Variational Methods in Banach Spaces for the Solution of Inverse Problems

Otmar Scherzer

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

Ill-Posed Problems

Operator equation

$$\boxed{Fu = y}$$

Setting:

- Available data y^δ of y is noisy
- Ill-Posed:

$$\boxed{y^\delta \to y \not\Rightarrow u^\delta \to u^\dagger}$$

- In our setting: F is a linear operator between infinite dimensional spaces (before discretization)

Examples of Ill–posed Problems

F=

1. Radon Transform: Computerized Tomography CT
2. Radon Transform squared: a problem of Schlieren Imaging (nonlinear problem)
3. Spherical Radon Transform: Photoacoustic Imaging
4. Circular Radon Transform: GPRadar and Photoacoustic Imaging

Three Kind of Variational Methods

ρ similarity functional for the data, Ψ an energy functional, δ an estimate for the amount of noise.

1. *Residual method ($\tau \geq 1$):*

$$\boxed{u_{Res} = \text{argmin}\Psi(u) \to \min \qquad \text{subject to } \rho(Fu, y^\delta) \leq \tau\delta}$$

2. *Tikhonov regularization with discrepancy principle ($\tau \geq 1$):*

$$\boxed{u_\alpha^\delta := \text{argmin}\left\{\rho^2(Fu, y^\delta) + \alpha\Psi(u)\right\} \ ,}$$

   where $\alpha > 0$ is chosen according to *Morozov's discrepancy principle*, i.e., the minimizer u_α^δ of the Tikhonov functional satisfies

$$\rho(Fu_\alpha^\delta, y^\delta) = \tau\delta$$

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

Relation between Methods

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

$$\boxed{\text{Residual Method} \equiv \text{Tikhonov with discrepancy principle}}$$

Parameter Dependent Regularization Method - Tikhonov

A method is a regularization method if the following holds:

- Stability for fixed α: $y^\delta \to y \Rightarrow u_\alpha^\delta \to u_\alpha$
- Convergence: There exists a parameter choice $\alpha = \alpha(\delta) > 0$ such that $y^\delta \to y \Rightarrow u_{\alpha(\delta)}^\delta \to u^\dagger$

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

$$\boxed{D(u_{\alpha(\delta)}^\delta, u^\dagger) \leq f(\delta)\,,}$$

where D is an appropriate distance measure, f rate ($f \to 0$ for $\delta \to 0$)

Note: In general, residual method is *not* stable but convergent.

FWF

Quadratic Regularization in Hilbert Spaces - Folklore

$$\boxed{u_\alpha^\delta = \operatorname{argmin}\left\{\|Fu - y^\delta\|^2 + \alpha\|u - u_0\|^2\right\}}$$

Assumptions:

- F is a bounded operator between Hilbert spaces U and Y
- $\left\|y - y^\delta\right\| \leq \delta$

Results:

- *Stability* $(\alpha > 0)$*:* $y^\delta \to y \Rightarrow u_\alpha^\delta \to u_\alpha$
- *Convergence:* Choose

$$\alpha = \alpha(\delta) \text{ such that } \delta^2/\alpha \to 0$$

If $\delta \to 0$, then $u_\alpha^\delta \to u^\dagger$, which solves $Fu^\dagger = y$

Convergence Rates

Assumptions:

- *Source Condition:* $\boxed{u^\dagger - u_0 \in F^*\eta}$
- $\alpha = \alpha(\delta) \sim \delta$

Result:

$$\boxed{\left\|u_\alpha^\delta - u^\dagger\right\|^2 = O(\delta) \text{ and } \left\|Fu_\alpha^\delta - y\right\| = O(\delta)}$$

Here F^* is the adjoint of F, i.e.,

$$\langle Fu, y\rangle = \langle u, F^* y\rangle$$

Reference: C. Groetsch (1984)

FWF

Non-Quadratic Regularization

$$\boxed{\frac{1}{2}\left\|Fu - y^\delta\right\|^2 + \alpha R(u) \to \min}$$

Examples:

- Total Variation regularization: $R(u) = \int |\nabla u|$ the total variation semi-norm
- ℓ^p regularization: $R(u) = \sum_i w_i \left|\langle u, \phi_i \rangle\right|^p, \quad 1 \leq p \leq 2$

  ϕ_i is an orthonormal basis of a Hilbert space with inner product $\langle \cdot, \cdot \rangle$, w_i are appropriate weights - we take $w_i \equiv 1$

Non-Quadratic Regularization

Assumptions:

- F is a bounded operator between Hilbert spaces U and Y with closed and convex domain $\mathcal{D}(F)$
- R is weakly lower semi-continuous

Results:

- *Stability:* $y^\delta \to y \Rightarrow u_\alpha^\delta \rightharpoonup u_\alpha$ and $R(u_\alpha^\delta) \to R(u_\alpha)$
- *Convergence:* $y^\delta \to y$ and $\alpha = \alpha(\delta)$ such that $\delta^2/\alpha \to 0$, then

$$\boxed{u_\alpha^\delta \rightharpoonup u^\dagger \text{ and } R(u_\alpha^\delta) \to R(u^\dagger)}$$

Note, for quadratic regularization in H-spaces weak convergence and convergence of the norm gives strong convergence

Convergence Rates, R convex

Assumptions:

- *Source Condition:* There exists η such that

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

- $\alpha \sim \delta$

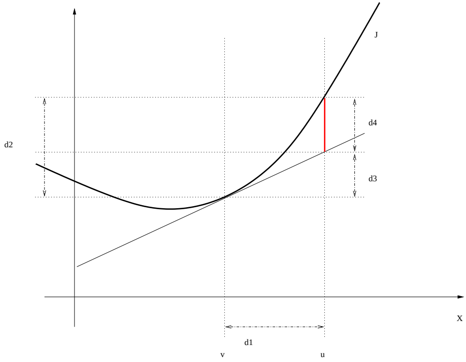
Result:

$$\boxed{D_\xi(u_\alpha^\delta, u^\dagger) = O(\delta) \text{ and } \left\| F u_\alpha^\delta - y \right\| = O(\delta)}$$

Comments:

1. $\partial R(v)$ is the subgradient of R at v, i.e., all elements ψ that satisfy $D_\psi(u,v) := R(u) - R(v) - \langle \psi, u - v \rangle \geq 0$ for all u
2. $D_\xi(u_\alpha^\delta, u^\dagger)$ is the Bregman distance
3. References: Burger & Osher, Hofmann et al

Bregman Distance



1. In general not a distance measure: It may be *non*-symmetric and may vanish for different elements.
2. If $R(\cdot) = \frac{1}{2}\|u - u_0\|^2$, then $D_\xi(u, v) = \frac{1}{2}\|u - v\|^2$. Thus generalizes the H-space results.

Constrained ℓ^1 - Compressed Sensing

U Hilbert space with inner product $\langle \cdot, \cdot \rangle$, ϕ_i orthonormal basis, Y H space, $F : U \to Y$ bounded.

Constrained optimization problem:

$$R(u) = \sum_i |\langle u, \phi_i \rangle| \to \min \qquad \text{such that } Fu = y \ .$$

Goal is to recover *sparse solutions*:

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

Comments:

1. References: e.g. Candes & Rhomberg & Tao
2. Here everything is infinite dimensional!
3. For noisy data

$$\boxed{R(u) \to \min \qquad \text{subject to } \left\| Fu - y^\delta \right\| \leq \tau\delta}$$

Sparsity Regularization

Unconstrained Optimization

$$\boxed{\left\| Fu - y^\delta \right\|^2 + \alpha R(u) \to \min}$$

General theory for sparsity regularization:

- *Stability:* $y^\delta \to y \Rightarrow u_\alpha^\delta \rightharpoonup u_\alpha$ and $\left\| u_\alpha^\delta \right\|_{\ell^1} \to \| u_\alpha \|_{\ell^1}$
- *Convergence:* $y^\delta \to y \Rightarrow u_\alpha^\delta \rightharpoonup u_\alpha^\delta$ and $\left\| u_\alpha^\delta \right\|_{\ell^1} \to \left\| u^\dagger \right\|_{\ell^1}$ if $\delta^2/\alpha \to 0$.

If α is chosen according to the discrepancy principle, then Sparsity Regularization $\equiv$ Constrained ℓ^1 - Compressed Sensing

Convergence Rates: Sparsity Regularization

Assumptions:

- *Source Condition:* There exists η such that

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

  Formally this means that $\xi_i = \text{sgn}(u_i^\dagger)$ and $u^\dagger$ is sparse (means in the domain of ∂R)

- $\alpha \sim \delta$

Result:

$$\boxed{D_\xi(u_\alpha^\delta, u^\dagger) = O(\delta) \text{ and } \left\| F u_\alpha^\delta - y \right\| = O(\delta)}$$

Comment: Rate is *optimal* for a choice $\alpha \sim \delta$.

Convergence Rates: Compressed Sensing

Assumption: Source condition

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

Then

$$\boxed{D_\xi(u^\delta, u^\dagger) \leq 2\|\eta\|\delta}$$

for every

$$u^\delta \in \operatorname{argmin}\left\{R(u) : \left\|Fu - y^\delta\right\| \leq \delta\right\}$$

Remark: Candes et al have rate δ with respect to the *finite dimensional norm* and not the Bregman distance.

What is the Bregman Distance of ℓ^1

Because ϕ_i is an orthonormal basis the Bregman distance simplifies to

$$
\begin{aligned}
D_\xi(u, u^\dagger) &= R(u) - R(u^\dagger) - \langle \xi, u - u^\dagger \rangle \\
&= R(u) - \langle \xi, u \rangle \\
&= \sum_i \Big(|\langle u, \phi_i \rangle| - \langle \xi, \phi_i \rangle \langle u, \phi_i \rangle \Big) \quad =: \sum_i |u_i| - \xi_i u_i
\end{aligned}
$$

Note, by the definition of the subgradient $|\langle \xi, \phi_i \rangle| \leq 1$

Rates with respect to the norm: On the big set

Recall source condition $\xi = F^*\eta \in \partial R(u^\dagger)$
Define

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

and the number (take into account that the coefficients of ζ are in ℓ^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

$$\left\| \pi_{\mathbb{N} \setminus \Gamma(\eta)}(u^\delta) - \underbrace{\pi_{\mathbb{N} \setminus \Gamma(\eta)}(u^\dagger)}_{0} \right\| \leq D_\xi(u, u^\dagger) \leq C\delta$$

Rates with respect to the Norm: On the small set

Additional Assumption: *Restricted injectivity:*

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

Thus on $\Gamma(\eta)$ the problem is well–posed and Consequently

$$\left\|\pi_{\Gamma(\eta)}(u^\delta) - \pi_{\Gamma(\eta)}(u^\dagger)\right\| \leq C\delta$$

Together with previous transparency:

$$\left\|u^\delta - u^\dagger\right\| \leq C\delta$$

Reference: Grasmair et al, Bredis & Lorenz

Restricted Isometry Property

Candes, Rhomberg, Tao: Key ingredient in proving linear convergence rates for the finite dimensional ℓ^1-residual method: The *s-restricted isometry constant* ϑ_s of F is defined as the smallest number $\vartheta \geq 0$ that satisfies

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

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

$$\left|\langle Fu, Fu'\rangle\right| \leq \vartheta \|u\| \left\|u'\right\|$$

for all s-sparse u and s'-sparse u' with $\mathrm{supp}(u) \cap \mathrm{supp}(u') = \emptyset$. The mapping F satisfies the s-restricted isometry property, if $\vartheta_s + \vartheta_{s,s} + \vartheta_{s,2s} < 1$

Linear Convergence of Candes & Rhomberg & Tao

Assumptions:

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

Result:

$$\left\| u^\delta - u^\dagger \right\| \leq c_s \delta$$

However: These condition imply the source condition and the restricted injectivity.

Conclusion on Sparsity

- Convergence rates results can be proven under standard assumptions of variational regularization
- The essential ingredient for rates is the Karush-Kuhn-Tucker condition (which in the inverse problems community is called source condition) from convex optimization
- Taking into account restricted injectivity gives a convergence rate $O(\delta)$. Otherwise: linear convergence in the Bregman distance

$0 < p < 1$: Nonconvex sparsity regularization

$$\boxed{\left\|Fu - y^\delta\right\|^2 + \sum |\langle u, \phi_i \rangle|^p \to \min}$$

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

- Zarzer: $O(\sqrt{\delta})$
- Grasmair + IP $\Rightarrow O(\delta)$

Recent developments: Replacement of subgradient by Clarke subdifferentials (nonconvex theory) (Grasmair 2010)

FWF

Thank you for your attention