Regularization of Inverse Problems

Otmar Scherzer

Computational Science Center, University Vienna, Austria
&
Radon Institute of Computational and Applied Mathematics (RICAM), Linz, Austria
Inverse Problems

desire to calculate or estimate causal factors from a set of observations
Inverse Problems are often Ill-Posed

Operator equation:

\[ Lu = y \]

Setting:

- Available data \( y^\delta \) of \( y \) are noisy
- Focus: \( L \) is a linear operator
- Ill-posed: Let \( u^\dagger \) be a solution:

\[ y^\delta \rightarrow y \nRightarrow u^\delta \rightarrow u^\dagger \]

---

A. N. Tikhonov
On the stability of inverse problems
*Doklady Akademii Nauk SSSR* 39. 1943
Examples of Ill–Posed Problems: \( L = \)

1. **Identity operator**: Measurements of noisy data. **Applications**: A.e.
2. **X-Ray transform**: Measurements of averages over lines. **Application**: Computerized Tomography (CT)
3. **Radon transform**: Measurements of averages over hyperplanes. **Application**: Cryo-EM
4. **Spherical Radon transform**: Measurements of averages over spheres. **Application**: Photoacoustic Imaging
5. **Circular Radon transform**: Measurements of averages over circles. **Applications**: Ground Penetrating Radar and Photoacoustics
An Application

GPR: Location of avalanche victims

Project with Wintertechnik AG and Alps
Various Philosophies

- **Continuous approach:** $L : H_1 \rightarrow H_2$. $H_i$ infinite dimensional spaces

- **Semi-continuous approach:** $L : H \rightarrow \mathbb{R}^n$. $H$ infinite dimensional space, finitely many measurements

- **Discrete Setting:** $L : \mathbb{R}^m \rightarrow \mathbb{R}^n$. Large scale inverse problems

- **Bayesian approach:** $L : \mathbb{R}^m \rightarrow \mathbb{R}^n$. Stochastic inverse problems

---

A. N. Tikhonov
Solution of incorrectly formulated problems and the regularization methods
*Soviet Mathematics. Doklady* 4. 1963

H. Engl, M. Hanke, and A. Neubauer
Regularization of inverse problems

M. Unser, J. Fageot, and J. P. Ward
Splines are universal solutions of linear inverse problems with generalized TV regularization
*SIAM Review* 59.4. 2017

C. R. Vogel
Computational Methods for Inverse Problems
SIAM, 2002

M. Hanke and P. C. Hansen
Regularization methods for large-scale problems
*Surveys on Mathematics for Industry* 3.4. 1994

J. Kaipio and E. Somersalo
Statistical and Computational Inverse Problems
Springer Verlag, 2005
Deterministic Setting

From **Continuous approach**: \( L : H_1 \rightarrow H_2 \). \( H_i \) infinite dimensional spaces.

- **Semi-Continuous approach**: \( P \circ L \) with \( P[f] = (f(x_i))_{i=1,...,n} \)
- **Discrete approach**: \( P \circ L \circ Q \) with \( Q[(c_i)_{i=1,...,m}](x) = \sum_{i=1}^{m} c_i \phi_i(x) \).

\((\phi_i)\) family of test functions.

---

A. Neubauer and O. Scherzer
Finite-dimensional approximation of Tikhonov regularized solutions of nonlinear ill-posed problems

C. Pöschl, E. Resmerita, and O. Scherzer
Discretization of variational regularization in Banach spaces
*Inverse Probl.* 26.10. 2010
Various Methods to Solve

- Backprojection formulas
- Iterative Methods for linear and nonlinear inverse problems
- Flow methods: Showalter’s methods and Inverse scale space methods
- Variational methods: Tikhonov type regularization.
- …

J. Radon
Über die Bestimmung von Funktionen durch ihre Integralwerte längs gewisser Mannigfaltigkeiten
Berichte über die Verhandlungen der Königlich-Sächsischen Gesellschaft der Wissenschaften zu Leipzig, Mathematisch-Physische Klasse 69. 1917

M. Hanke and P. C. Hansen
Regularization methods for large-scale problems
Surveys on Mathematics for Industry 3.4. 1994

B. Kaltenbacher, A. Neubauer, and O. Scherzer
Iterative regularization methods for nonlinear ill-posed problems
Walter de Gruyter, 2008


H. Engl, M. Hanke, and A. Neubauer
Regularization of inverse problems

O. Scherzer, M. Grasmair, H. Grossauer, M. Haltmeier, and F. Lenzen
Variational methods in imaging
Springer, 2009
Outline

1. Variational Methods
2. Numerical Differentiation
3. General Regularization
4. Sparsity and $\ell^1$-Regularization
5. TV-Regularization
6. Regularization of High-Dimensional Data
Numerical Differentiation as an Inverse Problem

- $y = y(x)$ is a smooth function on $0 \leq x \leq 1$
- **Given:** Noisy samples $y_i^\delta$ of $y(x_i)$ on a uniform grid

$$\Delta = \{0 = x_0 < x_1 < \cdots < x_n = 1\}, \quad h = x_{i+1} - x_i$$

satisfying

$$|y_i^\delta - y(x_i)| \leq \delta$$

Boundary data are known exactly: $y_0^\delta = y(0)$ and $y_n^\delta = y(1)$

- **Goal:** Find a smooth approximation $u'$ of $y'$

---

M. Hanke and O. Scherzer
Inverse problems light: numerical differentiation

M. Hanke and O. Scherzer
Error analysis of an equation error method for the identification of the diffusion coefficient in a quasi-linear parabolic differential equation
*SIAM J. Appl. Math.* 59.3. 1999
Strategy I: Constrained Minimization

Approach: Continuous to discrete

1. \[ \|u''\|^2_{L^2} = \int_0^1 (u'')^2 \, dx \rightarrow \min \] among smooth functions \( u \) satisfying
   - \( u(0) = y(0), \ u(1) = y(1) \),
   - Constraint:
     \[
     \frac{1}{n-1} \sum_{i=1}^{n-1} (y_i^{\delta} - u(x_i))^2 \leq \delta^2
     \]

2. Minimizer \( u_*: \ u'_* \approx y' \)
Let $\alpha > 0$. Minimization among smooth functions $u$ satisfying $u(0) = y(0), u(1) = y(1)$, of

$$
\begin{align*}
    u^{\delta}_\alpha &= \arg\min \Phi[u], \\
    \Phi[u] &= \frac{1}{n-1} \sum_{i=1}^{n-1} (y_i^{\delta} - u(x_i))^2 + \alpha \left\| u'' \right\|_{L^2}^2
\end{align*}
$$

$u^{\delta}_\alpha \approx y'$
Strategy II: Tikhonov Regularization + Discrepancy Principle

**Theorem**

If $\alpha$ is selected according to the discrepancy principle

$$\frac{1}{n-1} \sum_{i=1}^{n-1} (y_i^\delta - u_\alpha^\delta(x_i))^2 = \delta^2$$

Then Strategy I and II are equivalent: $u_\alpha^\delta = u^*$

---

V. A. Morozov
Methods for Solving Incorrectly Posed Problems
Springer, 1984
Analysis of Constrained Optimization

Let
1. \( y'' \in L^2(0, 1) \) (assumption on the data to be reconstructed) and
2. \( u_\ast \) be the minimizer of Strategy II

Then
\[
\| u'_\ast - y' \|_{L^2} \leq \sqrt{8} \left( h \| y'' \|_{L^2} + \sqrt{\delta \| y'' \|_{L^2}} \right)
\]

approx. error
noise influence

---

I. J. Schoenberg
Spline interpolation and the higher derivatives
*Proceedings of the National Academy of Sciences of the USA* 51.1. 1964

M. Unser
Splines: a perfect fit for signal and image processing
Textbook Example: Numerical Differentiation

Let \( y \in C^2[0, 1] \), then

\[
\frac{y_{i+1}^\delta - y_i^\delta}{h} - y'(x) \leq \mathcal{O}(h + \delta/h), \quad x_i \leq x \leq x_{i+1}
\]

\( y'' \in L^2 \): \( h \to h + \delta/h \) is minimal for \( h \sim \sqrt{\delta} \). Optimal rates for Strategy I, II and numerical differentiation is \( \mathcal{O}(\sqrt{\delta}) \)

The rate \( \mathcal{O}(\sqrt{\delta}) \) does not hold if \( y'' \notin L^2(0, 1) \)
Properties of \( u^* \)

**Theorem**

- A solution \( u^* \) of Strategy I exists
- \( u^* \) is a natural cubic spline, i.e.,
  - A function that is twice continuously differentiable over \([0, 1]\) with
  - \( u''^*(0) = u''^*(1) = 0 \), and coincides on each subinterval \([x_{i-1}, x_i] \) of \( \Delta \)
  with some cubic polynomial

Generalizations of the ideas to non-quadratic regularization and general inverse problems in Adcock and A. C. Hansen 2015; Unser, Fageot, and Ward 2017

B. Adcock and A. C. Hansen
Generalized sampling and the stable and accurate reconstruction of piecewise analytic functions from their Fourier coefficients
*Math. Comp.* 84.291. 2015

M. Unser, J. Fageot, and J. P. Ward
Splines are universal solutions of linear inverse problems with generalized TV regularization
*SIAM Review* 59.4. 2017
General Variational Methods: Setting

General:

- $H_1$ and $H_2$ are Hilbert spaces
- $L : H_1 \to H_2$ linear and bounded
- $\rho : H_2 \times H_2 \to \mathbb{R}_+$ similarity functional
- $\mathcal{R} : H_1 \to \mathbb{R}_+$ an energy functional
- $\delta$: estimate for the amount of noise

Numerical differentiation:

- $H_1 = W_0^2(0, 1) = \{w : w, w' \in L^2(0, 1)\}$ and $H_2 = \mathbb{R}^{n-1}$
- $L : W_0^2(0, 1) \to \mathbb{R}^{n-1}$, $u \mapsto (u(x_i))_{1 \leq i \leq n-1}$
- $\rho(\xi, \nu) = \frac{1}{n-1} \sum_{i=1}^{n-1} (\xi_i - \nu_i)^2$
- $\mathcal{R}[u] = \int_0^1 (u'')^2 \, dx$
- $\frac{1}{n-1} \sum_{i=1}^{n-1} (y_i - y_i^\delta)^2 \leq \delta^2$. 
Three Kind of Variational Methods ($\tau \geq 1$)

1. Residual method:

$$u^\delta_{\alpha} = \arg\min \mathcal{R}(u) \to \min \text{ subject to } \rho(Lu, y^\delta) \leq \tau \delta$$
Three Kind of Variational Methods ($\tau \geq 1$)

1. Residual method:

$$u^\delta_\alpha = \arg\min R(u) \rightarrow \min \text{ subject to } \rho(Lu, y^\delta) \leq \tau \delta$$

2. Tikhonov regularization with discrepancy principle:

$$u^\delta_\alpha := \arg\min \left\{ \rho^2(Lu, y^\delta) + \alpha R(u) \right\},$$

where $\alpha > 0$ is chosen according to Morozov’s discrepancy principle, i.e., the minimizer $u^\delta_\alpha$ of the Tikhonov functional satisfies

$$\rho(Lu^\delta_\alpha, y^\delta) = \tau \delta$$
Three Kind of Variational Methods ($\tau \geq 1$)

1. Residual method:

$$
u^{\delta}_{\alpha} = \text{argmin} \mathcal{R}(u) \rightarrow \min \quad \text{subject to } \rho(Lu, y^{\delta}) \leq \tau \delta$$

2. Tikhonov regularization with discrepancy principle:

$$
u^{\delta}_{\alpha} := \text{argmin} \left\{ \rho^2(Lu, y^{\delta}) + \alpha \mathcal{R}(u) \right\},$$

where $\alpha > 0$ is chosen according to Morozov’s discrepancy principle, i.e., the minimizer $\nu^{\delta}_{\alpha}$ of the Tikhonov functional satisfies

$$\rho(Lu^{\delta}_{\alpha}, y^{\delta}) = \tau \delta$$

3. Tikhonov regularization with a–priori parameter choice: $\alpha = \alpha(\delta)$
Relation between Methods

E.g. $\mathcal{R}$ convex and $\rho^2(a, b) = \|a - b\|^2$

Residual Method $\equiv$ Tikhonov with discrepancy principle

Note, this was exactly the situation in the spline example!
R-Minimal Solution

If $L$ has a null-space, we concentrate on a particular solution. The $\mathcal{R}$-Minimal Solution is denoted by $u^\dagger$ and satisfies:

$$\mathcal{R}(u^\dagger) = \inf \left\{ \mathcal{R}(u) : Lu = y \right\}$$

Uniqueness of $\mathcal{R}$-minimal solution: For instance if $\mathcal{R}$ is strictly convex
Regularization Method

A method is called a **regularization method** if the following holds:

- **Stability for fixed** $\alpha$: $y^\delta \to H_2 y \Rightarrow u^\delta_\alpha \to H_1 u_\alpha$

- **Convergence**: There exists a parameter choice $\alpha = \alpha(\delta) > 0$ such that $y^\delta \to H_2 y \Rightarrow u^\delta_{\alpha(\delta)} \to H_1 u^\dagger$
Regularization Method

A method is called a regularization method if the following holds:

- **Stability for fixed** \( \alpha \): \( y^\delta \rightarrow_{H_2} y \Rightarrow u^\delta_\alpha \rightarrow_{H_1} u_\alpha \)
- **Convergence**: There exists a parameter choice \( \alpha = \alpha(\delta) > 0 \) such that \( y^\delta \rightarrow_{H_2} y \Rightarrow u^\delta_{\alpha(\delta)} \rightarrow_{H_1} u^\dagger \)

It is an **efficient** regularization method if there exists a parameter choice \( \alpha = \alpha(\delta) \) such that

\[
D(u^\delta_{\alpha(\delta)}, u^\dagger) \leq f(\delta),
\]

where

- \( D \) is an appropriate distance measure
- \( f \) rate (\( f \rightarrow 0 \) for \( \delta \rightarrow 0 \))
Importance of Topologies

It is important to specify the topology of the convergence. Typically Sobolev or Besov spaces.

Example

Differentiation is well-posed from $W^1_0(0,1)$ into $L^2(0,1)$, but not from $L^2(0,1)$ into itself. Take

\[ x \rightarrow f_n(x) := \frac{1}{n} \sin(2\pi nx) \]

Then

\[ x \rightarrow f_n'(x) := 2\pi \cos(2\pi nx) \]

Note

\[ \|f_n\|^2_{W^1_0(0,1)} = \frac{1}{2} \frac{1}{n^2} + \pi \rightarrow \pi \sim \|f_n'\|^2_{L^2(0,1)} = \pi \text{ but } \|f_n\|^2_{L^2(0,1)} = \frac{1}{2} \frac{1}{n^2} \rightarrow 0 \]
Quadratic Regularization in Hilbert Spaces

\[ u_\alpha^\delta = \text{argmin} \left\{ \| Lu - y^\delta \|_{H_2}^2 + \alpha \| u - u_0 \|_{H_1}^2 \right\} \]

Results:

- **Stability** \((\alpha > 0)\): \(y^\delta \to_{H_2} y \Rightarrow u_\alpha^\delta \to_{H_1} u_\alpha\)
- **Convergence**: Choose \(\alpha = \alpha(\delta)\) such that \(\delta^2/\alpha \to 0\)

If \(\delta \to 0\), then \(u_\alpha^\delta \to u^\dagger\)

Note that \(u^\dagger\) is the \(R(\cdot) = \| \cdot - u_0 \|^2\) minimal solution
Convergence Rates (The Simplest Case)

Assumptions:
- **Source Condition:** $u^\dagger - u_0 \in L^* \eta$
- $\alpha = \alpha(\delta) \sim \delta$

Result:

\[
\|u_\alpha^\delta - u^\dagger\|^2 = \mathcal{O}(\delta) \quad \text{and} \quad \|L u_\alpha^\delta - y\| = \mathcal{O}(\delta)
\]

Here $L^*$ is the adjoint of $L$, i.e.,

\[
\langle Lu, y \rangle = \langle u, L^* y \rangle
\]

- If $L \in \mathbb{R}^{m \times n}$, then $L^* = L^T \in \mathbb{R}^{n \times m}$
- If $L = \text{Radon transform}$, the $L^*$ is backprojection operator
Convergence Rates for the Spline Example

Recall \( Lu = u(0.5) \) (just one sampling point) and \( \Delta = \{0, 0.5, 1\} \).

Adjoint operator of \( L : W_0^2(0, 1) \rightarrow \mathbb{R}, \ L^* : \mathbb{R} \rightarrow W_0^2(0, 1) \).

Let \( z \) be the solution of

\[
z^{(IV)}(x) = \delta_{0.5}(x)
\]

satisfying \( z(0) = z(1) = z''(0) = z''(1) = 0 \) and \( z(0.5) = 1 \) and \( C^2 \)-smoothness, i.e. it is a fundamental solution.

Then \( z \) is a natural cubic spline!  

\[1\]Note that a cubic spline is infinitely often differentiable between sampling point and the third derivative jumps. Thus fourth derivative is a \( \delta \)-distribution at the sampling points.
Adjoint for the Spline Example

Let \( v \in \mathbb{R} \)

\[
\langle Lu, v \rangle_{\mathbb{R}} = Luv = v \int_0^1 u(x)\delta_{0.5}(x) \, dx = v \int_0^1 u(x)z^{(IV)}(x) \, dx
\]

\[
= \int_0^1 u''(x) \left( vz''(x) \right) \, dx = \langle u, vz \rangle_{W^2_0(0,1)}
\]

Thus \( L^* v(x) = vz(x) \).

A convergence rate \( O(\sqrt{\delta}) \) holds if the solution is a natural cubic spline and \( u^{\dagger''} \in L^2(0,1) \) (integration by parts)
Classical Convergence Rates - Spectral Decomposition

First, let $L \in \mathbb{R}^{n \times m}$ be a **matrix**:

$$L = \Psi^T \Lambda \Phi$$

with $\Phi \in \mathbb{R}^{m \times m}$, $\Psi \in \mathbb{R}^{n \times n}$ orthogonal and $\Lambda$ diagonal with rank $\leq \min \{m, n\}$.

Then

$$L^* L = L^T L = \Phi^T \Lambda \Psi \Psi^T \Lambda D \Phi = \Phi^T \Lambda ^2 \Phi$$

which rewrites to

$$L^* Lu = \sum_{n=1}^{\min\{m,n\}} \lambda_n^2 \langle u, \phi_n \rangle \phi_n = \int_0^\infty \lambda^2 \langle u, \phi_n \rangle \delta_{\lambda_n} \, dx = de(\lambda) u$$
Classical Convergence Rates (Generalized)

Spectral Theory:

- $L^*L$ is a bounded, positive definitive, self-adjoint operator
- $L^*Lu = \int_0^\infty \lambda^2 e(\lambda)u$, where $e(\lambda)$ denotes the spectral measure of $L^*L$
- If $L$ is compact, then

$$L^*Lu = \sum_{n=0}^{\infty} \lambda_n^2 \langle u, \phi_n \rangle \phi_n,$$

where $(\lambda_n^2, \phi_n)$ are the spectral values of $L$
Classical Convergence Rates

- **Source Condition:** $u^\dagger - u_0 \in (L^*L)^\nu \eta, \nu \in (0, 1]$

- $\alpha = \alpha(\delta) \sim \delta^{\frac{2}{2\nu + 1}}$

**Result:**

$$\|u^{\delta}_{\alpha} - u^\dagger\| = O(\delta^{\frac{2\nu}{2\nu + 1}}) \quad \text{and} \quad \|Lu^{\delta}_{\alpha} - y\| = O(\delta)$$

Note, that for $\nu = 1/2$

$$\mathcal{R}((L^*L)^{1/2}) = \mathcal{R}(L^*)$$

---

Non-Quadratic Regularization

\[ \frac{1}{2} \| Lu - y^\delta \|^2 + \alpha R[u] \rightarrow \min \]

Examples:

- **Total Variation regularization:**
  \[ TV[u] = \sup \left\{ \int_\Omega u \nabla \cdot \phi \, dx : \phi \in C^\infty_0(\Omega; \mathbb{R}^m), \|\phi\|_{L^\infty} \leq 1 \right\} \]
  the total variation semi-norm.

- **\( \ell^p \) regularization:**
  \[ R[u] = \sum_i w_i |\langle u, \phi_i \rangle|^p, \quad 1 \leq p \leq 2 \]
  \( \phi_i \) is an orthonormal basis of a Hilbert space with inner product \( \langle \cdot, \cdot \rangle \), \( w_i \) are appropriate weights - we take \( w_i \equiv 1 \)
Functional Analysis, Basics I

- Let \((u_n)\) be a sequence in a Hilbert space \(H\), then \(u_n \rightharpoonup_H u\) iff
  \[ \langle u_n, \phi \rangle_H \to \langle u, \phi \rangle_H \quad \forall \phi \in H \]

- The set
  \[
  \{ u : u \in L^1(\Omega) \text{ and } TV[u] < \infty \}
  \]
  with the norm
  \[ \|u\|_{BV} := \|u\|_{L^1(\Omega)} + TV[u] \]
  is a Banach space and is called **Space of Functions of Bounded Variation**

- A sequence in \(BV \cap L^2(\Omega)\) is weak* convergent, \(u_n \rightharpoonup^* u\), iff
  \[ \langle u_n, \phi \rangle_{L^2(\Omega)} \to \langle u, \phi \rangle_{L^2(\Omega)} \quad \forall \phi \in L^2(\Omega) \text{ and } TV[u_n] \to TV[u] \]

- If \(u \in C^1(\Omega)\), then \(TV[u] = \int_\Omega |\nabla u| \, dx\)
Let $H$ be a Hilbert space

- $\mathcal{R} : H \rightarrow \mathbb{R} \cup \{+\infty\}$ is called **proper** if $\mathcal{R} \neq \infty$
- $\mathcal{R}$ is **weakly lower semi-continuous** if for $u_n \rightharpoonup_H u$

\[ \mathcal{R}[u] \leq \lim \inf \mathcal{R}[u_n] \]

---

R. T. Rockafellar  
Convex Analysis  
Princeton University Press, 1970
Non-Quadratic Regularization

Assumptions:
- $L$ is a bounded operator between Hilbert spaces $H_1$ and $H_2$ with closed and convex domain $\mathcal{D}(L)$
- $\mathcal{R}$ is weakly lower semi-continuous

Results:
- Stability: $y_\delta \rightarrow_{H_2} y \Rightarrow u_\alpha^\delta \rightarrow_{H_2} u_\alpha$ and $\mathcal{R}[u_\alpha^\delta] \rightarrow \mathcal{R}[u_\alpha]$
- Convergence: $y_\delta \rightarrow_{H_2} y$ and $\alpha = \alpha(\delta)$ such that $\delta^2/\alpha \rightarrow 0$, then
  \[ u_\alpha^\delta \rightarrow_{H_2} u^\dagger \text{ and } \mathcal{R}[u_\alpha^\delta] \rightarrow \mathcal{R}[u^\dagger] \]

Asplund property: For quadratic regularization in H-spaces weak convergence and convergence of the norm gives strong convergence
Some Convex Analysis: The Subgradient

Illustration of the function $f : (-1, 1) \rightarrow \mathbb{R}$, $f(x) = |x|$, and the graphs of two of its subgradients $p_1, p_2 \in \partial f(0) = \{ p \in \mathbb{R}^* \mid p(x) = cx, \ c \in [-1, 1] \}$.
Some Convex Analysis: The Bregman Distance

Illustration of the Bregman distance
\[ D_{f'}(x, y) = f(x) - f(y) - f'(y)(x - y) \] for the function \( f : \mathbb{R} \rightarrow \mathbb{R} \), \( f(x) = x^2 \), between the points \( x = 2 \) and \( y = 1 \).
Bregman Distance

1. We consider Bregman distance for functionals
2. If $\mathcal{R}[u] = \frac{1}{2} \| u - u_0 \|^2 \Rightarrow \partial \mathcal{R}[u^\dagger] = u - u^\dagger$
3. and $D_\xi(u, v) = \frac{1}{2} \| u - v \|^2$.
4. In general not a distance measure: It may be non-symmetric and may vanish for non-equal elements
5. Bregman distance can be a weak measure and difficult to interpret
Convergence Rates, $\mathcal{R}$ convex

**Assumptions:**

- **Source Condition:** There exists $\eta$ such that

$$\xi = F^* \eta \in \partial \mathcal{R}(u^\dagger)$$

- $\alpha \sim \delta$

**Result:**

$$D_{\xi}(u_\alpha^\delta, u^\dagger) = \mathcal{O}(\delta) \text{ and } \|Lu_\alpha^\delta - y\| = \mathcal{O}(\delta)$$

---

M. Burger and S. Osher  
Convergence rates of convex variational regularization  
*Inverse Problems* 20.5. 2004

B. Hofmann, B. Kaltenbacher, C. Pöschl, and O. Scherzer  
A convergence rates result for Tikhonov regularization in Banach spaces with non-smooth operators  
*Inverse Probl.* 23.3. 2007
Compressed Sensing

Let $\phi_i$ be an orthonormal basis of a Hilbert space $H_1$. $L : H_1 \to H_2$

Constrained optimization problem:

$$\mathcal{R}[u] = \sum_{i} |\langle u, \phi_i \rangle| \to \min \text{ such that } Lu = y$$

Goal is to recover sparse solutions:

$$\text{supp}(u) := \{ i : \langle u, \phi_i \rangle \neq 0 \} \text{ is finite}$$
Compressed Sensing

Let $\phi_i$ be an orthonormal basis of a Hilbert space $H_1$. $L : H_1 \rightarrow H_2$

**Constrained optimization problem:**

$$\mathcal{R}[u] = \sum_i |\langle u, \phi_i \rangle| \rightarrow \min \text{ such that } Lu = y$$

Goal is to recover *sparse solutions*:

$$\text{supp}(u) := \{ i : \langle u, \phi_i \rangle \neq 0 \} \text{ is finite}$$

For noisy data: Residual method

$$\mathcal{R}[u] \rightarrow \min \text{ subject to } \| Lu - y^\delta \| \leq \tau \delta$$

---

E. J. Candès, J. K. Romberg, and T. Tao
Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information
*IEEE Transactions on Information Theory* 52.2. 2006
Sparsity Regularization

Unconstrained Optimization

\[ \|Lu - y^\delta\|^2 + \alpha R[u] \to \min \]

General theory for sparsity regularization:

- **Stability:** \( y^\delta \to_{H^2} y \Rightarrow u^\delta_\alpha \to_{H^1} u_\alpha \) and \( \|u^\delta_\alpha\|_{\ell^1} \to \|u_\alpha\|_{\ell^1} \)

- **Convergence:** \( y^\delta \to_{H^2} y \Rightarrow u^\delta_\alpha \to_{H^1} u^\dagger \) and \( \|u^\delta_\alpha\|_{\ell^1} \to \|u^\dagger\|_{\ell^1} \) if \( \delta^2/\alpha \to 0 \).

If \( \alpha \) is chosen according to the discrepancy principle, then Sparsity Regularization \( \equiv \) Compressed Sensing.
Convergence Rates: Sparsity Regularization

Assumptions:

- **Source Condition**: There exists $\eta$ such that

\[ H_2 \ni \xi = L^* \eta \in \partial R[u^\dagger] = \partial \left( \sum_i \left| \langle u^\dagger, \phi_i \rangle \right| \right) = \sum_i \text{sgn}(\langle u^\dagger, \phi_i \rangle) \phi_i =: \xi_i \]

$\Rightarrow u^\dagger$ is sparse (means in the domain of $\partial R$

- $\alpha \sim \delta$

Result:

\[ D_{\xi}(u_\alpha^\delta, u^\dagger) = \mathcal{O}(\delta) \text{ and } \left\| Lu_\alpha^\delta - y \right\| = \mathcal{O}(\delta) \]
Analogous Convergence Rates: Compressed Sensing

**Assumption:** Source condition

\[ \xi = L^* \eta \in \partial \mathcal{R}[u^\dagger] \]

Then

\[ D_\xi(u_*, u^\dagger) \leq 2 \| \eta \| \delta \]

for every

\[ u_* \in \arg\min \left\{ \mathcal{R}[u] : \| Lu - y^\delta \| \leq \delta \right\} \]

Note: \( u_* \) is the constraint solution

---

E. J. Candès, J. K. Romberg, and T. Tao
Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information
*IEEE Transactions on Information Theory* 52.2. 2006

M. Grasmair, M. Haltmeier, and O. Scherzer
The residual method for regularizing ill-posed problems
What can we deduce from the Bregman Distance?

Because we assume $(\phi_i)_{i \in \mathbb{N}}$ to be an orthonormal basis, the Bregman distance simplifies to

$$D_\xi(u, u^\dagger) = \mathcal{R}[u] - \mathcal{R}[u^\dagger] - \langle \xi, u - u^\dagger \rangle$$

$$= \mathcal{R}[u] - \langle \xi, u \rangle$$

$$= \sum_i \left( |\langle u, \phi_i \rangle| - \langle \xi, \phi_i \rangle \langle u, \phi_i \rangle \right)$$

$$= \xi_i u_i$$

Note, by the definition of the subgradient $|\langle \xi, \phi_i \rangle| \leq 1$
Rates with respect to the norm: On the infinite set!

Recall source condition $\xi = L^* \eta \in \partial \mathcal{R}[u^\dagger]$

Define

$$\Gamma(\eta) := \{i : |\xi_i| = 1\} \text{ (which is finite – solution is sparse)}$$

and the number (take into account that the coefficients of $\zeta$ are in $\ell^2$)

$$m_\eta := \max \{|\xi_i| : i \notin \Gamma(\eta)\} < 1$$

Then

$$D_\xi(u_*, u^\dagger) = \sum_i |u_*,i| - \xi_i u_*,i \geq (1 - m_\eta) \sum_{i \notin \Gamma(\eta)} |u_*,i|$$

Consequently, since $\|\cdot\|_{\ell^1} \geq \|\cdot\|_{\ell^2}$, we get

$$\left\| \pi_N \setminus \Gamma(\eta)(u_*) - \pi_N \setminus \Gamma(\eta)(u^\dagger) \right\|_{H_1} = 0 \leq CD_\xi(u_*, u^\dagger) \leq C\delta$$
Rates with respect to the Norm: On the small Set

Additional Assumption: *Restricted injectivity:*

The mapping $L_{\Gamma(\eta)}$ is injective

Thus on $\Gamma(\eta)$ the problem is well–posed on the small set and consequently

$$\|\pi_{\Gamma(\eta)}(u_*) - \pi_{\Gamma(\eta)}(u^\dagger)\|_{H_1} \leq C\delta$$

Together with previous slide:

$$\|u_* - u^\dagger\|_{H_1} \leq C\delta$$
Restricted Isometry Property (RIP)

Candès, Romberg, and Tao 2006: Key ingredient in proving linear convergence rates for the finite dimensional $\ell_1$-residual method:
The $s$-restricted isometry constant $\vartheta_s$ of $L$ is defined as the smallest number $\vartheta \geq 0$ that satisfies

$$(1 - \vartheta) \|u\|^2 \leq \|Lu\|^2 \leq (1 + \vartheta) \|u\|^2$$

for all $s$-sparse $u \in X$. The $(s, s')$-restricted orthogonality constant $\vartheta_{s,s'}$ of $L$ is defined as the smallest number $\vartheta \geq 0$ such that

$$|\langle Lu, Lu' \rangle| \leq \vartheta \|u\| \|u'\|$$

for all $s$-sparse $u$ and $s'$-sparse $u'$ with $\text{supp}(u) \cap \text{supp}(u') = \emptyset$.
The mapping $L$ satisfies the $s$-restricted isometry property, if

$$\vartheta_s + \vartheta_{s,s} + \vartheta_{s,2s} < 1$$
Linear Convergence of Candes & Rhomberg & Tao

Assumptions:
1. $L$ satisfies the $s$-restricted isometry property
2. $u^\dagger$ is $s$-sparse

Result:
$$\|u_* - u^\dagger\|_{H_1} \leq c_s \delta$$

However: These condition imply the source condition and the restricted injectivity
0 < p < 1: Nonconvex sparsity regularization

\[ \|Lu - y^\delta\|^2 + \alpha \sum |\langle u, \phi_i \rangle|^p \rightarrow \min \]

is stable, convergent, and well–posed in the Hilbert-space norm

- Zarzer 2009: \( O(\sqrt{\delta}) \)
- Grasmair 2010b: \( \Rightarrow O(\delta) \)
An Application: Wintertechnik AG and Alps

Ground Penetrating Radar: Location of avalanche victims
GPR: $L$ ist the spherical mean operator

Assumption: GPR which focused radar wave

**Figure:** Simulations with noise free synthetic data: Left: Data. Middle: Reconstruction by Kirchhoff migration. Right: Reconstruction with sparsity regularization
GPR: Simulations with noisy data

Figure: Noisy data. Left: Data. Middle: Reconstruction by Kirchhoff migration. Right: Reconstruction with sparsity regularization
Reconstruction with real data

Figure: Reconstruction from real data. Left: Data. Middle: Reconstruction by Kirchhoff migration. Right: Reconstruction with sparsity regularization.
Let $\Omega, \Sigma$ two two open sets. TV minimization consists in calculating

$$ u_\alpha^\delta := \text{argmin}_{u \in L^2(\Omega)} \left\{ \frac{1}{2} \| Lu - y^\delta \|_{L^2(\Sigma)}^2 + \alpha \text{TV}[u] \right\} $$

---

L. I. Rudin, S. Osher, and E. Fatemi
Nonlinear total variation based noise removal algorithms
*Physica D. Nonlinear Phenomena* 60.1–4. 1992
TV-Regularization

- Assumption: $L$ is a bounded operator between $L^2(\Omega)$ and $L^2(\Sigma)$
- Fact: $TV$ is weakly lower semi-continuous on $L^2(\Omega)$

**Results:**

- **Stability:** $y^\delta \to_{L^2(\Sigma)} y \Rightarrow u^\delta_\alpha \to_{L^2(\Omega)} u_\alpha$ and $TV[u^\delta_\alpha] \to TV[u_\alpha]$
- **Convergence:** $y^\delta \to_{L^2(\Sigma)} y$ and $\alpha = \alpha(\delta)$ such that $\delta^2/\alpha \to 0$, then

\[
\begin{align*}
  u^\delta_\alpha &\to_{L^2(\Omega)} u^\dagger \\
  TV[u^\delta_\alpha] &\to TV[u^\dagger]
\end{align*}
\]
TV-Regularization: Source Condition

\( u^\dagger \) satisfies the source condition if there exist \( \xi \in L^2(\Omega) \) and \( \eta \in L^2(\Sigma) \) such that

\[
\xi = L^* \eta \in \partial TV[u^\dagger]
\]

Then for \( \alpha \sim \delta \)

\[
TV[u^{\delta}] - TV[u^\dagger] - \langle \xi, u^{\delta}_\alpha - u^\dagger \rangle_{L^2(\Omega)} = D_\xi TV(u^{\delta}_\alpha, u^\dagger) = O(\delta)
\]
Source Condition for the Circular Radon Transform

Notation: \( \Omega := B(0, 1) \subseteq \mathbb{R}^2 \) open, \( \varepsilon \in (0, 1) \). We consider the **Circular Radon transform**

\[
S_{\text{circ}}[u] := t \int_{\mathbb{S}^1} u(z + tw) d\mathcal{H}^1(w)
\]

for functions from

\[
L^2(B(0, 1 - \varepsilon)) := \left\{ u \in L^2(\mathbb{R}^2) : \operatorname{supp}(u) \subseteq \overline{B(0, 1 - \varepsilon)} \right\}
\]

- is well-defined
- bounded from \( L^2(B(0, 1 - \varepsilon)) \) into \( L^2(\mathbb{S}^1 \times (0, 1)) \)
- and \( \|S_{\text{circ}}\| \leq 2\pi \)
Finer Properties of the Circular Radon Transform

- There exists a constant $C_\varepsilon > 0$, such that
  \[
  C_\varepsilon^{-1} \| S_{circ} u \|_2 \leq \| i^*(u) \|_{1/2,2} \leq C_\varepsilon \| S_{circ} u \|_2 , \quad u \in L^2(B(0, 1 - \varepsilon))
  \]

  where $i^*$ is the adjoint of the embedding $i : W^{1/2,2}(B(0, 1)) \rightarrow L^2(B(0, 1))$.

- For every $\varepsilon \in (0, 1)$ we have
  \[
  W^{1/2,2}(B(0, 1 - \varepsilon)) = \mathcal{R}(S_{circ}^*) \cap L^2(B(0, 1 - \varepsilon))
  \]

---

O. Scherzer, M. Grasmair, H. Grossauer, M. Haltmeier, and F. Lenzen
Variational methods in imaging
Springer, 2009
Wellposedness of TV-minimization for $S_{\text{circ}}$

Minimization of the TV-functional with $L = S_{\text{circ}}$ is

- well-posed, stable, and convergent
- Let $\varepsilon \in (0, 1)$ and $u^\dagger$ the TV-minimizing solution. Moreover, if the Source Condition

$$\xi \in \partial TV[u^\dagger] \cap W^{1/2,2}(B(0, 1 - \varepsilon))$$

is satisfied, then

$$TV[u^\delta] - TV[u^\dagger] - \langle \xi, u^\delta - u^\dagger \rangle = O(\delta)$$
Functions that satisfy the Source Condition

- Let $\rho \in C_0^\infty(\mathbb{R}^2)$ be an adequate mollifier and $\rho_\mu$ the scaled function of $\rho$. Moreover, let $x_0 = (0.2, 0)$, $a = 0.1$, and $\mu = 0.3$. Then
  $$u^\dagger := 1_{B(x_0, a+\mu)} \ast \rho_\mu$$
  satisfies the source condition.

- Let $u^\dagger := 1_F$ be the indicator function of a bounded subset of $\mathbb{R}^2$ with smooth boundary.
Convergence of Level-Sets

\[ \Omega \subset \mathbb{R}^2! \]

\[
\frac{1}{2} \left\| Lu - y^\delta \right\|^2_{L^2(\Sigma)} + \alpha \text{TV}[u] \to \min
\]

for

\[ u \in L^2(\Omega) \cong \left\{ u \in L^2(\mathbb{R}^2) : \text{supp}(u) \subset \overline{\Omega} \right\} \]
Convergence of Level-Sets

$t$-super level-set of $u^\delta_\alpha$:

$$U^\delta_\alpha(t) := \{ x \in \Omega : u^\delta_\alpha(x) \geq t \} \quad \text{for } t \geq 0$$

$$U^\delta_\alpha(t) := \{ x \in \Omega : u^\delta_\alpha(x) \leq t \} \quad \text{for } t < 0$$

Theorem

Assume that source condition holds! Let $\delta_n, \alpha_n \to 0^+$ such that

$$\frac{\delta_n}{\alpha_n} \leq \sqrt{\frac{\pi}{2}}.$$

Then, up to a subsequence and for almost all $t \in \mathbb{R}$, denoting $U_n := U^\delta_{\alpha_n}$,

$$\lim_{n \to \infty} |U_n(t) \Delta U^\dagger(t)| = 0, \quad \text{and} \quad \lim_{n \to \infty} \partial U_n(t) = \partial U^\dagger(t).$$
A Deblurring Result

**Figure:** Deblurring of a characteristic function by total variation regularization with Dirichlet boundary conditions. First row: Input image blurred with a known kernel and with additive noise. Second row: numerical deconvolution results. Third row: some level lines of the results.
Image Registration: Model Problems

- Given: Images $I_1, I_2 : \Omega \subseteq \mathbb{R}^2 \to \mathbb{R}$
- Find $u : \Omega \to \Omega$ satisfying

$$L[u] := l_2 \circ u = l_1$$

$u$ should be a diffeomorphism (no twists)
Calculus of Variations: Notions of Convexity

\[ f : \mathbb{R}^m \times \mathbb{R}^n \times \mathbb{R}^{m\times n} \rightarrow \mathbb{R}, \]

\[ (x, u, v) \rightarrow f(x, u, v) \]

Hierarchy:

\[ f \text{ convex} \Rightarrow \text{polyconvex} \Rightarrow \text{quasi-convex} \Rightarrow \text{rank-one convex} \]

Up to quasi-convexity:

\[ u \rightarrow \int_{\mathbb{R}^m} f(x, u, \nabla u) \, dx \text{ is weakly lower semicontinuous on } \]

\[ H_1 := W^{1,p}(\Omega, \mathbb{R}^n) \text{ with } 1 \leq p \leq \infty \]

If \( m = 1 \) or \( n = 1 \), then all convexity definitions are equivalent

Polyconvex functionals are used in elasticity theory

---

C. B. Morrey
Multiple Integrals in the Calculus of Variations
Springer Verlag, 1966

B. Dacorogna
Direct Methods in the Calculus of Variations
Springer Verlag, 1989
Polyconvex Functions

For $A \in \mathbb{R}^{m \times n}$ and $1 \leq s \leq m \wedge n$

$\text{adj}_s(A)$ consists of all $s \times s$ minors of $A$ (subdeterminants)

$f : \mathbb{R}^{m \times n} \to \mathbb{R} \cup \{+\infty\}$ is polyconvex if

$$f = F \circ T,$$

where $F : \mathbb{R}^{\tau(m,n)} \to \mathbb{R} \cup \{+\infty\}$ is convex and

$$T : \mathbb{R}^{m \times n} \to \mathbb{R}^{\tau(m,n)}, \quad A \to (A, \text{adj}_2(A), \ldots, \text{adj}_{\tau(m,n)}(A))$$

Typical example:

$$f(A) = (\det[A])^2$$

---

J. M. Ball
Convexity conditions and existence theorems in nonlinear elasticity
*Archive for Rational Mechanics and Analysis* 63.
1977
Polyconvex Regularization

Assumptions:

- \( \mathcal{R}[u] := \int_\Omega F \circ T[u](x) \, dx. \)
- \( L \) is a **non-linear** continuous operator between \( W^{1,p}(\Omega, \mathbb{R}^n) \) and \( H_2 \) (sometimes needs to be a Banach space) with closed and convex domain of definition \( \mathcal{D}(L) \)

Results:

- **Stability:** \( y^\delta \rightharpoonup_{H_2} y \Rightarrow u^\delta_\alpha \rightharpoonup_{W^{1,p}} u_\alpha \) and \( \mathcal{R}[u^\delta_\alpha] \rightarrow \mathcal{R}[u_\alpha] \)
- **Convergence:** \( y^\delta \rightharpoonup_{H_2} y \) and \( \alpha = \alpha(\delta) \) such that \( \delta^2 / \alpha \rightarrow 0 \), then

\[ u^\delta_\alpha \rightharpoonup_{W^{1,p}} u^\dagger \text{ and } \mathcal{R}[u^\delta_\alpha] \rightarrow \mathcal{R}[u^\dagger] \]
Generalized Bregman Distances

Let $W$ be a family of functionals on $H_1 = W^{1,p}(\Omega, \mathbb{R}^n)$

- The $W$-subdifferential of a functional $\mathcal{R}$ is defined by

$$\partial_W \mathcal{R}[u] = \{ w \in W : \mathcal{R}[v] \geq \mathcal{R}[u] + w[v] - w[u], \forall v \in H_1 \}$$

- For $w \in \partial_W \mathcal{R}[u]$ the $W$-Bregman distance is defined by

$$D^W_w(v, u) = \mathcal{R}[v] - \mathcal{R}[u] - w[v] + w[u]$$
Bregman Distances of Polyconvex Integrands

Let \( p \in [1, \infty) \) and \( H_1 = \mathcal{W}^{1,p}(\Omega, \mathbb{R}^n) \).

\[
T(\nabla u) \in \prod_{s=2}^{m \wedge n} L^p_s(\Omega, \mathbb{R}^{\sigma(s)}) =: S_2.
\]

We define

\[
\mathcal{W}_{\text{poly}} := \{ w : H_1 \to \mathbb{R} : \exists (u^*, v^*) \in H_1^* \times S_2^* \text{ s.t.} \}
\]

\[
w[u] = \langle u^*, u \rangle_{H_1^*, H_1} + \langle v^*, T(\nabla u) \rangle_{S_2^*, S_2}
\]

Remark:
- \( \mathcal{W}_{\text{poly}} = (H_1 \times S_2)^* \). However, functionals \( w \) are non-linear
- \( \mathcal{W}_{\text{poly}} \)-Bregman distance:

\[
D_w^{\text{poly}}(u, \bar{u}) = \mathcal{R}[u] - \mathcal{R}(\bar{u}) - w[u] + w(\bar{u})
\]

\[
= \mathcal{R}[u] - \mathcal{R}(\bar{u}) - \langle u^*, u - \bar{u} \rangle_{H_1^*, H_1}
\]

\[
- \langle v^*, T(\nabla u) - T(\nabla \bar{u}) \rangle_{S_2^*, S_2}
\]
Polyconvex Subgradient

- $\Omega \subset \mathbb{R}^m$ and $H_1 = W^{1,p}(\Omega, \mathbb{R}^n)$
- For $x \in \Omega$, the map $(u, A) \mapsto F(x, u, A)$ is convex and differentiable
- $\mathcal{R}[u] = \int_{\Omega} F(x, u(x), T(\nabla u(x))) \, dx$

**Definition**

If $\mathcal{R}[\bar{v}] \in \mathbb{R}$ and the function $x \mapsto F'_{u,A}(x, \bar{v}(x), T(\nabla \bar{v}(x)))$ lies in

$$L^{p^*}(\Omega, \mathbb{R}^n) \times \prod_{s=1}^{m \wedge n} L^p_s(\Omega, \mathbb{R}^{\sigma(s)})$$

then this function is a $W_{\text{poly}}$-subgradient of $\mathcal{R}$ at $\bar{v}$
Rates result

Let $H_1 = W^{1,p}(\Omega, \mathbb{R}^n)$ and consider regularization by

$$u \rightarrow \| L[u] - y^\delta \|^2 + \alpha R[u]$$

**Assumptions:**

- $R$ has a $W_{\text{poly}}$-subgradient $w$ at $u^\dagger$
- Let $\alpha(\delta) \sim \delta$ and $\exists \beta_1 \in [0, 1), \beta_2$ such that in a neighborhood

$$w[u^\dagger] - w[u] \leq \beta_1 D^\text{poly}_w(u, u^\dagger) + \beta_2 \| L[u] - y \|$$

**Results:**

$$D^\text{poly}_w(u^\delta, u^\dagger) = O(\delta) \quad \text{and} \quad \| L[u] - y^\delta \| = O(\delta)$$

Note, that for polyconvex regularization one requires a stronger condition than for convex regularization.
Thank you for your attention