Topics in Mathematical Imaging
Lecture 4

Carola-Bibiane Schönlieb

Department for Applied Mathematics and Theoretical Physics
Cantab Capital Institute for the Mathematics of Information
EPSRC Centre for Mathematical Imaging in Healthcare
Alan Turing Institute
University of Cambridge, UK

Spring School SFB 1294, March 2018

Lecture plan

UNIVERSITY OF CAMBRIDGE

- Lecture 1: Variational models & PDEs for imaging by examples
- Lecture 2: Derivation of these models & analysis
- Lecture 3: Numerical solution
- Lecture 4: Some machine learning connections

The variational approach

General task: **restore** $\mathbf{u}$ from an **observed datum** $\mathbf{g}$ where

$$g = \underbrace{Tu}_{\text{forward model}} + \underbrace{n}_{\text{noise}}.$$

Variational approach: Compute u as a minimizer of

$$\mathcal{J}(u) = \alpha \underbrace{R(u)}_{\text{regularization}} + \underbrace{D(Tu, g)}_{\text{data fidelity}} \to \min_{u \in B},$$

where

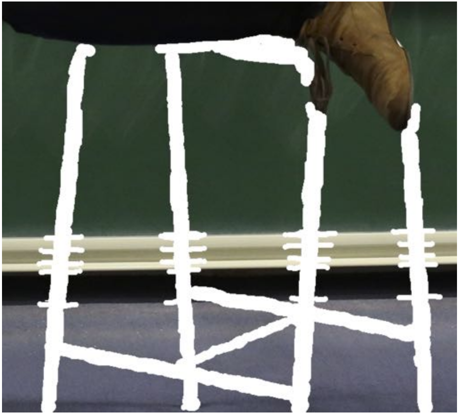
- $R(u)$ is a prior/regularizer that models a-priori information on u weighted by positive α, e.g., $R(u) = \|\nabla u\|_{L^1}$
- $D(\cdot, \cdot)$ is a distance function, e.g. $D(Tu, g) = \|Tu - g\|_2^2$ and B suitable Banach space, e.g., $B = BV(\Omega)$.

Engl, Hanke, Neubauer '96; Natterer, Wübbeling '01; Kaltenbacher, Neubauer, Scherzer '08; Schuster, Kaltenbacher, Hofmann, Kazimierski '12

Which model to choose?

Mathematics can make you fly! J. Grah, K. Papafitsoros, CBS, EPSRC Science Photo Award '14, Burger, He, CBS '09; CBS, Bertozzi '11; CBS, CUP '15; Chan, Shen '01; Bertalmio et al. '00; Masnou, Morel '98.

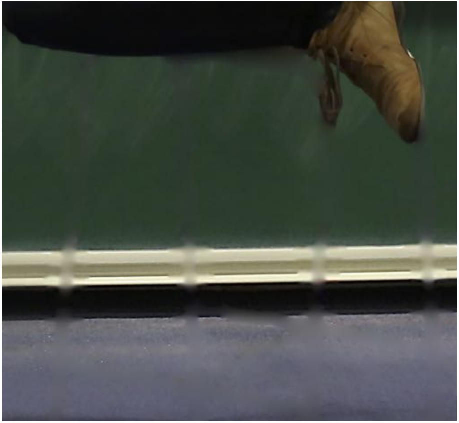
UNIVERSITY OF CAMBRIDGE



Input image

References: Bertalmio, Sapiro, Caselles, Ballester 2000; Telea 2004; Bornemann, Maerz 2007; Burger, He, CBS, SIAM Imaging Science '09; CBS, CUP '15.

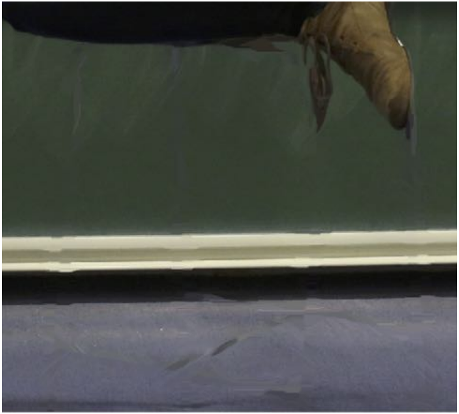
Diffusion versus transport inpainting



Diffusion

References: Bertalmio, Sapiro, Caselles, Ballester 2000; Telea 2004; Bornemann, Maerz 2007; Burger, He, CBS, SIAM Imaging Science '09; CBS, CUP '15.

Diffusion versus transport inpainting



Transport

References: Bertalmio, Sapiro, Caselles, Ballester 2000; Telea 2004; Bornemann, Maerz 2007; Burger, He, CBS, SIAM Imaging Science ’09; CBS, CUP ’15.

How to inpaint?

Image inpainting: create desired inpaintings.

‘Ecce mono’

How to inpaint?

Image inpainting: create desired inpaintings.

'Ecce homo'

'Ecce mono'

Image inpainting: create desired inpaintings.

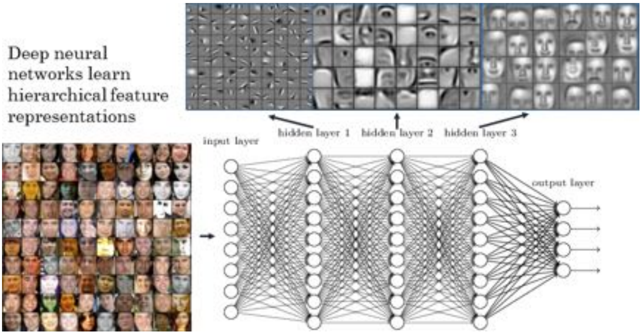
Image courtesy of R. Hocking.

References: Arias, Facciolo, Caselles, Sapiro ’09–

‘Ecce mono’

Deep image processing

UNIVERSITY OF CAMBRIDGE



Picture from strong analytics. LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. Nature, 521(7553), 436-444.

Learning variational models - one 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 α in admissible set $\mathcal{A}$

$$\min_{(R,\phi,\alpha,T)\in\mathcal{A}} \sum_k \mathrm{loss}(\bar{u}_k, u_k)$$

subject to

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

Learning by optimisation in imaging

Some 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; optimal image filters; optimal SVM; optimal reaction-diffusion . . .
- De Los Reyes, CBS '13 –: results on bilevel learning in function space and development of numerical optimisation.
- Fornasier, Naumova, Pereverzyev 14': parameter estimation in multipenalty regularisation.
- Hintermüller et al. '14 – : bilevel optimisation for blind deconvolution, and for adaptive TV denoising.
- Nikolova, Steidl, Weiss '15
- Fonseca, Liu et al. '16 –: bilevel model for higher-order TV type regularisation and Mumford-Shah; analysis in function space . . .

Learning by optimisation in imaging

UNIVERSITY OF CAMBRIDGE

Some 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; optimal image filters; optimal SVM; optimal reaction-diffusion . . .
- De Los Reyes, CBS '13 –: results on bilevel learning in function space and development of numerical optimisation.
- Fornasier, Naumova, Pereverzyev 14': parameter estimation in multipenalty regularisation.
- Hintermüller et al. '14 – : bilevel optimisation for blind deconvolution, and for adaptive TV denoising.
- Nikolova, Steidl, Weiss '15
- Fonseca, Liu et al. '16 –: bilevel model for higher-order TV type regularisation and Mumford-Shah; analysis in function space . . .

Analysis in function space & resolution independent optimisation.

Learning a parametrised model

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

$$\min_{(\lambda,\alpha)\in[0,\infty]^{M+N}} F(u_{\lambda,\alpha})$$

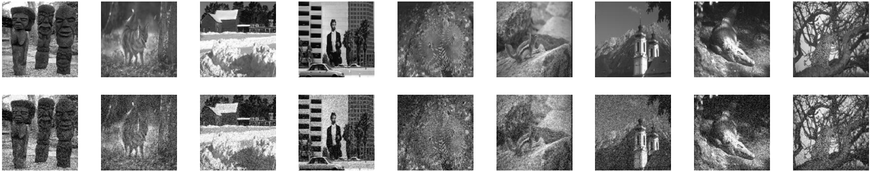
subject to

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

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.

TV regularisation

Cross-validated computations on the Berkeley database split into two halves (100 images each): Total variation regularisation with L^2 cost and fidelity. Noise variance $\sigma = 10$.



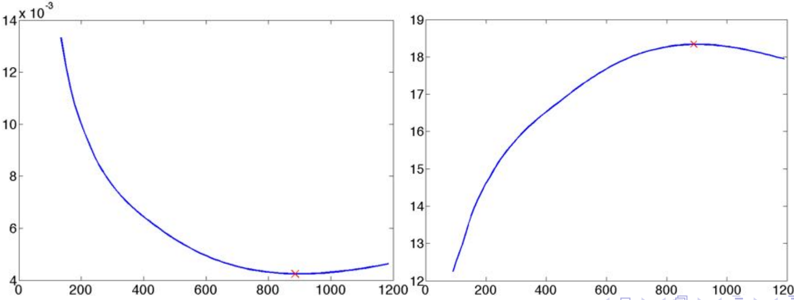
Validation	Learning	α	Average PSNR	Average SSIM
1	1	0.0190	31.3679	0.8885
1	2	0.0190	31.3672	0.8884
2	1	0.0190	31.2619	0.8851
2	2	0.0190	31.2612	0.8850

Parameter optimality?

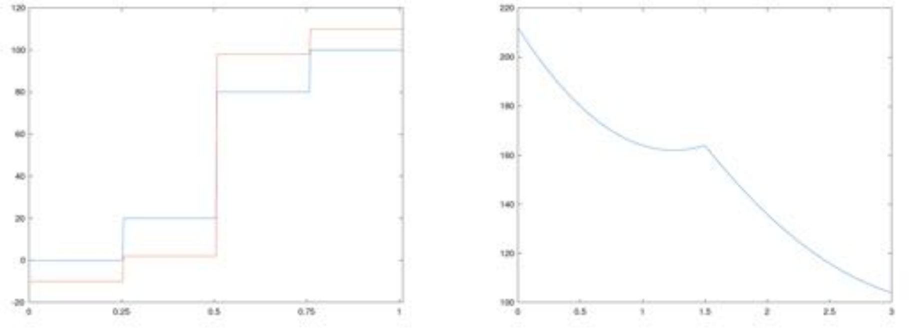
Quality measure

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

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



Parameter optimality?



(A) u_c in blue and u_η in red

(B) $\mathcal{I}(\alpha)$ is not quasi-convex

Courtesy of Pan Liu and Irene Fonseca using Strong, Chan, et al. '96.

And that is not all ...

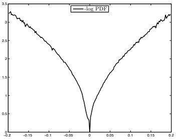
A few more examples of bringing together model-based imaging and learning ...

The nonconvex fields of experts model

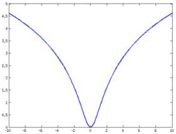
- Let us consider the following nonconvex model [Roth, Black '09], [Samuel, Tappen '09], called the "Fields of Experts" model:

$$\mathcal{R}(u) = \sum_{k=1}^{q} \sum_{i,j=1}^{m,n} \rho_k((K_k u)_{i,j})$$

- $\{K_k\}$ are arbitrary filter kernels, and $\{\rho_k\}$ are potential functions
- Has much more parameters compared to the ℓ_1 model (several thousands)
- Allows only to compute a stationary point (local minimum)
- Suitable potential functions ρ_l are derived from statistics of natural images [Huang and Mumford '99]:



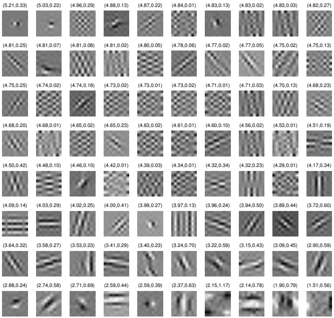
$$\rho_k(t) = \alpha_k \log(1 + \beta_k t^2)$$

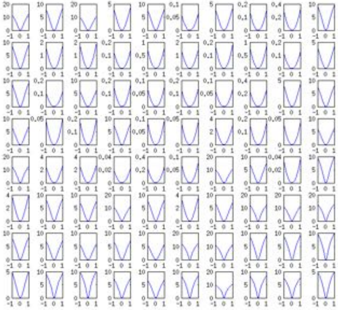


UNIVERSITY OF CAMBRIDGE

The learned filters and functions

- In [Chen, Ranftl, P. '14] we learned 80 filters of size 9×9 plus function parameters $\rightarrow$ 6480 parameters on a database of $\sim$ 200 images
- … two weeks later …





UNIVERSITY OF CAMBRIDGE

Evaluation

- Comparison with five state-of-the-art approaches: K-SVD [Elad and Aharon '06], FoE [Q. Gao and Roth '12], BM3D [Dabov et al. '07], GMM [D. Zoran et al. '12], LSSC [Mairal et al. '09]
- We report the average PSNR on 68 images of the Berkeley image data base [Chen, P. 14]

σ	KSVD	FoE	BM3D	GMM	LSSC	BL7x7	BL9x9
15	30.87	30.99	31.08	31.19	**31.27**	31.18	31.22
25	28.28	28.40	28.56	28.68	**28.70**	28.66	**28.70**
50	25.17	25.35	25.62	25.67	25.72	25.70	**25.76**

- Performs equally or better as the state-of-the-art

Variational networks

- Inspired by the conditional shrinkage fields (CSF) [Schmidt, Roth '14], we allow to change the parameters during the iterations:

$$\begin{cases} u^0 = f \\ u^{t+1} = u^t - \lambda^t \left(\sum_{k=1}^{q} (K_k^t)^\top (\rho_k^t)'(K_k^t u^t) + (u^t - f) \right), \quad t = 0 \dots T-1 \end{cases}$$

- In each step we perform one gradient descent on a learned variational energy
- Can be interpreted as one cycle of a block incremental gradient descent
- Can also be interpreted as learned non-linear diffusion, trying to "invert" the convolution $\int p(f|u)p(u)\mathrm{d}u$
- And it can be interpreted as a convolutional neural network with T layers

UNIVERSITY OF CAMBRIDGE

Quantitative evaluation

- We evaluated our learned models on a standard database of 68 images

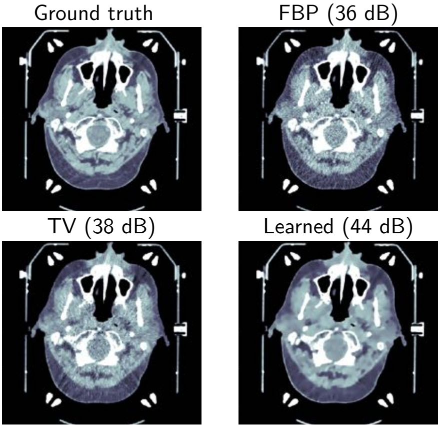
Method	σ = 15	σ = 25	St.	$\sigma = 15$: $TRD_{5\times5}$	$\sigma = 15$: $TRD_{7\times7}$
BM3D	31.08	28.56	2	31.14	31.30
LSSC	31.27	28.70	5	31.30	**31.42**
EPLL	31.19	28.68	8	31.34	**31.43**
opt-MRF	31.18	28.66		$\sigma = 25$: $TRD_{5\times5}$	$\sigma = 25$: $TRD_{7\times7}$
RTF_5	–	28.75			
WNNM	31.37	28.83	2	28.58	28.77
$CSF^5_{5\times5}$	31.14	28.60	5	28.78	**28.91**
$CSF^5_{7\times7}$	31.24	28.72	8	28.83	**28.95**

Learning to reconstruct

- Variational regularization: Iterative schemes
- Learned operators
- Data in $\rightarrow$ reconstruction out

Algorithm 1 Learned Gradient

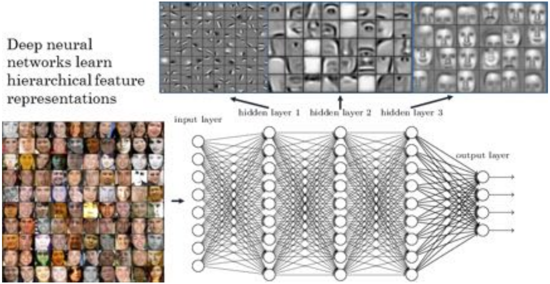
for $i = 1, \ldots$ **do**
  $\Delta f_i \leftarrow \Lambda_{\Theta}\big(f_i, \nabla\big[\mathcal{L}(\mathcal{T}(\cdot), g)\big](f_{i-1})\big)$
  $f_i \leftarrow f_{i-1} + \Delta f_i$



J. Adler and O. Öktem, *Solving ill-posed inverse problems using iterative deep neural networks*, to appear in Inverse Problems '17. See also M. Unser et al. 2017 forward

Deep image processing

UNIVERSITY OF CAMBRIDGE



This is unfeasible for many ill-posed inverse imaging problems

Picture from strong analytics. LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. Nature, 521(7553), 436-444.

Lecture plan

- Lecture 1: Variational models & PDEs for imaging by examples
- Lecture 2: Derivation of these models & analysis
- Lecture 3: Numerical solution
- Lecture 4: Some machine learning connections

Thank you very much for your attention!

More information see:

http://www.ccimi.maths.cam.ac.uk

http://www.cmih.maths.cam.ac.uk

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

Email: cbs31@cam.ac.uk