Data-driven approaches to inverse problems

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
University of Cambridge, UK

Solving Inverse Problems with Deep Learning - Autumn School, September 2023

UNIVERSITY OF CAMBRIDGE

CCIMI
CANTAB CAPITAL INSTITUTE FOR THE MATHEMATICS OF INFORMATION

m4DL

Cambridge Image Analysis

- Dr Angelica Aviles-Rivero
- Dr Debmita Bandyopadhyay
- Dr Anna Breger
- Dr Sören Dittmer
- Dr Carlos Esteve-Yagüe
- Dr Julian Gilbey
- Dr Yury Korolev
- Dr Chao Li
- Dr Subhadip Mukherjee
- Dr Malena Sabate Landman
- Dr Mike Roberts
- Dr Ferdia Sherry
- Dr Junqi Tang
- Georgios Batzolis
- Willem Diepeveen
- Tamara Grossmann
- Lihao Liu
- Christina Runkel
- Ben Schreiber
- Jan Stanczuk
- Hong Ye Tan

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

Sponsored by EPSRC, The Leverhulme Trust, The Royal Society, Cantab Capital Partners, NPL, Alan Turing Institute, Wellcome Trust, Horizon2020, Unilever, Aviva, GSK.

Advanced mathematical methods based on partial differential equations, variational models, optimisation and machine learning

Inverse imaging, image analysis and processing:

- Image reconstruction from indirect measurements
- Image restoration
- Image segmentation and classification
- Motion estimation and object tracking

Real-world applications:

- Photography
- Remote sensing
- Biomedical imaging (MRI, PET/SPECT, microscopy imaging)
- Arts restoration
- Forensics
- ...

Two paradigms in inverse imaging

Knowledge driven imaging, based on mathematical models & physical principles, e.g. non-linear PDEs

$$u_t = \mathrm{div}\left(\frac{\nabla u}{|\nabla u|}\right),$$

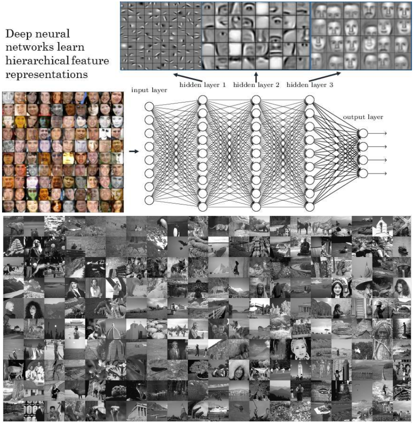
variational models

$$\min_\xi \{\int |D\xi| + \int \xi(c_1 - f)^2 + \int (1-\xi)(c_2 - f)^2\},$$

multi-resolution/sparsity

$$u \approx S u_\Lambda = \sum_{\lambda\in\Lambda} u_\lambda \psi_\lambda.$$

Data driven imaging, based on overparametrised models, e.g. support vector machines or neural networks.



Lecture plan

- Lecture 1: Variational models & PDEs for inverse imaging problems
- Lecture 2: Learned variational regularisers & plug-and-play denoising
- Lecture 3: Learned iterative reconstruction & perspectives.

Based on: Arridge, Maass, Öktem, CBS, Acta Numerica '19

An inverse problem example

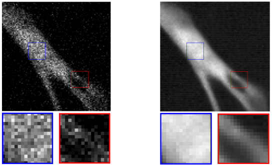
Making stuff disappear using applied maths - collaboration with Samuli Siltanen and Simone Parisotto.

Problems we want to solve

UNIVERSITY OF CAMBRIDGE

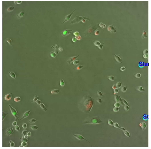
The common goal: Extract information from large, diverse and corrupted imaging data in an automated and robust way.

Image denoising



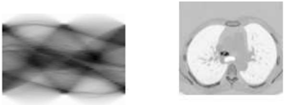
Given noisy $y = u + n$, compute denoised u^*

Image segmentation



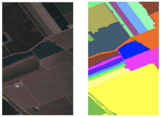
Given u on Ω, compute χ_S, $S \subset \Omega$.

Image reconstruction



Compute u from $y = T(u) + n$.

Image classification



Given images u_i assign labels y_i.

Ke, CBS, IEEE Transactions, PAMI '21; Grah, Burger, Reichelt, CBS et al. Methods 115, 15 February 2017; Benning, Burger, Acta Numerica '18; Arridge, Maass, Öktem, CBS, Acta Numerica '19; Aviles-Rivero, Papadakis, Li, Sellars, Fan, Tan, CBS, MICCAI '19, **Software:** `github.com/JoanaGrah/MitosisAnalyser`

Mathematical approaches

$$u_t = \mathrm{div}\left(\frac{\nabla u}{|\nabla u|}\right)$$

Problem formulation

Given measurements $\mathbf{y} = (y_1, \ldots, y_m) \in \mathbb{R}^m$ and the forward model

$$\mathbf{y} = T\mathbf{u} + \mathbf{n},$$

where T is a linear operator, compute physical quantity u (element in an infinite dimensional function space; discretisation renders state vector $\mathbf{u} = (u_1, \ldots, u_n)$).

Can we always compute a reliable answer $\mathbf{u}$?

Definition (Well-posed problem)

A generic problem is well-posed if

- a solution exists;
- a solution is unique;
- a solution continuously depends on the given data, that is small changes in the data amount to small changes in the solution.

UNIVERSITY OF CAMBRIDGE

Example: Blurring in the continuum

Given blurred function $y(x) = (G_\sigma * u)(x)$, $x \in (0,1)$, *where*

$$G_\sigma(x) := \frac{1}{2\pi\sigma^2}\, e^{-|x|^2/(2\sigma^2)} \quad = \textit{Gaussian kernel}$$

with standard deviation σ.

Goal: reconstruct u *from knowing* y.

Measurement y is a solution of the heat equation until time $t = \sigma^2/2$.

Retrieving u from y is like solving the heat equation backward in time!
Ill-posedness from lack of continuous dependence.

UNIVERSITY OF CAMBRIDGE

Example: Blurring in the continuum

From Fourier convolution theorem we can write:

$$y = \sqrt{2\pi}\mathcal{F}^{-1}(\mathcal{F}G_\sigma\ \mathcal{F}u)$$
$$u = \frac{1}{\sqrt{2\pi}}\mathcal{F}^{-1}\frac{\mathcal{F}y}{\mathcal{F}G_\sigma}.$$

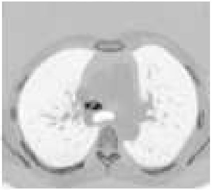
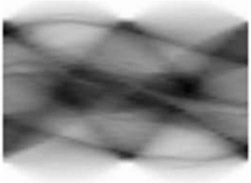
Now, assume instead of measuring blurry y we measure blurry and noisy $y_\delta = y + n^\delta$ with deblurred solution u_δ, then

$$\sqrt{2\pi}|u - u_\delta| = \left|\mathcal{F}^{-1}\frac{\mathcal{F}(y - y_\delta)}{\mathcal{F}G_\sigma}\right| = \left|\mathcal{F}^{-1}\frac{\mathcal{F}n_\delta}{\mathcal{F}G_\sigma}\right|$$

Now, for high-frequencies, $\mathcal{F}(n_\delta)$ will be large while $\mathcal{F}G_\sigma$ will tend to zero for high frequencies (since G_σ is a compact operator), hence the high frequencies in the error are amplified!

Problem formulation

Example: Computed Tomography: compute image $u \in X$ from $Y \ni y = Tu + n$, X, Y normed vector spaces. In computed tomography (CT) T is the Radon transform.

T  $+ n \;=$ 

Video clip from Samuli:

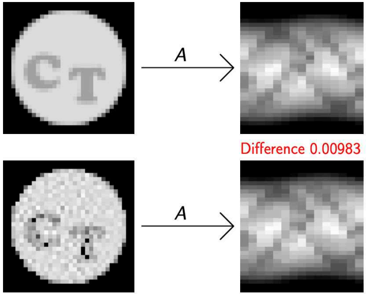
`https://www.youtube.com/watch?v=newxZbw7YAs`

Inverse problem: measuring (corrupted!) data y compute image u.

CT images from LUNA dataset `https://luna16.grand-challenge.org`.

Problem-formulation

Example: Computed Tomography:



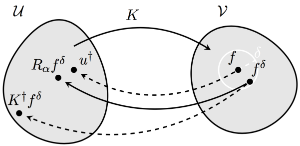
Courtesy of Samuli Siltanen

Ill-posedness: T^{-1} is not continuously invertible (unbounded or discontinuous)

Typical reasons: noise, undersampling, nonlinearity, ...

And there are many more examples of ill-posed inverse problems ... Engl, Hanke, Neubauer '96; Clason, lecture notes Inverse Problems, Duisburg '18; Benning, Ehrhardt, Lang, lecture notes in Inverse Problems, Cambridge '18.

From ill-posed to well-posed via regularisation



Reconstruct an approximation of $u^\dagger$ by solving

$$\min_u \left\{ \underbrace{\alpha\mathcal{R}(u)}_{\text{Regularisation}} + \|Tu - y_\delta\|_2^2 \right\},$$

where f_δ corresponds to noisy measurement. Under appropriate assumptions on $u^\dagger$ (source condition) and for appropriate choice of $\mathcal{R}$ and α we have

$$u_\alpha^\delta \to u^\dagger \text{ as } (\delta, \alpha) \to \mathcal{O} \quad \text{(convergence in appropriate sense)}$$

Engl, Hanke, Neubauer '96.

Bayes theorem and the MAP retrieval

Bayes theorem: for $\mathbf{y} \in \mathbb{R}^m, \mathbf{u} \in \mathbb{R}^n$

$$P(\mathbf{u}|\mathbf{y}) = \frac{P(\mathbf{y}|\mathbf{u})P(\mathbf{u})}{P(\mathbf{y})},$$

where

- $P(\mathbf{y}|\mathbf{u})$ is determined by the forward model and the statistics of the measurement error;
- $P(\mathbf{u})$ encodes our prior knowledge on $\mathbf{u}$;
- $P(\mathbf{y})$ in practice only a normalising factor - ignore it.

Maximum a posteriori (MAP) estimate: compute retrieval $\mathbf{u}^*$ for which

$$P(\mathbf{u}^*|\mathbf{y}) = \max_{\mathbf{u}} P(\mathbf{u}|\mathbf{y}) = \max_{\mathbf{u}} \{P(\mathbf{y}|\mathbf{u})P(\mathbf{u})\}$$

Example: independent Gaussian noise

Gven measurements $y \in \mathbb{R}^m$.

Two components for solving a general inverse problem:

- Data model: $y = Tu + n$, where $u \in \mathbb{R}^n$ original image (to be reconstructed), T linear transformation, n is the noise (simplest situation: n is Gaussian distributed with mean 0 and standard deviation σ)
- A-priori probability density: $P(u) = e^{-\mathcal{R}(u)}$. A-priori information on the original image.

Example: independent Gaussian noise

A posteriori probability for u knowing y given by Bayes:

$$P(u|y) = \frac{P(y|u)P(u)}{P(y)},$$

with

$$P(y|u) = e^{-\frac{1}{2\sigma^2}\sum_{i,j}|(Tu)_{i,j}-y_{i,j}|^2}, \quad P(u) = e^{-\mathcal{J}(u)}$$

Idea of maximum a posteriori" (MAP) image reconstruction: find the "best" image as the one which maximises this probability or equivalently, which solves the minimisation problem

$$\min_u \left\{ \mathcal{R}(u) + \frac{1}{2\sigma^2}\sum_{i,j}|y_{i,j} - (Tu)_{i,j}|^2 \right\}.$$

Extensions of this concept to the infinite dimensional setting Andrew Stuart, Acta Numerica 2010.

MAP estimate and regularisation

Go back to the continuous setting: $u \in L^2(\Omega)$, with Ω open and bounded with Lipschitz boundary (in practice a rectangle).
Transformation T is a bounded linear operator, here from $L^2(\Omega)$ to itself.

The minimisation problem to recover u from y reads

$$\min_u \alpha \mathcal{R}(u) + \frac{1}{2}\|Tu - y\|^2,$$

where $\mathcal{R}$ functional corresponding to a-priori information and $\alpha > 0$ weight balancing.

How to choose $\mathcal{R}$ for images u?

Choice of a-priori information

1. Classical Tychonov regularisation: $\mathcal{R}(u) = \frac{1}{2}\int_\Omega u^2\, dx$ or $\frac{1}{2}\int_\Omega |\nabla u|^2\, dx$. But then:

Reconstructed image $u \in H^1(\Omega)$ cannot present discontinuities across lines (such as edges or boundaries of objects in an image).

To see this, consider 1D situation: $u : [0,1] \to \mathbb{R}$, $u \in H^1(0,1)$. Then, for each $0 < s < t < 1$

$$u(t) - u(s) = \int_s^t u'(r)\, dr \leq \sqrt{t-s}\sqrt{\int_s^t |u'(r)|^2\, dr} \leq \sqrt{t-s}\|u\|_{H^1}^2,$$

and hence $u \in C^{1/2}(0,1)$.

Choice of a-priori information (cont)

Next 2D situation: If $u \in H^1((0,1)^2)$, then the map $x \mapsto u(x,y) \in H^1(0,1)$ for a.e. $y \in (0,1)$ since

$$\int_0^1 \left(\int_0^1 \left| \frac{\partial u(x,y)}{\partial x} \right|^2 \, dx \right) \, dy \leq \|u\|_{H^1}^2 < \infty,$$

so

u cannot jump across vertical boundaries in the image

A similar kind of regularity can be shown for any $u \in W^{1,p}(\Omega)$, $1 \leq p \leq +\infty$ (although a bit weaker for $p = 1$; still now "large" discontinuities are allowed).

This leads us to the total variation (TV)

UNIVERSITY OF CAMBRIDGE

2. Total variation regularisation: $R(u) = |Du|(\Omega)$. For $u \in L^1_{loc}(\Omega)$

$$V(u,\Omega) := \sup\left\{\int_\Omega u\nabla\cdot\varphi\, dx : \varphi \in \left[C_c^1(\Omega)\right]^2, \|\varphi\|_\infty \leq 1\right\}$$

is the variation of u. Further

$u \in BV(\Omega)$ **(the space of bounded variation functions)**

$$\Leftrightarrow$$

$$V(u,\Omega) < \infty.$$

In such a case,

$$|Du|(\Omega) = V(u,\Omega),$$

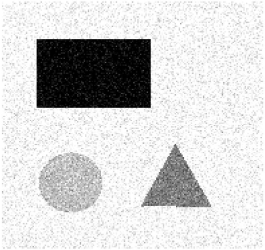
where $|Du|(\Omega)$ is the **total variation** of the finite Radon measure Du, the derivative of u in the sense of distributions.

See e.g. Ambrosio, Fusco, Pallara '99.

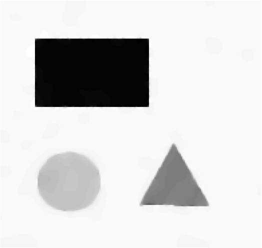
TV smoothing

Properties:

- BV images allow edges (in contrast to $W^{1,1}$, i.e., H^1 images).
- The total variation penalizes small irregularities/oscillations while respecting instrinsic image features such as edges.



Weighted isotropic tv denoising result with $\lambda = 0$



TV smoothing (cont.)

What does it measure?

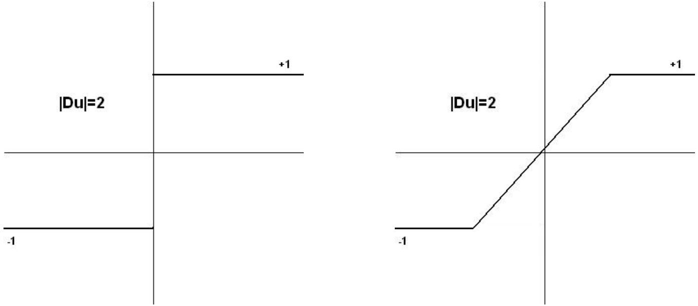


Figure: The total variation measures the size of the jump.

Total variation & H^1 regularisation

(a) original

(b) noisy

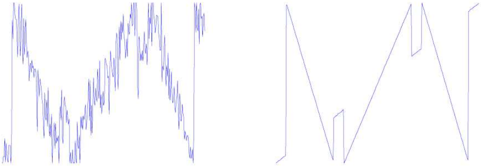
(c) $\|\nabla u\|_2^2$

(d) $\|\nabla u\|_1$

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

TV smoothing (cont.)

Total variation quantifies the oscillations of u:



(a) Large TV (b) Small TV

The total variation penalizes small irregularities/oscillations while respecting instrinsic image features such as edges.

Rudin, Osher, Fatemi, Physica D ’92

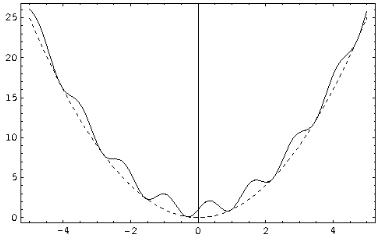
Typical examples of the total variation

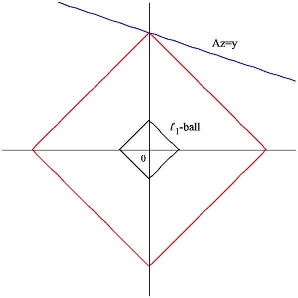
1. Compressed sensing: For $u \in W^{1,1}(\Omega)$ we have

$$|Du|(\Omega) = \|\nabla u\|_{L^1(\Omega)}.$$

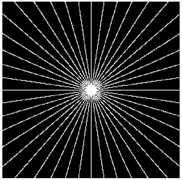
Convex relaxation of sparsity constraint $\|u\|_{L^0}$.

Reconstruct piecewise constant image with only a few discontinuities.

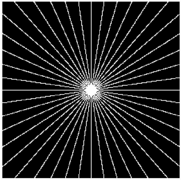




Sparsity promoting regularisation



Undersampled Fourier



Undersampled Fourier

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

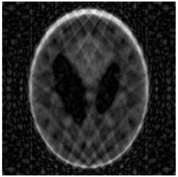
Sparsity in $\nabla u \Rightarrow$ NP-hard problem!

Convex relaxation $\Rightarrow$ ℓ_1- and total variation minimization

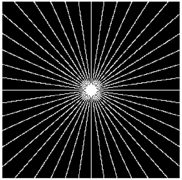
$$\alpha \|\nabla u\|_1 + \frac{1}{2}\|(\mathcal{F}u)_{|\Lambda} - g\|^2 \to \min_u,$$

Rudin, Osher, Fatemi, Physica D '92; Candes, Romberg, Tao, IEEE Trans Inf Theory '06; Lustig et al. IEEE SPM '08

Sparsity promoting regularisation



Zero-filling solution



Undersampled Fourier

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

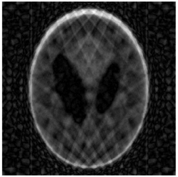
$$\alpha \|\nabla u\|_1 + \frac{1}{2}\|(\mathcal{F}u)_{|\Lambda} - g\|^2 \to \min_u,$$

Rudin, Osher, Fatemi, Physica D '92; Candes, Romberg, Tao, IEEE Trans Inf Theory '06; Lustig et al. IEEE SPM '08

Shepp-Logan Phantom

Sparsity promoting regularisation

Reconstruction via Fourier-Transformation with 0.11% of the data



Zero-filling solution

Reconstructed image

Total variation solution

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

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

Rudin, Osher, Fatemi, Physica D '92; Candes, Romberg, Tao, IEEE Trans Inf Theory '06; Lustig et al. IEEE SPM '08

Typical examples of the total variation (cont)

2. Sets of finite perimeter: Let D be a set with $C^{1,1}$ boundary and $u = \chi_D$ (characteristic function of D, i.e. equals 1 in D and 0 outside), then

$$|Du|(\Omega) = \mathcal{H}^1(\partial D \cap \Omega),$$

the perimeter of D in Ω.

More general Co-area formula: For $u \in BV(\Omega)$ we have

$$|Du|(\Omega) = \int_{-\infty}^{+\infty} \mathrm{Per}(\{u > s\}; \Omega)\, ds,$$

where

$$\mathrm{Per}(\{u > s\}; \Omega) = \|D\chi_{\{u>s\}}\|(\Omega)$$

is the total variation of the characteristic functions of the upper level set of u corresponding to the level s.

UNIVERSITY OF CAMBRIDGE

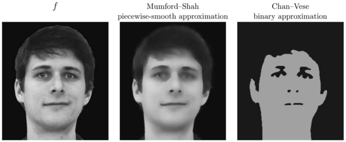
Mumford-Shah segmentation under piece-constancy assumption:

$$\min_{\chi, c_1, c_2} = \alpha |D\chi|(\Omega) + \int_\Omega (y - c_1)^2 \chi + \int_\Omega (y - c_2)^2 (1 - \chi)$$

with $\chi \in \{0, 1\}$ or its convex relaxation (with given c_1 and c_2)

$$\min_v = \alpha |Dv|(\Omega) + \int_\Omega (y - c_1)^2 v + \int_\Omega (y - c_2)^2 (1 - v),$$

with $v \in [0, 1]$ and segmentation is thresholded v.



Mumford, Shah '89; Chan, Vese '01; Pock, Chambolle, Cremers '09; IPOL demo Pascal Getreuer '12

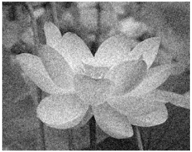
Connection to nonlinear PDEs

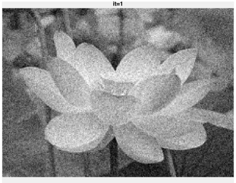
Nonlinear image smoothing with total variation regularisation (ROF model)

$$\alpha \|\nabla u\|_1 + \frac{1}{2}\|u - y\|^2 \to \min_u$$

with steepest descent

$$u_t = \alpha\, p + (u - y), \quad p \in \partial\|\nabla u\|_1, \quad \text{in } \Omega,$$





TV scale space

Perona, Malik '90; Rudin, Osher, Fatemi, Physica D '92; and various contributions from Ambrosio, Caselles, Chambolle, Kuijper, Lions, Morel, Novaga, ...

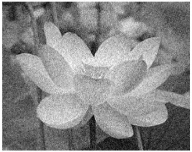
Connection to nonlinear PDEs

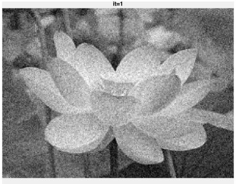
Nonlinear image smoothing with total variation regularisation (ROF model)

$$\alpha\|\nabla u\|_1 + \frac{1}{2}\|u-y\|^2 \to \min_u$$

with steepest descent for $|\nabla u| \neq 0$

$$u_t = \alpha \operatorname{div}\left(\frac{\nabla u}{|\nabla u|}\right) + (u-y), \quad \text{in } \Omega,$$





TV scale space

Perona, Malik '90; Rudin, Osher, Fatemi, Physica D '92; and various contributions from Ambrosio, Caselles, Chambolle, Kuijper, Lions, Morel, Novaga, ...

Come with mathematical foundation

Analytic properties:

- Space $BV(\Omega)$ is non-reflexive; instead derive and use other compactness properties [Ambrosio '90s-; De Giorgi '80-90s]
- Novel metric for stability estimates, e.g. Bregman distances [Burger, Resmerita, He, Computing '07 ; Hofmann, Kaltenbacher, Poeschl, Scherzer '07; Burger, He, CBS '09]
- TV not differentiable; classical derivation of Euler-Lagrange not possible; convex analysis: non-unique sub gradients;
- Analysis of TV flow via differential inclusions or viscosity solutions; explicit solutions using coarea formula. [Chen, Giga, Goto, '91; Ambrosio, Soner, '94, Caselles, Chambolle, Novaga, Paolini, '00-10s].; geometric measure theory [Federer '96; Allard '07]

Numerical properties:

- Standard framework for strictly convex smooth minimization fails; new analysis based on generalized Lagrange multipliers, Douglas-Rachford splitting [Lions, Mercier], iterative thresholding algorithms [Combettes; Chambolle; Daubechies, Pock, ...]; semi-smooth Newton [Hintermüller, Stadler, ...]
- Large-scale nonsmooth convex minimizations: scale badly; accelerations through preconditioning, splitting approaches, partial smoothness, stochastic optimisation; [Chan '89 –; Fornasier, CBS, et al. '08-'11; Figueiredo '07; Beck et al. 09'; Cevher et al. '14; Liang et al. '14; Bredies, Sun '16; Chambolle, Ehrhardt, Richtarik, CBS '18; ...]
- Nonsmooth and nonlinear [Kaltenbacher '08; Bachmayr et al. '09; Attouch et al. '10; Valkonen '13; Bolte et al. '14; Bach, Chizat '19; Driggs, et al. '20; Benning, CBS et al '21; ...].

Regulariser zoo

- Multi-resolution analysis, wavelets (e.g. Daubechies, Dragotti, Foucart, Kutyniok, Langer, Mallat, Rauhut, Unser, . . .).
- Other Banach-space norms, e.g. Sobolev norms, Besov norms, etc. (e.g. Lassas, Siltanen 09)
- Higher-order total variation regularisation (Osher, Sole, Vese '03, Infimal convolution Chambolle, Lions 97; Setzer, Steidl, Teuber 11, Total Generalised Variation Bredies, Kunisch, Pock 10, . . .); Higher-order PDEs, Euler elastica Morel, Masnou '98; Shen, Chan, Kang '03; Bertozzi, Esedoglu, Gillette '07; . . .
- Non-local regularisation (non-local TV Gilboa, Osher, . . . ; non-local means Morel . . .)
- Anisotropic regularisation Weickert 98
- Free-discontinuity problems Mumford, Shah '89; Tomarelli et al.
- and mixtures of the above . . . and probably more which I have forgotten

Introductory books to variational & PDE imaging Chan & Shen 05; Bredies & Lorenz 18.

General variational approach

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

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

Variational approach: Compute u as a minimizer of

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

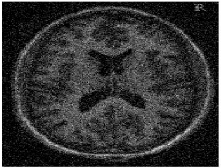
- $T = Id$ for image denoising,
- $T = 1_{\Omega \setminus D}$ for image inpainting,
- $T = *k$ for image deconvolution,
- T is (a possibly undersampled) Radon/Fourier transform for tomography.

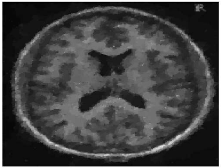
Engl, Hanke, Neubauer '96; Rudin, Osher, Fatemi, Physica D '92; Natterer, Wübbeling '01; Candes, Romberg, Tao, IEEE Trans Inf Theory '06; Kaltenbacher, Neubauer, Scherzer '08; Schuster, Kaltenbacher, Hofmann, Kazimierski '12

Choice of D depends on data statistics

Gaussian

$$\phi(Tu, f) = \|Tu - f\|_2^2$$

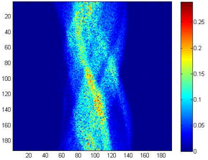


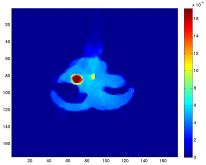


MRI

Poisson

$$\phi(Tu, f) = \int Tu - f \log(Tu)\, dx$$





PET[1]

Impulse

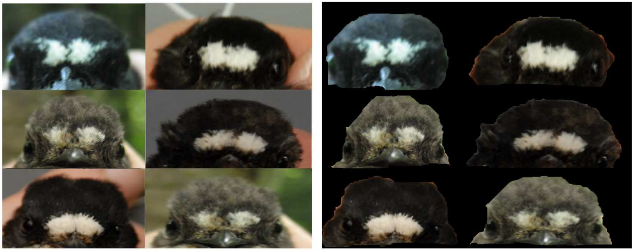
$$\phi(Tu, f) = \|Tu - f\|_1$$

sparse noise.

References: see works by Hohage and Werner '12–

[1] Data courtesy of EIMI, Münster.

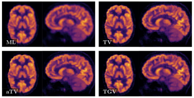
Applications



Calatroni, van Gennip, Rowland, CBS, Flenner, JMIV 2016

Biomedical research and clinical support

PET imaging

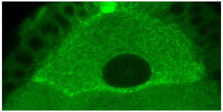


Ehrhardt, Markiewicz, CBS, Physics in Medicine & Biology '19; Chambolle, Ehrhardt, Richtarik, CBS, SIAM Optim '18

Spatio-temporal MRI

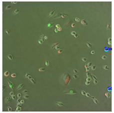
Aviles-Rivero, Graves, Williams, CBS, ISMRM '18; Aviles-Rivero, Debroux, Williams, Graves, CBS, MedIA '21.

Estimating dynamics



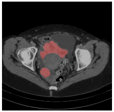
Drechsler, Lang, Dirks, Frerking, Burger, CBS, Palacios, Molecular biology of the cell '20

Mitosis analysis



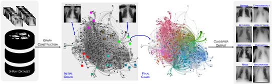
Grah, Burger, Reichelt, CBS et al. Methods '17

Tumour seg



Buddenkotte, . . . , CBS, Radiology AI '21.

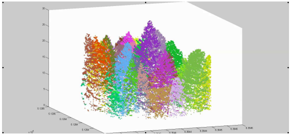
Diagnosis/prognosis



Aviles-Rivero, Papadakis, Li, Sellars, Fan, Tan, CBS, MICCAI '19

Conservation and sustainability

Tree monitoring from LiDAR

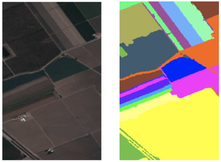


Lee, Cai, CBS, Coomes, IEEE TGRS ’15; Lee, Cai, Lellmann, Dalponte, Malhi, Butt, Morecroft, CBS, Coomes, IEEE Applied Earth Observation and Remote Sensing ’16.

Analysing traffic volume / type

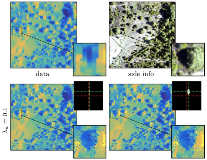
EPSRC GCRF project on Robust and Efficient Analysis Approaches of Remote Imagery for Assessing Population and Forest Health in India

Landcover analysis



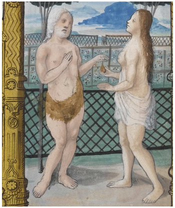
Sellars, Aviles-Rivero, CBS, IEEE TGRS ’19.

Multi-modal image fusion



Bungert, Coomes, Ehrhardt, Rasch, Reisenhofer, CBS, Inverse Problems ’18

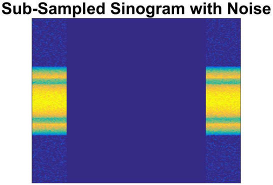
Virtual art restoration and interpretation



Mathematics for Applications in Cultural Heritage (MACH) https://mach.maths.cam.ac.uk funded by the Leverhulme Trust; L. Calatroni, M. d'Autume, R. Hocking, S. Panayotova, S. Parisotto, P. Ricciardi, CBS, Heritage Science '18; Parisotto, Calatroni, Caliari, CBS, Weickert, Inverse Problems '19; Parisotto, L. Calatroni, A. Bugeau, N. Papadakis, CBS, IEEE TIP '20

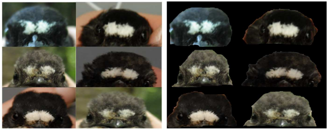
And many more . . .

Material Sciences



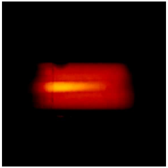
Tovey, Benning, Brune, Lagerwerf, Collins, Leary, Midgley, CBS, Inverse Problems '18.

Zoology



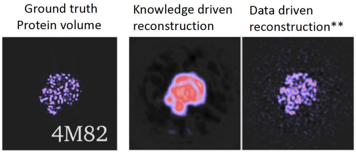
Calatroni, van Gennip, Rowland, CBS, Flenner, JMIV '16

CFD



Benning, Gladden, Holland, CBS, Valkonen, J. Magnetic Resonance '14

Molecular biology



**Collaboration on cryo-EM with Sjors Scheres (LMB, Cambridge)
Appeared in the Journal of the International Union of Crystallography 2020

Numerical realisation

UNIVERSITY OF CAMBRIDGE

$$
\begin{aligned}
u^{k+1} &= \operatorname*{argmin}_{u \in \mathbb{R}^{n\times m}} \frac{1}{2}\|u_0 - Tu\|_2^2 + \frac{\lambda}{2}\|b_1^k \\
&\quad + \nabla u - v^k\|_2^2 + \frac{\lambda}{2}\|b_2^k \\
&\quad + \nabla^2 u - w^k\|_2^2, \\
v^{k+1} &= \operatorname*{argmin}_{v \in \left(\mathbb{R}^{n\times m}\right)^2} \alpha\|v\|_1 \\
&\quad + \frac{\lambda}{2}\|b_1^k + \nabla u^{k+1} - v\|_2^2, \\
w^{k+1} &= \operatorname*{argmin}_{w \in \left(\mathbb{R}^{n\times m}\right)^4} \beta\|w\|_1 \\
&\quad + \frac{\lambda}{2}\|b_2^k + \nabla^2 u^{k+1} - w\|_2^2,
\end{aligned}
$$

Highly nonlinear iterative schemes

Numerical realisation

Consider for $u \in \mathbb{R}^n$

$$\min \mathcal{J}(u) + \mathcal{H}(u),$$

where $\mathcal{J}$ and $\mathcal{H}$ are proper and convex, and (possibly) $\mathcal{J}$ and / or $\mathcal{H}$ is Lipschitz differentiable.

Example: ROF problem for $y \in \mathbb{R}^n$ solve

$$\min_u \alpha \|\nabla u\|_{2,1} + \frac{1}{2}\|u - y\|_2^2,$$

where $\|\nabla u\|_{2,1} = \sum_{i,j} |(\nabla u)_{i,j}|_2 = \sum_{i,j} \sqrt{(u_x)_{i,j}^2 + (u_y)_{i,j}^2}$.
A quick overview of main approaches for minimising such functionals …

Reference for this part: Chambolle, Pock, Acta Numerica 2016.

Classical methods for smooth problems

...algorithms which attempt to compute minimisers of the regularised ROF problem

$$\min_u \left\{ \alpha \sum \sqrt{u_x^2 + u_y^2 + \epsilon} + \frac{1}{2}\|u - g\|_2^2 \right\}$$

for a small $0 < \epsilon \ll 1$.

Since in this case the regularised TV is differentiable in the classical sense we can apply classical numerical algorithms to compute a minimiser, e.g. gradient descent, conjugate gradient etc.

In what follows: convex algorithms which look at the non-regularised problem. For this, let X be an Euclidean space.

Subdifferential

UNIVERSITY OF CAMBRIDGE

Definition

For a convex function $J : X \to \mathbb{R}$, we define the *subdifferential* of J at $x \in V$, as $\partial J(x) = \emptyset$ if $J(x) = \infty$, otherwise

$$\partial J(x) := \{p \in X' : \langle p, y - x\rangle + J(x) \le J(y) \quad \forall y \in V\},$$

where X' denotes the dual space of X. It is obvious from this definition that $0 \in \partial J(x)$ if and only if x is a minimizer of J.

Example: Let $X = \ell_1(\Lambda)$ and $J(x) := \|x\|_1$ is the ℓ_1-norm. We have

$$\partial\|\cdot\|_1(x) = \{\xi \in \ell_\infty(\Lambda) : \ \xi_\lambda \in \partial|\cdot|(x_\lambda), \lambda \in \Lambda\} \tag{1}$$

where $\partial|\cdot|(z) = \{\operatorname{sign}(z)\}$ if $z \neq 0$ and $\partial|\cdot|(0) = [-1, 1]$.

The Legendre-Fenchel transform

For J being one-homogeneous

that is, $J(\lambda u) = \lambda J(u)$ for every u and $\lambda > 0$,

it is a standard fact in convex analysis that the Legendre-Fenchel transform

that is $J^*(v) = \sup_u \langle u, v \rangle_X - J(u)$ (with $\langle u, v \rangle_X = \sum_{i,j} u_{i,j} v_{i,j}$)

is the characteristic function of a closed convex set K:

$$J^*(v) = \chi_K(v) = \begin{cases} 0 & \text{if } v \in K \\ +\infty & \text{otherwise.} \end{cases}$$

Since $J^{**} = J$, we recover

$$J(u) = \sup_{v \in K} \langle u, v \rangle_X \, .$$

Proximal map

Let $\mathcal{J}$ convex, proper and l.s.c., then for any y there is a unique minimiser

$$u^* = \mathrm{argmin}_u \mathcal{J}(u) + \frac{1}{2\tau}\|u-y\|_2^2$$

We call $u^* = \mathrm{prox}_{\tau\mathcal{J}}(y)$ the proximal map of $\mathcal{J}$ at y. With optimality condition

$$0 \in \partial\mathcal{J}(u^*) + \frac{u^*-y}{\tau}$$

this reads

$$u^* = (I + \tau\partial J)^{-1} y.$$

Rockafellar 1997

Moreau's identity

One can show

$$y = \mathrm{prox}_{\tau \mathcal{J}}(y) + \tau \mathrm{prox}_{\frac{1}{\tau}\mathcal{J}^*}(\frac{y}{\tau}),$$

which shows:

> *If we know how to compute* $\mathrm{prox}_{\mathcal{J}}$ *we also know how to compute* $\mathrm{prox}_{\mathcal{J}^*}$*.*

Convex duality

Consider

$$\min_{u \in X} \mathcal{J}(Ku) + \mathcal{H}(u),$$

where $\mathcal{J} : Y \to (-\infty, +\infty]$, $\mathcal{H} : X \to (-\infty, +\infty]$ convex, l.s.c., $K : X \to Y$ linear and bounded. Then (under mild appropriate assumptions on $\mathcal{J}, \mathcal{H}$)

$$\begin{aligned}
&\min_{u \in X} \mathcal{J}(Ku) + \mathcal{H}(u) \\
&\underbrace{=}_{\mathcal{J}^{**}=\mathcal{J}} \min_{u \in X} \sup_{p \in Y} \langle p, Ku \rangle - \mathcal{J}^*(p) + \mathcal{H}(u) \\
&= \max_p \inf_u \langle p, Ku \rangle - \mathcal{J}^*(p) + \mathcal{H}(u) \\
&= \max_p -\mathcal{J}^*(p) - \mathcal{H}^*(-K^* p).
\end{aligned}$$

The latter is the dual problem. Under above assumptions there exists at least one solution p^*. Book, Ekeland, Temam 1999; Survey article by Borwein, Luke 2015

Saddle-point problem

If u^* solves primal problem and p^* dual problem, then (u^*, p^*) is a saddle-point of primal-dual problem

$$\forall (u,p) \in X \times X' \text{ we have } \mathcal{L}(u^*, p) \leq \mathcal{L}(u^*, p^*) \leq \mathcal{L}(u, p^*)$$

where

$$\mathcal{L}(u,p) := \langle p, Ku \rangle - \mathcal{J}^*(p) + \mathcal{H}(u),$$

the Lagrangian. Moreover, we can define the primal-dual gap

$$\begin{aligned} \mathcal{G}(u,p) &:= \sup_{(u',p')} \mathcal{L}(u,p') - \mathcal{L}(u',p) \\ &= \mathcal{J}(Ku) + \mathcal{H}(u) + \mathcal{J}^*(p) + \mathcal{H}^*(-K^*p), \end{aligned}$$

which vanishes iff (u,p) is a saddle point.

Example: dual ROF

$K = \nabla$, $\mathcal{J} = \alpha\|\cdot\|_{2,1}$, $\mathcal{H} = \|\cdot - y\|_2^2/2$. Then, the dual is

$$\begin{aligned}&\max_p -\mathcal{J}^*(p) - \left(\frac{1}{2}\|\nabla^* p\|_2^2 - \langle \nabla^* p, y\rangle\right)\\ &= -\min_p \left(\mathcal{J}^*(p) + \frac{1}{2}\|\nabla^* p - y\|_2^2\right) + \frac{1}{2}\|y\|^2\end{aligned}$$

where $p \in \mathbb{R}^{m\times n\times 2}$. Here

$$\mathcal{J}^*(p) = \chi_{\{\|\cdot\|_{2,\infty}\leq\alpha\}}(p) = \begin{cases} 0 & \text{if } |p_{i,j}|_2 \leq \alpha \forall i,j \\ +\infty & \text{otherwise,}\end{cases}$$

and therefore the dual ROF problem is

$$\min_p\{\|\nabla^* p - y\|_2^2 : \; |p_{i,j}|_2 \leq \alpha \; \forall i,j\}.$$

From optimality conditions of saddle-point problem we have relationship between u and p: $u = y - \nabla^* p$.

Implicit gradient descent

Let J be differential. A more 'advanced' version of gradient descent is implicit gradient descent: Initial guess u^0, then iterate for $k = 0, 1, 2, \ldots$

$$u^{k+1} = u^k - \tau \nabla J(u^{k+1}).$$

If u^{k+1} exists then it is a critical point of

$$\mathcal{J}(u) + \frac{\|u - u^k\|^2}{2\tau},$$

and if $\mathcal{J}$ is convex and l.s.c. then $u^{k+1} = \mathrm{prox}_{\tau\mathcal{J}}(u^k)$. If prox is easy to calculate we call $\mathcal{J}$ simple.

The prox can also make sense for non-differentiable J and the above can be generalised to subgradient descent.

Proximal point algorithm

Define Moreau-Yosida regularisation of $\mathcal{J}$ with parameter τ:

$$\mathcal{J}_\tau(\bar{u}) := \min_u \mathcal{J}(u) + \frac{\|u - \bar{u}\|^2}{2\tau}.$$

One can show

$$\nabla \mathcal{J}_\tau(\bar{u}) = \frac{\bar{u} - \mathrm{prox}_{\tau\mathcal{J}}(\bar{u})}{\tau},$$

and so, implicit gradient descent on $\mathcal{J}$

$$\begin{aligned} u^{k+1} &= \mathrm{prox}_{\tau\mathcal{J}}(u^k) \\ &= (I + \tau\partial J)^{-1}(u^k) \\ &= u^k - \tau\nabla J_\tau(u^k), \end{aligned}$$

is explicit gradient descent on $\mathcal{J}_\tau$. This is a special case of the proximal point algorithm.

Martinet 1970. Convergence rates and accelerations Bertsekas 2015; Nesterov 1983, 2004.

Forward-backward descent

Consider

$$\min_u \mathcal{J}(u) + \mathcal{H}(u),$$

with

- $\mathcal{J}$ is convex, l.s.c. and simple.
- $\mathcal{H}$ is convex with Lipschitz gradient.

Idea: Explicit descent in $\mathcal{H}$ and implicit descent in $\mathcal{J}$. That is

$$u^{k+1} = T_\tau u^k,$$

with

$$T_\tau u = \mathrm{prox}_{\tau \mathcal{J}}(u - \tau \nabla \mathcal{H}(u)).$$

Note, if u is a fixed point of T_τ then it satisfies $0 \in \nabla \mathcal{H}(u) + \partial \mathcal{J}(u)$. If $\tau \leq 1/L$ then u^k converge to a minimiser.

Accelerated version FISTA Nesterov 2004, Beck & Teboulle 2009

Primal-dual hybrid gradient

Consider

$$\min_u \mathcal{J}(Ku) + \mathcal{H}(u),$$

where $\mathcal{J}, \mathcal{H}$ are convex, l.s.c. and simple, K bounded and linear.
Then, solve corresponding saddle-point problem

$$\max_p \inf_u \langle p, Ku \rangle - \mathcal{J}^*(p) + \mathcal{H}(u)$$

via

Alternate proximal descent in u and ascent in p:

$$u^{k+1} = \operatorname{prox}_{\tau\mathcal{H}}(u^k - \tau K^* p^k)$$
$$p^{k+1} = \operatorname{prox}_{\sigma\mathcal{J}^*}(p^k + \sigma K u^{k+1})$$

Arrow,Hurwicz,Uzawa 1958; Pock, Cremers, Bischof, Chambolle 2009; Esser et al. 2010
Linked to other approaches such as augmented Lagrangian and ADMM (alternating direction method of multipliers).

Primal-dual hybrid gradient

Consider

$$\min_u \mathcal{J}(Ku) + \mathcal{H}(u),$$

where $\mathcal{J}, \mathcal{H}$ are convex, l.s.c. and simple, K bounded and linear. Then, solve corresponding saddle-point problem

$$\max_p \inf_u \langle p, Ku \rangle - \mathcal{J}^*(p) + \mathcal{H}(u)$$

via

Alternate proximal descent in u and ascent in p:

$$u^{k+1} = \mathrm{prox}_{\tau\mathcal{H}}(u^k - \tau K^* p^k)$$
$$p^{k+1} = \mathrm{prox}_{\sigma\mathcal{J}^*}(p^k + \sigma K u^{k+1})$$

Not immediately clear that this converges

Arrow,Hurwicz,Uzawa 1958; Pock, Cremers, Bischof, Chambolle 2009; Esser et al. 2010

Linked to other approaches such as augmented Lagrangian and ADMM (alternating direction method of multipliers).

Primal-dual hybrid gradient

Consider

$$\min_u \mathcal{J}(Ku) + \mathcal{H}(u),$$

where $\mathcal{J}, \mathcal{H}$ are convex, l.s.c. and simple, K bounded and linear. Then, solve corresponding saddle-point problem

$$\max_p \inf_u \langle p, Ku \rangle - \mathcal{J}^*(p) + \mathcal{H}(u)$$

via

Primal-dual hybrid gradient:

$$\begin{aligned} u^{k+1} &= \mathrm{prox}_{\tau\mathcal{H}}(u^k - \tau K^* p^k) \\ p^{k+1} &= \mathrm{prox}_{\sigma\mathcal{J}^*}(p^k + \sigma K(2u^{k+1} - u^k)) \end{aligned}$$

Arrow,Hurwicz,Uzawa 1958; Pock, Cremers, Bischof, Chambolle 2009; Esser et al. 2010
Linked to other approaches such as augmented Lagrangian and ADMM (alternating direction method of multipliers).

Lecture plan

- Lecture 1: Variational models & PDEs for inverse imaging problems
- Lecture 2: Learned variational regularisers & plug-and-play denoising
- Lecture 3: Learned iterative reconstruction & perspectives.

Based on: Arridge, Maass, Öktem, CBS, Acta Numerica ’19

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

Temporary page!

LaTeX was unable to guess the total number of pages correctl
there was some unprocessed data that should have been ad
final page this extra page has been added to receive it.
If you rerun the document (without altering it) this surplus pa
away, because LaTeX now knows how many pages to expect f
document.