Topics in Mathematical Imaging
Lecture 2

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

- 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
Problem formulation and possible traps

Given measurements \( f = (f_1, \ldots, f_m) \) and the forward model

\[ f = T u + n, \]

compute physical quantity \( u \) (element in an infinite dimensional function space; discretisation renders state vector \( u = (u_1, \ldots, u_n) \).

Can we always compute a reliable answer \( u \)?

Definition (Well-posed problem)

A generic problem is well-posed if

- there exists a solution;
- 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.
Example: Blurring in the continuum

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

\[
G_\sigma(x) := \frac{1}{2\pi \sigma^2} e^{-|x|^2/(2\sigma^2)} = \text{Gaussian kernel}
\]

with standard deviation \( \sigma. \)

Goal: reconstruct \( u \) from knowing \( f. \)

Measurement \( f \) is a solution of the heat equation until time \( t = \sigma^2/2. \)

Retrieving \( u \) from \( f \) is like solving the heat equation backward in time! Ill-posedness from lack of continuous dependence.
Problem formulation and possible traps

Example: Blurring in the continuum

From Fourier convolution theorem we can write:

$$f = \sqrt{2\pi} F^{-1} (FG_\sigma F u)$$

$$u = \frac{1}{\sqrt{2\pi}} F^{-1} \frac{Ff}{FG_\sigma}.$$

Now, assume instead of measuring blurry $f$ we measure blurry and noisy $f_\delta = f + n_\delta$ with deblurred solution $u_\delta$, then

$$\sqrt{2\pi} \left| u - u_\delta \right| = \left| F^{-1} \frac{F(f - f_\delta)}{FG_\sigma} \right| = \left| F^{-1} \frac{Fn_\delta}{FG_\sigma} \right|$$

Now, for high-frequencies, $F(n_\delta)$ will be large while $FG_\sigma$ will tend to zero for high frequencies (since $G_\sigma$ is a compact operator), hence the high frequencies in the error are amplified!
and many more such as matrix inversion if the condition number of the matrix is large; differentiation; computed tomography which involves the reconstruction of a function from its line integrals . . .

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

$$\min_u \left\{ \alpha J(u) + \| Tu - f_\delta \|_2^2 \right\},$$

where $f_\delta$ corresponds to noisy measurement. Under appropriate assumptions on $u^\dagger$ (source condition) and for appropriate choice of $J$ and $\alpha$ we have

$$u_\alpha^\delta \to u^\dagger \text{ as } (\delta, \alpha) \to \mathcal{O}.$$

Engl, Hanke, Neubauer '96.
Bayes theorem and the MAP retrieval

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

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

where

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

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

$$P(\mathbf{u}^*|\mathbf{f}) = \max_u P(\mathbf{u}|\mathbf{f}) = \max_u \{P(\mathbf{f}|\mathbf{u})P(\mathbf{u})\}$$
Example: independent Gaussian noise

Discrete setting: given image \( f \in \mathbb{R}^N \times \mathbb{R}^N \).

Two components for solving a general inverse problem:

- **Data model:** \( g = Tu + n \), where \( u \in \mathbb{R}^N \times \mathbb{R}^N \) original image (to be reconstructed), \( T \) linear transformation, \( n \) is the noise (simplest situation: \( n \) is Gaussian distributed with mean 0 and standard deviation \( \sigma \)).

- **A-priori probability density:** \( P(u) = e^{-\mathcal{J}(u)} du \). A-priori information on the original image.
Example: independent Gaussian noise

A posteriori probability for $u$ knowing $f$ given by Bayes:

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

with

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

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

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

Extensions of this concept to the infinite dimensional setting Andrew Stuart, Acta Numerica 2010.
The minimisation problem to recover $u$ from $f$ reads

$$\min_u \alpha \mathcal{J}(u) + \frac{1}{2} \| u(x) - f(x) \|^2,$$

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

How to choose $\mathcal{J}$ for images $u$?
Choice of a-priori information

Go back to the continuous setting: \( g, u : \Omega \rightarrow \mathbb{R} \), image domain \( \Omega \) open and bounded with Lipschitz boundary (in practice a rectangle). Transformation \( T \) is a bounded linear operator, here from \( L^2(\Omega) \) to itself. Today, for simplicity \( T = Id \).

Then the minimisation problem to recover \( u \) from \( g \) reads

\[
\min_{u \in L^2(\Omega)} \alpha J(u) + \frac{1}{2} \int_{\Omega} |u(x) - g(x)|^2 \, dx,
\]

where \( J \) functional corresponding to a-priori information and \( \alpha > 0 \) weight balancing.

How to choose \( J \)?
1. Classical Tychonov regularisation: \[ J(u) = \frac{1}{2} \int_{\Omega} u^2 \, dx \] 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 \( u \in C^{1/2}(0, 1) \).
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
\[
\int_0^1 \left( \int_0^1 \left| \frac{\partial u(x, y)}{\partial x} \right|^2 \, dx \right) \, dy \leq \| u \|^2_{H^1} < \infty,
\]
so
\[
u \text{ cannot jump across vertical boundaries in the image}
\]

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) measure

For $u \in L^1_{loc}(\Omega)$

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

is the variation of $u$. Further

$$u \in BV(\Omega) \text{ (the space of bounded variation functions)}$$

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

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

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

the perimeter of $D$ in $\Omega$.

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

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

where

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

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

Chan-Vese segmentation

Mumford-Shah segmentation under piece-constancy assumption:

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

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

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

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

Mumford, Shah ’89; Chan, Vese ’01; Pock, Chambolle, Cremers ’09; IPOL demo Pascal Getreuer ’12
From linear to nonlinear diffusion

References: Perona, Malik, Pattern Analysis and Machine Intelligence ’90; Rudin, Osher, Fatemi, Physica D ’92; Chambolle, Lions, Numerische Mathematik ’97; Vese, Applied Mathematics and Optimization ’01, Ambrosio, Belletini, Caselles, March, Novaga, …
From linear to nonlinear diffusion

\[ u_t = \Delta u, \; u(x, t = 0) = f(x). \]

Solution \( u(x, t) = (G \sqrt{2t} \ast f)(x), \; t > 0 \)

**References:** Perona, Malik, Pattern Analysis and Machine Intelligence ’90; Rudin, Osher, Fatemi, Physica D ’92; Chambolle, Lions, Numerische Mathematik ’97; Vese, Applied Mathematics and Optimization ’01, Ambrosio, Belletini, Caselles, March, Novaga, . . .
From linear to nonlinear diffusion

\[-\Delta t \Delta u + u(\Delta t) - u(0) = 0\]

\[\iff \quad u_t = \text{div} \left( g(|\nabla u|) \nabla u \right), \quad u(x, t = 0) = f(x).\]

\[u(\Delta t) = \text{argmin}_v \left\{ \Delta t \|\nabla v\|_2^2 + \|v - u(0)\|_2^2 \right\} \quad \text{e.g. } g(s) = 1/|\nabla u|, \quad |\nabla u| \neq 0.\]

References: Perona, Malik, Pattern Analysis and Machine Intelligence ’90; Rudin, Osher, Fatemi, Physica D ’92; Chambolle, Lions, Numerische Mathematik ’97; Vese, Applied Mathematics and Optimization ’01, Ambrosio, Belletini, Caselles, March, Novaga, …
From linear to nonlinear diffusion

\[-\Delta t \Delta u + u(\Delta t) - u(0) = 0\]

\[\Longleftrightarrow\]

\[u(\Delta t) = \arg\min_v \left\{ \Delta t \| \nabla v \|_2^2 + \| v - u(0) \|_2^2 \right\} \]

\[\Longleftrightarrow\]

\[\arg\min_v \left\{ \Delta t \int |\nabla u| + \| v - u(0) \|_2^2 \right\} \]

References: Perona, Malik, Pattern Analysis and Machine Intelligence ’90; Rudin, Osher, Fatemi, Physica D ’92; Chambolle, Lions, Numerische Mathematik ’97; Vese, Applied Mathematics and Optimization ’01, Ambrosio, Belletini, Caselles, March, Novaga, …
From linear to nonlinear diffusion

\[-\Delta t \Delta u + u(\Delta t) - u(0) = 0 \quad \iff \quad 0 \in -\Delta t \partial R(u) + u(\Delta t) - u(0)\]

\[u(\Delta t) = \arg\min_v \{\Delta t \|\nabla v\|_2^2 + \|v - u(0)\|_2^2\} \quad \iff \quad (R \text{ is convex})\]

\[\arg\min_v \{\Delta t R(v) + \|v - u(0)\|_2^2\}\]

References: Perona, Malik, Pattern Analysis and Machine Intelligence ‘90; Rudin, Osher, Fatemi, Physica D ‘92; Chambolle, Lions, Numerische Mathematik ‘97; Vese, Applied Mathematics and Optimization ‘01, Ambrosio, Belletini, Caselles, March, Novaga, …
The nonlinear spectral transform

![Image/Signal] = ![Low frequency component] + ![High frequency component]

©Wikimedia commons

The nonlinear spectral transform

- Let $J$ be a proper, convex, lower semi-continuous and absolutely one-homogeneous functional, i.e. $J(cu) = |c|J(u)$ for all $u \in \text{dom}(J)$ and $c \in \mathbb{R}$

- Let $\partial J(u) = \{ p \mid J(v) - J(u) - \langle p, v - u \rangle \geq 0, \forall v \in \text{dom}(J) \}$ denote the subdifferential of $J$

The inverse scale space method is defined as

$$\partial_t p(x, t) = f(x) - u(x, t)$$

for $p(x, t) \in \partial J(u(x, t))$, $p(x, 0) = 0$, $u(x, 0) = \bar{f}(x)$

and $\bar{f}(x) := \text{argmin}_{v \in \ker(J)} \| v - f \|

Property (for $J$ that satisfies the conditions above): $\lim_{t \to \infty} u(x, t) = f(x)$

Total Variation ISS

Example: \( J(u) = TV(u) = \int |Du| \).

Input image \( f \)  
Iterates of ISS \( u^{k+1} \)

Courtesy of Martin Benning.
Motivation: inverse scale space

Variational framework:

\[
\hat{u} \in \arg\min_{u \in \text{dom}(J)} \left\{ \frac{1}{2} \| T u - f \|_{\mathcal{H}}^2 + \alpha J(u) \right\}
\]

Corresponding gradient flow (PDE)

\[
u_t + T^*(f - T u) = \alpha \, p, \quad p \in \partial J(u).
\]

Penalization of \( J \) introduces bias in the solution (e.g. loss of contrast).

Engl, Hanke, Neubauer, Springer ’96; Ambosio, Gigli, Savare, Springer ’08
Bregman iteration:

\[ u^{k+1} \in \arg \min_{u \in \text{dom}(J)} \left\{ \frac{1}{2} \| Tu - f \|_H^2 + \alpha(J(u) - \langle p^k, u \rangle) \right\} \]

with \( p^{k+1} \in \partial J(u^{k+1}) \) satisfying \( p^0 \equiv 0 \) and

\[ p^{k+1} = p^k + \frac{1}{\alpha} T^*(f - Tu^{k+1}). \]

Reintroduces the contrast by solving the constrained problem \( \min J(u) \) s.t. \( Tu = f \) in the limit.

Osher, Burger, Goldfarb, Xu, Yin, SMMS, 4(2), 460-489, '05; Burger, Gilboa, Osher, Xu, Comm. in Math. Sci. 4(1), 179-212, '06.
Computation of Approximate Solutions

(a) Ground truth

(b) Line profile

(c) Tikhonov reconstruction

(d) Bregman iteration based reconstruction
Computation of Approximate Solutions

Special cases of Bregman iteration:

- Linearized Bregman iteration leads to generalized Landweber iteration:

\[ p^{k+1} = p^k + \frac{1}{\alpha} T^* (f - Tu^k), \quad p^{k+1} \in \partial J(u^{k+1}). \]

- Setting \( J(u) = \|u\|_U^2 \) for \( U \) being a Hilbert space leads to iterative Tikhonov regularization:

\[ u^{k+1} \in \arg \min_{u \in \text{dom}(J)} \left\{ \frac{1}{2} \|Tu - f\|_H^2 + \alpha \|u - u^k\|_U^2 \right\}. \]

Darbon, Osher ’07; Cai, Osher, Shen, Mathematics of Computation 78.267, 1515-1536, ’09.
The Inverse Scale Space (ISS) flow is a continuous version of Bregman iteration. It is obtained by letting $\alpha \to \infty$ and interpreting $\Delta t = 1/\alpha$ as a time step, $t_k = k \Delta t$, $p(t_k) = p^k$ and $u(t_k) = u^k$:

$$\frac{\partial t}{\partial t} p(t) = T^*(f - Tu(t)), \quad p(t) \in \partial J(u(t)),$$

with $p(0) = 0$ and $u(0) = u_0 \in \ker(J)$. W.l.o.g. we can assume $u(0) = 0$.

One can show $\lim_{t \to \infty} Tu(x, t) = f(x)$.

Burger, Gilboa, Osher, Xu, Comm. in Math. Sci. 4(1), 179–212, ’06
The nonlinear spectral transform

By defining $\varphi(x, t) = \partial_t u(x, t) = -\partial_t^2 p(x, t)$, we observe

$$\int_0^\infty \varphi(x, t) \, dt = \int_0^\infty \partial_t u(x, t) \, dt =$$

$$\lim_{t \to \infty} u(x, t) - u(x, 0) = f(x) - \bar{f}(x)$$

Hence, we can define a linear inverse transform of the inverse scale space flow as follows:

$$f(x) = \int_0^\infty \varphi(x, t) \, dt + \bar{f}(x)$$
The nonlinear spectral transform

In the discrete setting, we simply replace the discrete spectral transform with

\[ \varphi^k = \begin{cases} u^1, & k = 1 \\ u^k - u^{k-1}, & \text{else} \end{cases} \]

Based on the analytical formula, we define the discrete inverse transform as

\[ f = \sum_{k=1}^{m} \varphi^k + \tilde{f} \]

where \( m \) is the number of iterations, and \( f \) simply defined as the residual, i.e.

\[ \tilde{f} = f - \sum_{k=1}^{m} \varphi^k \]
The nonlinear spectral transform

Example:

\[ J(u) = TV(u) \quad \text{with} \quad TV(u) = \sup_{g \in C_0^\infty(\Omega; \mathbb{R}^n)} \int_{\Omega} u(x)(\text{div } g)(x) dx \]

\[ \|g\|_\infty \leq 1 \]

©Wikimedia commons
The nonlinear spectral transform

This non-linear spectral decomposition allows to filter the original image

\[ \tilde{f}(x) := \int_0^\infty H(t) \varphi(x, t) \, dt + \bar{f}(x) \]

Example: bandpass filter \( H(t) = \begin{cases} 1 & t \in [t_1, t_2] \\ 0 & \text{else} \end{cases} \)
Nonlinear spectral image fusion

Courtesy of Martin Benning
Nonlinear spectral image fusion

Hey Abe, I know how to make your memorial great again...

©Wikimedia commons

Courtesy of Martin Benning
Nonlinear spectral image fusion

Courtesy of Martin Benning
Nonlinear spectral image fusion

Courtesy of Martin Benning
Nonlinear spectral image fusion

Before the US election…

©Wikimedia commons

Courtesy of Martin Benning
Nonlinear spectral image fusion

Before the US election...

Not even the American people would seriously consider playing the Trump card

Courtesy of Martin Benning
Nonlinear spectral image fusion

After the US election...

Good joke 😊

©Wikimedia commons
©Pixabay commons

Courtesy of Martin Benning
Automated (facial) image fusion

1. Face detection
2. Landmark detection
3. Registration
4. Face segmentation
5. Spectral decomposition
6. Image fusion
Automated (facial) image fusion

\[ \hat{f}(x): = \int_0^\infty H_1(x, t) \varphi_1(x, t) + R(H_2(x, t) \varphi_2(x, t)) \, dt + \bar{f}_1(x) \]

- Fusion image
- Spectral filters & segmentation for spectrum of first image
- Non-rigid registration operator
- Spectral filters & segmentation for spectrum of second image
Total variation regularisation model

Compute the restored image $u$ as a minimizer of

$$
J(u) = \left\{ \alpha \left\| \nabla u \right\|_1 + \frac{1}{2} \| Tu - f \|_2^2 \right\} \rightarrow \min_{u \in BV(\Omega)}.
$$

Eliminates corruptions while preserving discontinuities / edges in the image data.

Examples:
- $T = Id$: image denoising
- $T = K_\sigma^*$: image deblurring
- $T = SF$: magnetic resonance tomography
- $T = \chi_{\Omega \setminus D}$: image inpainting.
Beyond TV regularisation

- Multi-resolution analysis, wavelets (e.g. Daubechies, Mallat, Unser, Kutyniok, Foucart & Rauhut, ...).
- Other Banach-space norms, e.g. Sobolev norms, Besov norms, etc. (e.g. Lassas, Siltanen 09)
- Higher-order total variation regularisation (Infimal convolution Chambolle, Lions 97; Setzer, Steidl, Teuber 11, Total Generalised Variation Bredies, Kunisch, Pock 10, ...)
- Non-local regularisation (non-local TV Osher, Gilboa, ...; non-local means Morel ...)
- Anisotropic regularisation Weickert98
- Free-discontinuity problems Mumford, Shah; 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 11 – currently only in German.
Different data statistics influences how $Tu - f$ is quantified, say with generic distance function $\phi(Tu, f)$
Choice of $\phi$ depends on type of corruptions:

- **Gaussian**
  \[ \phi(Tu, f) = \| Tu - f \|^2_2 \]

- **Poisson**
  \[ \phi(Tu, f) = \int Tu - f \log(Tu) \, dx \]

- **Impulse**
  \[ \phi(Tu, f) = \| Tu - f \|_1 \]

Referenced images:

- **MRI**
- **PET$^1$**
- Sparse noise.

**References:** see works by Hohage and Werner ’12–

---

$^1$Data courtesy of EIMI, Münster.
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