A Double Regularization Approach for Inverse Problems with Noisy Data and Inexact Operator

Ismael Rodrigo Bleyer

Prof. Dr. Ronny Ramlau

Johannes Kepler Universität - Linz

Helsinki - April 18, 2011.

supported by

Der Wissenschaftsfonds.

Overview

- Introduction
- 1st Case: noisy data
- 2nd Case: inexact operator and noisy data
- Proposed method
- Computational aspects
- Numerical illustration
- Conclusions and future work

Overview

Inverse of what?

We call *two problems inverses of one another* if the formulation of each involves all or part of the solution of the other. [J. Keller]

Consider a linear or non-linear operator equation

$$Ax = y \quad \text{or} \quad F(x) = y$$

- **Direct:** given x, compute $y = Ax$ (studied ealier or the easier one)
- **Inverse:** given y, solve $Ax = y$ (if there is solution)

"Inverse problems are concerned with determining causes for a desired or an observed effect" [Engl, Hanke, Naubauer]

Inverse problems most oft do not fulfill **Hadamard**'s postulate of well posedness.

It is called well-posed [Hadamard, 1902] if

- **existence:** for all admissible data, a solution exists;
- **uniqueness:** for all admissible data, a solution is unique;
- **stability:** the solution depends continuously on the data.

Hadamard (1865 – 1963)

Computational issues: observed effect has measurement errors or perturbations caused by noise

Overview

Regularization

Solve $Ax = y_0$ out of the measurement y_δ with $\|y_0 - y_\delta\| \leq \delta$.
Need apply some **regularization** technique

$$\underset{x}{\text{minimize}} \|Ax - y_\delta\|^2 + \alpha \|Lx\|^2.$$

Tikhonov regularization (Tik-R)

- fidelity term (based on LS);
- stabilization term (Hilbert space);
- regularization parameter α.

[Tihonov, 1963 & Phillips, 1962]

Tikhonov (1906 – 1993)

Tikhonov-type regularization

Exchange the quadract term by a general functional $\mathcal{R}$, namely a *proper*, *convex* and *weakly lower semicontinuous functional*:

$$\underset{x}{\text{minimize}} \left\| Ax - y_\delta \right\|^2 + \alpha \mathcal{R}(x)$$

Also called: non-quadract regularization, convex regularization or generalized Tikhonov regularization.

Convergence rates (wrt Bregman distances)

2004	linear	SC type I	Banach – Hilbert	Burger and Osher
2005	linear	SC type II	Banach – Hilbert	Resmerita
2006	nonlinear	SC type I and II	Banach – Banach	Resmerita and Scherzer

Def. Bregman Source Conditions

Subgradient

The *Fenchel subdifferential* of a functional $\mathcal{R} : \mathcal{U} \to [0, +\infty]$ at $\bar{u} \in \mathcal{U}$ is the set

$$\partial^F \mathcal{R}\,(\bar{u}) = \{\xi \in \mathcal{U}^* \mid \mathcal{R}(v) - \mathcal{R}(\bar{u}) \geq \langle \xi\,,\, v - \bar{u} \rangle \;\forall v \in \mathcal{U}\}.$$

First in 1960 by Moreau and Rockafellar and extended by Clark 1973.

Optimality condition:

If $\bar{u}$ minimizes $\mathcal{R}$ then

$$0 \in \partial^F \mathcal{R}\,(\bar{u})$$

Example

Consider the function $\mathcal{R}(u) = |u|$

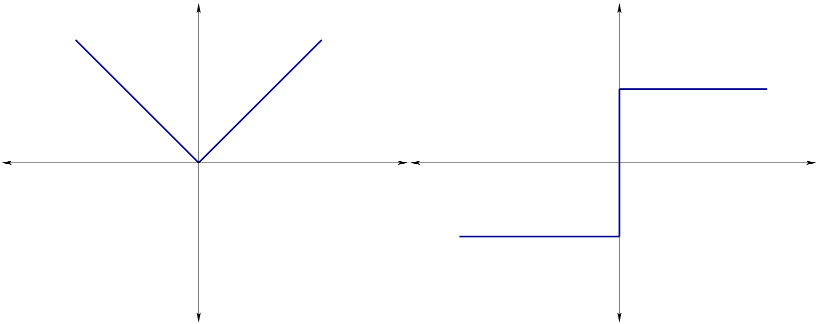


Figure: Function (left) and its subdifferential (right).

Iterative Soft-Shrinkage

Therefore we look the following minimization problem

$$J(k) = \underset{k}{\text{minimize}} \left\| \tilde{F}k - g_\delta \right\|^2 + \beta \mathcal{R}(k).$$

Regularization term: weighted l_p-norm of k wrt an orthonormal basis $\{\phi_\lambda\}_\lambda$ of $L_2(\Omega^2)$,

$$\left\| k \right\|_{w,p}^p = \sum_\lambda w_\lambda |k_\lambda|^p,$$

where $k_\lambda = |\langle k \ , \ \phi_\lambda \rangle|$.

Idea: apply a surrogate functional that removes the term $\tilde{F}^* \tilde{F} k$ Daubechies et al. [2004], adding a functional which depends of an auxiliary element u,

$$\Xi(k; u) = C \left\| k - u \right\|^2 - \left\| \tilde{F}k - \tilde{F}u \right\|^2.$$

Remark: for a suitable choice of C the whole functional is strictly convex.
Therefore the **surrogate functional** - extended functional is

$$\begin{aligned}
J^{\mathrm{Sur}}(k;u) &= J(k) + \Xi(k;u) \\
&= \|\tilde{F}k - g_\delta\|^2 + \beta\|k\|_{w,p}^p + C\|k-u\|^2 - \|\tilde{F}k - \tilde{F}u\|^2 \\
&= \|\tilde{F}k\|^2 + \|g_\delta\|^2 - 2\langle \tilde{F}k \, , \, g_\delta\rangle + \beta\|k\|_{w,p}^p + C\|k\|^2 + C\|u\|^2 \\
&\quad -2C\langle k \, , \, u\rangle - \|\tilde{F}k\|^2 - \|\tilde{F}u\|^2 + 2\langle \tilde{F}k \, , \, \tilde{F}u\rangle \\
&= C\|k\|^2 - 2\langle k \, , \, Cu - \tilde{F}^*(\tilde{F}u - g_\delta)\rangle + \beta\|k\|_{w,p}^p + c_1
\end{aligned}$$

Writing k as a linear combination of an ONB $\{\phi_\lambda\}_\lambda$

$$J^{\mathrm{Sur}}(k;u) = \sum_\lambda C(k_\lambda)^2 - 2k_\lambda\left(Cu - \tilde{F}^*(\tilde{F}u - g_\delta)\right)_\lambda + \beta w_\lambda |k_\lambda|^p + c_1$$

Compute the minimizer of $J^{\text{Sur}}(k; u)$ wrt k for a given u. For a choice $p = 1$ the optimality condition (derivative) is

$$2Ck_\lambda = 2\left(Cu - \tilde{F}^*(\tilde{F}u - g_\delta)\right)_\lambda - \beta w_\lambda \operatorname{sgn}(k_\lambda).$$

Under definition of **soft-shrinkage** operator

$$\mathcal{S}_\beta(x) = \max\{\|x\| - \beta, 0\}\frac{x}{\|x\|}$$

or equivalent

$$\mathcal{S}_\beta(x) = \begin{cases} x - \beta\frac{x}{\|x\|} & \text{if } \|x\| > \beta \\ 0 & \text{if } \|x\| \leq \beta \end{cases} \tag{1}$$

we end up

$$k_\lambda = \mathcal{S}_{\frac{w_\lambda}{C}\frac{\beta}{2}}\left(u - \frac{1}{C}[\tilde{F}^*(\tilde{F}u - g_\delta)]_\lambda\right)$$

An iterative approach can be done setting $u = k^n$ and so

$$k^{n+1} = \arg\min_k J^{\text{Sur}}(k; k^n)$$

for a initial guess k^0.

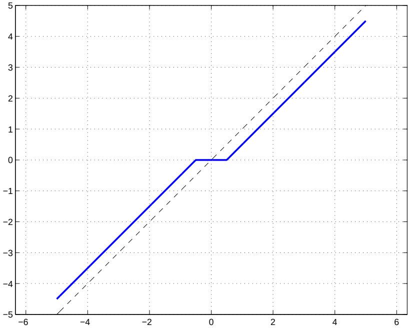


Figure: Soft Shrinkage operator.

Overview

Solve $A_0 x = y_0$ under the assumptions

(i) noisy data $\left\| y_0 - y_\delta \right\| \leq \delta$.

(ii) inexact operator $\left\| A_0 - A_\epsilon \right\| \leq \epsilon$.

What have been done so far?

- linear case:
  - **TLS**: Total least squares by Golub and Van Loan [1980];
  - **R-TLS**: Regularized TLS by Golub et al. [1999];
  - **D-RTLS**: Dual R-TLS by Lu et al. [2007].
- nonlinear case: no publication (?)

LS: y_δ and A_0

$$\begin{array}{ll} \text{minimize}_y & \left\| y - y_\delta \right\|_2 \\ \text{subject to} & y \in \mathscr{R}(A_0) \end{array}$$

TLS: y_δ and A_ϵ

$$\begin{array}{ll} \text{minimize} & \left\| [A, y] - [A_\epsilon, y_\delta] \right\|_F \\ \text{subject to} & y \in \mathscr{R}(A) \end{array}$$

Solve $A_0 x = y_0$ under the assumptions

(i) noisy data $\left\| y_0 - y_\delta \right\| \le \delta$.

(ii) inexact operator $\left\| A_0 - A_\epsilon \right\| \le \epsilon$.

What have been done so far?

- linear case:
  - **TLS**: Total least squares by Golub and Van Loan [1980];
  - **R-TLS**: Regularized TLS by Golub et al. [1999];
  - **D-RTLS**: Dual R-TLS by Lu et al. [2007].
- nonlinear case: no publication (?)

LS: y_δ and A_0

$$\begin{array}{ll} \text{minimize}_y & \left\| y - y_\delta \right\|_2 \\ \text{subject to} & y \in \mathscr{R}(A_0) \end{array}$$

TLS: y_δ and A_ϵ

$$\begin{array}{ll} \text{minimize} & \left\| [A, y] - [A_\epsilon, y_\delta] \right\|_F \\ \text{subject to} & y \in \mathscr{R}(A) \end{array}$$

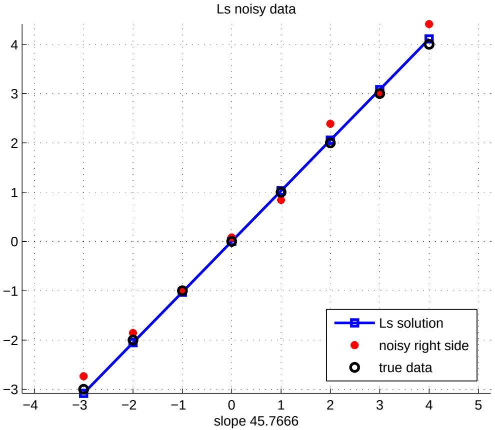
Illustration

Solve 1D problem: $am = b$, find the slope m.

Cases:

1. b_δ
2. a_ϵ
3. b_δ, a_ϵ

Solution:
$m = 1$



Example: $\arctan(1) = 45^o$ (Van Huffel and Vandewalle [1991]).

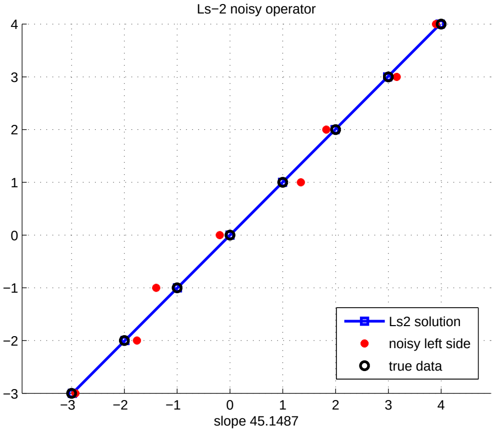
Illustration

Solve 1D problem: $am = b$, find the slope m.

Cases:

1. b_δ
2. a_ϵ
3. b_δ, a_ϵ

Solution:
$m = 1$



Example: $\arctan(1) = 45^o$ (Van Huffel and Vandewalle [1991]).

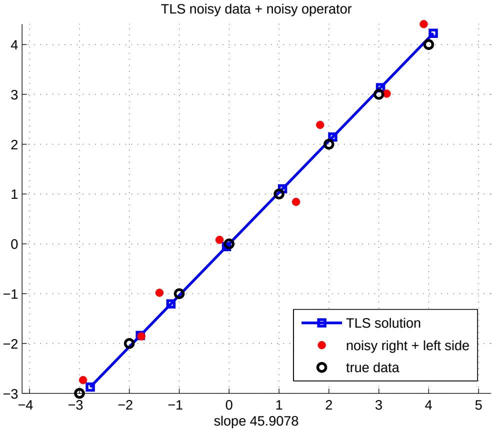
Illustration

Solve 1D problem: $am = b$, find the slope m.

Cases:

1. b_δ
2. a_ϵ
3. b_δ, a_ϵ

Solution:
$m = 1$



Example: $\arctan(1) = 45^o$ (Van Huffel and Vandewalle [1991]).

R-TLS

The R-TLS method [Golub, Hansen, O'Leary, 1999]

$$\text{minimize} \quad \|A - A_\epsilon\|^2 + \|y - y_\delta\|^2$$
$$\text{subject to} \quad \begin{cases} Ax = y \\ \|Lx\|^2 \leq M \,. \end{cases}$$

If the inequality constraint is active, then

$$(A_\epsilon^T A_\epsilon + \alpha L^T L + \beta I)\hat{x} = A_\epsilon^T y_\delta \text{ and } \|L\hat{x}\| = M$$

with $\alpha = \mu(1 + \|\hat{x}\|^2)$, $\beta = -\dfrac{\|A_\epsilon \hat{x} - y_\delta\|^2}{1 + \|\hat{x}\|^2}$ and $\mu > 0$ is the Lagrange multiplier.

Difficulty: requires a reliable bound M for the norm $\|Lx^\dagger\|^2$.

DR-TLS

The DR-TLS method [Lu et al., 2007]:

$$\begin{aligned} &\text{minimize} && \|Lx\|^2 \\ &\text{subject to} && \begin{cases} Ax = y \\ \|y - y_\delta\|^2 \le \delta \\ \|A - A_\epsilon\|^2 \le \epsilon\,. \end{cases} \end{aligned}$$

side condition

If the inequalities constraints are active, then

$$\left(A_\epsilon^T A_\epsilon + \alpha L^T L + \beta I\right)\tilde{x} = A_\epsilon^T y_\delta$$

with $\alpha = \dfrac{\nu + \mu\|\tilde{x}\|^2}{\nu\mu}$, $\beta = -\dfrac{\mu\|A_\epsilon\tilde{x} - y_\delta\|^2}{\nu + \mu\|\tilde{x}\|^2}$ and $\nu, \mu > 0$ are Langrange multipliers. Moreover, $\|A_\epsilon\tilde{x} - y_\delta\| = \delta + \epsilon\|\tilde{x}\|$.

Overview

Consider the operator equation

$$B(k, f) = g_0$$

where B is a bilinear operator (nonlinear)

$$\begin{aligned} B : \quad \mathcal{U} \times \mathcal{V} &\longrightarrow \mathcal{H} \\ (k, f) &\longmapsto B(k, f) \end{aligned}$$

and B is characterized by a function k_0.

- $K \cdot = B(\tilde{k}, \cdot)$ compact linear operator for a fixed $\tilde{k} \in \mathcal{U}$
- $F \cdot = B(\cdot, \tilde{f})$ linear operator for a fixed $\tilde{f} \in \mathcal{V}$
- $\|B(k_0, \cdot)\|_{\mathcal{V} \to \mathcal{H}} \leq \|k_0\|_{\mathcal{U}}$;
- $\|B(k, f)\|_{\mathcal{H}} \leq \|k\|_{\mathcal{U}} \|f\|_{\mathcal{V}}$;

Example:

$$B(k, f)(s) := \int_\Omega k(s, t) f(t) dt \,.$$

Consider the operator equation

$$B(k,f) = g_0$$

where B is a bilinear operator (nonlinear)

$$\begin{array}{rccc} B: & \mathcal{U}\times\mathcal{V} & \longrightarrow & \mathcal{H} \\ & (k,f) & \longmapsto & B(k,f) \end{array}$$

and B is characterized by a function k_0.

- $K\cdot = B(\tilde{k},\cdot)$ compact linear operator for a fixed $\tilde{k}\in\mathcal{U}$
- $F\cdot = B(\cdot,\tilde{f})$ linear operator for a fixed $\tilde{f}\in\mathcal{V}$
- $\|B(k_0,\cdot)\|_{\mathcal{V}\to\mathcal{H}} \le \|k_0\|_{\mathcal{U}}$;
- $\|B(k,f)\|_{\mathcal{H}} \le \|k\|_{\mathcal{U}}\|f\|_{\mathcal{V}}$;

Example:

$$B(k,f)(s) := \int_\Omega k(s,t)f(t)dt\,.$$

"Some mathematicians still have a kind of fear whenever they encounter a Fredholm integral equation of the first kind".

We want to solve

$$B(k_0, f) = g_0$$

out of the measurements k_ϵ and g_δ with

(i) noisy data $\|g_0 - g_\delta\|_{\mathcal{H}} \leq \delta$.

(ii) inexact operator $\|k_0 - k_\epsilon\|_{\mathcal{U}} \leq \epsilon$.

Francesco Tricomi (1897-1978)

"Some mathematicians still have a kind of fear whenever they encounter a Fredholm integral equation of the first kind".

We want to solve

$$B(k_0, f) = g_0$$

out of the measurements k_ϵ and g_δ with

(i) noisy data $\|g_0 - g_\delta\|_{\mathcal{H}} \leq \delta$.

(ii) inexact operator $\|k_0 - k_\epsilon\|_{\mathcal{U}} \leq \epsilon$.

Francesco Tricomi (1897-1978)

We introduce the DBL-RTLS

$$\underset{k,f}{\text{minimize}} \;\; J(k,f) := T(k,f,k_\epsilon,g_\delta) + R(k,f) \qquad (2)$$

where

$$T(k,f,k_\epsilon,g_\delta) = \frac{1}{2}\|B(k,f) - g_\delta\|_{\mathcal{H}}^2 + \frac{\gamma}{2}\|k - k_\epsilon\|_{\mathcal{U}}^2$$

$$R(k,f) = \frac{\alpha}{2}\|Lf\|_{\mathcal{V}}^2 + \beta\mathcal{R}(k)$$

- T is based on TLS method, measures the discrepancy on both data and operator;
- α, β are the regularization parameters and γ is a scaling parameter;
- $L : \mathcal{V} \to \mathcal{V}$ is a linear bounded operator;
- **double regularization** was proposed by You and Kaveh [1996], $\mathcal{R} : U \to [0,+\infty]$ is proper **convex** function and **w-lsc**.

We introduce the DBL-RTLS

$$\underset{k,f}{\text{minimize}} \; J(k,f) := T(k,f,k_\epsilon,g_\delta) + R(k,f) \tag{2}$$

where

$$T(k,f,k_\epsilon,g_\delta) = \frac{1}{2}\|B(k,f) - g_\delta\|^2_{\mathcal{H}} + \frac{\gamma}{2}\|k - k_\epsilon\|^2_{\mathcal{U}}$$

$$R(k,f) = \frac{\alpha}{2}\|Lf\|^2_{\mathcal{V}} + \beta\mathcal{R}(k)$$

- T is based on TLS method, measures the discrepancy on both data and operator;
- α, β are the regularization parameters and γ is a scaling parameter;
- $L : \mathcal{V} \to \mathcal{V}$ is a linear bounded operator;
- **double regularization** was proposed by You and Kaveh [1996], $\mathcal{R} : U \to [0, +\infty]$ is proper **convex** function and **w-lsc**.

Main theoretical results

Assumption:

(A1) B is strongly continuous, ie, if $(k^n, f^n) \rightharpoonup (\bar{k}, \bar{f})$ then $B(k^n, f^n) \to B(\bar{k}, \bar{f})$

Proposition

Let J be the functional defined on (2) *and L be a* **bounded** *and* **positive** *operator. Then J is* **positive**, **weak lower semi-continuous** *and* **coercive** *functional.*

Theorem (existence)

Let the assumptions of Proposition 1 hold. Then there exists a **global minimum** *of*

$$\text{minimize} \;\; J(k, f) \, .$$

Theorem (stability)

- $\delta_j \to \delta$ *and* $\epsilon_j \to \epsilon$
- $g_{\delta_j} \to g_\delta$ *and* $k_{\epsilon_j} \to k_\epsilon$
- $\alpha, \beta > 0$
- (k^j, f^j) *is a minimizer of* J *with* g_{δ_j} *and* k_{ϵ_j}
- *Then there* **exists** *a convergent subsequence of* $(k^j, f^j)_j$

$$(k^{j_m}, f^{j_m}) \longrightarrow (\bar{k}, \bar{f})$$

where $(\bar{k}, \bar{f})$ *is a minimizer of* J *with* $g_\delta, k_\epsilon, \alpha$ *and* β.

Theorem (stability)

- $\delta_j \to \delta$ *and* $\epsilon_j \to \epsilon$
- $g_{\delta_j} \to g_\delta$ *and* $k_{\epsilon_j} \to k_\epsilon$
- $\alpha, \beta > 0$
- (k^j, f^j) *is a minimizer of* J *with* g_{δ_j} *and* k_{ϵ_j}
- *Then there* **exists** *a convergent subsequence of* $(k^j, f^j)_j$

$$(k^{j_m}, f^{j_m}) \longrightarrow (\bar{k}, \bar{f})$$

where $(\bar{k}, \bar{f})$ *is a minimizer of* J *with* $g_\delta, k_\epsilon, \alpha$ *and* β.

Consider the convex functional

$$\Phi(k,f) := \frac{1}{2}\left\|Lf\right\|^2 + \eta\mathcal{R}(k)$$

where the parameter η represents the different scaling of f and k.

For convergence results we need to define

Definition

We call $(k^\dagger, f^\dagger)$ a Φ-**minimizing solution** if

$$(k^\dagger, f^\dagger) = \underset{(k,f)}{\arg\min}\left\{\Phi(k,f) \mid B(k,f) = g_0\right\}.$$

Theorem (convergence)

- $\delta_j \to 0$ *and* $\epsilon_j \to 0$
- $\|g_{\delta_j} - g_0\| \leq \delta_j$ *and* $\|k_{\epsilon_j} - k_0\| \leq \epsilon_j$
- $\alpha_j = \alpha(\epsilon_j, \delta_j)$ *and* $\beta_j = \beta(\epsilon_j, \delta_j)$, *s.t.* $\alpha_j \to 0$, $\beta_j \to 0$,

$$\lim_{j\to\infty} \frac{\delta_j^2 + \gamma\epsilon_j^2}{\alpha_j} = 0 \quad \text{and} \quad \lim_{j\to\infty} \frac{\beta_j}{\alpha_j} = \eta$$

- (k^j, f^j) *is a minimizer of* J *with* g_{δ_j}, k_{ϵ_j}, α_j *and* β_j
- *Then there* **exists** *a convergent subsequence of* $(k^j, f^j)_j$

$$(k^{j_m}, f^{j_m}) \longrightarrow (k^\dagger, f^\dagger)$$

where $(k^\dagger, f^\dagger)$ *is a* Φ-**minimizing solution**.

Theorem (convergence)

- $\delta_j \to 0$ *and* $\epsilon_j \to 0$
- $\left\| g_{\delta_j} - g_0 \right\| \leq \delta_j$ *and* $\left\| k_{\epsilon_j} - k_0 \right\| \leq \epsilon_j$
- $\alpha_j = \alpha(\epsilon_j, \delta_j)$ *and* $\beta_j = \beta(\epsilon_j, \delta_j)$, *s.t.* $\alpha_j \to 0$, $\beta_j \to 0$,

$$\lim_{j\to\infty} \frac{\delta_j^2 + \gamma\epsilon_j^2}{\alpha_j} = 0 \quad \text{and} \quad \lim_{j\to\infty} \frac{\beta_j}{\alpha_j} = \eta$$

- (k^j, f^j) *is a minimizer of* J *with* g_{δ_j}, k_{ϵ_j}, α_j *and* β_j
- *Then there* **exists** *a convergent subsequence of* $(k^j, f^j)_j$

$$(k^{j_m}, f^{j_m}) \longrightarrow (k^\dagger, f^\dagger)$$

where $(k^\dagger, f^\dagger)$ *is a* Φ-**minimizing solution**.

Overview

Optimality condition

If the pair $(\bar{k}, \bar{f})$ is a minimizer of $J(k, f)$, then $(0,0) \in \partial J\left(\bar{k}, \bar{f}\right)$.

Theorem

Let $J : \mathcal{U} \times \mathcal{V} \to \mathbb{R}$ *be a nonconvex functional,*

$$J(u, v) = \varphi(u) + Q(u, v) + \psi(v)$$

where Q *is a nonlinear differentiable term and* φ, ψ *are lsc convex functions. Then*

$$\begin{aligned} \partial J(u, v) &= \{\partial\varphi(u) + D_u Q(u, v)\} \times \{\partial\psi(v) + D_v Q(u, v)\} \\ &= \{\partial_u J(u, v)\} \times \{\partial_v J(u, v)\} \end{aligned}$$

Remark:

- is difficult to solve wrt both (k, f)
- J is bilinear and biconvex (linear and convex to each one)
- applied **alternating minimization** method.

Alternating minimization algorithm

Require: $g_\delta, k_\epsilon, L, \gamma, \alpha, \beta$

1: $n = 0$
2: **repeat**
3: $f^{n+1} \in \arg\min_f J(k, f | k^n)$
4: $k^{n+1} \in \arg\min_k J(k, f | f^{n+1})$
5: **until** convergence

Remark:

- is difficult to solve wrt both (k, f)
- J is bilinear and biconvex (linear and convex to each one)
- applied **alternating minimization** method.

Alternating minimization algorithm

Require: $g_\delta, k_\epsilon, L, \gamma, \alpha, \beta$

1: $n = 0$
2: **repeat**
3: $\quad f^{n+1} \in \arg\min_f J(k, f | k^n)$
4: $\quad k^{n+1} \in \arg\min_k J(k, f | f^{n+1})$
5: **until** convergence

Proposition

The sequence generated by the function $J(k^n, f^n)$ *is non-increasing,*

$$J(k^{n+1}, f^{n+1}) \leq J(k^n, f^{n+1}) \leq J(k^n, f^n).$$

Assumptions:

(A1) B is strongly continuous, ie, if $(k^n, f^n) \rightharpoonup (\bar{k}, \bar{f})$ then $B(k^n, f^n) \to B(\bar{k}, \bar{f})$

(A2) B is weakly sequentially closed, ie, if $(k^n, f^n) \rightharpoonup (\bar{k}, \bar{f})$ and $B(k^n, f^n) \rightharpoonup g$ then $B(\bar{k}, \bar{f}) = g$

(A3) the adjoint of B' is strongly continuous, ie, if $(k^n, f^n) \rightharpoonup (\bar{k}, \bar{f})$ then $B'(k^n, f^n)^* z \to B'(\bar{k}, \bar{f})^* z$, $\forall z \in \mathscr{D}(B')$

Proposition

The sequence generated by the function $J(k^n, f^n)$ is non-increasing,

$$J(k^{n+1}, f^{n+1}) \leq J(k^n, f^{n+1}) \leq J(k^n, f^n).$$

Assumptions:

(A1) B is strongly continuous, ie, if $(k^n, f^n) \rightharpoonup (\bar{k}, \bar{f})$ then $B(k^n, f^n) \to B(\bar{k}, \bar{f})$

(A2) B is weakly sequentially closed, ie, if $(k^n, f^n) \rightharpoonup (\bar{k}, \bar{f})$ and $B(k^n, f^n) \rightharpoonup g$ then $B(\bar{k}, \bar{f}) = g$

(A3) the adjoint of B' is strongly continuous, ie, if $(k^n, f^n) \rightharpoonup (\bar{k}, \bar{f})$ then $B'(k^n, f^n)^* z \to B'(\bar{k}, \bar{f})^* z$, $\forall z \in \mathscr{D}(B')$

Theorem

Given regularization parameters $0 < \underline{\alpha} \leq \alpha$ and β, compute AM algorithm. The sequence $\{(k^{n+1}, f^{n+1})\}_{n+1}$ has a weakly convergent subsequence, namely $(k^{n_j+1}, f^{n_j+1}) \rightharpoonup (\bar{k}, \bar{f})$ and the limit has the property

$$J(\bar{k}, \bar{f}) \leq J(\bar{k}, f) \quad \text{and} \quad J(\bar{k}, \bar{f}) \leq J(k, \bar{f})$$

for all $f \in \mathcal{V}$ and for all $k \in \mathcal{U}$.

Proposition

Let $\{(k^n, f^n)\}_n$ be a weakly convergent sequence generated by AM algorithm, where $k^n \rightharpoonup \bar{k}$ and $f^n \rightharpoonup \bar{f}$. Then there exists a subsequence $\{k^{n_j}\}_{n_j}$ such that $k^{n_j} \to \bar{k}$ and there exists $\{\xi_k^{n_j}\}_{n_j}$ with $\xi_k^{n_j} \in \partial_k J(k^{n_j}, f^{n_j})$ such that $\xi_k^{n_j} \to 0$.

Theorem

Given regularization parameters $0 < \underline{\alpha} \leq \alpha$ *and* β*, compute AM algorithm. The sequence* $\{(k^{n+1}, f^{n+1})\}_{n+1}$ *has a weakly convergent subsequence, namely* $(k^{n_j+1}, f^{n_j+1}) \rightharpoonup (\bar{k}, \bar{f})$ *and the limit has the property*

$$J(\bar{k}, \bar{f}) \leq J(\bar{k}, f) \quad \text{and} \quad J(\bar{k}, \bar{f}) \leq J(k, \bar{f})$$

for all $f \in \mathcal{V}$ *and for all* $k \in \mathcal{U}$*.*

Proposition

Let $\{(k^n, f^n)\}_n$ *be a weakly convergent sequence generated by AM algorithm, where* $k^n \rightharpoonup \bar{k}$ *and* $f^n \rightharpoonup \bar{f}$*. Then there exists a subsequence* $\{k^{n_j}\}_{n_j}$ *such that* $k^{n_j} \to \bar{k}$ *and there exists* $\{\xi_k^{n_j}\}_{n_j}$ *with* $\xi_k^{n_j} \in \partial_k J(k^{n_j}, f^{n_j})$ *such that* $\xi_k^{n_j} \to 0$*.*

Proposition

Let $\{n\}$ *be a subsequence of* $\mathbb{N}$ *such that the sequence* $\{(k^n, f^n)\}_n$ *generated by AM algorithm satisfies* $k^n \to \bar{k}$ *and* $f^n \rightharpoonup \bar{f}$*. Then* $f^{n_j} \to \bar{f}$ *and there exists* $\{\xi_f^{n_j}\}_{n_j}$ *with* $\xi_f^{n_j} \in \partial_f J(k^{n_j}, f^{n_j})$ *such that* $\xi_f^{n_j} \to 0$*.*

Remark: Graph of subdifferential mapping is sw-closed, ie, if $v_n \to \bar{v}$ and $\xi_n \rightharpoonup \bar{\xi}$ with $\xi_n \in \partial\varphi(v_n)$, then $\xi \in \partial\varphi(\bar{v})$.

Theorem

Let $\{(k^n, f^n)\}_n$ *be the sequence generated by the AM algorithm, then there exists a subsequence converging towards to a critical point of* J*, i.e.,*

$$(0,0) \in \partial J\left(\bar{k}, \bar{f}\right).$$

Proposition

Let $\{n\}$ be a subsequence of $\mathbb{N}$ such that the sequence $\{(k^n, f^n)\}_n$ generated by AM algorithm satisfies $k^n \to \bar{k}$ and $f^n \rightharpoonup \bar{f}$. Then $f^{n_j} \to \bar{f}$ and there exists $\{\xi_f^{n_j}\}_{n_j}$ with $\xi_f^{n_j} \in \partial_f J(k^{n_j}, f^{n_j})$ such that $\xi_f^{n_j} \to 0$.

Remark: Graph of subdifferential mapping is sw-closed, ie, if $v_n \to \bar{v}$ and $\xi_n \rightharpoonup \bar{\xi}$ with $\xi_n \in \partial\varphi(v_n)$, then $\xi \in \partial\varphi(\bar{v})$.

Theorem

Let $\{(k^n, f^n)\}_n$ be the sequence generated by the AM algorithm, then there exists a subsequence converging towards to a critical point of J, i.e.,

$$(0,0) \in \partial J\left(\bar{k}, \bar{f}\right).$$

Short comments

For minimization on k we follow Daubechies et al. [2004].

- penalty term: $\mathcal{R}(k) = \sum_\lambda \omega_\lambda |k_\lambda|$ where $k_\lambda = |\langle k \, , \, \phi_\lambda \rangle|$
- apply **surrogate functional** - extended functional

$$\tilde{J}^{Sur}(k,u) = \tilde{J}(k) + C\left\|k-u\right\| - \left\|\tilde{F}k - \tilde{F}u\right\|$$

- $\tilde{B}(k,f) : (k,f) \longmapsto (B(k,f),k)$ and $\left\|(x,y)\right\|_\gamma = \left\|x\right\| + \gamma\left\|y\right\|$
- combine with **soft-shrinkage** operator

$$\mathcal{S}_\beta\left(x\right) = \max\{\left\|x\right\| - \beta, 0\}\frac{x}{\left\|x\right\|}$$

- $k_\lambda^{n+1} = \mathcal{S}_{\frac{\omega_\lambda}{2}\frac{\beta}{\gamma C}}\left(k_\lambda^n - \frac{1}{C}(k_\lambda^n - k_{\epsilon\lambda}) - \frac{1}{C\alpha}[F^*(Fk^n - g_\delta)]_\lambda\right)$

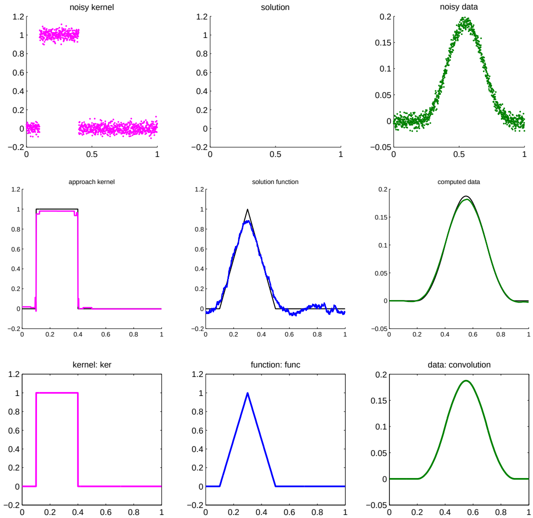
Overview

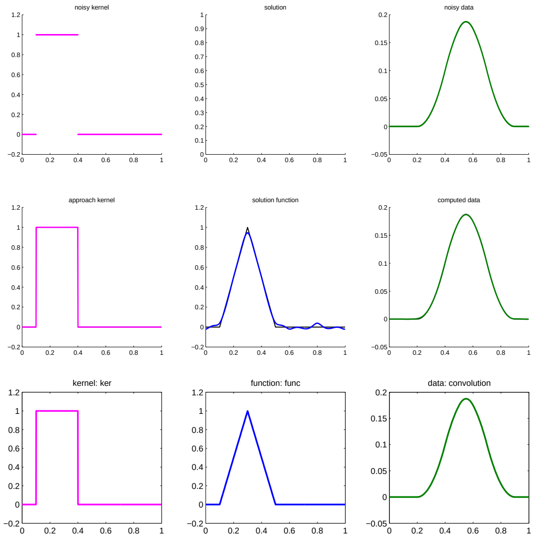
First numerical result

Convolution in 1D

$$\int_{\Omega} k(s-t)f(t)dt = g(s)$$

- characteristic kernel and gaussian function;
- space: $\Omega = [0,1]$, discretization: $N = 2048$ points;
- Haar wavelet for $\{\phi\}_\lambda$ and $J = 10$;
- initial guess: $k^0 = k_\epsilon$, $\tau = 1.0$;
- A. relative error: 10% and 10%.
- B. relative error: 0.1% and 0.1%.


noisy kernel
solution
noisy data
approach kernel
solution function
computed data
kernel: ker
function: func
data: convolution


noisy kernel
solution
noisy data
approach kernel
solution function
computed data
kernel: ker
function: func
data: convolution

Overview

Conclusions and future work

So far:

- introduced a method for nonlinear equation (bilinear operator) with noisy data and inexact operator;
- proved existence, stability and convergence;
- suggested an iterative implementation;
- proved convergence of AM algorithm to a critical point;

For further work:

- study of source conditions;
- prove convergence rates (k and f);
- how to choose the best regularization parameter?
- *a priori* and *a posteriori* choice;
- implementations and numerical experiments (2D);

Conclusions and future work

So far:

- introduced a method for nonlinear equation (bilinear operator) with noisy data and inexact operator;
- proved existence, stability and convergence;
- suggested an iterative implementation;
- proved convergence of AM algorithm to a critical point;

For further work:

- study of source conditions;
- prove convergence rates (k and f);
- how to choose the best regularization parameter?
- *a priori* and *a posteriori* choice;
- implementations and numerical experiments (2D);

I. Daubechies, M. Defrise, and C. De Mol. An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. *Comm. Pure Appl. Math.*, 57(11):1413–1457, 2004. ISSN 0010-3640.

G. H. Golub and C. F. Van Loan. An analysis of the total least squares problem. *SIAM J. Numer. Anal.*, 17(6): 883–893, 1980. ISSN 0036-1429.

G. H. Golub, P. C. Hansen, and D. P. O'leary. Tikhonov regularization and total least squares. *SIAM J. Matrix Anal. Appl*, 21:185–194, 1999.

S. Lu, S. V. Pereverzev, and U. Tautenhahn. Regularized total least squares: computational aspects and error bounds. Technical Report 30, Ricam, Linz, Austria, 2007. URL http://www.ricam.oeaw.ac.at/publications/reports/07/rep07-30.pdf.

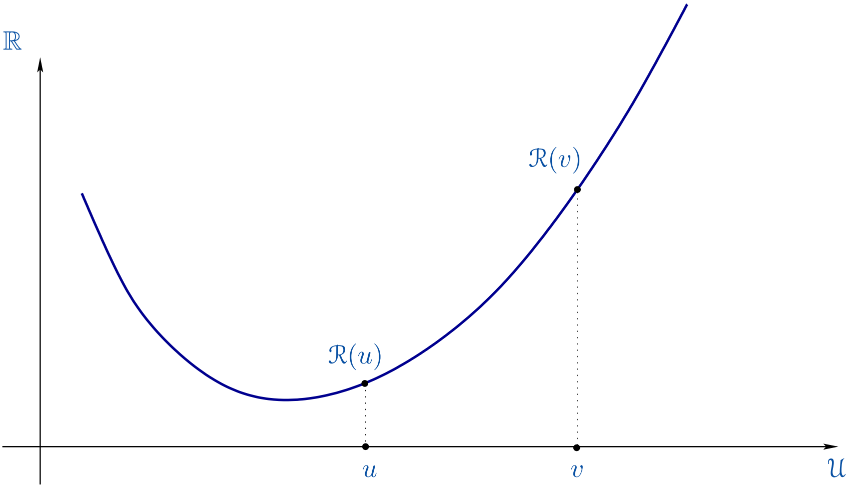
S. Van Huffel and J. Vandewalle. *The total least squares problem*, volume 9 of *Frontiers in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1991. ISBN 0-89871-275-0. Computational aspects and analysis, With a foreword by Gene H. Golub.

Y.-L. You and M. Kaveh. A regularization approach to joint blur identification and image restoration. *Image Processing, IEEE Transactions on*, 5(3):416 –428, mar 1996. ISSN 1057-7149.

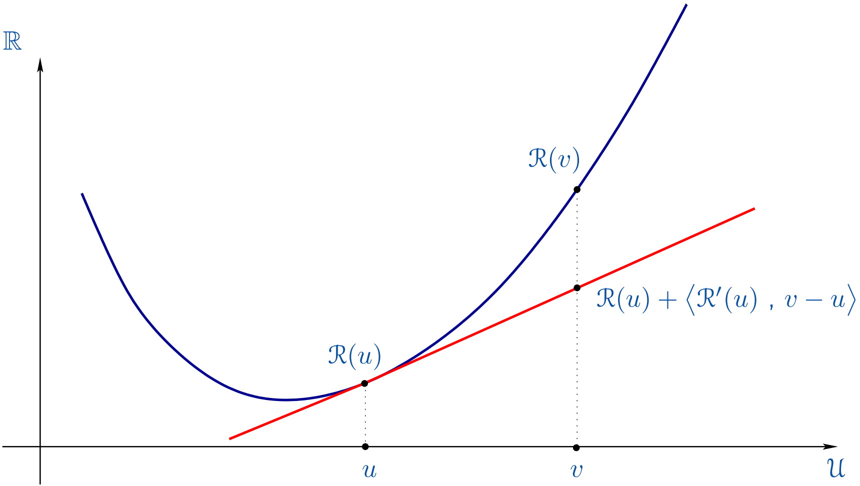
Thank you for your kind attention!

Questions?

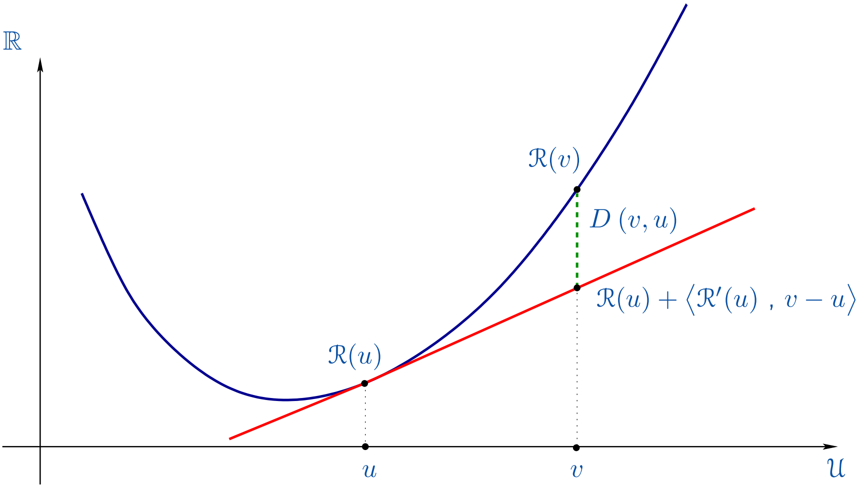
Reminder Bregman distance



Reminder Bregman distance



Reminder Bregman distance



Bregman distance: $\xi = \{\mathcal{R}'(u)\}$

$$D\,(v,u) = \mathcal{R}(v) - \mathcal{R}(u) - \langle \xi\;,\,v-u\rangle.$$

Generalized Bregman distances: subgradient $\xi \in \mathcal{U}^*$

$$D\,(v,u) = \left\{D^{\xi}\,(v,u) := \mathcal{R}(v) - \mathcal{R}(u) - \left\langle \xi\;,\,v-u\right\rangle \;|\; \xi \in \partial\mathcal{R}\,(u)\right\}.$$

Back

Source condition

Consider a nonlinear operator $F : X \to Y$ and the the nonlinear equation

$$F(x) = y.$$

Measurement y_δ with $\|y - y_\delta\| \leq \delta$.

Study of Source conditions: how fast a solution of the Tikhonov-type functional

$$J_\alpha(x) = \|F(x) - y_\delta\|^2 + \alpha\Psi(x)$$

converges to the Ψ-minimizing solution $x^\dagger$.

Theorem *[Schock, '84]*

Without any further assumptions, the convergence

$$x_\alpha^\delta \to x^\dagger \quad \text{as} \quad \delta \to 0$$

can (and will) be arbitrarily slow.

The way out...

Source and Nonlinearity conditions

Assume that there is $\xi \in \partial\Psi(x^\dagger)$ and $w \in Y^*$ such that

$$\xi = F'(x^\dagger)^* w, \qquad (S)$$

and that – locally near $x^\dagger$ – we have

$$\|F(x) - F(x^\dagger) - F'(x^\dagger)(x - x^\dagger)\| \leq c D_\Psi^\xi(x, x^\dagger), \qquad (NL)$$

where $c < 1/\|w\|$.

Selected convergence rate results for MDP

MDP: Morozov's discrepancy principle.

dist	rate	lin	sparse	ℓ_p	src/nl cond	due to
D_Ψ	$\mathcal{O}(\delta)$	✓		–	(S)	*[Bonesky '09]*
D_Ψ	$\mathcal{O}(\delta)$			–	(S) & (NL)	*[Aneznguber, Ramlau '10]*
$\|\|.\|\|$	$\mathcal{O}(\delta^{1/p})$			$2 \leq p$	(S) & (NL)	*[Grasmair, Haltmeier,*
$\|\|.\|\|$	$\mathcal{O}(\delta^{1/2})$	✓		$p \in (1,2)$	(S)	*Scherzer '09]*
$\|\|.\|\|$	$\mathcal{O}(\delta^{1/p})$	✓	✓	$p \in [1,2)$	(S)	*(for Residual Method)*
$\|\|.\|\|$	$\mathcal{O}(\delta^{1/r})$			–	(VIE)	*[Anzengruber, Ramlau '11]*

Back