Regularization in Banach Space

Barbara Kaltenbacher, Alpen-Adria-Universität Klagenfurt

joint work with

Uno Hämarik, University of Tartu
Bernd Hofmann, Technical University of Chemnitz
Urve Kangro, University of Tartu
Christiane Pöschl, Alpen-Adria-Universität Klagenfurt
Elena Resmerita, Alpen-Adria-Universität Klagenfurt
Otmar Scherzer, University of Vienna
Frank Schöpfer, University of Oldenburg
Thomas Schuster, Saarland University
Ivan Tomba, University of Bologna

PICOF 2014, Hammamet, Tunisis, 8 May 2014

Outline

- motivation: parameter identification in PDEs
  - examples
  - motivation for using Banach spaces
- Regularization in Banach space
  - variational methods
  - iterative methods

Motivation: Parameter Identification in PDEs

Parameter identification in PDEs: Examples (I)

- e.g., electrical impedance tomography (EIT)

$$-\nabla(\sigma\nabla\phi) = 0 \text{ in } \Omega\,.$$

Identify conductivity σ from measurements of the Dirichlet-to-Neumann map Λ_σ, i.e., all possible pairs $(\phi\ ,\ \sigma\partial_{\boldsymbol{n}}\phi)$ on $\partial\Omega$.

Parameter identification in PDEs: Examples (I)

- e.g., electrical impedance tomography (EIT)

$$-\nabla(\sigma\nabla\phi) = 0 \text{ in } \Omega\,.$$

Identify conductivity σ from measurements of the Dirichlet-to-Neumann map Λ_σ, i.e., all possible pairs $(\phi\ ,\ \sigma\partial_n\phi)$ on $\partial\Omega$.

- e.g., quantitative thermoacoustic tomography (qTAT):

$$\nabla\times\left(\mu^{-1}\nabla\times\mathbf{E}\right)+\sigma\frac{\partial}{\partial t}\mathbf{E}+\varepsilon\frac{\partial^2}{\partial t^2}\mathbf{E}=\mathbf{J} \text{ in } \Omega\,.$$

Identify σ from measurements of the deposited energy $\sigma|\mathbf{E}|^2$ in Ω.

ALPEN-ADRIA UNIVERSITÄT KLAGENFURT | WIEN GRAZ

Parameter identification in PDEs: Examples (II)

- e.g. "a-example" (transmissivity in groundwater modelling)

$$-\nabla(a\nabla u) = 0 \text{ in } \Omega\,.$$

Identify a from measurements of u in Ω.

Parameter identification in PDEs: Examples (II)

- e.g. "a-example" (transmissivity in groundwater modelling)

$$-\nabla(a\nabla u) = 0 \text{ in } \Omega\,.$$

Identify a from measurements of u in Ω.

- e.g. "c-example" (potential in stat. Schrödinger equation)

$$-\Delta u + c\, u = 0 \text{ in } \Omega\,.$$

Identify c from measurements of u in Ω.

ALPEN-ADRIA UNIVERSITÄT KLAGENFURT | WIEN GRAZ

Abstract Formulation

Identify parameter x in PDE

$$A(x, u) = f$$

from measurements of the state u

$$C(u) = y\,,$$

where $x \in X$, $u \in V$, $y \in Y$, $X, V, Y \ldots$ Banach spaces
$A : X \times V \to W^* \ldots$ differential operator
$C : V \to Y \ldots$ observation operator

Abstract Formulation

Identify parameter x in PDE

$$A(x,u) = f$$

from measurements of the state u

$$C(u) = y\,,$$

where $x \in X$, $u \in V$, $y \in Y$, $X, V, Y \ldots$ Banach spaces
$A : X \times V \to W^* \ldots$ differential operator
$C : V \to Y \ldots$ observation operator

(a) reduced approach: operator equation for x

$$F(x) = y,$$

$F = C \circ S$ with $S : X \to V$, $x \mapsto u$ parameter-to-state map

ALPEN-ADRIA UNIVERSITÄT KLAGENFURT | WIEN GRAZ

Abstract Formulation

Identify parameter x in PDE

$$A(x,u) = f$$

from measurements of the state u

$$C(u) = y\,,$$

where $x \in X$, $u \in V$, $y \in Y$, $X, V, Y \ldots$ Banach spaces
$A : X \times V \to W^* \ldots$ differential operator
$C : V \to Y \ldots$ observation operator

(a) reduced approach: operator equation for x

$$F(x) = y,$$

$F = C \circ S$ with $S : X \to V$, $x \mapsto u$ parameter-to-state map

(b) all-at once approach: measurements and PDE as system for (x,u)

$$\begin{array}{rcl} C(u) & = & y \text{ in } Y \\ A(x,u) & = & f \text{ in } W^* \end{array} \quad \Leftrightarrow \quad \mathbf{F}(x,u) = \mathbf{y}$$

ALPEN-ADRIA UNIVERSITÄT KLAGENFURT | WIEN GRAZ

Abstract Formulation

Identify parameter x in PDE

$$A(x, u) = f$$

from measurements of the state u

$$C(u) = y\,,$$

where $x \in Y$, $u \in V$, $y \in Y$, $X, V, Y \ldots$ Banach spaces
$A : X \times V \to W^* \ldots$ differential operator
$C : V \to Y \ldots$ observation operator

(a) reduced approach: operator equation for x

$$F(x) = y,$$

$F = C \circ S$ with $S : X \to V$, $x \mapsto u$ parameter-to-state map

(b) all-at once approach: measurements and PDE as system for (x, u)

$$\begin{array}{rcl} C(u) & = & y \text{ in } Y \\ A(x, u) & = & f \text{ in } W^* \end{array} \quad \Leftrightarrow \quad \mathbf{F}(x, u) = \mathbf{y}$$

Nonlinear ill-posed operator equation

$$F(x) = y$$

$F : \mathcal{D}(F)(\subseteq X) \to Y$ ... nonlinear operator;
F not continuously invertible;

X, Y ... Banach spaces;

$y^\delta \approx y$ ... noisy data, $\|y^\delta - y\| \leq \delta$... noise level.

$\rightsquigarrow$ regularization necessary

Motivation for working in Banach space

- $X = L^P$ with $P = 1$ $\rightsquigarrow$ sparse solutions
- $Y = L^R$ with $R = 1$ $\rightsquigarrow$ impulsive noise
- $Y = L^R$ with $R = \infty$ $\rightsquigarrow$ uniform noise

ALPEN-ADRIA UNIVERSITÄT KLAGENFURT | WIEN GRAZ

Motivation for working in Banach space

- $X = L^P$ with $P = 1$ $\rightsquigarrow$ sparse solutions
- $Y = L^R$ with $R = 1$ $\rightsquigarrow$ impulsive noise
- $Y = L^R$ with $R = \infty$ $\rightsquigarrow$ uniform noise
- $X = L^P$ with $P = \infty$ $\rightsquigarrow$ ellipticity and boundedness in the context of parameter id. in PDEs (e.g. $\nabla(a\nabla u) = 0$); avoid artificial increase of ill-posedness, that would result from a Hilbert space choice $X = H^{d/2+\epsilon}$

ALPEN-ADRIA UNIVERSITÄT KLAGENFURT | WIEN GRAZ

Motivation for working in Banach space

- $X = L^P$ with $P = 1$ $\rightsquigarrow$ sparse solutions
- $Y = L^R$ with $R = 1$ $\rightsquigarrow$ impulsive noise
- $Y = L^R$ with $R = \infty$ $\rightsquigarrow$ uniform noise
- $X = L^P$ with $P = \infty$ $\rightsquigarrow$ ellipticity and boundedness in the context of parameter id. in PDEs (e.g. $\nabla(a\nabla u) = 0$); avoid artificial increase of ill-posedness, that would result from a Hilbert space choice $X = H^{d/2+\epsilon}$
- $Y = L^R$ with $R = \infty$ $\rightsquigarrow$ realistic measurement noise model; avoid artificial increase of ill-posedness, that would result from a Hilbert space choice $Y = L^2$

ALPEN-ADRIA UNIVERSITÄT KLAGENFURT | WIEN GRAZ

Some Banach space tools (I)

Smoothness

- X ... smooth $\iff$ norm Gâteaux differentiable on $X \setminus \{0\}$;
- X ... uniformly smooth $\iff$ norm Fréchet differentiable on unit sphere;

Convexity

- X ... strictly convex $\iff$ boundary of unit ball contains no line segment;
- X ... uniformly convex $\iff$ modulus of convexity $\delta_X(\epsilon) > 0 \ \forall \epsilon \in (0, 2]$;

$\delta_X(\epsilon) = \inf\{1 - \left\|\frac{1}{2}(x+y)\right\| : \|x\| = \|y\| = 1\,,\ \|x-y\| \geq \epsilon\}$

$L^P(\Omega)$, $P \in (1, \infty)$ is uniformly convex (Hanner's ineq.) and uniformly smooth

Some Banach space tools (II)

- **Dual space:**

  $X^* = L(X, \mathbb{R}) \dots$ bounded linear functionals on X
  $x^* : x \mapsto \langle x^*, x \rangle$
- X uniformly smooth $\Leftrightarrow$ X^* uniformly convex
- X reflexive: X smooth $\Leftrightarrow$ X^* strictly convex

Some Banach space tools (III)

- **Duality mapping:**

  $J_p : \ X \to 2^{X^*}$,
  $J_p(x) = \{x^* \in X^* \ : \ \langle x^*, x\rangle = \|x^*\| \, \|x\| = \|x\|^p\}$
  J_p set valued;
  j_p ... single valued selection of J_p;
- $J_p = \partial \frac{1}{p} \|\cdot\|^p$ (Asplund)
- X smooth $\Leftrightarrow$ J_p single valued
- X reflexive, smooth, strictly convex $\Rightarrow$ $J_p^{-1} = J^*_{\frac{p}{p-1}}$

Some Banach space tools (III)

- **Duality mapping:**

  $J_p: \ X \to 2^{X^*}$,
  $J_p(x) = \{x^* \in X^* \ : \ \langle x^*, x\rangle = \|x^*\| \, \|x\| = \|x\|^p\}$
  J_p set valued;
  j_p ... single valued selection of J_p;
- $J_p = \partial \frac{1}{p} \|\cdot\|^p$ (Asplund)
- X smooth $\Leftrightarrow$ J_p single valued
- X reflexive, smooth, strictly convex $\Rightarrow$ $J_p^{-1} = J^*_{\frac{p}{p-1}}$

$L^P(\Omega)$: $\quad J_p(x) = \|x\|_{L_P}^{p-P} |x|^{P-1} \text{sign}(x)$

$W^{1,P}(\Omega)$: $\quad J_p(x) = \|x\|_{W^{1,P}}^{p-P} \left(-\nabla(|\nabla x|^{P-2}\nabla x) + |x|^{P-1}\text{sign}(x)\right)$

$P \in (1,\infty)$, $\frac{\partial x}{\partial n} = 0$ on $\partial\Omega$

Some Banach space tools (III)

- **Duality mapping:**

  $J_p : \; X \to 2^{X^*}$,
  $J_p(x) = \{x^* \in X^* \; : \; \langle x^*, x\rangle = \|x^*\| \, \|x\| = \|x\|^p\}$
  J_p set valued;
  j_p ... single valued selection of J_p;
- $J_p = \partial \frac{1}{p} \|\cdot\|^p$ (Asplund)
- X smooth $\Leftrightarrow$ J_p single valued
- X reflexive, smooth, strictly convex $\Rightarrow$ $J_p^{-1} = J^*_{\frac{p}{p-1}}$

$L^P(\Omega)$: $\quad J_p(x) = \|x\|_{L_P}^{p-P} |x|^{P-1} \mathrm{sign}(x)$

$W^{1,P}(\Omega)$: $\quad J_p(x) = \|x\|_{W^{1,P}}^{p-P} \left(-\nabla(|\nabla x|^{P-2} \nabla x) + |x|^{P-1} \mathrm{sign}(x) \right)$

$P \in (1, \infty)$, $\frac{\partial x}{\partial n} = 0$ on $\partial\Omega$

$\rightsquigarrow$ J_P possibly nonlinear and nonsmooth

Some Banach space tools (IV)

Bregman distance:

$$D_p(x,\tilde{x}) = \frac{1}{p}\|x\|^p - \frac{1}{p}\|\tilde{x}\|^p - \inf\{\langle \xi, \tilde{x}-x\rangle \; : \; \xi \in J_p(\tilde{x})\}$$

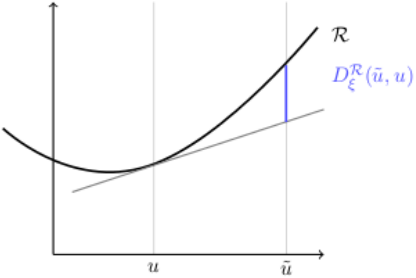
smooth and uniformly convex X:
convergence in D_p $\Leftrightarrow$ convergence in $\|\cdot\|$

Hilbert space case: $D_2(x,\tilde{x}) = \frac{1}{2}\|x-\tilde{x}\|^2$

Bregman distance
wrt general convex functional
$\mathcal{R} : X \to [-\infty, +\infty]$:

$$D_\xi(x,\tilde{x}) = \mathcal{R}(x) - \mathcal{R}(\tilde{x}) - \langle \xi, \tilde{x}-x\rangle$$

with $\xi \in \partial\mathcal{R}(\tilde{x})$.



Regularization in Banach Space: Variational Methods

Tikhonov Regularization in Banach space

$$F(x) = y$$

$$x_\alpha \in \operatorname{argmin}_{x \in \mathcal{D}(F)} \left\| F(x) - y^\delta \right\|^r + \alpha \left\| x - x_0 \right\|^p ,$$

$p, r \in (1, \infty)$, and x_0 ... initial guess;

more generally

$$x_\alpha \in \operatorname{argmin}_{x \in \mathcal{D}(F)} \mathcal{S}(F(x), y^\delta) + \alpha \mathcal{R}(x)$$

with convex functionals $\mathcal{S}$, $\mathcal{R}$.

Assumptions

- X, Y Banach spaces
- additionally to norm topologies: weaker topologies τ_X, τ_Y
- $\mathcal{S}$ symmetric and satisfies generalized triangle inequality $\mathcal{S}(y_1, y_3)^{1/r} \leq C_{\mathcal{S}}(\mathcal{S}(y_1, y_2) + \mathcal{S}(y_2, y_3))$
- For all $(y_n)_{n\in\mathbb{N}} \subseteq Y$, $y, \tilde{y} \in Y$:
  - $\mathcal{S}(y, y_n) \overset{n\to\infty}{\to} 0 \Rightarrow y_n \overset{n\to\infty}{\to} y$ wrt τ_Y
  - $\mathcal{S}(y, y_n) \overset{n\to\infty}{\to} 0 \ \Rightarrow \mathcal{S}(\tilde{y}, y_n) \overset{n\to\infty}{\to} \mathcal{S}(\tilde{y}, y)$
- $(y, \tilde{y}) \mapsto \mathcal{S}(y, \tilde{y})$ is lower semicontinuous wrt. τ_Y^2
- $F : \mathcal{D}(F) \subset X \to Y$ is continuous wrt τ_X, τ_Y.
- $\mathcal{R} : X \to [0, +\infty]$ is proper, convex, τ_X lower semicont.
- $\mathcal{D} := \mathcal{D}(F) \cap \mathcal{D}(\mathcal{R}) \neq \emptyset$ is closed wrt τ_X.
- $\forall \alpha > 0, M > 0$, the sublevel set $\mathcal{M}_\alpha(M)$ $= \{x \in \mathcal{D} : \mathcal{T}_\alpha(x) \leq M\}$ is τ_X compact.

ALPEN-ADRIA UNIVERSITÄT KLAGENFURT | WIEN GRAZ

Assumptions

- X, Y Banach spaces
- additionally to norm topologies: weaker topologies τ_X, τ_Y
- $\mathcal{S}$ symmetric and satisfies generalized triangle inequality $\mathcal{S}(y_1,y_3)^{1/r} \leq C_\mathcal{S}(\mathcal{S}(y_1,y_2)+\mathcal{S}(y_2,y_3))$
- For all $(y_n)_{n\in\mathbb{N}} \subseteq Y$, $y,\tilde{y} \in Y$:
  - $\mathcal{S}(y,y_n) \overset{n\to\infty}{\to} 0 \Rightarrow y_n \overset{n\to\infty}{\to} y$ wrt τ_Y
  - $\mathcal{S}(y,y_n) \overset{n\to\infty}{\to} 0 \ \Rightarrow \mathcal{S}(\tilde{y},y_n) \overset{n\to\infty}{\to} \mathcal{S}(\tilde{y},y)$
- $(y,\tilde{y}) \mapsto \mathcal{S}(y,\tilde{y})$ is lower semicontinuous wrt. τ_Y^2
- $F:\mathcal{D}(F)\subset X \to Y$ is continuous wrt τ_X, τ_Y.
- $\mathcal{R}: X \to [0,+\infty]$ is proper, convex, τ_X lower semicont.
- $\mathcal{D} := \mathcal{D}(F)\cap\mathcal{D}(\mathcal{R}) \neq \emptyset$ is closed wrt τ_X.
- $\forall\alpha>0, M>0$, the sublevel set $\mathcal{M}_\alpha(M)$ $=\{x\in\mathcal{D} : \mathcal{T}_\alpha(x)\leq M\}$ is τ_X compact.

. . . are satisfied, e.g., for $\mathcal{S}(y,\tilde{y}) = \|y-\tilde{y}\|^r$, $\mathcal{R}(x) = \|x-x_0\|^p$

ALPEN-ADRIA UNIVERSITÄT KLAGENFURT | WIEN GRAZ

Well-posedness

Theorem (Well-definedness; Hofmann&BK&Pöschl&Scherzer'07)

For any $\alpha > 0$, $y^\delta \in Y$ there exists a minimizer of $T_\alpha = \mathcal{S}(F(\cdot), y^\delta) + \alpha\mathcal{R}(\cdot)$.

Theorem (Stability; HKPS07)

For any $\alpha > 0$, minimizers $x_\alpha(y^\delta)$ of T_α are subsequentially τ_X stable wrt the data $y^\delta \in Y$, i.e.

$$\mathcal{S}(y^\delta, y_n) \overset{n\to\infty}{\to} 0 \;\Rightarrow\; x_\alpha(y_n) \overset{n\to\infty}{\to} x_\alpha(y^\delta) \text{ wrt } \tau_X$$

if $x_\alpha(y^\delta)$ is unique.
Moreover, $\mathcal{R}(x_\alpha(y_n)) \overset{n\to\infty}{\to} \mathcal{R}(x_\alpha(y^\delta))$.

ALPEN-ADRIA UNIVERSITÄT KLAGENFURT | WIEN GRAZ

Theorem (Convergence; HKPS'07)

Assume that a solution x of $F(x) = y$ exists. Then there exists an $\mathcal{R}$-minimizing solution $x^\dagger$.

Let the regularization parameter α be chosen in dependence of the noise level $\delta \geq \mathcal{S}(y, y^\delta)^{1/r}$, $r > 1$, according to

$$\alpha \to 0 \text{ and } \frac{\delta^r}{\alpha} \to 0 \text{ as } \delta \to 0$$

Then minimizers $x_\alpha^\delta = x_\alpha(y^\delta)$ of T_α converge τ_X-subsequentially to $x^\dagger$, i.e.

$$\mathcal{S}(y_n^\delta, y) \overset{n\to\infty}{\to} 0 \;\Rightarrow\; x_\alpha^\delta \overset{n\to\infty}{\to} x^\dagger \text{ wrt } \tau_X$$

if $x^\dagger$ is unique.

ALPEN-ADRIA UNIVERSITÄT KLAGENFURT | WIEN GRAZ

Theorem (Convergence Rates; HKPS07)

Assume existence of an $\mathcal{R}$-minimizing solution $x^\dagger$ of $F(x) = y$, satisfying the source condition $\partial\mathcal{R}(x^\dagger) \cap \text{range}(F'(x^\dagger)^*) \neq \emptyset$, i.e.,

$$\exists \xi \in \partial\mathcal{R}(x^\dagger) \;\; w \in Y^* \; : \; \xi = F'(x^\dagger)^* w$$

and F' satisfies the Lipschitz type condition

$$\|F'(x^\dagger)(x - x^\dagger)\| - c_2 \mathcal{S}(F(x), F(x^\dagger))^{1/r} \leq c_1 D_\xi(x, x^\dagger) \quad x \in \mathcal{D}(F)$$

for some $c_1, c_2 > 0$, $\xi \in \partial\mathcal{R}(x^\dagger)$, $r > 1$ with $c_1\|w\| < 1$.
Let the regularization parameter α be chosen in dependence of the noise level $\delta \geq \mathcal{S}(y, y^\delta)^{1/r}$ according to

$$\alpha \sim \delta^{r-1}$$

Then minimizers x_α^δ of T_α satisfy

$$\mathcal{S}(F(x_\alpha^\delta), y)^{1/r} = O(\delta), \quad D_\xi(x_\alpha^\delta, x^\dagger) = O(\delta)$$

Idea of proof: $s_\alpha^\delta := \mathcal{S}(F(x_\alpha^\delta), y^\delta)$, $d_\alpha^\delta = D_\xi(x_\alpha^\delta, x^\dagger)$

Idea of proof: $s_\alpha^\delta := \mathcal{S}(F(x_\alpha^\delta), y^\delta)$, $d_\alpha^\delta = D_\xi(x_\alpha^\delta, x^\dagger)$

- minimality of $x_\alpha(y^\delta)$ $\Rightarrow$

$$\delta^r + \alpha\mathcal{R}(x^\dagger) = T_\alpha(x^\dagger) \geq T_\alpha(x_\alpha^\delta) = s_\alpha^\delta + \alpha\mathcal{R}(x_\alpha^\delta)$$

Idea of proof: $s_\alpha^\delta := \mathcal{S}(F(x_\alpha^\delta), y^\delta)$, $d_\alpha^\delta = D_\xi(x_\alpha^\delta, x^\dagger)$

- minimality of $x_\alpha(y^\delta)$ $\Rightarrow$

$$\delta^r + \alpha\mathcal{R}(x^\dagger) = T_\alpha(x^\dagger) \geq T_\alpha(x_\alpha^\delta) = s_\alpha^\delta + \alpha\mathcal{R}(x_\alpha^\delta)$$

- definition of Bregman distance and source condition

$$\begin{aligned}\delta^r \geq & s_\alpha^\delta + \alpha(\mathcal{R}(x_\alpha^\delta) - \mathcal{R}(x^\dagger)) \\ = & s_\alpha^\delta + \alpha(d_\alpha^\delta + \langle \xi, x^\dagger - x_\alpha(y^\delta)\rangle) \\ = & s_\alpha^\delta + \alpha(d_\alpha^\delta + \langle F'(x^\dagger)^* w, x^\dagger - x_\alpha^\delta\rangle)\end{aligned}$$

ALPEN-ADRIA UNIVERSITÄT KLAGENFURT | WIEN GRAZ

Idea of proof: $s_\alpha^\delta := \mathcal{S}(F(x_\alpha^\delta), y^\delta)$, $d_\alpha^\delta = D_\xi(x_\alpha^\delta, x^\dagger)$

- minimality of $x_\alpha(y^\delta)$ $\Rightarrow$

$$\delta^r + \alpha\mathcal{R}(x^\dagger) = T_\alpha(x^\dagger) \geq T_\alpha(x_\alpha^\delta) = s_\alpha^\delta + \alpha\mathcal{R}(x_\alpha^\delta)$$

- definition of Bregman distance and source condition

$$\begin{aligned}\delta^r \geq & s_\alpha^\delta + \alpha(\mathcal{R}(x_\alpha^\delta) - \mathcal{R}(x^\dagger))\\ = & s_\alpha^\delta + \alpha(d_\alpha^\delta + \langle \xi, x^\dagger - x_\alpha(y^\delta)\rangle)\\ = & s_\alpha^\delta + \alpha(d_\alpha^\delta + \langle F'(x^\dagger)^* w, x^\dagger - x_\alpha^\delta\rangle)\end{aligned}$$

- use condition on F, triangle and Young's inequality

$$\begin{aligned}\delta^r \geq & s_\alpha^\delta + \alpha(d_\alpha^\delta + \langle w, F'(x^\dagger)(x^\dagger - x_\alpha^\delta)\rangle)\\ \geq & s_\alpha^\delta + \alpha d_\alpha^\delta - \alpha\|w\|\left(c_1 d_\alpha^\delta + c_2\mathcal{S}(F(x_\alpha^\delta), F(x^\dagger))^{1/r}\right)\\ \geq & s_\alpha^\delta + (1 - c_1\|w\|)\alpha d_\alpha^\delta - \tfrac{1}{2}s_\alpha^\delta - C_1\alpha^{r^*} - C_2\alpha\delta\end{aligned}$$

ALPEN-ADRIA UNIVERSITÄT KLAGENFURT I WIEN GRAZ

Idea of proof: $s_\alpha^\delta := \mathcal{S}(F(x_\alpha^\delta), y^\delta)$, $d_\alpha^\delta = D_\xi(x_\alpha^\delta, x^\dagger)$

- minimality of $x_\alpha(y^\delta)$ $\Rightarrow$

$$\delta^r + \alpha\mathcal{R}(x^\dagger) = T_\alpha(x^\dagger) \geq T_\alpha(x_\alpha^\delta) = s_\alpha^\delta + \alpha\mathcal{R}(x_\alpha^\delta)$$

- definition of Bregman distance and source condition

$$\begin{aligned}\delta^r \geq & s_\alpha^\delta + \alpha(\mathcal{R}(x_\alpha^\delta) - \mathcal{R}(x^\dagger)) \\ = & s_\alpha^\delta + \alpha(d_\alpha^\delta + \langle \xi, x^\dagger - x_\alpha(y^\delta)\rangle) \\ = & s_\alpha^\delta + \alpha(d_\alpha^\delta + \langle F'(x^\dagger)^* w, x^\dagger - x_\alpha^\delta\rangle)\end{aligned}$$

- use condition on F, triangle and Young's inequality

$$\begin{aligned}\delta^r \geq & s_\alpha^\delta + \alpha(d_\alpha^\delta + \langle w, F'(x^\dagger)(x^\dagger - x_\alpha^\delta)\rangle) \\ \geq & s_\alpha^\delta + \alpha d_\alpha^\delta - \alpha\|w\|\left(c_1 d_\alpha^\delta + c_2\mathcal{S}(F(x_\alpha^\delta), F(x^\dagger))^{1/r}\right) \\ \geq & s_\alpha^\delta + (1 - c_1\|w\|)\alpha d_\alpha^\delta - \tfrac{1}{2}s_\alpha^\delta - C_1\alpha^{r^*} - C_2\alpha\delta\end{aligned}$$

- combine with $\alpha \sim \delta^{r-1} = \delta^{r/r^*}$ to

$$C\delta^r \geq \tfrac{1}{2}s_\alpha^\delta + (1 - c_1\|w\|)\delta^{r/r^*} d_\alpha^\delta$$

Theorem (Convergence Rates; HKPS07)

Assume existence of an $\mathcal{R}$-minimizing solution $x^\dagger$ of $F(x) = y$, satisfying the source condition $\partial\mathcal{R}(x^\dagger) \cap \text{range}(F'(x^\dagger)^*) \neq \emptyset$, i.e.,

$$\exists \xi \in \partial\mathcal{R}(x^\dagger) \;\; w \in Y^* \; : \; \xi = F'(x^\dagger)^* w$$

and F' satisfies the Lipschitz type condition

$$\|F'(x^\dagger)(x - x^\dagger)\| - c_2 \mathcal{S}(F(x), F(x^\dagger))^{1/r} \leq c_1 D_\xi(x, x^\dagger) \quad x \in \mathcal{D}(F)$$

for some $c_1, c_2 > 0$, $\xi \in \partial\mathcal{R}(x^\dagger)$, $r > 1$ with $c_1\|w\| < 1$.
Let the regularization parameter α be chosen in dependence of the noise level $\delta \geq \mathcal{S}(y, y^\delta)^{1/r}$ according to

$$\alpha \sim \delta^{r-1}$$

Then minimizers x_α^δ of T_α satisfy

$$\mathcal{S}(F(x_\alpha^\delta), y)^{1/r} = O(\delta)\,, \quad D_\xi(x_\alpha^\delta, x^\dagger) = O(\delta)$$

Variational methods in Banach space: Further References

- case $Y = X$:
  [Plato'92,'94,'95, Bakushinskii&Kokurin'04]
- general X, Y:
  [Burger&Osher'04, Resmerita&Scherzer'06, Pöschl'08, Flemming'11, Hofmann&BK&Pöschl&Scherzer'07, Grasmair&Haltmeier&Scherzer'08,'10, Neubauer&Hein&Hofmann&Kindermann&Tautenhahn'10, Grasmair'10,'11, ...]

Regularization in Banach Space: Iterative Methods

Assumptions on pre-image and image space X, Y

- X smooth, uniformly convex
  $\Rightarrow$ X reflexive (Milman-Pettis) and strictly convex
  J_p single valued, norm-to-weak-continuous, bijective
- Y arbitrary Banach space
- some conditions on F

Iteratively Regularized Gauss-Newton Method (IRGNM) in Banach space

$$F(x) = y$$

$$x_{k+1} \in \mathrm{argmin}_{x \in \mathcal{D}(F)} \left\| F'(x_k)(x - x_k) + F(x_k) - y^\delta \right\|^r + \alpha_k \left\| x - x_0 \right\|^p ,$$

$p, r \in (1, \infty)$, and x_0 ... initial guess;

convex minimization problem:
efficient solution see, e.g., [Bonesky, Kazimierski, Maass, Schöpfer, Schuster'07]

for comparison: IRGNM in Hilbert space, $r = p = 2$:

$$x_{k+1} = x_k - (F'(x_k)^* F'(x_k) + \alpha_k I)^{-1} F'(x_k)^* \Big(F(x_k) - y^\delta + \alpha_k (x_k - x_0) \Big)$$

ALPEN-ADRIA UNIVERSITÄT KLAGENFURT | WIEN GRAZ

Stopping rule

discrepancy principle

$$k_*(\delta) = \min\{k \in \mathbb{N} \ : \ \left\| F(x_k) - y^\delta \right\| \leq C_{dp}\delta\}$$

$C_{dp} > 1, \ \ \|y^\delta - y\| \leq \delta \ldots$ noise level

Trade off between stability and approximation:

Stop as early as possible (stability) such that
the resudual is lower than the noise level (approximation)

ALPEN-ADRIA UNIVERSITÄT KLAGENFURT | WIEN GRAZ

Choice of α_k

discrepancy type principle:

$$\underline{\theta} \left\| F(x_k) - y^\delta \right\| \leq \left\| F'(x_k)(x_{k+1}(\alpha) - x_k) + F(x_k) - y^\delta \right\| \leq \overline{\theta} \left\| F(x_k) - y^\delta \right\|$$

$$0 < \underline{\theta} < \overline{\theta} < 1$$

Trade off between stability and approximation:

Choose α_k as large as possible (stability) such that the predicted residual is smaller than the old one (approximation)

see also: inexact Newton (for inverse problems: [Hanke'97, Rieder'99,'01])

Convergence of the IRGN

Theorem (BK&Schöpfer&Schuster'09)

For all $k \leq k_*(\delta) - 1$ the iterates

$$x_{k+1} \in \operatorname{argmin} \left\| F'(x_k)(x - x_k) + F(x_k) - y^\delta \right\|^r + \alpha_k \|x - x_0\|^p$$

$$\alpha_k \text{ s.t. } \left\| F'(x_k)(x_{k+1} - x_k) + F(x_k) - y^\delta \right\| \sim \theta \left\| F(x_k) - y^\delta \right\|$$

are well-defined and

$$x_{k_*(\delta)} \to x^\dagger \quad \text{solution to } F(x) = y \quad \text{as } \delta \to 0$$

if $x^\dagger$ unique, (and along subsequences otherwise).

Convergence rates for the IRGN

Theorem (BK&Hofmann'10)

Under the source condition $J_p(x^\dagger - x_0) \cap range(F'(x^\dagger)^*) \neq \emptyset$, *i.e.* ,

$$\exists \xi \in J_p(x^\dagger - x_0), \; w \in Y^* \; : \; \xi = F'(x^\dagger)^* w$$

x_{k_*} *converges at optimal rates*

$$D_p(x_{k_*} - x_0, x^\dagger - x_0) = O(\delta),$$

for comparison:
Hilbert space case: $\|x_{k_*} - x^\dagger\| = O(\sqrt{\delta})$, i.e., $D_p(x_{k_*} - x_0, x^\dagger - x_0) = O(\delta)$

ALPEN-ADRIA UNIVERSITÄT KLAGENFURT | WIEN GRAZ

Remarks

- rates result can be extended to $D_p^{x_0}(x_{k_*}, x^\dagger) = O(\delta^{\frac{4\nu}{2\nu+1}})$ with $\nu \in (0, \frac{1}{2})$ under regularity assumptions (variational source conditions);

$$\exists\, \beta > 0\ \forall x \in \mathcal{B} :$$
$$|\langle J_p(x^\dagger - x_0), x - x^\dagger \rangle| \le \beta D_p^{x_0}(x^\dagger, x)^{\frac{1-2\nu}{2}} \|F'(x^\dagger)(x - x^\dagger)\|^{2\nu} .$$

- logarithmic rates under logarithmic source conditions
- rates can be shown alternatively with *a priori* choice of α_k and k_* instead of the discrepancy principle; needs a priori information on smoothness index ν of $x^\dagger$, though

ALPEN-ADRIA UNIVERSITÄT KLAGENFURT | WIEN GRAZ

Newton-Landweber iterations

Combine outer Newton iteration with
inner reg. Landweber iteration (in place of Tikhonov)

For $k = 0, 1, 2 \ldots k_*$ do (Newton)

$$u_{k,0} = 0$$

$$z_{k,0} = x_k$$

For $n = 0, 1, 2 \ldots, n_k$ do (Landweber)

$$u_{k,n+1} = u_{k,n} - \alpha_{k,n}(J_p^X(x_k - x_0) + u_{k,n})$$
$$-\omega_{k,n} F'(x_k)^* j_r^Y(F'(x_k)(z_{k,n} - x_k) - y^\delta + F(x_k))$$

$$z_{k,n+1} = x_0 + J_p^{X^{-1}}\left(u_{k,n+1} + J_p^X(x_k - x_0)\right)$$

$$x_{k+1} = z_{k,n_k} ,$$

Newton-Landweber iterations

Combine outer Newton iteration with
inner reg. Landweber iteration (in place of Tikhonov)

Hilbert space case:

For $k = 0, 1, 2 \ldots k_*$ do (Newton)

$$u_{k,0} = 0$$

$$z_{k,0} = x_k$$

For $n = 0, 1, 2 \ldots, n_k$ do (Landweber)

$$u_{k,n+1} = u_{k,n} - \underbrace{\alpha_{k,n}(u_{k,n} + x_k - x_0)}_{\text{additional regularization}}$$

$$-\omega_{k,n} F'(x_k)^* (\underbrace{F'(x_k) u_{k,n} - y^\delta + F(x_k)}_{\text{Newton step residual}})$$

$$x_{k+1} = x_k + u_{k,n_k} \,,$$

ALPEN-ADRIA UNIVERSITÄT KLAGENFURT | WIEN GRAZ

Newton-Landweber iterations

Combine outer Newton iteration with
inner reg. Landweber iteration (in place of Tikhonov)

For $k = 0, 1, 2 \dots k_*$ do (Newton)

$$u_{k,0} = 0$$

$$z_{k,0} = x_k$$

For $n = 0, 1, 2 \dots, n_k$ do (Landweber)

$$u_{k,n+1} = u_{k,n} - \alpha_{k,n}(J_p^X(x_k - x_0) + u_{k,n}) \\ -\omega_{k,n} F'(x_k)^* j_r^Y (F'(x_k)(z_{k,n} - x_k) - y^\delta + F(x_k))$$

$$z_{k,n+1} = x_0 + {J_p^X}^{-1}\left(u_{k,n+1} + J_p^X(x_k - x_0)\right)$$

$$x_{k+1} = z_{k,n_k} \,,$$

Newton-Landweber iterations

Combine outer Newton iteration with
inner reg. Landweber iteration (in place of Tikhonov)

For $k = 0, 1, 2 \ldots k_*$ do (Newton)

$$u_{k,0} = 0$$

$$z_{k,0} = x_k$$

For $n = 0, 1, 2 \ldots, n_k$ do (Landweber)

$$u_{k,n+1} = u_{k,n} - \alpha_{k,n}(J_p^X(x_k - x_0) + u_{k,n}) - \omega_{k,n} F'(x_k)^* j_r^Y (F'(x_k)(z_{k,n} - x_k) - y^\delta + F(x_k))$$

$$z_{k,n+1} = x_0 + {J_p^X}^{-1} \Big(u_{k,n+1} + J_p^X(x_k - x_0) \Big)$$

$$x_{k+1} = z_{k,n_k} \, ,$$

choice of parameters $\alpha_{k,n}$, $\omega_{k,n}$, n_k, k_* ?

ALPEN-ADRIA UNIVERSITÄT KLAGENFURT | WIEN GRAZ

IRGNM-Landweber Algorithm

Algorithm (Newton – it. reg. Landweber method)

Choose $C_{dp}, \tilde{\tau}, C_\alpha$ *sufficiently large,* x_0 *sufficiently close to* $x^\dagger$,
$\alpha_{00} \leq 1$, $\vartheta > 0$ *sufficiently small,* $\overline{\omega} > 0$, $(a_n)_{n \in \mathbb{N}_0}$ *s. t.* $\sum_{n=0}^\infty a_n < \infty$
For $k = 0, 1, 2 \ldots$ *until* $\mathrm{r}_k \leq C_{dp}\delta$ *do*

$$u_{k,0} = 0$$
$$z_{k,0} = x_k$$
$$\alpha_{k,0} = \alpha_{k-1,n_{k-1}} \text{ if } k > 0$$

For $n = 0, 1, 2 \ldots$ *until* $\left\{ \begin{array}{ll} n = n_k - 1 = a_k \mathrm{r}_k^{-r} & \text{if } \mathrm{r}_k > C_{dp}\delta \\ \alpha_{k_*, n^*_{k_*}} \leq C_\alpha (\mathrm{r}_{k_*} + \delta)^{\frac{r}{1+\theta}} & \text{if } \mathrm{r}_k \leq C_{dp}\delta \end{array} \right\}$

$$\omega_{k,n} = \vartheta \min\{\mathrm{t}_{k,n}^{r(s-1)} \tilde{\mathrm{t}}_{k,n}^{-s}, \; \mathrm{t}_{k,n}^{r(p-1)} \tilde{\mathrm{t}}_{k,n}^{-p}, \; \overline{\omega}\}$$
$$u_{k,n+1} = u_{k,n} - \alpha_{k,n} J_p^X (z_{k,n} - x_0) - \omega_{k,n} F'(x_k)^* j_r^Y (F'(x_k)(z_{k,n} - x_k) + F(x_k) - y^\delta)$$
$$z_{k,n+1} = x_0 + J_p^{X\,-1} \Big(J_p^X (x_k - x_0) + u_{n,k+1} \Big)$$
$$\alpha_{k,n+1} = \max\{\check{\alpha}_{k,n+1}, \hat{\alpha}_{k,n+1}\} \text{ with } \check{\alpha}_{k,n+1}, \hat{\alpha}_{k,n+1}$$

$$x_{k+1} = z_{k,n_k} .$$

$$
\begin{aligned}
\mathrm{t}_{n,k} &= \|F'(x_k)(z_{n,k} - x_k) + F(x_k) - y^\delta\| \\
\tilde{\mathrm{t}}_{n,k} &= \|F'(x_k)^* j_r^Y (F'(x_k)(z_{n,k} - x_k) + F(x_k) - y^\delta)\| \\
\mathrm{r}_n &= \|F(x_k) - y^\delta\| \\
\theta &= \frac{4\nu}{r(1+2\nu) - 4\nu} \\
\check{\alpha}_{n,k} &= \tilde{\tau}\left(\mathrm{t}_{n,k} + \eta \mathrm{r}_n + (1+\eta)\delta\right)^{\frac{r}{1+\theta}} \\
\hat{\alpha}_{n,k+1} &= \alpha_{n,k}\Big(1 - (1-q)\alpha_{n,k}\Big)^{1/\theta}
\end{aligned}
$$

ν. . . smoothness index according to (approximate) source condition

$\rightsquigarrow$ a priori parameter choice

a posteriori choice of n_k might spoil strong convergence

ALPEN-ADRIA UNIVERSITÄT KLAGENFURT | WIEN GRAZ

Convergence of the IRGNM-Landweber

Theorem (BK&Tomba'12)

Assume: X smooth and s-convex with $r \geq s \geq p$, $s \leq 1 + \frac{1}{\theta}$ Then the iterates according to the above algorithm are well-defined and converge strongly

$$x_{k_*(\delta)} \to x^\dagger \quad \text{solution to } F(x) = y \quad \text{as } \delta \to 0$$

if $x^\dagger$ unique, (and along subsequences otherwise).
Optimal rates under variational source conditions.

Iterative methods in Banach space: Further References

- Newton-Landweber with $\alpha_{n,k} \equiv 0$
  [Q.Jin'12]
- Landweber (without outer Newton loop):
  [Schöpfer&Louis&Schuster'06, Hein&Kazimierski'10,
  BK&Schöpfer&Schuster'09, [Hein&Kazimierski'10], [BK'12]
- iterated Tikhonov regularization:
  [Q.Jin&Stals'12]
- ...

Numerical tests with IRGNM-Landweber

c-example:

$$-\Delta u + c\, u = 0 \text{ in } \Omega\,.$$

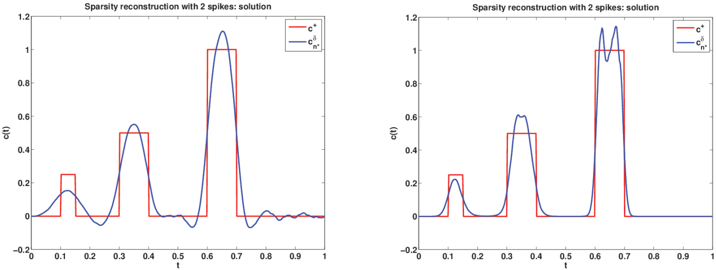
Identify c from measurements of u in Ω.

$\Omega = (0,1) \subseteq \mathbb{R}$, $\quad \Omega = (0,1)^2 \subseteq \mathbb{R}^2$,

sparse c, Gaussian noise
smooth c, impulsive noise

ALPEN-ADRIA UNIVERSITÄT KLAGENFURT | WIEN GRAZ

1-d, sparse c, Gaussian noise



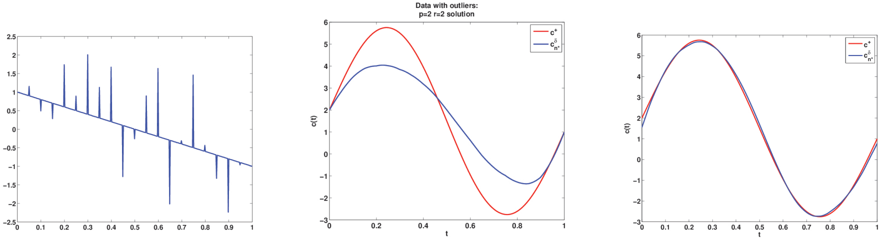
(a) $X = L^2$, $Y = L^2$ (b) $X = L^{1.1}$, $Y = L^2$

Figure: $\delta = 1.e - 3$;

(a): 4141 its (26303 its with $\alpha_{k,n} \equiv 0$), rel.err.:1.e-1
(b): 3110 its (5701 its with $\alpha_{k,n} \equiv 0$), rel.err.:0.5e-1

ALPEN-ADRIA UNIVERSITÄT KLAGENFURT | WIEN GRAZ

1-d, smooth c, impulsive noise

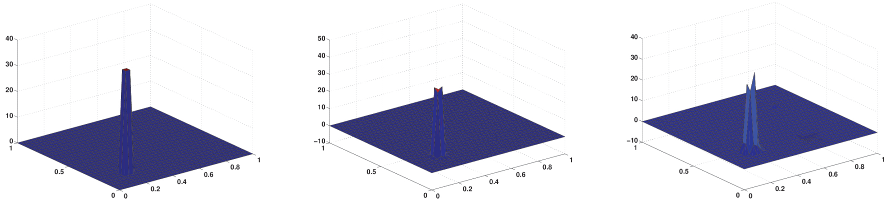


(a) noisy data (b) $X = L^2$, $Y = L^2$ (c) $X = L^2$, $Y = L^{1.1}$

Figure: impulsive noise

(c): 278 its (3285 its with $\alpha_{k,n} \equiv 0$), rel.err.:0.5e-1

2-d, sparse c, Gaussian noise



(a) Exact solution (b) $\delta = 1.e-3$ (c) $\delta = 1.e-2$

Figure: $X = L^{1.1}$, $Y = L^2$

Summary and Outlook

- motivation for solving inverse problems in Banach spaces: more natural norms, possible reduction of ill-posedness, sparsity
- variational methods: generalization of Tikhonov regularization
- iterative methods: Newton, Newton-Landweber,...

→ Regularization by discretization in Banach space [Hämarik&BK&Kangro&Resmerita'14]

Thank you for your attention!