Methods for Inverse Problems:
I. Regularization methods for linear problems

Barbara Kaltenbacher, University of Klagenfurt, Austria

ALPEN-ADRIA UNIVERSITÄT KLAGENFURT | WIEN GRAZ

Overview

1. Inverse problems as operator equations
2. Linear Problems
3. Compact operators and singular system
4. Generalized inverse and ill-posedness
5. A class of regularization methods
   - Some auxiliary results
   - Error representation
   - Some examples of methods
   - Convergence
   - Convergence rates

Inverse problems as operator equations

- Often, inverse problems can be formulated as operator equations
$$F(x) = y\,, \tag{1}$$
where $F : \mathcal{D}(F) \to \mathcal{Y}$ with domain $\mathcal{D}(F) \subset \mathcal{X}$, $\mathcal{X}, \mathcal{Y}$ Hilbert spaces.
- Measurements are usually contaminated with noise, therefore, we assume that noisy data y^δ with
$$\|y^\delta - y\| \leq \delta\,. \tag{2}$$
are given.
- Example: EIT: $F : a \mapsto \Lambda_a$, where Λ_a is the Dirichlet-Neumann operator for
$$\nabla(a\nabla u) = 0 \quad \text{in } \Omega$$

Linear Problems

We consider an operator equation

$$Tx = y \tag{3}$$

where $T \in L(X, Y)$ and X and Y are Hilbert spaces.
$\mathcal{R}(T) \subseteq Y$ ... range of T
$\mathcal{N}(T) \subseteq Y$ ... nullspace of T

$$Q = \mathrm{Proj}_{\overline{\mathcal{R}(T)}}, \quad P = \mathrm{Proj}_{\mathcal{N}(T)},$$

$^{\perp}$ ... orthogonal complement of linear subspace $M \subseteq Z$:

$$M^{\perp} = \{z \in Z \mid \langle z, m \rangle_Z = 0 \ \forall m \in M\}$$

T^*... adjoint operator

$$\langle Tx, y \rangle_Y = \langle x, T^* y \rangle_X \quad \forall x \in X, y \in Y$$

Compact operators and singular system

Theorem

A compact operator $T \in L(X, Y)$ has a singular system $(\sigma_j; u_j, v_j)_{j\in\mathbb{N}}$: $(u_j)_{j\in\mathbb{N}} \subseteq X$ and $(v_j)_{j\in\mathbb{N}} \subseteq Y$ orthonormal systems

$$\begin{aligned} Tu_j &= \sigma_j v_j\,, \quad span(u_j)_{j\in\mathbb{N}} = \mathcal{N}(T)^\perp = \overline{\mathcal{R}(T^*)}\,, \\ T^* v_j &= \sigma_j u_j\,, \quad span(v_j)_{j\in\mathbb{N}} = \overline{\mathcal{R}(T)} = \mathcal{N}(T^*)^\perp\,, \end{aligned}$$

$$\sigma_j \to 0 \text{ as } j \to \infty\,. \tag{4}$$

$$Tx = \sum_{j=1}^{\infty} \sigma_j \langle x, u_j\rangle v_j\,, \quad T^* y = \sum_{j=1}^{\infty} \sigma_j \langle y, v_j\rangle u_j\,. \tag{5}$$

$$x = \sum_{j=1}^{\infty} \langle x, u_j\rangle u_j \ + Px\,, \quad y = \sum_{j=1}^{\infty} \langle y, v_j\rangle v_j \ + (I - Q)y\,.$$

...generalized Fourier series.

Generalized inverse and ill-posedness

$T^\dagger$ ... generalized inverse of T:

$$\forall y \in \mathcal{D}(T^\dagger) = \mathcal{R}(T) + \mathcal{R}(T)^\perp \ : \quad T^\dagger y = (T|_{\mathcal{N}(T)^\perp \to \mathcal{R}(T)})^{-1} Q y \,.$$

Compact T with singular system $(\sigma_j; u_j, v_j)_{j\in\mathbb{N}}$:

$$T^\dagger y = \sum_{j=1}^{\infty} \frac{1}{\sigma_j} \langle y, v_j \rangle u_j \,,$$

provided this sum converges:

$$y \in \mathcal{D}(T^\dagger) \iff \sum_{j=1}^{\infty} \frac{\langle y, v_j \rangle^2}{\sigma_j^2} < \infty \quad \text{Picard criterion} \qquad (6)$$

Note that in general only $\sum_{j=1}^{\infty} \langle y^\delta, v_j \rangle^2 + \|(I-Q)y\|^2 < \infty$ and on the other hand $\sigma_j \to 0$ as $j \to \infty$. $\rightsquigarrow$ ill-posedness:
Noise in the jth generalized Fourier coefficient $\langle y, v_j \rangle$ is amplified by $\frac{1}{\sigma_j}$ $\rightsquigarrow$ stronger amplification of high frequent noise.

An Example (1-d source identification):

$$\begin{aligned} -\Delta u &= f \quad \text{in } \Omega \\ u &= 0 \quad \text{on } \partial\Omega \end{aligned}$$

Identify $f \in L^2(\Omega)$ from given measurements of u
equivalently: solve $Tx = y$ with $y = u$, $x = f$, $T = (-\Delta)^{-1}$
$X = L^2(\Omega)$, $Y = L^2(\Omega)$ (We measure values but not derivatives!)
$\mathcal{R}(T) \subseteq H^2(\Omega) \hookrightarrow L^2(\Omega) \;\Rightarrow\; T$ compact.

1-d case $\Omega = (0,1)$: singular system
$((\pi j)^{-2}; \frac{1}{\sqrt{\pi}} \sin(\pi j \cdot), \frac{1}{\sqrt{\pi}} \sin(\pi j \cdot))$,

$$y \in \mathcal{R}(T) + \mathcal{R}(T)^{\perp} \iff \sum_{j=1}^{\infty} j^4 \left(\int_{\Omega} y(\xi) \sin(\pi j \xi)\, d\xi \right)^2 < \infty$$

Generalized inverse and best approximate solution

- The generalized inverse $T^\dagger$ satisfies the Moore-Penrose equations
$$\begin{aligned} TT^\dagger T &= T \\ T^\dagger TT^\dagger &= T^\dagger \\ T^\dagger T &= I - P = \mathrm{Proj}_{\mathcal{N}(T)^\perp} \\ TT^\dagger &= Q = \mathrm{Proj}_{\overline{\mathcal{R}(T)}}. \end{aligned}$$
- A *least squares solution* x_{lss} of $Tx = y$ is defined by
$$x_{lss} \in \mathrm{argmin}\{\|Tx - y\| \,:\, x \in X\}$$
It exists $\Leftrightarrow y \in \mathcal{D}(T^\dagger) = \mathcal{R}(T) + \mathcal{R}(T)^\perp$.
x_{lss} is least-squares-solution $\Leftrightarrow T^*Tx = T^*y$ (normal equation)
- A best approximate solution $x^\dagger$ of $Tx = y$ is defined by
$$x^\dagger \in \mathrm{argmin}\{\|x\| \,:\, x \text{ is least squares solution of } Tx = y\}$$
It exists $\Leftrightarrow y \in \mathcal{D}(T^\dagger) = \mathcal{R}(T) + \mathcal{R}(T)^\perp$.
$x^\dagger$ is best approximate solution $\Leftrightarrow x = T^\dagger y$

A class of regularization methods: Definition

$$T^\dagger y = (T^*T)^\dagger T^* y \quad \rightsquigarrow \quad R_\alpha y^\delta := q_\alpha(T^*T)T^*y^\delta \qquad (7)$$

- $q_\alpha \in C([0, \|T^*T\|])$ depending on some regularization parameter $\alpha > 0$.
- Definition of $f(A)$ by spectral theory for
  $f \ldots$ piecewise continuous function
  $A \ldots$ selfadjoint nonnegative definite operator.
- Case A compact with eigensystem $(\sigma_j^2; u_j)_{j\in\mathbb{N}}$:

$$Ax = \sum_{j=1}^{\infty} \sigma_j^2 \langle x, u_j\rangle u_j \;\rightsquigarrow\; f(A)x = \sum_{j=1}^{\infty} f(\sigma_j^2)\langle x, u_j\rangle u_j \,.$$

- Notation: $x_\alpha := R_\alpha y$, $x_\alpha^\delta := R_\alpha y^\delta$, $x^\dagger := T^\dagger y$

Some auxiliary results

Lemma

f piecewise continuous function, $A \in L(X,X)$ selfadjoint. Then

$$(\forall \lambda \in [0, \|A\|] : |f(\lambda)| \leq C) \implies \|f(A)\| \leq C$$

Lemma

f piecewise continuous function, $T \in L(X,Y)$. Then

$$f(T^*T)T^* = T^*f(TT^*)$$

Lemma

$T \in L(X,Y)$, $x \in X$, $\rho > 0$. Then

$$\exists w \in X \;\; \|w\| \leq \rho : \quad x = (T^*T)^{1/2} w$$

$$\iff$$

$$\exists v \in Y \;\; \|v\| \leq \rho : \quad x = T^* v$$

A class of regularization methods: Error representation

$$x_\alpha^\delta := R_\alpha y^\delta := q_\alpha(T^*T)T^*y^\delta \tag{8}$$

Reconstruction error for exact data:

$$x^\dagger - x_\alpha = (I - q_\alpha(T^*T)T^*T)x^\dagger = r_\alpha(T^*T)x^\dagger \tag{9}$$

$$r_\alpha(\lambda) := 1 - \lambda q_\alpha(\lambda), \qquad \lambda \in [0, \|T^*T\|]. \tag{10}$$

Total error:

$$x^\dagger - x_\alpha^\delta = \underbrace{r_\alpha(T^*T)x^\dagger}_{\text{approximation error}} + \underbrace{R_\alpha(y - y^\delta)}_{\text{propagated noise}}$$

A class of regularization methods: Some examples

- Tikhonov regularization (Tikh)
  min $\left\{\|Tx - y^\delta\|^2 + \alpha\|x - x_0\|^2\right\}$, which is equivalent to

$$x_\alpha^\delta = (T^*T + \alpha I)^{-1}(T^*y^\delta + \alpha x_0).$$

- iterated Tikhonov regularization (itTikh)

$$x_{\alpha,0}^\delta := 0\,, \qquad x_{\alpha,n+1}^\delta := (T^*T+\alpha I)^{-1}(T^*y^\delta+\alpha x_{\alpha,n}^\delta), \quad n \geq 0$$

- Landweber iteration (LW):
  gradient method for min $\|Tx - y^\delta\|^2$: $\quad \alpha = \frac{1}{n}$

$$x_0^\delta = 0\,, \qquad x_{n+1}^\delta = x_n^\delta - T^*(Tx_n^\delta - y^\delta), \quad n \geq 0,$$

- Truncated Singular Value Decomposition (TSVD):

replace $x = \sum_{j\in\mathbb{N}} \frac{1}{\sigma_j}\langle y, v_j\rangle u_j$ by finite sum $x_\alpha^\delta = \sum_{\sigma_j^2 \geq \alpha} \frac{1}{\sigma_j}\langle y^\delta, v_j\rangle u_j$

Tikh	$q_\alpha(\lambda) = \frac{1}{\lambda+\alpha}$	$r_\alpha(\lambda) = \frac{\alpha}{\lambda+\alpha}$
itTikh	$q_\alpha(\lambda) = \frac{(\lambda+\alpha)^n - \alpha^n}{\lambda(\lambda+\alpha)^n}$	$r_\alpha(\lambda) = \left(\frac{\alpha}{\lambda+\alpha}\right)^n$
LW	$q_n(\lambda) = \sum_{j=0}^{n-1}(1-\lambda)^j$	$r_n(\lambda) = (1-\lambda)^n$
TSVD	$q_\alpha(\lambda) = \begin{cases} \lambda^{-1}, & \lambda \geq \alpha \\ 0, & \lambda < \alpha \end{cases}$	$r_\alpha(\lambda) = \begin{cases} 0, & \lambda \geq \alpha \\ 1, & \lambda < \alpha \end{cases}$

A class of regularization methods: Convergence

In all these examples the functions r_α, q_α satisfy

$$\lim_{\alpha \to 0} r_\alpha(\lambda) = \begin{cases} 0, & \lambda > 0 \\ 1, & \lambda = 0 \end{cases} \tag{11}$$

$$|r_\alpha(\lambda)| \leq C_{\mathrm{r}} \qquad \text{for } \lambda \in [0, \|T^*T\|] \tag{12}$$

$$|q_\alpha(\lambda)| \leq \frac{C_{\mathrm{q}}}{\alpha} \qquad \text{for } \lambda \in [0, \|T^*T\|] \tag{13}$$

Theorem

If (11) *and* (12) *hold true, then the operators* R_α *defined by* (8) *converge pointwise to* $T^\dagger$ *on* $\mathcal{D}(T^\dagger)$ *as* $\alpha \to 0$*. With the additional assumption* (13) *the norm of the regularization operators can be estimated by*

$$\|R_\alpha\| \leq \sqrt{\frac{(C_\mathrm{r}+1)C_\mathrm{q}}{\alpha}}. \tag{14}$$

If $\overline{\alpha}(\delta, y^\delta)$ *is a parameter choice rule satisfying*

$$\overline{\alpha}(\delta, y^\delta) \to 0, \qquad \text{and} \qquad \delta/\sqrt{\overline{\alpha}(\delta, y^\delta)} \to 0 \qquad \text{as } \delta \to 0, \tag{15}$$

then $(R_\alpha, \overline{\alpha})$ *is a regularization method in the sense that*

$$\lim_{\delta\to 0} \sup \left\{ \|R_{\overline{\alpha}(\delta,y^\delta)} y^\delta - T^\dagger y\| : y^\delta \in Y, \|y^\delta - y\| \leq \delta \right\} = 0 \tag{16}$$

for all $y \in \mathcal{D}(T^\dagger)$*..*

Convergence rates under source conditions

source wise representation condition

$$x^\dagger = (T^*T)^\mu w, \qquad w \in X, \|w\| \le \rho. \tag{17}$$

T …smoothing operator $\Rightarrow$ (17) is abstract smoothness condition.
For the above methods (Tikh, itTikh, LW, TSVD), there exist $\mu_0 \in (0, \infty]$ (*qualification*), $C_\mu > 0$ such that

$$\sup_{\lambda \in [0, \|T^*T\|]} |\lambda^\mu r_\alpha(\lambda)| \le C_\mu \alpha^\mu \qquad \text{for } 0 \le \mu \le \mu_0. \tag{18}$$

Theorem

Assume that (17) *and* (18) *hold. Then the approximation error and its image under* T *satisfy*

$$\begin{aligned} \|x^\dagger - x_\alpha\| &\leq C_\mu \alpha^\mu \rho, && \text{for } 0 \leq \mu \leq \mu_0, \\ \|Tx^\dagger - Tx_\alpha\| &\leq C_{\mu+1/2} \alpha^{\mu+1/2} \rho, && \text{for } 0 \leq \mu \leq \mu_0 - \frac{1}{2}. \end{aligned}$$

If the regularization parameter α *is chosen according to*

$$\alpha^{\mu+\frac{1}{2}} \sim \delta \tag{19}$$

then the optimal convergence rate

$$\|x_\alpha^\delta - x^\dagger\| \leq \tilde{C}_\mu \delta^{\frac{2\mu}{2\mu+1}} \text{ for } 0 \leq \mu \leq \mu_0 \tag{20}$$

is obtained.

Remarks

- a posteriori regularization parameter choice rules ("μ-free")
  - Morozov's discrepancy principle:
    $\alpha = \max \; s.t. \; \| T x_\alpha^\delta - y^\delta \| \leq \delta \,,$:
    optimal rates (20) for $\mu \leq \mu_0 - \frac{1}{2}$ [Morozov 1968];
    mod.vers.: (20) for $\mu \leq \mu_0$: [Raus 1988, Engl&Gfrerer 1988]
  - balancing principle (or Lepskii rule) [Goldenshluger&Perverzev 2000, Bauer&Perverzev 2005]:
    optimal rates (20), also stochastic setting
  - generalized cross validation [Wahba 1977, Lukas 2006] for stochastic setting
  - L-curve [Hansen 1992] "δ-free" (Bakushinski - veto)
- logarithmic source conditions for severely ill-posed problems [Hohage 1999]
- alternative choice of regularization term in Tikhonov:
  TV, L^1 to enhance sparsity $\rightsquigarrow$ analysis in Banach spaces
  [Burger&Osher 2004, Schöpfer&Louis&Schuster 2006, Schuster&Hofmann&BK&Kazimierski 2012]