On the Choice of the Tikhonov Regularization Parameter and the Discretization Level: A Discrepancy-Based Strategy

New Trends in Parameter Identification for Mathematical Models

Adriano De Cezaro - FURG

Joint work with Vinicius Albani and Jorge P. Zubelli

in Inverse Problems and Imaging, 2016.

Inverse Problem

- In theory formulated as an operator equation

$$F(x) = y. \tag{1}$$

defined in the reflexive Banach spaces X and Y, with a convex domain $\mathcal{D}(F) \subset X$.
infinite dimensional framework

Inverse Problem

- In theory formulated as an operator equation

$$F(x) = y. \tag{1}$$

  defined in the reflexive Banach spaces X and Y, with a convex domain $\mathcal{D}(F) \subset X$.
  infinite dimensional framework

- In many practical inverse problems:

1- solved under a finite-dimensional and discrete setup

2- access only to noisy data y^δ $\quad\quad ||y^\delta - y|| \leq \delta.$

3- sometimes the data are sparse (very fill measurements are accessible)

Inverse Problem

- In theory formulated as an operator equation

$$F(x) = y. \tag{1}$$

defined in the reflexive Banach spaces X and Y, with a convex domain $\mathcal{D}(F) \subset X$.
infinite dimensional framework

- In many practical inverse problems:
1- solved under a finite-dimensional and discrete setup
2- access only to noisy data y^δ $\qquad ||y^\delta - y|| \leq \delta.$
3- sometimes the data are sparse (very fill measurements are accessible)
- Thus, the relation between the finite- and the infinite-dimensional descriptions of the same problem should be well-understood.

Addressed questions

- Is possible to state a criterion to find appropriately the domain discretization level in terms of the available data, in order to find a reliable solution of the inverse problem, which is in general ill-posed?

Addressed questions

- Is possible to state a criterion to find appropriately the domain discretization level in terms of the available data, in order to find a reliable solution of the inverse problem, which is in general ill-posed?
  YES - many authors have already addressed this issue

Addressed questions

- Is possible to state a criterion to find appropriately the domain discretization level in terms of the available data, in order to find a reliable solution of the inverse problem, which is in general ill-posed?
  YES - many authors have already addressed this issue ex. Kirsch's book for a very specific problem!!!

Addressed questions

- Is possible to state a criterion to find appropriately the domain discretization level in terms of the available data, in order to find a reliable solution of the inverse problem, which is in general ill-posed?
  YES - many authors have already addressed this issue ex. Kirsch's book for a very specific problem!!! Regularization by discretization !!!

Our contribution

- Our contribution: Under the context of Tikhonov-type regularization, i.e.,

Problem

Find a minimizer for the Tikhonov functional

$$\mathcal{F}_{\alpha,x_0}^{y^\delta}(x) = \|F(x) - y^\delta\|_Y^p + \alpha f_{x_0}(x), \tag{2}$$

with $\alpha > 0$ and $1 < p < +\infty$.

we propose a discrepancy-based rule for choosing appropriately a regularization parameter and a domain discretization level. We also establish the corresponding regularizing properties of this rule under fairly general assumptions.

like a 2 regularization parameter choice!!!

Related works

Some strongly related works on this subject

1. Anzengruber, S., Hofmann, B. and Mathé, P. Regularization properties of the sequential discrepancy principle for Tikhonov regularization in Banach spaces. Appl. Anal., 93(7):1382 - 1400, 2013.

2. Anzengruber, S. and Ramlau, R. Morozov's discrepancy principle for Tikhonov-type functionals with nonlinear operators. Inverse Problems, 26(2), February 2010.

3. Anzengruber, S. and Ramlau, R. Convergence rates for Morozov's discrepancy principle using variational inequalities. Inverse Problems, 27(10), 2011.

4. Bonesky, T. Morozov's discrepancy principle and Tikhonov-type functionals. Inverse Problems, 25(1), 2009.

What we add on the field (Conclusions!!!)

- We propose Morozov's discrepancy principle in the same spirit of [1], [2], [3]. [4] in the context of nonlinear operators in a discrete setting.

- We also state that the continuous case, presented in these references, can be recovered from the discrete one, when the discretization level goes to infinity. noting really new

- We use this discrepancy principle as a rule to find appropriately the discretization level in the domain and the regularization parameter in Tikhonov regularization. Some how a 2 regularization parameter choice rule

- regularizing properties and (+ existence of source conditions) convergence rates results. as expected

Assumptions

Assumption

The regularizing functional $f_{x_0} : \mathcal{D}(f_{x_0}) \to \mathbb{R}_+$ is weakly lower semi-continuous, convex, coercive, and proper. We also assume that $\mathcal{D}(F)$ is in the interior of $\mathcal{D}(f_{x_0})$.

Assumption

The forward operator F is continuous under the strong topologies of X and Y. We also assume that the level sets

$$\mathcal{M}_\alpha(\rho) = \{x \in \mathcal{D}(F) \ : \ \mathcal{F}_{\alpha,x_0}^{y^\delta}(x) \leq \rho\}$$

are weakly pre-compact and weakly closed. Moreover, the restriction of F to $\mathcal{M}_\alpha(\rho)$ is weakly continuous under the weak topologies of X and Y.

Definition

An element $x^\dagger$ of $\mathcal{D}(F)$ is called a least-square f_{x_0}-minimizing solution or simply an f_{x_0}-minimizing solution of Problem 1 if it is a least-square solution, i.e.,

$$x^\dagger \in \mathcal{LS} := \{x \in \mathcal{D}(F) \,:\, \|F(x) - y\| = 0\}$$

and minimizes f_{x_0} in $\mathcal{LS}$, i.e.,

$$x^\dagger \in \mathcal{L} := argmin\{f_{x_0}(x) \,:\, x \in \mathcal{LS}\}.$$

We always assume that $\mathcal{L} \neq \emptyset$.
Note that the sets $\mathcal{LS}$ and $\mathcal{L}$ depend on the noiseless data y.

Assumptions

Assumption

Let $x^{\dagger}$ be an f_{x_0}-minimizing solution for Problem 1 and $x_0 \in \mathcal{D}(F)$ be fixed. We assume that:

$$\liminf_{t \to 0^+} \frac{\|F((1-t)x^{\dagger} + tx_0) - y\|^p}{t} = 0 \tag{3}$$

Note that Assumption 3 is satisfied by many classes of operators, such as the class of locally Hölder continuous functions with exponent greater than $1/2$, with $p = 2$.

We also need to consider the sequence $\{X_m\}_{m\in\mathbb{N}}$ of finite-dimensional subspaces of X satisfy:

$$X_m \subset X_{m+1}, \text{ for } m \in \mathbb{N}, \text{ and } \overline{\bigcup_{m\in\mathbb{N}} X_m} = X. \tag{4}$$

Definition

Define the finite-dimensional sets:

$$\mathcal{D}_m = \mathcal{D}(F) \cap X_m, \text{ for } m \in \mathbb{N}. \tag{5}$$

The set $\mathcal{D}_m$ is convex since it is the intersection of a subspace of X with a convex set.

Note that, if we had chosen $\mathcal{D}_m$ as the orthogonal projection of $\mathcal{D}(F)$ onto the finite-dimensional subspace X_m, we could possibly have that $\mathcal{D}_m \cap X - \mathcal{D}(F) \neq \emptyset$, since F is not necessarily linear and $\mathcal{D}(F)$ is not necessarily a subspace of X. Therefore, this definition ensures that $\mathcal{D}_m \subset \mathcal{D}(F)$ for every $m \in \mathbb{N}$.

For now on, we assume that $\mathcal{D}_m \neq \emptyset$, for every m. Thus, we want to find $x_{m,\alpha}^{\delta} \in \mathcal{D}_m$ minimizing (2), with m and α appropriately chosen.

The analysis that follows depends on how fast the restriction of the operator F to $\mathcal{D}_m$ converges to F as $m \to \infty$. Thus, we have the following definition:

Definition

Let $P_m : X \to \mathcal{D}_m$ be the projection of X onto $\mathcal{D}_m$, $x^\dagger$ be a least-square f_{x_0}-minimizing solution. Define:

$$\gamma_m := \sup_{x \in B(x_0, \eta) \cap \mathcal{D}(F)} \|F(x^\dagger) - F(P_m x^\dagger)\| \text{ and} \tag{6}$$

$$\phi_m := \sup_{x \in B(x_0, \eta) \cap \mathcal{D}(F)} \|x^\dagger - P_m x^\dagger\|.$$

Lemma

For every $x \in \mathcal{D}(F)$, $\|F(x) - F(P_m x)\| \to 0$ when $m \to \infty$.

Proof: From (4) it follows that $\|x - P_m x\| \to 0$ as $m \to \infty$ for every $x \in \mathcal{D}(F)$. Since the operator F is continuous, the assertion follows.

Existence

We consider the following optimization problem:

Problem

Find an element of

$$argmin\{\|F(x) - y^\delta\|^p + \alpha f_{x_0}(x)\}, \quad subject\ to \quad x \in \mathcal{D}_m. \quad (7)$$

Theorem (Existence)

Let $m \in \mathbb{N}$ and $\delta > 0$ be fixed. Moreover, let the Assumptions be satisfied. Then, for any $y^\delta \in Y$, it follows that Problem (7) has a solution.

Stability

Definition

For given data y^δ, we call a solution of Problem (7) stable if for a strongly convergent sequence $\{y_k\}_{k\in\mathbb{N}} \subset Y$, with limit y^δ, the corresponding sequence $\{x_k\}_{k\in\mathbb{N}} \subset X$ of solutions of Problem (7), where y^δ is replaced by y_k in the functional of Problem (7), has a weakly convergent subsequence $\{x_{k_l}\}_{l\in\mathbb{N}}$, with limit $\tilde{x}$, a solution of Problem (7) with data y^δ.

Theorem (Stability)

For each $m \in \mathbb{N}$, the solutions of Problem (7) are stable in the sense of Definition 4. Moreover, the convergent subsequence $\{x_{k_l}\}_{l\in\mathbb{N}}$ with limit $\tilde{x}$ from Definition 4 satisfies the limit $f_{x_0}(x_{k_l}) \to f_{x_0}(\tilde{x})$.

Convergence

The following theorem shows that the finite-dimensional Tikhonov minimizers converge to some f_{x_0}-minimizing solution of Problem ((1)).

Theorem (Convergence)

Let $m \in \mathbb{N}$ and $\delta > 0$ be fixed. Assume that $\alpha = \alpha(\delta, \gamma_m) > 0$ satisfies the limits:

$$\lim_{\delta, \gamma_m \to 0} \alpha(\delta, \gamma_m) = 0 \quad \text{and} \quad \lim_{\delta, \gamma_m \to 0} \frac{(\delta + \gamma_m)^p}{\alpha(\delta, \gamma_m)}. \tag{8}$$

*Let $\{x_k\}_{k \in \mathbb{N}}$ be a sequence of solutions of Problem (7) with $x_k = x_{m_k, \alpha_k}^{\delta_k}$ and $\delta_k, \gamma_{m_k} \to 0$ when $k \to \infty$. Then, it has a weakly convergent subsequence $\{x_{k_l}\}_{l \in \mathbb{N}}$ with weak limit $x^\dagger$, an f_{x_0}-minimizing solution of Problem (**??**) with $f_{x_0}(x_{k_l}) \to f_{x_0}(x^\dagger)$.*

The discrepancy principle

Definition

Let $\delta > 0$ and y^δ be fixed. For $\lambda > \tau > 1$, we choose $m \in \mathbb{N}$ and $\alpha > 0$, with $m = m(\delta, y^\delta)$ and $\alpha = \alpha(\delta, y^\delta)$, such that

$$\tau\delta \leq \|F(x_{m,\alpha}^\delta) - y^\delta\| \leq \lambda\delta, \tag{9}$$

holds for $x_{m,\alpha}^\delta$ a solution of (7) with these same m and α.

Proposition

There exist $m \in \mathbb{N}$ and $\alpha > 0$ satisfying (9).

Proof: Uses similar arguments as in [1]–[4].

Discrete Morozov's Principle

Definition (Discrete Morozov's Principle)

Let δ, y^δ and the domain discretization level m be fixed. Define $\tau_1 := \tau$ and let τ_2 be such that $1 < \tau_1 \leq \tau_2 < \lambda$. Then, find $\alpha = \alpha(\delta, y^\delta, m) > 0$ such that

$$\tau_1(\delta + \gamma_m) \leq \|F(x^\delta_{m,\alpha}) - y^\delta\| \leq \tau_2(\delta + \gamma_m), \tag{10}$$

holds for $x^\delta_{m,\alpha}$, a solution of Problem 7.

Idea behind: Diagonal argument. We have to choose $m \in \mathbb{N}$ such that γ_m satisfies a modified version of (9). For this same m, we choose $\alpha > 0$ through (6), given that it is well-posed. Then, these α and m satisfy the same discrepancy principle, as required.

The *a priori* choice for the parameters

Under the present setup, if we choose $m \in \mathbb{N}$ sufficiently large and such that

$$\gamma_m \leq \left(\frac{\lambda}{\tau_2} - 1 \right) \delta \qquad (11)$$

is satisfied with $\lambda > \tau_2 > 1$. Then, for this same $m \in \mathbb{N}$, it follows that, when α is chosen through Definition 6, the discrepancy

$$\tau_1 \delta \leq \|F(x_{m,\alpha}^\delta) - y^\delta\| \leq \lambda \delta, \qquad (12)$$

is satisfied with $x_{m,\alpha}^\delta$ a solution of (7). This follows since,
$\tau_1 \delta \leq \tau_1 (\delta + \gamma_m)$ and $\tau_2 (\delta + \gamma_m) \leq \lambda \delta$.

Replace the continuous forward operator by a finite-dimensional approximation

Let us consider a sequence of finite-dimensional subspaces $\{Y_n\}_{n\in\mathbb{N}}$ of the space Y, such that

$$Y_n \subset Y_{n+1} \subset ... \subset Y \quad \text{and} \quad \overline{\cup_{n\in\mathbb{N}} Y_n} = Y.$$

In the present discrete setting, we consider the following alternative discrepancy principle:

Definition

Let $\delta > 0$ and y^δ be fixed. For $\lambda > \tau > 1$, we choose $m, n \in \mathbb{N}$ and $\alpha > 0$, with $m = m(\delta, y^\delta)$, $n = n(\delta, y^\delta)$ and $\alpha = \alpha(\delta, y^\delta)$, such that

$$\tau\delta \leq \|F_n(x_{m,n}^{\delta,\alpha}) - y^\delta\| \leq \lambda\delta, \tag{13}$$

holds for $x_{m,n}^{\delta,\alpha}$, a solution of

$$\min\{\|F_n(x) - y^\delta\|^p + \alpha f_{x_0}(x)\} \quad \text{subject to} \quad x \in \mathcal{D}_m. \tag{14}$$

Replace the continuous forward operator by a finite-dimensional approximation

In the present context, all the results of the previous sections hold?

Replace the continuous forward operator by a finite-dimensional approximation

In the present context, all the results of the previous sections hold?
Yes!!!

However, some additional calculations should be done when F is replaced by F_n. The main argument in the convergence analysis is based on the existence of a diagonal subsequence converging (weakly) to an f_{x_0}-minimizing solution of Problem 1, when the limits $\delta \to 0$, $m, n \to \infty$ are taken.

Convergence rates: Two situations

Case 1 m and α chosen by the corresponding discrepancy principle.

Case 2 m, n and α chosen by the corresponding discrepancy principle.

Assumptions and notations

Definition

Let U denote a Banach space and

$$f : \mathcal{D}(f) \subset U \to \mathbb{R} \cup \{\infty\}$$

be a convex functional with sub-differential $\partial f(u)$ at $u \in \mathcal{D}(f)$. The Bregman distance (or divergence) of f at $u \in \mathcal{D}(f)$ and $\xi \in \partial f(u) \subset U^*$ is defined by

$$D_\xi(\tilde{u}, u) = f(\tilde{u}) - f(u) - \langle \xi, \tilde{u} - u \rangle, \tag{15}$$

for every $\tilde{u} \in U$, where $\langle \cdot, \cdot \cdot \rangle$ is the dual product of U^* and U. Moreover, the set

$$\mathcal{D}_B(f) = \{x \in \mathcal{D}(f) \ : \ \partial f(u) \neq \emptyset\}$$

is called the Bregman domain of f.

Case 1: Convergence Rates

Assume the variational source condition

Assumption

There exist $\beta_1 \in [0,1)$, $\beta_2 \geq 0$ *and* $\xi^\dagger \in \partial f_{x_0}(x^\dagger)$ *such that*

$$\langle \xi^\dagger, x^\dagger - x \rangle \leq \beta_1 D_{\xi^\dagger}(x, x^\dagger) + \beta_2 \|F(x) - F(x^\dagger)\| \tag{16}$$

for $x \in \mathcal{M}_{\alpha_{max}}(\rho)$, *where* $\alpha_{max}, \rho > 0$ *satisfy* $\rho > \alpha_{max} f_{x_0}(x^\dagger)$.

Theorem

Let m and α be chosen through the discrepancy principle (9) *and let Assumption 4 be satisfied. Up to some technical additional hypotheses* (*to long for one page slide, we have:*

$$\|F(x_{m,\alpha}^\delta) - y^\delta\| \leq \lambda\delta \quad and \quad D_{\xi^\dagger}(x_{m,\alpha}^\delta, x^\dagger) \leq \frac{\beta_2(1+\lambda)}{1-\beta_1}\delta. \tag{17}$$

Case 2: Convergence Rates

Define the estimate $\eta_m := D_{\xi^\dagger}(P_m x^\dagger, x^\dagger)$.

Theorem (Convergence Rates)

Assume that $x_{m,\alpha}^\delta$ is a minimizer of the functional in Equation (2)and the regularization parameter $\alpha = \alpha(\delta, y^\delta, \gamma_m)$ satisfies the discrepancy principle (6). Then, we have the following estimates

$$\|F(x_{m,\alpha}^\delta) - y^\delta\| = O(\delta + \gamma_m + \eta_m + \phi_n) \quad \text{and} \quad D_{\xi^\dagger}(x_{m,\alpha}^\delta, x^\dagger) = O(\delta + \gamma_m + \eta_m \tag{18}$$

with $\xi^\dagger \in \partial f_{x_0}(x^\dagger)$.

Numerical Examples

Assuming that the data *u* was generated by the following parabolic problem (Black-Scholes equation):

$$
\begin{cases}
\dfrac{\partial u}{\partial \tau} - a(\tau, y)\left(\dfrac{\partial^2 u}{\partial y^2} - \dfrac{\partial u}{\partial y}\right) - b\dfrac{\partial u}{\partial y} &=& 0 & \tau > 0,\ y \in \mathbb{R} \\
u(\tau = 0, K) &=& \max\{0, 1 - e^y\}, & \text{for } y \in \mathbb{R}, \\
\text{dacay conditions}
\end{cases}
$$

(19)

Numerical Examples

Assuming that the data u was generated by the following parabolic problem (Black-Scholes equation):

$$
\begin{cases}
\dfrac{\partial u}{\partial \tau} - a(\tau, y)\left(\dfrac{\partial^2 u}{\partial y^2} - \dfrac{\partial u}{\partial y}\right) - b\dfrac{\partial u}{\partial y} &= 0 \qquad\qquad \tau > 0,\ y \in \mathbb{R} \\
u(\tau = 0, K) &= \max\{0, 1 - e^y\}, \quad \text{for } y \in \mathbb{R}, \\
\text{dacay conditions}
\end{cases}
$$

(19)

The inverse problem is to find the diffusion parameter $a \in Q := \{a \in a_0 + H^{1+\varepsilon}(\mathbb{R}_+ \times \mathbb{R}) : a_1 \leq a \leq a_2\}$ for given sparse data u.

We define the forward operator by:

$$F : Q \subset H^{1+\varepsilon}(\mathbb{R}_+ \times \mathbb{R}) \longrightarrow L^2(\mathbb{R}_+ \times \mathbb{R})$$
$$a \longmapsto u(a) - u(a_0),$$

with $a_0 \in Q$ fixed and *a priori* chosen.

In the calibration we take as true (known) diffusion coefficient the following:

$$\sigma(\tau, y) = \begin{cases} \dfrac{2}{5} - \dfrac{4}{25} e^{-\tau/2} \cos\left(\dfrac{4\pi y}{5}\right), & \text{if } -2/5 \le y \le 2/5 \\ \\ 2/5, & \text{otherwise}, \end{cases} \tag{20}$$

and set $a = \sigma^2/2$. We also assume that $b = 0.03$ in Equation (**??**).

The data is generated with step sizes $\Delta\tau = 0.0025$ and $\Delta y = 0.01$ and the coarser grid is given by the step lengths $\Delta\tau = 0.02$ and $\Delta y = 0.1$. In the numerical solution of the inverse problem, Equation (**??**) is numerically solved in the same mesh we interpolate the data, i.e., we use $\Delta\tau = 0.02$ and $\Delta y = 0.1$ in both cases. We vary the mesh used to evaluate the diffusion coefficient in order to highlight the discrepancy principle (**??**). The step sizes used in the tests were the following:

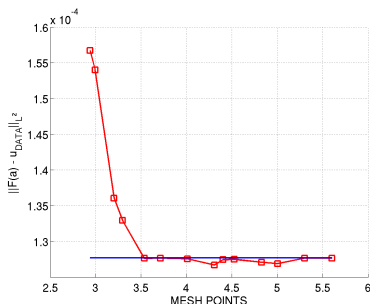$$\Delta\tau = 0.1, 0.08, 0.07, 0.06, 0.05, 0.04, 0.03, 0.02, 0.01, 0.0075, 0.005, 0.0025$$

Figura: Evolution of the residual as a function of the number of mesh points. We choose the regularization parameter presenting lower residual. In the presence of noise, some discretization levels in the domain satisfy the discrepancy principle. Compare it to the error estimation in Figure 2. The horizontal line corresponds to $\lambda\delta$.
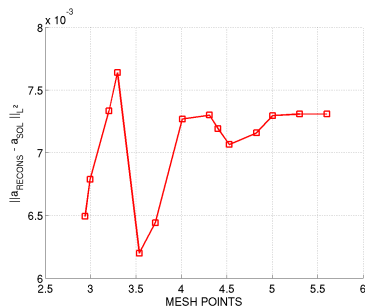
Figura: Evolution of the L^2-error. In the presence of noise, its minimum is attained for a coarser mesh satisfying the discrepancy principle of Equation (**??**).
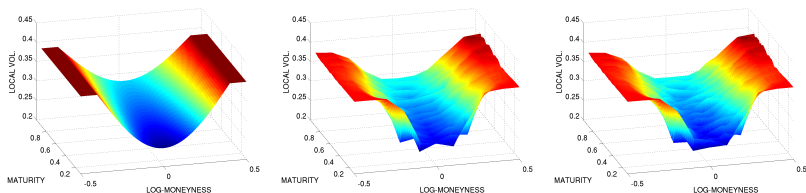
Figura: Left: original surface. Center and right: reconstructions corresponding to the first and second points satisfying the discrepancy principle of Figure 1, respectively.
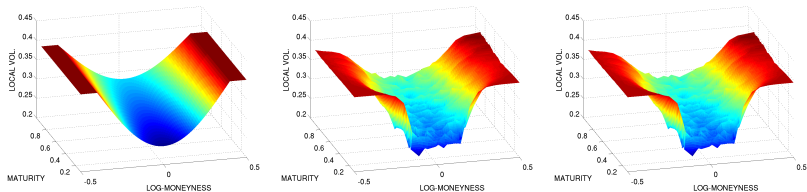
Figura: Left: original surface. Center and right: reconstructions satisfying the discrepancy principle of Figure 1.

Conclusions

- We propose Morozov's discrepancy principle in the same spirit of [1], [2], [3]. [4] in the context of nonlinear operators in a discrete setting.

- We also state that the continuous case, presented in these references, can be recovered from the discrete one, when the discretization level goes to infinity. noting really new

- We use this discrepancy principle as a rule to find appropriately the discretization level in the domain and the regularization parameter in Tikhonov regularization. Some how a 2 regularization parameter choice rule

- regularizing properties and (+ existence of source conditions) convergence rates results. as expected

Something more

- What we can say about iterative regularization????

Something more

- What we can say about iterative regularization????
  We know results in Hilbert Spaces and Linear Operators
- Not much for iterative regularization + nonlinear operators +
  discretization

Obrigado pela Atenção