Optimising image reconstruction by nonlinear PDE constrained optimisation

Carola-Bibiane Schönlieb

Department for Applied Mathematics and Theoretical Physics
University of Cambridge, UK

Chicheley Hall - July, 1st 2014

UNIVERSITY OF CAMBRIDGE

Optimise the optimisers?
Optimise our model choice?

- Image reconstruction model customised to data.
- Combine physical modelling with data learning.

Outline

1. Mathematical imaging models

Outline

1. Mathematical imaging models
2. Optimal TV denoising via PDE-constrained optimisation

Outline

1. Mathematical imaging models
2. Optimal TV denoising via PDE-constrained optimisation
3. Glimps into optimising $\mathrm{TGV}^2_{\beta,\alpha}$ regularisation

Outline

1. Mathematical imaging models
2. Optimal TV denoising via PDE-constrained optimisation
3. Glimps into optimising $\mathrm{TGV}^2_{\beta,\alpha}$ regularisation
4. Conclusions and outlook

Outline

A generic inverse problem in imaging

The problem

Given data f, find the image information u which solves

$$f = T(u) + n$$

where T is a linear (or nonlinear) forward operator that models the relation between u and f and n is a noise component.

If T has an unbounded inverse, the problem is ill-posed. Causes: non-uniqueness, unstable inversion, noise, under sampling, …

The problem has to be regularised by adding a-priori information on u…

The variational approach..

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

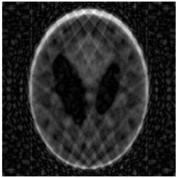
$$\mathcal{J}(u) = \underbrace{R(u)}_{\text{Prior}} + \lambda \underbrace{\phi(T(u), f)}_{\text{Data model}} \to \min_u,$$

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 $u \in L^2$.
- $\phi(T(u), f)$ is a 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.

Nonsmooth convex minimization from real-life problems

Reconstruction via Fourier-Transformation with 0.11% of the data



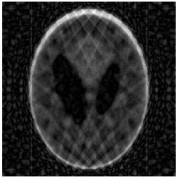
Reconstructed image

MRI: measured datum is a sampled Fourier transform; nonadaptive compressed acquisition, i.e., compressed sensing

$$f = \mathcal{S}\mathcal{F}u + n$$

Nonsmooth convex minimization from real-life problems

Reconstruction via Fourier-Transformation with 0.11% of the data



Reconstructed image

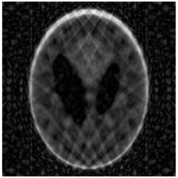
MRI: measured datum is a sampled Fourier transform; nonadaptive compressed acquisition, i.e., compressed sensing

$$f = \mathcal{S}\mathcal{F}u + n$$

Goal: identify a piecewise constant function u consistent to the datum

Nonsmooth convex minimization from real-life problems

Reconstruction via Fourier-Transformation with 0.11% of the data



Reconstructed image

MRI: measured datum is a sampled Fourier transform; nonadaptive compressed acquisition, i.e., compressed sensing

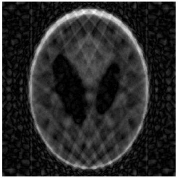
$$f = \mathcal{S}\mathcal{F}u + n$$

Goal: identify a piecewise constant function u consistent to the datum

Sparsity in $\nabla u \Rightarrow$ NP-hard problem!

Nonsmooth convex minimization from real-life problems

Reconstruction via Fourier-Transformation with 0.11% of the data



Reconstructed image

MRI: measured datum is a sampled Fourier transform; nonadaptive compressed acquisition, i.e., compressed sensing

$$f = \mathcal{S}\mathcal{F}u + n$$

Goal: identify a piecewise constant function u consistent to the datum

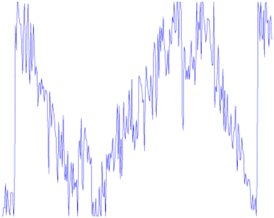
Sparsity in $\nabla u \Rightarrow$ NP-hard problem!

Convex relaxation $\Rightarrow$ ℓ_1- and total variation (TV) minimization

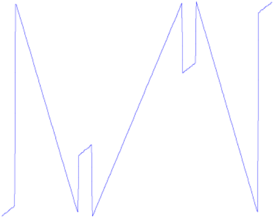
$$\alpha\|\nabla u\|_1 + \frac{1}{2}\|\mathcal{S}\mathcal{F}u - f\|_2^2 \to \min_u,$$

TV $= \|\nabla u\|_1$?

Total variation quantifies the oscillations of u:



(a) Large TV



(b) Small TV

Modelling

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

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).

What difference does it make? A few examples . . .

H^1 versus TV regularisation

(a) original

(b) noisy

(c) $R(u) = \|\nabla u\|_2^2$

References: Rudin, Osher, Fatemi, Physica D '92; Chambolle, Lions, Numerische Mathematik '97; Vese, Applied Mathematics and Optimization '01, ...

H^1 versus TV regularisation

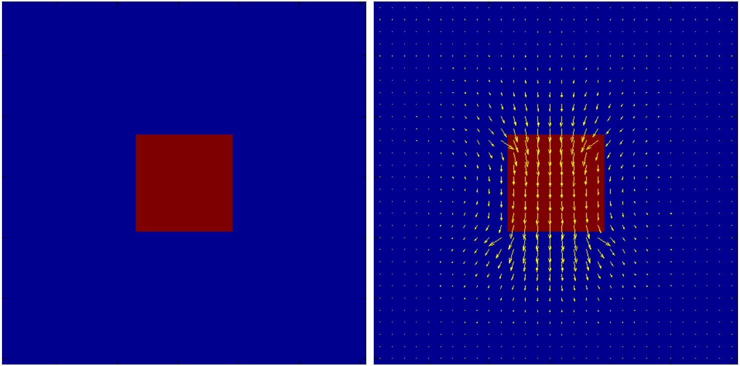
(d) original

(e) noisy

(f) $R(u) = TV(u)$

References: Rudin, Osher, Fatemi, Physica D '92; Chambolle, Lions, Numerische Mathematik '97; Vese, Applied Mathematics and Optimization '01, ...

H^1 versus TV regularisation

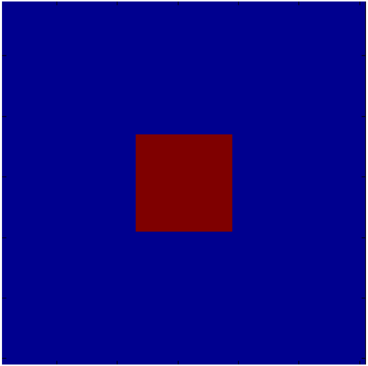


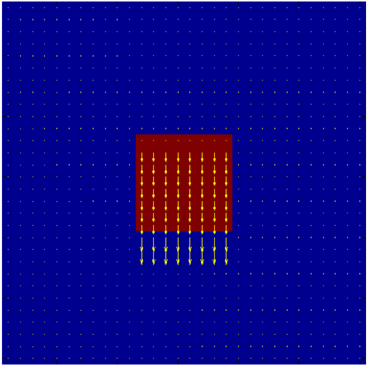
(g) $\|\nabla v\|_2^2$-flow estimation

Image courtesy of H. Dirks

References: Horn, Schunk, Artifical Intelligence '81; Aubert, Kornprobst, SIAM Journal on Mathematical Analysis '99; Cremers, Pock et al. '09; . . . Burger, Dirks, CBS '14.

H^1 versus TV regularisation



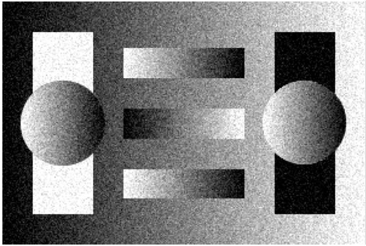


(h) $TV(v)$-flow estimation

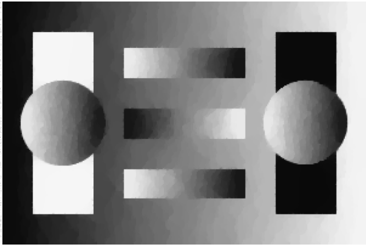
Image courtesy of H. Dirks

References: Horn, Schunk, Artifical Intelligence ’81; Aubert, Kornprobst, SIAM Journal on Mathematical Analysis ’99; Cremers, Pock et al. ’09; . . . Burger, Dirks, CBS ’14.

TV versus 2nd order TV regularisation



Noisy image

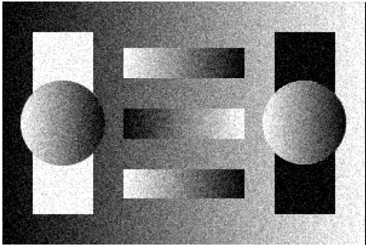


TV denoised image

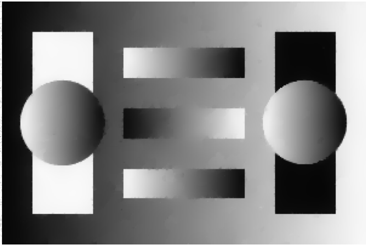
Image courtesy of K. Papafitsoros

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



Noisy image

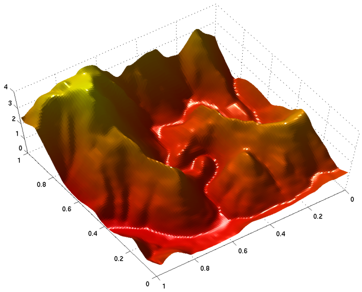


TGV^2 denoised image

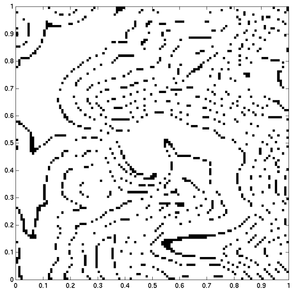
Image courtesy of K. Papafitsoros

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

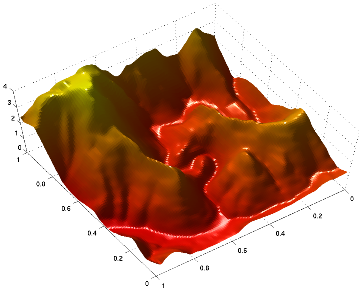


Input contours

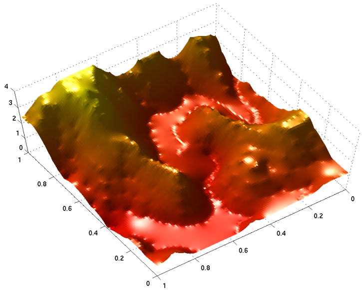
Image courtesy of J. Lellmann

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

Anisotropic TV^3 interpolation



Ground truth

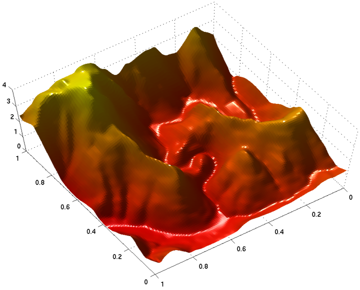


TV^2

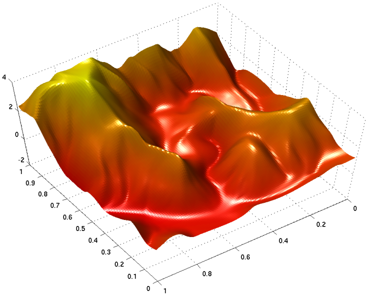
Image courtesy of J. Lellmann

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

Anisotropic TV^3 interpolation



Ground truth



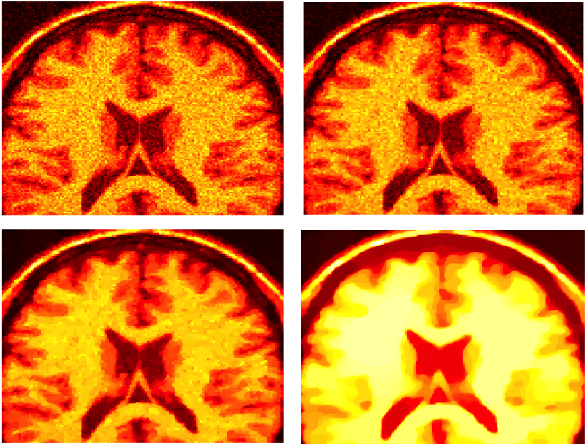
Anisotropic TV^3

Image courtesy of J. Lellmann

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

Weight λ between image and data model

Total variation denoising for Gaussian noise

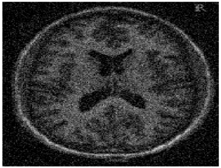


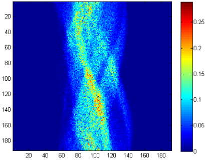
with increasing regularisation (from left to right).

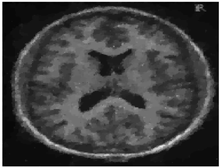
Effect of regulariser is complemented by effect of data term . . .

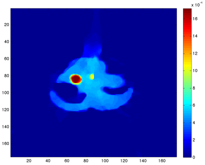
Choice of discrepancy ϕ depends on data

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] | shot noise.

References: see recent works by Hohage and Werner '12–

[1] Data courtesy of EIMI, Münster.

Optimise the choice of image and data model

State of the art in optimal model design

- Using a-priori information such as the noise level, e.g. Morozov '93; Engl, Hanke, Neubauer '96; . . . ; Baus, Nikolova, Steidl, JMIV '14;
- Adaptive parameter choice rules Hintermüller et al. '11; Frick, Marnitz, Munk '12–; Naumova, Pereverzyev, Sivananthan '12; Fornasier, Naumova, Pereverzyev '13.
- Regulariser: Chung, O'Leary et al. '11– (optimal spectral filters); Sapiro et al. (dictionary learning); Peyré, Fadili '11; (learning sparsity priors).
- Bayesian statistics, e.g. Hero et al. Dobigeon, Hero, Tourneret '09; Park, Dobigeon, Hero '14.
- Examples of machine learning approaches: support vector machines, e.g. Tong, Chang '01, reproducible kernel Hilbert spaces Quang, Kang, Le '10, Gaussian mixture models Pedemonte, Bousse, Hutton, Arridge, Ourselin '11, learning by shape priors Cremers, Rousson '07, Markov random fields Tappen '07; Domke '12–, non-smooth priors and noise models Aujol, Gilboa '06; Kunisch, Pock '13; De Los Reyes, CBS '13.
- Statistical optimal design, e.g. Haber and Horesh et al. Haber, Tenorio '03; Huang, Haber, Horesh, Seo '12; Brune,Osher, Sapiro '14.

State of the art in optimal model design

- Using a-priori information such as the noise level, e.g. Morozov '93; Engl, Hanke, Neubauer '96; . . . ; Baus, Nikolova, Steidl, JMIV '14;
- Adaptive parameter choice rules Hintermüller et al. '11; Frick, Marnitz, Munk '12–; Naumova, Pereverzyev, Sivananthan '12; Fornasier, Naumova, Pereverzyev '13.
- Regulariser: Chung, O'Leary et al. '11– (optimal spectral filters); Sapiro et al. (dictionary learning); Peyré, Fadili '11; (learning sparsity priors).
- Bayesian statistics, e.g. Hero et al. Dobigeon, Hero, Tourneret '09; Park, Dobigeon, Hero '14.
- Examples of machine learning approaches: support vector machines, e.g. Tong, Chang '01, reproducible kernel Hilbert spaces Quang, Kang, Le '10, Gaussian mixture models Pedemonte, Bousse, Hutton, Arridge, Ourselin '11, learning by shape priors Cremers, Rousson '07, Markov random fields Tappen '07; Domke '12–, non-smooth priors and noise models Aujol, Gilboa '06; Kunisch, Pock '13; De Los Reyes, CBS '13.
- Statistical optimal design, e.g. Haber and Horesh et al. Haber, Tenorio '03; Huang, Haber, Horesh, Seo '12; Brune,Osher, Sapiro '14.

Bilevel optimisation

Outline

1. Mathematical imaging models
2. Optimal TV denoising via PDE-constrained optimisation
3. Glimps into optimising $\mathrm{TGV}^2_{\beta,\alpha}$ regularisation
4. Conclusions and outlook

A first toy model: TV denoising

$$\mathcal{J}(u) = |Du|(\Omega) + \lambda \int_\Omega \phi(u,f)\, dx,$$

with

$$|Du|(\Omega) = TV(u) = \sup_{\mathbf{g} \in C_0^\infty(\Omega;\mathbb{R}^2), \|g\|_\infty \leq 1} \int_\Omega u\, \nabla \cdot \mathbf{g}\, dx$$

$\longrightarrow$ the total variation of u in Ω

$\longrightarrow$ $\lambda > 0$ positive parameter

$\longrightarrow$ ϕ a suitable distance function called the data fidelity term.

A first toy model: TV denoising

$$\mathcal{J}(u) = |Du|(\Omega) + \lambda \int_\Omega \phi(u, f)\, dx,$$

with

$$|Du|(\Omega) = TV(u) = \sup_{\mathbf{g} \in C_0^\infty(\Omega;\mathbb{R}^2), \|g\|_\infty \leq 1} \int_\Omega u \, \nabla \cdot \mathbf{g}\, dx$$

$\longrightarrow$ the total variation of u in Ω

$\longrightarrow$ $\lambda > 0$ positive parameter

$\longrightarrow$ ϕ a suitable distance function called the data fidelity term.

Joint with Juan Carlos De Los Reyes.

Selection of the noise model with a bilevel optimisation approach!

A generic TV denoising model

$$\min_u \left(|Du|(\Omega) + \sum_{i=1}^{d} \lambda_i \int_\Omega \phi_i(u,f)\, dx \right).$$

where

$\longrightarrow$ ϕ_i, $i = 1, \ldots, d$, convex & differentiable functions in u,

$\longrightarrow$ $\lambda_i \geq 0$

Choices for data fidelities ϕ_i's

- Gaussian noise: $\phi_1(u,f) = (u-f)^2$, ROF,Chambolle & Lions, Vese 1990's, ...
- Impulse noise: $\phi_2(u,f) = |u-f|$, Aujol, Gousseau, Nikolova, Osher, 2000's, ...
- Poisson noise: $\phi_3(u,f) = u - f\log u$, Burger et al. 2009-12
- Other possible choices, e.g. multiplicative noise, Rician noise Getreuer, Tong, Vese '11, ...

...weighted against each other with weights λ_i, which depend on the amount and strength of noise of different distributions in f.

Combinations of noise models also in Nikolova, Wen Chan '12; Hintermüller, Langer '13

Assumptions

Measure noisy data (images) with fixed device & fixed settings

...

interested to learn more about the noise properties of measurements and the optimal setup of the total variation denoising model.

Supervised learning optimal noise model

Assumptions

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

- f_k noisy images
- u_k represent the ground truth

Supervised learning optimal noise model

Assumptions

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

- f_k noisy images
- u_k represent the ground truth

Determine the optimal weights λ_i

$$\min_{\lambda_i \geq 0,\ i=1,\ldots,d} \sum_k \|\bar{u}_k - u_k\|^2_{L^2(\Omega)}$$

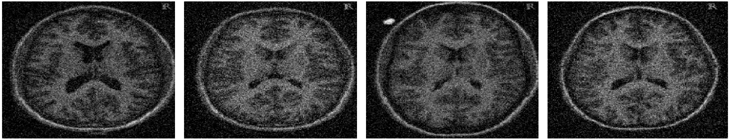
subject to

$$\bar{u}_k = \mathrm{argmin} u \left\{ |Du|(\Omega) + \sum_{i=1}^{d} \lambda_i \int_\Omega \phi_i(u, f_k)\, dx \right\}$$

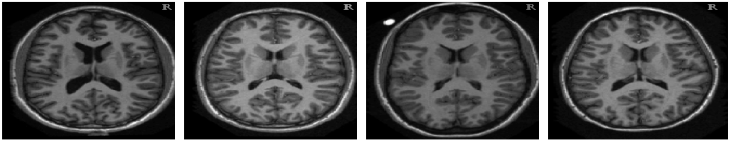
Learning from training sets

Image denoising training sets such as

Low resolution MRI scan

f_k  $\dots$

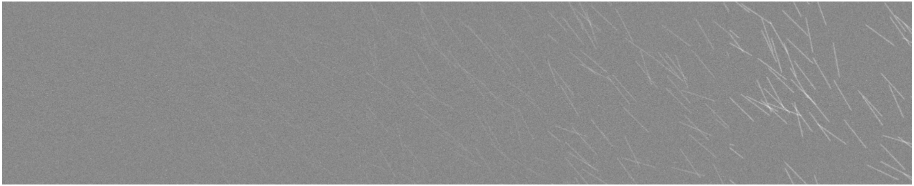
High resolution MRI scan

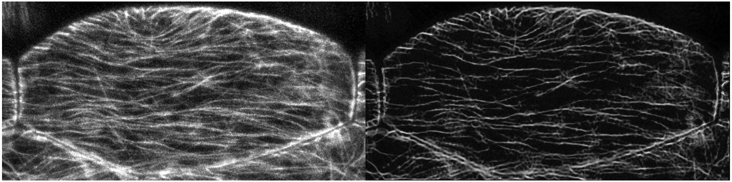
u_k  $\dots$

Simulated data from OASIS online database. Arridge, Kaipio, Kolehmainen, Schweiger, Somersalo, Tarvainen, Vauhkonen '06; Benning, Gladden, Holland, CBS,Valkonen '14

Learning from training sets

Training sets for enhancement of line-like structures such as

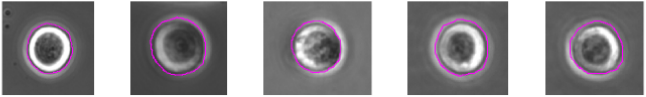




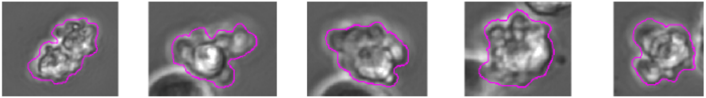
Simulated and real polymer data from Sidney Shaw (Uni Illinois).

Learning from training sets

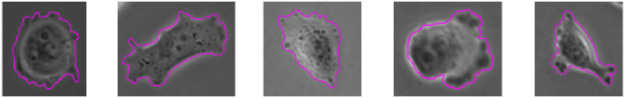
Image segmentation training sets such as



(a) Mitotic cells



(b) Apoptotic cells



(c) Flat cells

Figure 5.9.: Manually segmented test set of mitotic, apoptotic and flat cells
(Courtesy of Light Microscopy Core Facility, Cancer Research UK Cambridge Institute)

Master thesis Joana Grah '14; in preparation Burger, Grah, Reichelt, Schreiner, CBS '14

Learning from training sets

Image inpainting: create desired inpaintings. **Joint with C. Brune**

‘Ecce homo’

‘Ecce mono’

Learning from training sets

Image inpainting: create desired inpaintings. **Joint with C. Brune**

Image courtesy of R. Hocking from CIA. **References:** Arias, Facciolo, Caselles, Sapiro '09–

'Ecce mono'

Bilevel learning in imaging

Related contributions

- 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 (also for ℓ_1 minimisation).
- Kunisch and Pock '13: results for finite dimensional case; semismooth Newton method

Bilevel learning in imaging

Related contributions

- 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 (also for ℓ_1 minimisation).
- Kunisch and Pock '13: results for finite dimensional case; semismooth Newton method

No results in function spaces

Learning by optimization in imaging

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

- Lagrange multipliers and optimality condition
- Compute optimal weights λ_i with a fast derivative-based and mesh independent optimization method (second-order method), e.g. Hintermüller, Stadler '06; resolution independent imaging Viola, Fitzgibbon, Cipolla '12.

Analysis and numerical optimisation

How to approach this constrained optimisation problem? Some hints:

- State of the art optimisation with non-smooth constraints in Hilbert space (elliptic constraints) $\Rightarrow$ derive optimality system (sharp?).
  Barbu (1984, 1993), Tiba (1990), Bonnans-Tiba (1991), Wenbin-Rubio (1991), Bonnans-Casas (1995), Bergounioux (1998).
- Use derivative-based iterative methods for numerical solution of our model (quasi-Newton). So need unique gradient of the total variation denoising model $\Rightarrow$ optimality system (unique).
- Relationship of smoothed problem to original one? Convergence?
  De Los Reyes 2012.

$\Rightarrow$ more on this in De Los Reyes 2012; CBS, De Los Reyes 2013; Calatroni, CBS, De Los Reyes 2014.

Regularized optimization problems

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

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

subject to

$$\varepsilon \int_\Omega \nabla \bar{u} \cdot \nabla (v - \bar{u}) \, dx + (h_\gamma(\nabla \bar{u}), \nabla (v - \bar{u})) = - \sum_{i=1}^{d} \lambda_i \int_\Omega \phi_i'(\bar{u}, f)(v - \bar{u}) \, dx, \; \forall v \in H_0^1(\Omega),$$

Optimization with PDE constraints

In this setting we can prove . . .

- existence of an optimal solution.
- differentiability of solution operator and derivation of sharp optimality system.
- convergence as Huber regularisation $\gamma \to \infty$ to sharp optimality system for non-smooth problem.

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

Optimality system for the regularized probl

There exist Lagrange multipliers $(p_\gamma, \mu) \in H_0^1(\Omega) \times \mathbb{R}^d$ such that

$$\epsilon(Du_\gamma, Dv) + (h_\gamma(Du_\gamma), Dv) + \sum_{i=1}^{d} \int_\Omega \bar{\lambda}_i \, \phi_i'(u_\gamma, f) v \, dx = 0, \forall v \in H_0^1(\Omega),$$

$$\begin{aligned}\epsilon(Dp_\gamma, Dv) &+ (h_\gamma'(Du_\gamma)^* Dp_\gamma, Dv) \\ &+ \sum_{i=1}^{d} \int_\Omega \lambda_i \, \phi_i''(u_\gamma, f) \, p_\gamma \, v \, dx = -2(u_\gamma - u_k, v), \forall v \in H_0^1(\Omega),\end{aligned}$$

$$\mu_i = 2\beta\bar{\lambda}_i + \int_\Omega p_\gamma \phi_i'(u_\gamma, f) \, dx, \qquad i = 1, ..., d,$$

$$\mu_i \geq 0, \; \lambda_i \geq 0, \; \mu_i \lambda_i = 0, \qquad i = 1, ..., d.$$

Optimality system for the regularized probl

There exist Lagrange multipliers $(p_\gamma, \mu) \in H_0^1(\Omega) \times \mathbb{R}^d$ such that

$$\epsilon(Du_\gamma, Dv) + (h_\gamma(Du_\gamma), Dv) + \sum_{i=1}^{d} \int_\Omega \bar{\lambda}_i \, \phi_i'(u_\gamma, f) v \, dx = 0, \forall v \in H_0^1(\Omega),$$

$$\epsilon(Dp_\gamma, Dv) + (h_\gamma'(Du_\gamma)^* Dp_\gamma, Dv) + \sum_{i=1}^{d} \int_\Omega \lambda_i \, \phi_i''(u_\gamma, f) \, p_\gamma \, v \, dx = -2(u_\gamma - u_k, v), \forall v \in H_0^1(\Omega),$$

$$\mu_i = 2\beta\bar{\lambda}_i + \int_\Omega p_\gamma \phi_i'(u_\gamma, f) \, dx, \qquad i = 1, ..., d,$$

$$\mu_i \geq 0, \ \lambda_i \geq 0, \ \mu_i \lambda_i = 0, \qquad i = 1, ..., d.$$

Characterization of the gradient

Optimality system for bilevel problem

Passing to the limit as $\gamma \to \infty$ we are able to derive a sharp optimality system:

$$\epsilon(D\bar{u}, Dv) + (q, Dv) + \sum_{i=1}^{d} \int_\Omega \bar{\lambda}_i \, \phi_i'(\bar{u}) v \, dx = 0, \forall v \in H_0^1(\Omega),$$

$$\langle q, \nabla \bar{u} \rangle_{\mathbb{R}^2} = |\nabla \bar{u}| \qquad \text{a.e. in } \Omega,$$

$$\epsilon(Dp, Dv) + \langle \xi, Dv \rangle_{(\nabla H_0^1(\Omega))'} + \sum_{i=1}^{d} \int_\Omega \bar{\lambda}_i \, \phi_i''(\bar{u}) p \, v \, dx = -2(\bar{u} - u_k, v), \forall v \in H_0^1(\Omega),$$

$$\langle \xi, Dp \rangle_{(\nabla H_0^1(\Omega))'} \geq 0, \langle \xi, D\bar{u} \rangle_{(\nabla H_0^1(\Omega))'} = 0,$$

$$\mu_i = 2\beta \bar{\lambda}_i + \int_\Omega p \phi_i'(\bar{u}, f) \, dx, \qquad i = 1, ..., d,$$

$$\mu_i \geq 0, \ \lambda_i \geq 0, \ \mu_i \lambda_i = 0, \qquad i = 1, ..., d.$$

Numerical strategy

Solve

$$\min_{\lambda_i \geq 0,\ i=1,\dots,d} \|\bar{u}_k - u_k\|^2_{L^2(\Omega)} + \beta \sum_{i=1}^{d} \|\lambda_i\|^2$$

subject to

$$\varepsilon(Du, D(v-u))_{L^2} + \sum_{i=1}^{d} \int_\Omega \lambda_i\, \phi_i'(u, f_k)(v-u)\, dx$$
$$+ \int_\Omega |Dv|\, dx - \int_\Omega |Du|\, dx \geq 0, \forall v \in H_0^1(\Omega),$$

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.

Optimal parameter for Gaussian model

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

subject to:

$$\min_{u \geq 0} \left\{ \frac{\varepsilon}{2} \|Du\|_{L^2}^2 + \|Du\|_\gamma + \frac{\lambda_1}{2} \|u - f_k\|_{L^2}^2 \right\}$$

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

Optimal parameter for Gaussian model

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

subject to:

$$\min_{u \geq 0} \left\{ \frac{\varepsilon}{2} \|Du\|_{L^2}^2 + \|Du\|_\gamma + \frac{\lambda_1}{2} \|u - f_k\|_{L^2}^2 \right\}$$

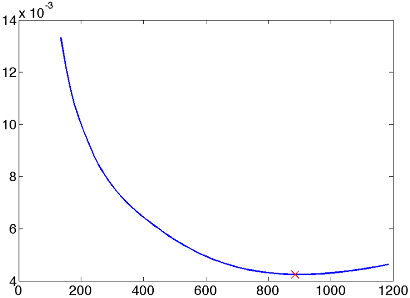
Noise $n \in N(0, 0.02)$ (optimal parameter $\lambda^* = 1770.9$)

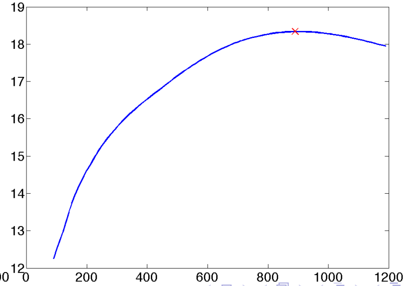
Optimality?

Quality measure

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

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





Mixed Gauss & Poisson noise

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

subject to:

$$\min_{u \geq 0} \left\{ \frac{\varepsilon}{2}\|Du\|_{L^2}^2 + \|Du\|_\gamma + \frac{\lambda_1}{2}\|u - f_k\|_{L^2}^2 + \lambda_2 \int_\Omega (u - f_k \log u)\, dx \right\}.$$

Poisson noise and Gaussian noise $n \in N(0, 0.001)$. Optimal parameters $\lambda_1^* = 1847.75$ and $\lambda_2^* = 73.45$.

Impulse noise

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

subject to:

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

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

And it does not stop there! . . .

Learning the optimal

- fidelity
- operator
- sampling pattern
- (sparse) basis expansion
- the regulariser
- . . .

Outline

Optimising TGV^2 regularisation

Joint work with Juan Carlos De Los Reyes and Tuomo Valkonen

$$TGV^2_{\beta,\alpha}(u) = \sup\left\{\int_\Omega u \operatorname{div}^2 v : \; v \in C^2_c(\Omega, S^{d\times d}), \; \|v\|_\infty \le \beta, \right.$$
$$\left. \|\operatorname{div} v\|_\infty \le \alpha\right\}$$

Regularisation heavily depends on the choice of α and β

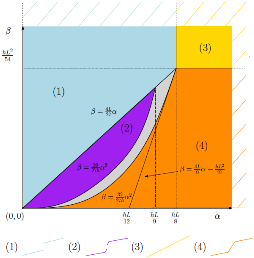


Photo courtesy of Kostas Papafitsoros.

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

Joint with De Los Reyes, Valkonen.

(a) Original image (b) Noisy image

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

(e) TV denoising, $L_1\nabla_\gamma$ cost functional

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

Optimal TV

Optimal TGV^2

Outline

Data learning versus physical modelling?

Data learning versus physical modelling?

Physical modelling	Classical data learning
physical model	non-physical
non-adaptive to data	adaptive to data
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 physical modelling?

Happy marriage of physical modelling and data learning?

Data learning versus physical modelling?

Happy marriage of physical modelling and data learning?

Hope: adaptive physical models.

Conclusions and outlook

Conclusions:

- Optimise physical image and data model by bilevel optimisation.
- Example: Choice of optimal weights of denoising models by optimization approach in function space;
- Preview: TGV optimisation.

Conclusions and outlook

Conclusions:

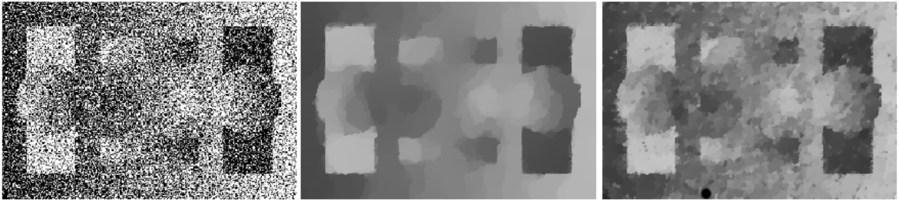
- Optimise physical image and data model by bilevel optimisation.
- Example: Choice of optimal weights of denoising models by optimization approach in function space;
- Preview: TGV optimisation.

Outlook:

- Alternative cost functionals. How to measure optimality?
- More complex (realistic!) noise models.
- General linear operator T (for other applications in MRI, inpainting, segmentation, . . .)
- Spatial dependent weights λ.
- Optimising other elements in the model, e.g. regularisation procedure, acquisition (sampling), . . .

Image quality measures?

Noisy Image | Best SSIM TV image | Best PSNR TV image



Which one do you prefer?[2]

[2] Z. Wang and A. Bovik, *Mean squared error: Love it or leave it? a new look at signal fidelity measures* Signal Processing Magazine, '09.

Conclusions and outlook

Conclusions:

- Optimise physical image and data model by bilevel optimisation.
- Example: Choice of optimal weights of denoising models by optimisation approach in function space;
- Preview: TGV optimisation.

Outlook:

- Alternative cost functionals. How to measure optimality?
- More complex (realistic!) noise models.
- General linear operator T (for other applications in MRI, inpainting, segmentation, ...)
- Spatial dependent weights λ.
- Optimising other elements in the model, e.g. regularisation procedure, acquisition (sampling), ...

Mixed noise models

Example: Gaussian noise on top of Poisson noise

$$f = \underbrace{Tu}_{\text{Poisson distributed}} + \underbrace{n}_{\text{Gaussian noise}}$$

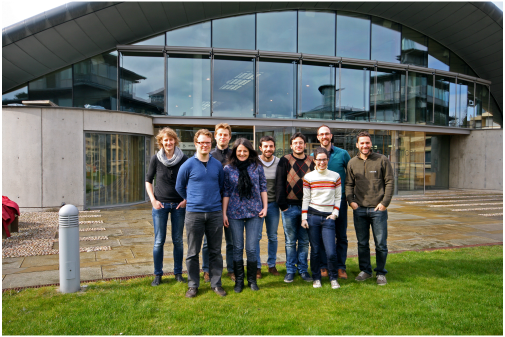
Results into infimal-convolution type fidelity term

$$\min_{u,v} \left\{ R(u) + \frac{1}{2}\|f - v\|_2^2 + \int_\Omega Tu - v \log Tu \, dx \right\}$$

Cambridge Image Analysis

- Dr Martin Benning
- Dr Xiaohao Cai
- Dr Jan Lellmann
- Dr Tuomo Valkonen
- Luca Calatroni
- Milana Gataric
- Rob Hocking
- Juheon Lee
- Kostas Papafitsoros
- Evangelos Papoutsellis
- Joana Grah

Visit us at `http://www.damtp.cam.ac.uk/research/cia/`
and check out our IMAGES network at
`http://www.images.group.cam.ac.uk`



Sponsored by EPSRC, Isaac Newton Trust, LMS, MATLAB and The Royal Society.

And thanks to

Juan Carlos De Los Reyes - Escuela Politécnica National de Quito

http://www.modemat.epn.edu.ec

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*, to appear in IFIP TC7-2013 proceedings.
- J. C. De Los Reyes, and C.-B. Schönlieb, and T. Valkonen, *Optimal parameter learning for higher-order regularisation models*, in preparation.

More information see:
`http://www.damtp.cam.ac.uk/research/cia/`
Email: `cbs31@cam.ac.uk`