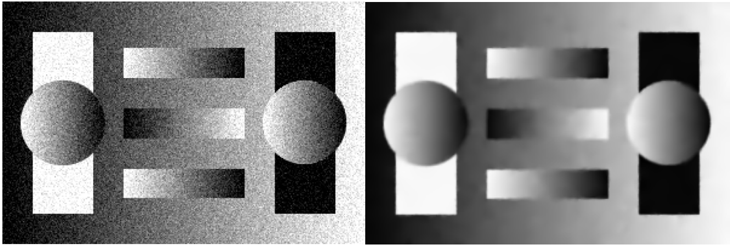
TV denoised image

Image courtesy of K. Papafitsoros

References: Chambolle, Lions, Numerische Mathematik '97; Chan, Marquina, Mulet, SSC '01; Chan, Kang, Shen, SIAM Applied Math '02; Hinterberger, Scherzer, Computing '06; Lysaker, Tai, IJCV '06; Setzer, Steidl, Approximation XII '08; Dal Maso, Fonseca, Leoni, Morini, SIAM Math. Anal. '09; Bergounioux, Piffet, Set Valued and Variational Analysis '10; Bredies, Kunisch, Pock, SIAM Imaging '10; Setzer, Steidl, Teuber, CMS '11; Lefkimmiatis, Bourquard, Unser, '12; Papafitsoros, CBS, J. Math. Imaging & Vision, '13 ...

TV versus 2nd order TV regularisation

$$\min_u \left\{\alpha \int_\Omega d|D^2u|(x) + \|u-f\|_2^2\right\}$$



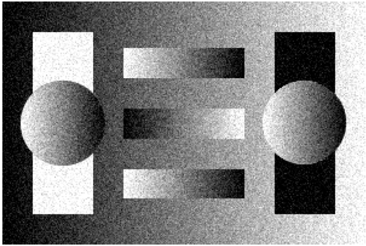
Noisy image TV^2 denoised image

Image courtesy of K. Papafitsoros

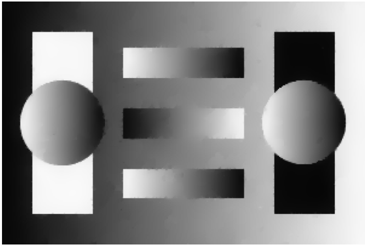
References: Chambolle, Lions, Numerische Mathematik '97; Chan, Marquina, Mulet, SSC '01; Chan, Kang, Shen, SIAM Applied Math '02; Hinterberger, Scherzer, Computing '06; Lysaker, Tai, IJCV '06; Setzer, Steidl, Approximation XII '08; Dal Maso, Fonseca, Leoni, Morini, SIAM Math. Anal. '09; Bergounioux, Piffet, Set Valued and Variational Analysis '10; Bredies, Kunisch, Pock, SIAM Imaging '10; Setzer, Steidl, Teuber, CMS '11; Lefkimmiatis, Bourquard, Unser, '12; Papafitsoros, CBS, J. Math. Imaging & Vision, '13 . . .

TV versus 2nd order TV regularisation

$$\min_u \left\{ \min_w \left\{ \alpha_1 \int_\Omega d|Du - w|(x) + \alpha_2 \int_\Omega d|Ew|(x) \right\} + \|u - f\|_2^2 \right\}$$



Noisy image

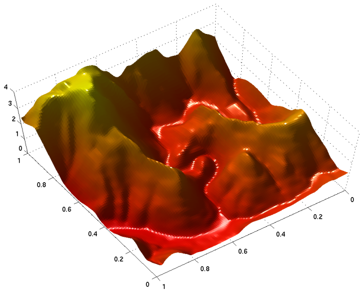


TGV^2 denoised image

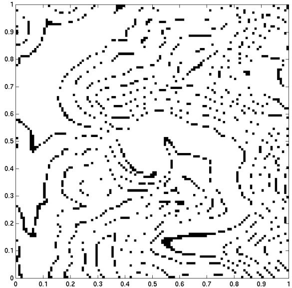
Image courtesy of K. Papafitsoros

References: Chambolle, Lions, Numerische Mathematik '97; Chan, Marquina, Mulet, SSC '01; Chan, Kang, Shen, SIAM Applied Math '02; Hinterberger, Scherzer, Computing '06; Lysaker, Tai, IJCV '06; Setzer, Steidl, Approximation XII '08; Dal Maso, Fonseca, Leoni, Morini, SIAM Math. Anal. '09; Bergounioux, Piffet, Set Valued and Variational Analysis '10; Bredies, Kunisch, Pock, SIAM Imaging '10; Setzer, Steidl, Teuber, CMS '11; Lefkimmiatis, Bourquard, Unser, '12; Papafitsoros, CBS, J. Math. Imaging & Vision, '13 ...

Anisotropic TV^3 interpolation



Ground truth

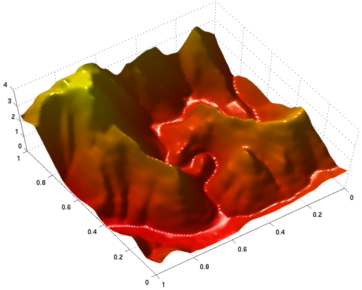


Input contours

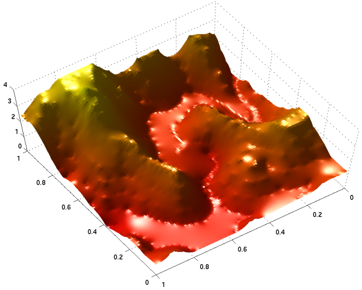
Image courtesy of J. Lellmann

References: Lellmann, Morel, CBS, Scale Space Var. Meth. Comp. Vis. '13; T. Meyer '11; Lellmann, Masnou, Parisotto, CBS, in preparation

Anisotropic TV^3 interpolation



Ground truth

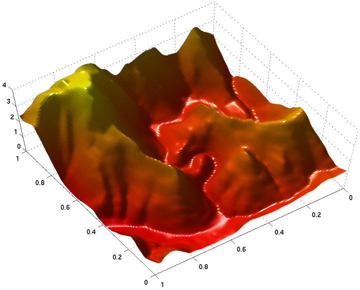


TV^2

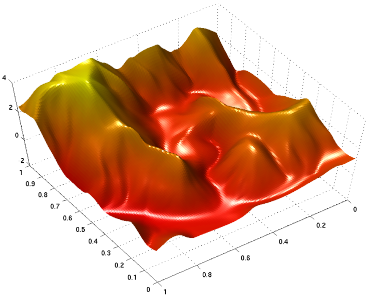
Image courtesy of J. Lellmann

References: Lellmann, Morel, CBS, Scale Space Var. Meth. Comp. Vis. ’13; T. Meyer ’11; Lellmann, Masnou, Parisotto, CBS, in preparation

Anisotropic TV^3 interpolation



Ground truth



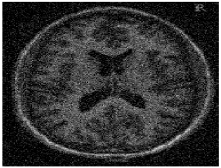
$\mathrm{TV}_v^3(u) = |D^3 u(v, v, v)|)$

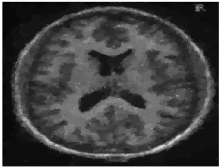
Image courtesy of J. Lellmann

References: Lellmann, Morel, CBS, Scale Space Var. Meth. Comp. Vis. '13; T. Meyer '11; Lellmann, Masnou, Parisotto, CBS, in preparation

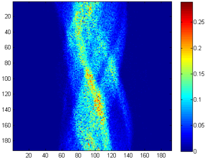
Choice of ϕ depends on type of corruption

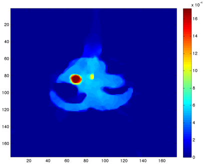
Gaussian	Poisson	Impulse
$\phi(Tu,f) = \|Tu - f\|_2^2$	$\phi(Tu,f) = \int Tu - f\log(Tu)\,dx$	$\phi(Tu,f) = \|Tu - f\|_1$





MRI





PET[1]

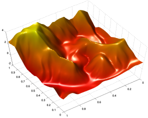
sparse noise.

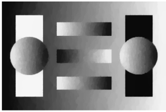
References: see works by Hohage and Werner '12–

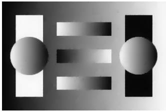
[1] Data courtesy of EIMI, Münster.

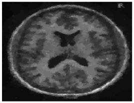
How to develop a variational model?

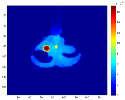
Which regularisation, which data fit to choose?











And how much of it? . . .

Bilevel learning - the idea

Assumptions

Training set of pairs $(f_k, u_k),\ k = 1, \ldots, N$ with

- f_k imperfect data
- u_k represent the ground truth

Determine optimal regulariser R, data model ϕ, and λ in admissible set $\mathcal{A}$

$$\min_{(R,\phi,\lambda)\in\mathcal{A}} \sum_k \text{Cost}(\bar{u}_k, u_k)$$

subject to

$$\bar{u}_k = \text{argmin}_u \left\{ R(u) + \int_\Omega \lambda\, \phi(Tu, f_k)\, dx \right\}$$

Learning by optimisation in imaging

Some related contributions

- Odone '05–, Tappen et al. '07, '09; Domke '11–: Markov Random Field models; stochastic descent method
- Lui, Lin, Zhang and Su '09: optimal control approach, no analytical justification; promising numerical results.
- Horesh, Tenorio, Haber et al. '03–: optimal design; ℓ_1 minimisation.
- Kunisch and Pock '13, Pock 13' –: results for finite dimensional case; semismooth Newton method; optimal image filters; optimal SVM;
- Chung et al. '14: finite dimensional; bounded operator T.
- Hintermüller et al. '14 – : bilevel optimisation for blind deconvolution, and for adaptive TV denoising.
- ...

Learning by optimisation in imaging

Some related contributions

- Odone '05–, Tappen et al. '07, '09; Domke '11–: Markov Random Field models; stochastic descent method
- Lui, Lin, Zhang and Su '09: optimal control approach, no analytical justification; promising numerical results.
- Horesh, Tenorio, Haber et al. '03–: optimal design; ℓ_1 minimisation.
- Kunisch and Pock '13, Pock 13' –: results for finite dimensional case; semismooth Newton method; optimal image filters; optimal SVM;
- Chung et al. '14: finite dimensional; bounded operator T.
- Hintermüller et al. '14 – : bilevel optimisation for blind deconvolution, and for adaptive TV denoising.
- ...

No results in function spaces

Learning in function space

Our goal: State and treat a nonsmooth optimization problem in function space (stick to the physical model).

- Infinite dimensional models more amenable to analysis of image features, e.g. edges.
- Lagrange multipliers and optimality condition.
- Compute optimal weights λ_i with a fast derivative-based and mesh independent optimisation method (second-order method), e.g. Hintermüller, Stadler '06; resolution independent imaging Viola, Fitzgibbon, Cipolla '12.

Learning for TV-type regularisation models . . .

Learning TV-type regularisation

Look for $\lambda = (\lambda_1, \ldots, \lambda_M)$ and $\alpha = (\alpha_1, \ldots, \alpha_N)$ solving

$$\min_{(\lambda,\alpha)\in\mathcal{Q}^+} F(u_{\lambda,\alpha})$$

subject to

$$u_{\lambda,\alpha} \in \operatorname{argmin}_{u\in X} \sum_{i=1}^{M} \int_\Omega \lambda_i(x)\phi_i([Tu](x))\,dx + \sum_{j=1}^{N} \int_\Omega \alpha_j(x)\,d|A_j u|(x).$$

Here $T : X \to Y \subset L^1(\Omega;\mathbb{R}^d)$ with X, Y Banach spaces, $A_j : X \to \mathcal{M}(\Omega;\mathbb{R}^{m_j})$, ($j = 1, \ldots, N$) are appropriate linear operators, $|A_j u|$ total variation measure, F is cost function.

Example I: Learning λ's in TV denoising

- Take $\alpha = 1$, $\lambda \in \mathbb{R}_+$.
- Take $X = BV(\Omega) \cup L^2(\Omega)$, $Y = L^2(\Omega)$, and set

$$T(u) = u, \quad A_1 = D$$

$$u_\lambda \in \text{argmin}_{u \in X} \frac{\lambda}{2} \|u - f\|_Y^2 + \int_\Omega d|Du|(x).$$

Example II: Learning (β, α) in $\mathrm{TGV}^2_{\beta,\alpha}$

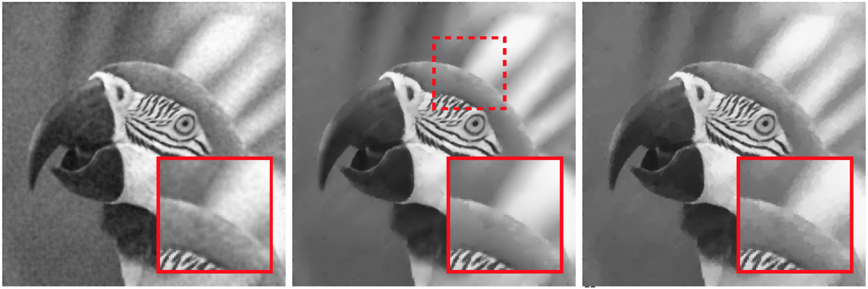
- Take $\alpha = (\alpha_1, \alpha_2) \in \mathbb{R}^2_+$, $\lambda = 1$
- Take $u = (v, w)$ and $\phi(u) = \int_\Omega \phi(u(x))\, dx = \frac{1}{2}\|v - f\|^2_Y$, $Y = L^2(\Omega)$
- Take $X = BV(\Omega) \cup L^2(\Omega) \times BD(\Omega)$ and set

$$T(v, w) = v, \quad A_1 u = Dv - w, \quad \text{and} \quad A_2 u = Ew$$

for E the symmetrised differential.

$$u_{\alpha_1,\alpha_2} \in \operatorname{argmin}_{(v,w)\in X} \frac{1}{2}\|v - f\|^2_{L^2(\Omega)} + \alpha_1 \int_\Omega d|Dv - w|(x) + \alpha_2 \int_\Omega d|Ew|(x).$$

Choice of β in $\mathrm{TGV}^2_{\beta,\alpha}$



(a) Too low β / High oscillation (b) Optimal β (c) Too high β / almost TV

Choice of α in $\mathrm{TGV}^2_{\beta,\alpha}$

(a) Too low α, low β. Good match to noisy data

(b) Too low α, optimal β. optimal TV^2-like behaviour

(c) Too high α, high β. Bad TV^2-like behaviour

Cost function: how to measure optimality?

Cost functions: For noise free data $\tilde{u}$ we take either PSNR

$$F_{L^2}(v) = \frac{1}{2}\|\tilde{u} - v\|_2^2,$$

or Huberised TV cost

$$F_{L^1\nabla_\gamma}(v) = |D(\tilde{u} - v)|_\gamma(\Omega).$$

Existence and beyond . . .

Existence of an optimal solution (under appropriate assumptions optimal parameter lies in the interior!) ✓

Would like to use derivate-based method for the numerical solution of this problem ⇒ need gradient of solution map, encoded in adjoint equation of ***optimality system****.*

Smoothed optimization problem

Given one training pair (f, u_{org})

$$\min \; \|\bar{u} - u_{org}\|^2_{L^2(\Omega)}$$

subject to Total Variation denoising

$$\underbrace{\mu \int_\Omega \nabla \bar{u} \cdot \nabla (v - \bar{u}) \, dx}_{\text{Elliptic regularisation}} + \underbrace{(h_\gamma(\nabla \bar{u}), \nabla (v - \bar{u}))}_{\text{Huber type smoothing}} =$$

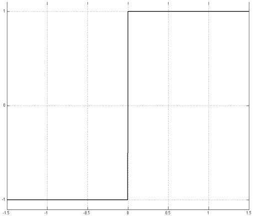
$$- \lambda \int_\Omega (\bar{u} - f)(v - \bar{u}) \, dx, \; \forall v \in H^1_0(\Omega),$$

Optimization with PDE constraints

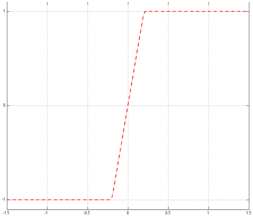
References: Barbu (1984, 1993), Tiba (1990), Bonnans-Tiba (1991), Wenbin-Rubio (1991), Bonnans-Casas (1995), Bergounioux (1998), De Los Reyes (2012)

Huber regularization

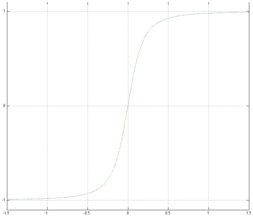
⇒ smoothing of TV measure needed



Subdifferential of $|\cdot|$

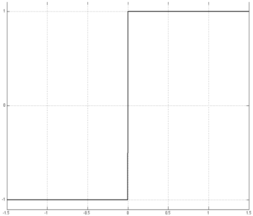


Huber type function

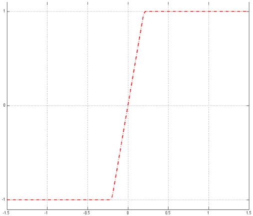


Huber regularization

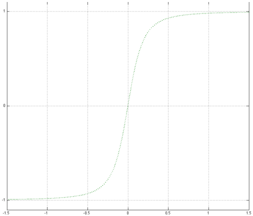
$\Rightarrow$ smoothing of TV measure needed



Subdifferential of $|\cdot|$



Huber type function

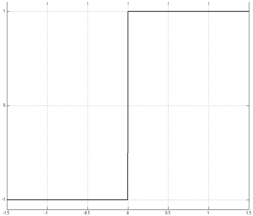


$\frac{x}{\sqrt{x^2+\epsilon^2}}$

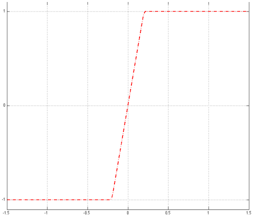
$$h_\gamma(z) = \begin{cases} \frac{z}{|z|} & \text{if } \gamma|z| \geq 1 + \frac{1}{2\gamma} \\ \frac{z}{|z|}(1 - \frac{\gamma}{2}(1 - \gamma|z| + \frac{1}{2\gamma})^2) & \text{if } 1 - \frac{1}{2\gamma} \leq \gamma|z| \leq 1 + \frac{1}{2\gamma} \\ \gamma z & \text{if } \gamma|z| \leq 1 - \frac{1}{2\gamma}, \end{cases}$$

Huber regularization

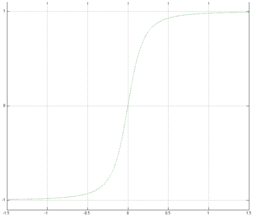
$\Rightarrow$ smoothing of TV measure needed



Subdifferential of $|\cdot|$



Huber type function ✓

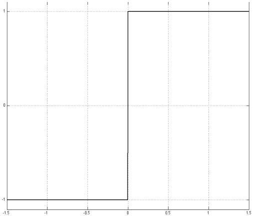


~~$\frac{x}{\sqrt{x^2+\epsilon^2}}$~~

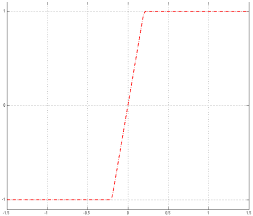
$$h_\gamma(z) = \begin{cases} \frac{z}{|z|} & \text{if } \gamma|z| \geq 1 + \frac{1}{2\gamma} \\ \frac{z}{|z|}(1 - \frac{\gamma}{2}(1 - \gamma|z| + \frac{1}{2\gamma})^2) & \text{if } 1 - \frac{1}{2\gamma} \leq \gamma|z| \leq 1 + \frac{1}{2\gamma} \\ \gamma z & \text{if } \gamma|z| \leq 1 - \frac{1}{2\gamma}, \end{cases}$$

Huber regularization

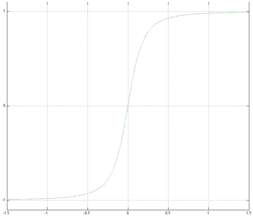
$\Rightarrow$ smoothing of TV measure needed



Subdifferential of $|\cdot|$



Huber type function ✓



~~$\frac{x}{\sqrt{x^2+\epsilon^2}}$~~

$$h_\gamma(z) = \begin{cases} \frac{z}{|z|} & \text{if } \gamma|z| \geq 1 + \frac{1}{2\gamma} \\ \frac{z}{|z|}(1 - \frac{\gamma}{2}(1 - \gamma|z| + \frac{1}{2\gamma})^2) & \text{if } 1 - \frac{1}{2\gamma} \leq \gamma|z| \leq 1 + \frac{1}{2\gamma} \\ \gamma z & \text{if } \gamma|z| \leq 1 - \frac{1}{2\gamma}, \end{cases}$$

Smoothed optimization problem

Given one training pair (f, u_{org})

$$\min \ \|\bar{u} - u_{org}\|^2_{L^2(\Omega)}$$

subject to Total Variation denoising

$$\underbrace{\mu \int_\Omega \nabla \bar{u} \cdot \nabla (v - \bar{u})\, dx}_{\text{Elliptic regularisation}} + \underbrace{(h_\gamma(\nabla \bar{u}), \nabla (v - \bar{u}))}_{\text{Huber type smoothing}} =$$

$$- \lambda \int_\Omega (\bar{u} - f)(v - \bar{u})\, dx, \ \forall v \in H^1_0(\Omega),$$

Optimization with PDE constraints

References: Barbu (1984, 1993), Tiba (1990), Bonnans-Tiba (1991), Wenbin-Rubio (1991), Bonnans-Casas (1995), Bergounioux (1998), De Los Reyes (2012)

In this setting we can prove . . .

- existence of an optimal solution, also in the case $\mu = 0$ (under appropriate assumptions optimal parameter lies in the interior!) ✓
- convergence of optimal parameters and corresponding reconstructions to solution of original, non-smooth optimisation problem as $\gamma \to +\infty$ and $\mu \to 0$ ✓
- differentiability of solution operator and derivation of sharp optimality system ✓ De Los Reyes '12
- convergence of optimality system as Huber regularisation $\gamma \to +\infty$ to sharp optimality system for non-smooth problem De Los Reyes '12.

. . . and in the numerics the parameters $0 < \mu \ll 1$ and $\gamma \gg 1$.

. . . open: limit of optimality system for $\mu \to 0$???

References: De Los Reyes 2012; CBS, De Los Reyes 2013; Calatroni, CBS, De Los Reyes 2014; **De Los Reyes, CBS, Valkonen 2015**

Numerical strategy

Solve

$$\min_{(\lambda,\alpha)\in\mathcal{Q}^+} F(u_{\lambda,\alpha})$$

subject to

$$u_{\lambda,\alpha} = \operatorname{argmin}_u$$

$$\frac{\mu}{2}\sum_{j=1}^{N}\|\nabla A_j\|_2^2 + \sum_{i=1}^{M}\int_\Omega \lambda_i(x)\phi_i(x,[Tu](x))\,dx + \sum_{j=1}^{N}\int_\Omega \alpha_j(x)\,d|A_j u|_\gamma(x).$$

by quasi-Newton method (BFGS)

- state equation is solved by Newton type algorithm (varies with ϕ)
- evaluation of the gradient of the cost functional with adjoint information
- Armijo line search with curvature verification.
- For large training set we use dynamic sampling technique for constraints á la Byrd et al.

Parameters: we typically choose $10^{-10} \approx \mu \ll 1$, $100 \approx \gamma \gg 1$.

Some examples

Optimal parameter for TV

$$\min_{\lambda \geq 0} \|u - u_k\|_{L^2}^2$$

subject to:

$$\min_{u \geq 0} \left\{ \frac{\mu}{2}\|Du\|_{L^2}^2 + \|Du\|_\gamma + \frac{\lambda}{2}\|u - f_k\|_{L^2}^2 \right\}$$

Noise $n \in N(0, 0.002)$ (optimal parameter $\lambda^* = 2980$)

Optimal parameter for TV

$$\min_{\lambda \geq 0} \ \|u - u_k\|_{L^2}^2$$

subject to:

$$\min_{u \geq 0} \left\{ \frac{\mu}{2} \|Du\|_{L^2}^2 + \|Du\|_\gamma + \frac{\lambda}{2} \|u - f_k\|_{L^2}^2 \right\}$$

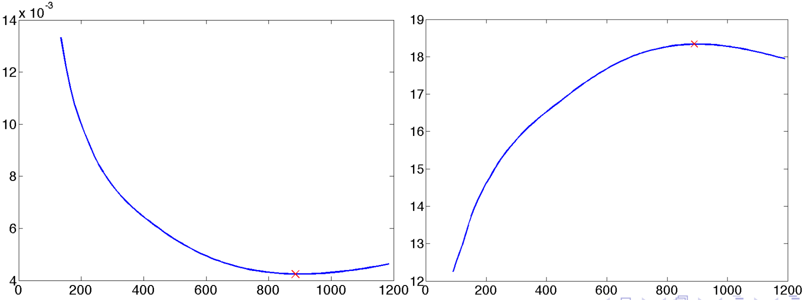

Noise $n \in N(0, 0.02)$ (optimal parameter $\lambda^* = 1770.9$)

Optimality?

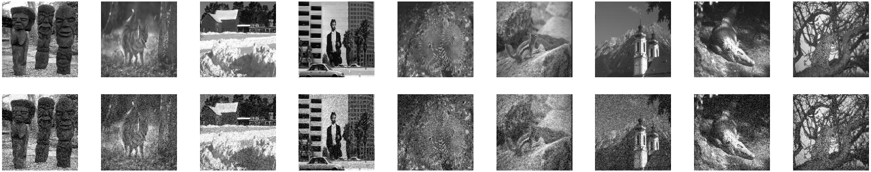
Quality measure

- Original cost functional (left figure) $\|u - u_k\|_{L^2}^2$
- Signal to noise ratio (right figure)

$$SNR = 20 \times \log_{10} \left(\frac{\|u_k\|_{L^2}}{\|u - u_k\|_{L^2}} \right),$$



Robustness and efficiency



N	10	20	30	40
λ^*	2732.15	2766.32	2170.23	2292.51

Learning (β, α) in $\text{TGV}^2_{\beta,\alpha}$

(a) Original image (b) Noisy image

Optimal $\text{TGV}^2_{\beta,\alpha}$

(c) TGV^2 denoising, $L_1\nabla_\gamma$ cost functional

(d) TGV^2 denoising, L_2^2 cost functional

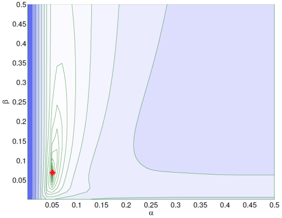
$L^1\nabla_\gamma$ cost

$$(\alpha, \beta) = (0.069/n^2, 0.051/n)$$

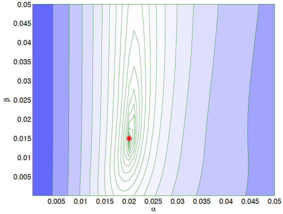
L^2 cost

$$(\alpha, \beta) = (0.058/n^2, 0.041/n)$$

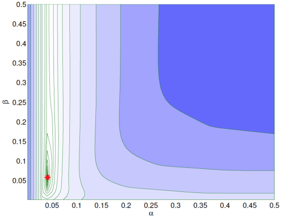
Optimal (β, α) in $\mathrm{TGV}^2_{\beta,\alpha}$?



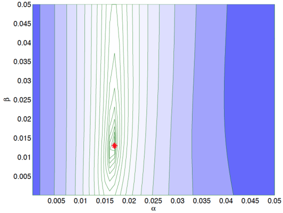
(a) Parrot, TGV^2, $L_1\nabla_\gamma$ cost functional



(b) Uplands, TGV^2, $L_1\nabla_\gamma$ cost functional



(c) Parrot, TGV^2, L_2^2 cost functional



(d) Uplands, TGV^2, L_2^2 cost functional

For TGV a good initialisation is important!

TV versus TGV^2 versus ICTV

(g) TV denoising, $L^1_\eta \nabla$ cost

(e) ICTV denoising, $L^1_\eta \nabla$ cost

(c) TGV^2 denoising, $L^1_\eta \nabla$ cost

TV versus TGV^2 versus ICTV

Test performance of TV versus TGV versus ICTV regularisation on 200 images from Berkeley image database; for noise levels $\sigma^2 = 2, 10$ and 20; and for the two cost functionals.

Evaluate performance wrt PSNR, SIIM, and cost functional (L^2 and L$^1\nabla$).

Perform statistical 95% one-tailed paired t-test on each of criteria, and pair of regularisers, to see whether any pair of regularisers can be ordered.

TV versus TGV^2 versus ICTV

Noise level $\sigma^2 = 20$; cost functional is $L^1\nabla$.

Colour coding: TV best; ICTV best; TGV best

TV versus TGV^2 versus ICTV

Overall result:

- Overall, studying the t-test and other data, the ordering of the regularisers appears to be

$$ICTV > TGV^2 > TV$$

- For high noise TGV^2 and ICTV performance are comparable.
- TGV^2 is better than ICTV for images with large smooth areas.
- L^2 cost corresponds to high PSNR; $\text{L}^1\nabla$ cost seems to relate to high SIIM.

Impulse noise

$$\min \ \frac{1}{2}\|u - u_{org}\|^2_{L^2}$$

subject to:

$$\min_{u \geq 0} \left\{ \frac{\mu}{2}\|Du\|^2_{L^2} + \|Du\|_\gamma + \lambda\|u - f\|_\gamma \right\}$$

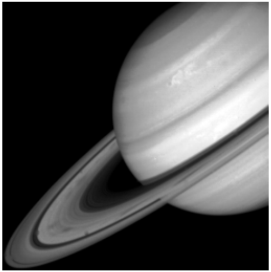
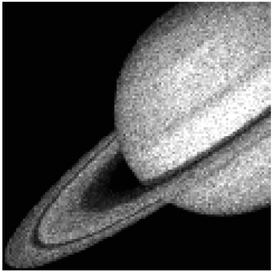
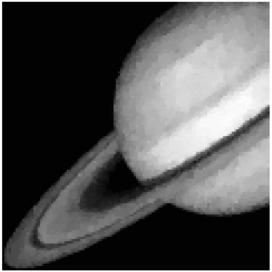
Impulse noise with 5% corrupted pixels; optimal parameter $\lambda^* = 45.88$

Poisson noise

$$\min_{\lambda \geq 0} \frac{1}{2}\|u - u_{org}\|_{L^2}^2$$

subject to:

$$\min_{u \geq 0} \left\{ \frac{\mu}{2}\|Du\|_{L^2}^2 + \|Du\|_\gamma + \lambda \int_\Omega (u - f \log u)\, dx \right\}.$$

Optimal parameter $\lambda^* = 1013.76$.

Spatial dependent noise

Gaussian noise with $\sigma = 0.04$ outside of region outlined in red and $\sigma = 0.06$ inside.

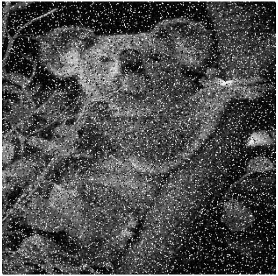
A bit beyond the theory . . .

Mixed Impulse & Gaussian noise

$$\min_{\lambda_1,\lambda_2 \geq 0} \frac{1}{2}\|u - u_{org}\|_{L^2}^2$$

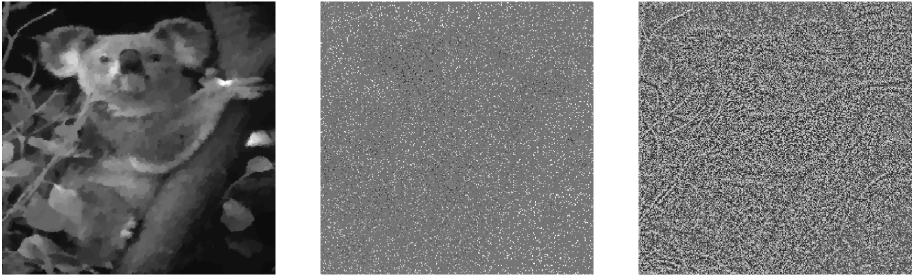
where u is the solution of the optimisation problem:

$$\min_{\substack{v \in BV \\ n \in L^2}} \left\{ \frac{\mu}{2}\|\nabla v\|_{L^2}^2 + \|Dv\|_\gamma + \lambda_1\|n\|_\gamma + \lambda_2\|f - v - n\|_{L^2}^2 \right\},$$



Original image (left) and noisy image (right) corrupted by impulse

Mixed Impulse & Gaussian noise



From left to right: Denoised image, impulse noise residuum and Gaussian noise residuum. Optimal parameters: $\lambda_1^* = 351.23$ and $\lambda_2^* = 5200.1$.

Space-time regularisation

Consider regularisers of the form

$$G\left(u;\alpha_1,\alpha_2,\kappa\right)=\inf_{u=v+w}\alpha_1\|\nabla_\kappa v\|_{2,1}+\alpha_2\|\nabla_{1-\kappa}w\|_{2,1}$$

with

$$\nabla_\kappa=\left(\kappa\frac{\partial}{\partial x},\kappa\frac{\partial}{\partial y},(1-\kappa)\frac{\partial}{\partial t}\right)$$

Nonlinear dependence on learned parameter $\Rightarrow$ can prove things in the elliptic setting.

References: M. Holler and K. Kunisch, SIIMS '14.

Space-time regularisation

Outline

Learning versus modelling?

Data learning versus variational modelling

Variational modelling	Recent data learning
physical model	non-physical
non-adaptive to data	adaptive to data
insight in structure of problem	blackbox
reconstruction guarantees stability, error analysis, . . .	in general no guarantees guarantees optimality?
heavily relies on a-priori model	learns the model from the data.

Data learning versus modelling?

Happy marriage of physical modelling and data learning?

Data learning versus modelling?

Happy marriage of physical modelling and data learning?

Hope: adaptive models with theoretical guarantees.

Conclusions and outlook

Conclusions:

- Bilevel learning of variational models: learning regularisation and learning data fit for image de-noising
- Analysis of bilevel model in function space
- Example: Choice of optimal TV-based regularisation and optimal data fit for noise model.

Conclusions and outlook

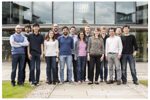
Conclusions:

- Bilevel learning of variational models: learning regularisation and learning data fit for image de-noising
- Analysis of bilevel model in function space
- Example: Choice of optimal TV-based regularisation and optimal data fit for noise model.

Outlook:

- Alternative cost functionals. How to measure optimality? Non-reference quality measures.
- Model learning for inverse problems: general linear/nonlinear operator T (MRI, PET, ET, ...)
- Learning other elements in the model, e.g. acquisition (sampling), inpainting procedure, segmentation (Leaci, Tomarelli, CBS) ...

Thank you very much for your attention!



- J. C. De Los Reyes, and C.-B. Schönlieb, *Image denoising: Learning noise distribution via PDE-constrained optimisation*, Inverse Problems and Imaging, Vol. 7, 1183-1214, 2013.
- L. Calatroni, J. C. De Los Reyes, and C.-B. Schönlieb, *Dynamic sampling schemes for optimal noise learning under multiple nonsmooth constraints*, IFIP TC7-2013 proceedings.
- L. Calatroni, C. Cao, J. C. De Los Reyes, C.-B. Schönlieb, and T. Valkonen, *Bilevel approaches for learning of variational imaging models*, to appear in Radon book series 2015.
- J. C. De Los Reyes, and C.-B. Schönlieb, and T. Valkonen, *The structure of optimal parameters for image restoration problems*, Journal of Mathematical Analysis and Applications 434 (2016), 464-500.
- J. C. De Los Reyes, and C.-B. Schönlieb, and T. Valkonen, *Optimal parameter learning for higher-order total variation regularisation models*, submitted.
- M. Benning, C.-B. Schönlieb, T. Valkonen, and V. Vladic, *Explorations on anisotropic regularisation of dynamic inverse problems*, tbs.
- L. Calatroni, J. C. De Los Reyes, and C.-B. Schönlieb, *A variational model for total variation mixed noise removal*, tbs.

More information see:

`http://www.damtp.cam.ac.uk/research/cia/`

Email: `cbs31@cam.ac.uk`