Learning variational models in imaging by bilevel and quotient optimisation

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

Dictionary learning on manifolds, Nice, 6 September 2017
Outline

1. Variational regularisation for imaging
2. Learning of variational models in imaging
   - Bilevel learning
   - Quotient learning & nonlinear eigenproblems
3. Conclusions and outlook
1. Variational regularisation for imaging

2. Learning of variational models in imaging
   - Bilevel learning
   - Quotient learning & nonlinear eigenproblems

3. Conclusions and outlook
The variational approach

General task: **restore** \( u \) from an **observed datum** \( g \) where

\[
g = \underbrace{Tu} + \underbrace{n}. \tag{1}
\]

\( T \) is the forward model and \( n \) is the noise.

**Variational approach:** Compute \( u \) as a minimizer of

\[
\mathcal{J}(u) = \alpha \underbrace{R(u)} + \underbrace{D(Tu, g)} \rightarrow \min_{u \in B},
\]

where

- \( R(u) \) is a prior/regularizer that models a-priori information on \( u \) weighted by positive \( \alpha \), 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., \( B = BV(\Omega) \).
Sparsity promoting regularisation

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

\[ g = (F u)_{|\Lambda} + n \]

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

Sparsity in \( \nabla u \) \( \Rightarrow \) NP-hard problem!

Convex relaxation \( \Rightarrow \ell_1 \)- and total variation minimization

\[ \alpha \| \nabla u \|_1 + \frac{1}{2} \| (F u)_{|\Lambda} - g \|^2 \rightarrow \min_u \]

Rudin, Osher, Fatemi, Physica D '92; Candes, Romberg, Tao, IEEE Trans Inf Theory '06
Variational regularisation for imaging

Shepp-Logan Phantom
Variational regularisation for imaging

Sparsity promoting regularisation

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

\[ g = (\mathcal{F}u)|_{\Lambda} + n \]

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

Sparsity in \( \nabla u \) \( \Rightarrow \) NP-hard problem!

Convex relaxation \( \Rightarrow \ell_1 \)- and total variation minimization

\[
\alpha \|\nabla u\|_1 + \frac{1}{2} \| (\mathcal{F}u)|_{\Lambda} - g \|^2 \rightarrow \min_u,
\]

Rudin, Osher, Fatemi, Physica D '92; Candes, Romberg, Tao, IEEE Trans Inf Theory '06
4D reconstruction from sub-sampled data

4D Bregman-TV Reconstruction from 20% line-sampling

Acquire $SF(u) \Rightarrow$ reconstruct $u(x,t) = \arg\min_v \alpha \|\nabla v\|_1 + \|SF(v) - g\|_2^2$

4D reconstruction from sub-sampled data

Fourier inversion from 20% line-sampling

Acquire $S\mathcal{F}(u) \Rightarrow$ reconstruct $u(x, t)$

M. Benning, A. Sederman, CBS, L. Gladden, et al.
4D Bregman-TV Reconstruction from 20% line-sampling

Acquire $\mathcal{S}\mathcal{F}(u)$ ⇒ reconstruct $u(x, t) = \arg\min_v \alpha \|\nabla v\|_1 + \|\mathcal{S}\mathcal{F}(v) - g\|_2^2$

What is the right sparsity?

\[
\min_u \left\{ \alpha \| \nabla u \|_1 + \| u - f \|_2^2 \right\}
\]

Noisy image

TV denoised image

Image courtesy of K. Papafitsoros

References: Rudin, Osher, Fatemi '92; Hinterberger, Scherzer, Computing '06; Bredies, Kunisch, Pock, SIAM Imaging '10; Papafitsoros, CBS, J. Math. Imaging & Vision, '13
What is the right sparsity?

\[
\min_u \left\{ \alpha \| \nabla^2 u \|_1 + \| u - f \|_2^2 \right\}
\]

Image courtesy of K. Papafitsoros

References: Rudin, Osher, Fatemi ’92; Hinterberger, Scherzer, Computing ’06; Bredies, Kunisch, Pock, SIAM Imaging ’10; Papafitsoros, CBS, J. Math. Imaging & Vision, ’13
Variational regularisation for imaging

What is the right sparsity?

\[
\min_u \left\{ \min_w \{ \alpha_1 \|\nabla u - w\|_1 + \alpha_2 \|Ew\| \} + \|u - f\|^2_2 \right\}
\]

Noisy image  
TGV\(^2\) denoised image

Image courtesy of K. Papafitsoros

Variational regularisation for imaging

Learning with structure


$$\min_{\lambda} \text{loss}(u(\lambda))$$

s.t. $u(\lambda) = \arg\min_v J(v, \lambda)$
Outline

1  Variational regularisation for imaging

2  Learning of variational models in imaging
   • Bilevel learning
   • Quotient learning & nonlinear eigenproblems

3  Conclusions and outlook
Outline

1. Variational regularisation for imaging

2. Learning of variational models in imaging
   - Bilevel learning
   - Quotient learning & nonlinear eigenproblems

3. Conclusions and outlook
Bilevel learning

Joint work with Juan Carlos De Los Reyes, Martin Benning, Luca Calatroni, Matthias Ehrhardt, Georg Maierhofer, Ferdia Sherry, Tuomo Valkonen . . .
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 \(\phi\), and \(\alpha\) in admissible set \(A\)

\[
\min_{(R, \phi, \alpha, T) \in A} \sum_k \text{Cost}(\bar{u}_k, u_k)
\]

subject to

\[
\bar{u}_k = \arg\min_u \left\{ \alpha R(u) + \int_{\Omega} \phi(Tu, f_k) \, dx \right\}
\]
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; $\ell_1$ 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.
- **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 . . .
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; $\ell_1$ 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.
- **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 of variational models in imaging

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 $T : X \rightarrow Y \subset L^1(\Omega; \mathbb{R}^d)$ with $X, Y$ Banach spaces, $A_j : X \rightarrow \mathcal{M}(\Omega; \mathbb{R}^{m_j})$, $(j = 1, \ldots, N)$ are appropriate linear operators, $|A_j u|$ total variation measure, $F$ is cost function.
Cost functions: For noise free data \( \tilde{u} \) we take either \textbf{PSNR}

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

or \textbf{Huberised TV cost}

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

For a training set, take average, e.g.

\[ F_{L^2}(v_1, \ldots, v_K) = \sum_{i=1}^{K} \| v_i - \tilde{u}_i \|_2^2. \]
Theorem 1: **Existence** of an optimal solution for non-smooth problem (under appropriate assumptions optimal parameter lies in the interior!) ✓

Theorem 2: **Existence and continuity** of optimal solution for smoothed problem with Huber-regularisation & elliptic term (Hilbert space) ✓

Theorem 3: **Derivation of optimality system and gradient of solution map** for smoothed problem ✓

Theorem 4: **Convergence of solutions** of regularised problem to non-smooth problem in Hilbert space ✓

Learning of variational models in imaging

Numerical strategy

Solve

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

subject to

$$u_{\lambda, \alpha} = \arg\min_u$$

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

by quasi-Newton method (BFGS)

- state equation is solved by Newton type algorithm (varies with $\phi$)
- 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 à la Byrd et al.

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

<table>
<thead>
<tr>
<th>Validation</th>
<th>Learning</th>
<th>$\alpha$</th>
<th>Average PSNR</th>
<th>Average SSIM</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>0.0190</td>
<td>31.3679</td>
<td>0.8885</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>0.0190</td>
<td>31.3672</td>
<td>0.8884</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>0.0190</td>
<td>31.2619</td>
<td>0.8851</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>0.0190</td>
<td>31.2612</td>
<td>0.8850</td>
</tr>
</tbody>
</table>
Parameter optimality?

**Quality measure**

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

\[
SNR = 20 \times \log_{10} \left( \frac{\| u_k \|_{L^2}}{\| u - u_k \|_{L^2}} \right),
\]
Cross-validated computations on the Berkeley database split into two halves (100 images each):

TGV$^2$ regularisation with $L^2$ cost and fidelity. Noise variance $\sigma = 10$.

<table>
<thead>
<tr>
<th>Validation</th>
<th>Learning</th>
<th>$(\beta, \alpha)$</th>
<th>Average PSNR</th>
<th>Average SSIM</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>$(0.0187, 0.0198)$</td>
<td>31.4325</td>
<td>0.8901</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>$(0.0186, 0.0191)$</td>
<td>31.4303</td>
<td>0.8899</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>$(0.0186, 0.0191)$</td>
<td>31.3281</td>
<td>0.8869</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>$(0.0187, 0.0198)$</td>
<td>31.3301</td>
<td>0.8870</td>
</tr>
</tbody>
</table>
Learning $(\beta, \alpha)$ in $\text{TGV}^2_{\beta,\alpha}$
Learning of variational models in imaging

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

$L^1 \nabla_\gamma \text{ cost} \quad (\alpha, \beta) = (0.069/n^2, 0.051/n)$

$L^2 \text{ cost} \quad (\alpha, \beta) = (0.058/n^2, 0.041/n)$

Schönlieb (DAMTP)
Optimal $(\beta, \alpha)$ in $\text{TGV}^2_{\beta,\alpha}$?

For TGV a good initialisation is important!
Learning of variational models in imaging

Mixed Impulse & Gaussian noise

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

where \( u \) is the solution of the optimisation problem:

\[
\min_{\begin{subarray}{l} v \in BV \\ n \in L^2 \end{subarray}} \left\{ \frac{\mu}{2} \| \nabla v \|_{L^2}^2 + \| Dv \|_\gamma + \lambda_1 \| n \|_\gamma + \lambda_2 \| f - v - n \|_{L^2}^2 \right\},
\]

Original image (left) and noisy image (right) corrupted by impulse noise.
From left to right: Denoised image, impulse noise residuum and Gaussian noise residuum. Optimal parameters: $\lambda_1^* = 351.23$ and $\lambda_2^* = 5200.1$. 
Optimal MR sampling

Question
What is an ideal (sparse) sampling pattern for a given (2d)-object?

Mathematical formulation
For \( p = (s, \alpha) \in [0, 1]^m \times \mathbb{R}_{\geq 0}, \)

\[
\min_{p} \frac{1}{2} \| u(p) - \hat{u} \|_2^2 + \beta \| s \|_1 + \tilde{c} \alpha , \text{ subject to }
\]

\[
u(p) \in \arg \min_{u \in \mathbb{R}_\geq 0^m} \left\{ \frac{1}{2} \sum_{j=1}^{m} s_j |(F u)_j - \tilde{y}_j|^2 + \alpha \| \nabla u \|_1 + \frac{\epsilon}{2} \| u \|_{H^1}^2 \right\}
\]

where \( \hat{u} \in \mathbb{R}_\geq 0^m \) is some ground truth, and \( \epsilon, \tilde{c} \ll \alpha, \beta. \)

M. Benning, M. Ehrhardt, G. Maierhofer, F. Sherry, CBS, 2017
Consider $\tilde{f} = f + \epsilon$

$$u^* \in \arg \min_{u \in \mathbb{R}^m_{\geq 0}} \frac{1}{2} \| S\mathcal{F}u - S\tilde{f} \|_2^2 + \text{TV}(u)$$

where $S : \mathbb{C}^m \rightarrow \mathbb{C}^m$ is a linear sampling operator.
Consider $\tilde{f} = f + \epsilon$

$$u^* \in \arg \min_{u \in \mathbb{R}^m_{\geq 0}} \frac{1}{2} \|S\mathcal{F}u - S\tilde{f}\|_2^2 + TV(u)$$

where $S : \mathbb{C}^m \rightarrow \mathbb{C}^m$ is a linear sampling operator.

(c) Sampling pattern    (d) Reconstructed image
Optimal MR sampling

Consider $\tilde{f} = f + \epsilon$

$$u^* \in \arg\min_{u \in \mathbb{R}^m_{\geq 0}} \frac{1}{2} ||SFu - S\tilde{f}||^2_2 + TV(u)$$

where $S : \mathbb{C}^m \rightarrow \mathbb{C}^m$ is a linear sampling operator.

Compressed sensing theory: Adcock, Hansen et al. 11--; R. Ward et al. 12--; P. Weiss et al. 13--
Learning of variational models in imaging

Optimal MR sampling

Figure: Discrete 2d bump

M. Benning, M. Ehrhardt, G. Maierhofer, F. Sherry, CBS, 2017
Optimal MR sampling

(a) Original data: $\log |y|$

(b) Noisy data: $\log |\tilde{y}|$

M. Benning, M. Ehrhardt, G. Maierhofer, F. Sherry, CBS, 2017
Optimal MR sampling

(c) Learned sampling pattern  (d) Largest 2.76% Fourier Coefficients

M. Benning, M. Ehrhardt, G. Maierhofer, F. Sherry, CBS, 2017
Optimal MR sampling

(e) Learned sampling pattern

(f) Largest 2.76% Fourier Coefficients

M. Benning, M. Ehrhardt, G. Maierhofer, F. Sherry, CBS, 2017
Optimal MR sampling

(g) Learned sampling pattern
(h) Largest 2.76% Fourier Coefficients

M. Benning, M. Ehrhardt, G. Maierhofer, F. Sherry, CBS, 2017
Optimal MR sampling

(i) Original image

(j) Learned sampling pattern vs. FC

M. Benning, M. Ehrhardt, G. Maierhofer, F. Sherry, CBS, 2017
Optimal MR sampling

(k) Learned sampling pattern

(\textit{Similarity}_a) = 57.8690

(l) Largest 6.5\% Fourier Coefficients

(\textit{Similarity}_b) = 58.5188

M. Benning, M. Ehrhardt, G. Maierhofer, F. Sherry, CBS, 2017
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 \( \ell_1 \) model (several thousands)
- Allows only to compute a stationary point (local minimum)
- Suitable potential functions \( \rho_1 \) are derived from statistics of natural images [Huang and Mumford '99]:

\[ \rho_k(t) = \alpha_k \log(1 + \beta_k t^2) \]
The learned filters and functions

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

<table>
<thead>
<tr>
<th>$\sigma$</th>
<th>KSVD</th>
<th>FoE</th>
<th>BM3D</th>
<th>GMM</th>
<th>LSSC</th>
<th>BL7x7</th>
<th>BL9x9</th>
</tr>
</thead>
<tbody>
<tr>
<td>15</td>
<td>30.87</td>
<td>30.99</td>
<td>31.08</td>
<td>31.19</td>
<td><strong>31.27</strong></td>
<td>31.18</td>
<td>31.22</td>
</tr>
<tr>
<td>50</td>
<td>25.17</td>
<td>25.35</td>
<td>25.62</td>
<td>25.67</td>
<td>25.72</td>
<td>25.70</td>
<td><strong>25.76</strong></td>
</tr>
</tbody>
</table>

- 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:

\[
\begin{align*}
    u^0 &= f \\
    u^{t+1} &= u^t - \lambda^t \left( \sum_{k=1}^{q} (K_k^t)^\top (\rho_k^t)'(K_k^t u^t) + (u^t - f) \right), \ t = 0 \ldots T - 1
\end{align*}
\]

- 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 \( \int p(f|u)p(u)du \)

- And it can be interpreted as a convolutional neural network with \( T \) layers
Quantitative evaluation

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

<table>
<thead>
<tr>
<th>Method</th>
<th>$\sigma$ = 15</th>
<th>$\sigma$ = 25</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\sigma$ = 15</td>
<td>$\sigma$ = 25</td>
</tr>
<tr>
<td>BM3D</td>
<td>31.08</td>
<td>28.56</td>
</tr>
<tr>
<td>LSSC</td>
<td>31.27</td>
<td>28.70</td>
</tr>
<tr>
<td>EPLL</td>
<td>31.19</td>
<td>28.68</td>
</tr>
<tr>
<td>opt-MRF</td>
<td>31.18</td>
<td>28.66</td>
</tr>
<tr>
<td>RTF$_5$</td>
<td>-</td>
<td>28.75</td>
</tr>
<tr>
<td>WNNM</td>
<td>31.37</td>
<td>28.83</td>
</tr>
<tr>
<td>CSF$_{5 \times 5}$</td>
<td>31.14</td>
<td>28.60</td>
</tr>
<tr>
<td>CSF$_{7 \times 7}$</td>
<td>31.24</td>
<td>28.72</td>
</tr>
</tbody>
</table>
Learning to reconstruct

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

**Algorithm 1** Learned Gradient

1. for $i = 1, \ldots$ do
2. $\Delta f_i \leftarrow \Lambda(\Theta, \nabla L(\mathcal{T}(\cdot), g))(f_{i-1})$
3. $f_i \leftarrow f_{i-1} + \Delta f_i$

See also M. Unser et al. 2017
Learning to reconstruct

- Variational regularization:
  - Iterative schemes
- Learned operators
- Data in → reconstruction out

**Algorithm 2** Learned Primal-Dual

1: for $i = 1, \ldots, l$ do
2: $h_i \leftarrow \Gamma_{\Theta_i}^q(h_{i-1}, \mathcal{T}(\bar{f}_{i-1}), g)$
3: $f_i, \bar{f}_i \leftarrow \Lambda_{\Theta_i}^p(f_{i-1}, \mathcal{T}^*(h_i))$
4: $\mathcal{T}_{\Theta}^\dagger(g) \leftarrow f_i^{(1)}$
Outline

1. Variational regularisation for imaging

2. Learning of variational models in imaging
   - Bilevel learning
   - Quotient learning & nonlinear eigenproblems

3. Conclusions and outlook
Learning of variational models in imaging

Learning from negative examples

Joint work with Martin Benning, Guy Gilboa and Joana Grah

Motivation

**Idea:** Learning parametrised regularisation functions by *quotient minimisation*, where both wanted and unwanted outcomes are incorporated in the model by integrating the former in the numerator and the latter in the denominator.

\[
\hat{h} \in \arg \min_{\|h\|_2=1} \frac{J(u^+; h)}{J(u^-; h)}, \quad J(u; h) := \|u \ast h\|_1,
\]

**Aim:** Learn convolution kernel \( h \) that makes \( u^+ \) sparse! Hope that \( u^+ \) can be represented by just a few building components, cf. works on dictionary learning, e.g.

Eigenproblems

**Linear Eigenproblem**

\[ Af - \lambda f = 0 \]

\[ F(f) = \frac{\langle f, Af \rangle}{\| f \|_2^2} \]

**Non-linear Eigenproblem**

\[ \partial R(f) - \lambda \partial S(f) \ni 0 \]

assuming \( R, S : \mathbb{R}^n \rightarrow \mathbb{R}_{\geq 0} \) convex, Lipschitz continuous, positively one-homogeneous and

\[ S(f) = 0 \Leftrightarrow f = 0 \]

\[ F(f) = \frac{R(f)}{S(f)} \]
Methods to Solve Eigenproblems

Linear case: Inverse Power Method

\[ Af^{k+1} = f^k \]
\[ \iff Af^{k+1} - f^k = 0 \]

\[ f^{k+1} = \arg \min_u \frac{1}{2} \langle u, Au \rangle - \langle u, f^k \rangle \]

Non-linear case: Generalised Inverse Power Method

\[ r(f^{k+1}) - s(f^k) = 0 \]
\[ r(f^{k+1}) \in \partial R(f^{k+1}), \ s(f^k) \in \partial S(f^k) \]

Optimisation problem:

\[ f^{k+1} = \arg \min_u R(u) - \langle u, s(f^k) \rangle \]
Algorithm: Generalised Inverse Power Method

Initialisation: Randomly generalised $f^0$ with $\|f^0\|_2 = 1$

while $|\lambda^{k+1} - \lambda^k| < \varepsilon$ do

$$f^{k+1} = \arg \min \left\{ R(u) - \lambda^k (u, s(f^k)) \right\}, \quad s(f^k) \in \partial S(f^k)$$

$$\lambda^{k+1} = R(f^{k+1}) / S(f^{k+1})$$

end while

Output: Eigenvalue $\lambda^{k+1}$, Eigenfunction $f^{k+1}$

Algorithm: Regularisation Learning Model

\[ \hat{h} \in \arg \min_{\|h\|_2=1} \frac{J(u^+;h)}{J(u^-;h)} \]

\[
\begin{align*}
    h^{k+\frac{1}{2}} &= \arg \min \left\{ J(u^+;h) - \mu^k \langle h, p^k \rangle \right\} \\
    h^{k+1} &= \frac{h^{k+\frac{1}{2}}}{\| h^{k+\frac{1}{2}} \|_2} \\
    p^{k+1} &\in \partial J(u^-;h^{k+1}) \\
    \mu^{k+1} &= \frac{J(u^+;h^{k+1})}{J(u^-;h^{k+1})}
\end{align*}
\]
Infimal Convolution Model

Learning

\[
\min_{h_i} \ \frac{1}{n-1} \sum_{j \neq i} \| h_i * u_j \|_1, \quad i = 1, \ldots, n
\]
Learning of variational models in imaging

Infimal Convolution Model

Learning

\[
\min_{h_i} \frac{1}{n-1} \sum_{j \neq i} \| h_i * u_j \|_1, \quad i = 1, \ldots, n
\]

Reconstruction

\[
\inf_{u = u_1 + \cdots + u_n} \sum_{j=1}^{n} \| h_j * u_j \|_1
\]
Infimal Convolution Model

Additive test image

Learning

ground truth

Reconstruction - without noise

Learning variational models

Nice - 06/09/2017
Infimal Convolution Model

Additive test image: Diagonal stripes (noise-free)
To conclude this talk we want to discuss the incorporation of misfit data in the context of a basic digit classification task.

\[
\begin{array}{cccccccc}
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
2 & 2 & 2 & 2 & 2 & 2 & 2 & 2 \\
3 & 3 & 3 & 3 & 3 & 3 & 3 & 3 \\
4 & 4 & 4 & 4 & 4 & 4 & 4 & 4 \\
5 & 5 & 5 & 5 & 5 & 5 & 5 & 5 \\
6 & 6 & 6 & 6 & 6 & 6 & 6 & 6 \\
7 & 7 & 7 & 7 & 7 & 7 & 7 & 7 \\
8 & 8 & 8 & 8 & 8 & 8 & 8 & 8 \\
9 & 9 & 9 & 9 & 9 & 9 & 9 & 9 \\
\end{array}
\]
**MNIST classification**

**Setup:** we pick two digit classes, let’s say 0 and 3, and learn the four filter functions

\[
h_0 = \arg \min_{\|h\|_2 = 1} \|h \ast U_0\|_1 \quad \text{and} \quad h_3 = \arg \min_{\|h\|_2 = 1} \|h \ast U_3\|_1
\]

as well as

\[
\tilde{h}_0 = \arg \min_{\|h\|_2 = 1} \frac{\|h \ast U_0\|_1}{\|h \ast U_3\|_1} \quad \text{and} \quad \tilde{h}_3 = \arg \min_{\|h\|_2 = 1} \frac{\|h \ast U_3\|_1}{\|h \ast U_0\|_1}.
\]

Here \(U_0 \in \mathbb{R}^{784 \times 5923}\) and \(U_3 \in \mathbb{R}^{784 \times 6131}\) are training-data matrices that contain images of hand-written zeros, respectively threes, only.
MNIST classification

Classification results:

\[
\min \left( \| h_0 \ast u_{\text{test}} \|_1, \| h_3 \ast u_{\text{test}} \|_1 \right) \rightarrow 77.7387 \% \text{ success rate}
\]

\[
\min \left( \| \tilde{h}_0 \ast u_{\text{test}} \|_1, \| \tilde{h}_3 \ast u_{\text{test}} \|_1 \right) \rightarrow 95.1759 \% \text{ success rate}
\]
Conclusions and outlook

1. Variational regularisation for imaging

2. Learning of variational models in imaging
   - Bilevel learning
   - Quotient learning & nonlinear eigenproblems

3. Conclusions and outlook
## Learning versus variational modelling?

<table>
<thead>
<tr>
<th>Variational modelling</th>
<th>Recent data learning</th>
</tr>
</thead>
<tbody>
<tr>
<td>physical model</td>
<td>non-physical</td>
</tr>
<tr>
<td>high generalizability</td>
<td>very limited generalizability</td>
</tr>
<tr>
<td>limited adaptivity to data</td>
<td>adaptive to data</td>
</tr>
<tr>
<td>insight in structure of problem</td>
<td>blackbox</td>
</tr>
<tr>
<td>reconstruction guarantees</td>
<td>often no guarantees</td>
</tr>
<tr>
<td>stability, error analysis, ...</td>
<td>guarantees stability?</td>
</tr>
<tr>
<td>heavily relies on a-priori model</td>
<td>learns the model from the data</td>
</tr>
</tbody>
</table>
Happy marriage of variational modelling and data learning?

Hope: adaptive, physical models with theoretical guarantees.
Take away messages

- Variational regularisation models in imaging – rigorous and generalisable tools in imaging
- Learning of variational models – towards rigorous, generalisable and adaptive imaging models
- The aim: learning structured but adaptive models with guarantees.

There are still many things to do: richer parametrisation, multi-level & multi-scale; better optimisation schemes for non-convex and high-dimensional learning problems with stronger guarantees and more efficiency.
Conclusions and outlook

Cambridge Image Analysis

- Dr Martin Benning
- Dr Matthias Ehrhardt
- Dr Lukas Lang
- Dr Jingwei Liang
- Veronica Corona
- Joana Grah
- Rob Hocking
- Lisa Kreusser
- Simone Parisotto
- Erlend Riis
- Ferdia Sherry
- Rob Tovey
- Jon Williams

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

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