On projective SGD type methods for solving large scale systems of ill-posed equations: Applications to machine learning

Antonio Leitão
acgleitao@gmail.com

XI Bienal de Matemática

São Carlos, 29/Jul a 2/Ago de 2024.

Outline of the talk:

(1) Motivation / Proposed approach

(2) The method under consideration: pSGD

(3) Exact data case

(4) Noisy data case

(5) Numerics: Two applications related to machine learning

Outline of the talk:

Motivation: Large scale systems of ill-posed equations

The inverse problem we are interested in consists of determining an unknown quantity $x \in X$ from the set of data $(y_0, \ldots, y_{N-1}) \in Y^N$, where X, Y are Hilbert spaces and $N >> 1$ is large.

In practical situations, the exact data are not known. Instead, only approximate measured data $y_i^\delta \in Y$ are available s.t.

$$\left\| y_i^\delta - y_i \right\| \leq \delta_i, \quad i = 0, \ldots, N-1, \tag{1}$$

with noise level $\delta_i > 0$ (notation $\delta := (\delta_0, \ldots, \delta_{N-1})$).

The finite set of data is obtained by indirect measurements of the parameter x, this process being described by the model $y_i = F_i(x)$, where $F_i : D(F_i) \subset X \to Y$ are nonlinear ill-posed operators.

The abstract formulation of the inverse problems under consideration reads: given the data y_i^δ and the levels of noise δ_i as in (1), find an approximate solution to the large scale linear system

$$F_i(x) = y_i^\delta, \quad i = 0, \ldots, N-1. \tag{2}$$

Applications: Big data analysis, Machine learning.

Proposed approach: A projective SGD type method (1/3)

Starting point of our approach: PLW and PLWK methods:

• The PLW method was originally proposed in [A.L. Svaiter, 2018] for solving nonlinear operator equations.

• Goal: to solve (1), (2) with $N = 1$ (i.e., $A_0 x = y_0$, $\|y_0 - y_0^\delta\| \leq \delta$).

• (x_k^δ) is generated as follows: at each iteration k, a half space

$$H_{x_k^\delta} := \left\{ z \in X \mid \langle z - x_k^\delta, A_0^*(y_0^\delta - A_0 x_k^\delta) \rangle \geq \|y_0^\delta - A_0 x_k^\delta\| (\|y_0^\delta - A_0 x_k^\delta\| - \delta) \right\}$$

separating the current iterate x_k^δ from the solution set $A_0^{-1}(y_0)$ is defined;

I.e., the next iterate x_{k+1}^δ is defined as a (relaxed) orthogonal projection of x_k^δ onto this set.

• PLW method summarized:

$$x_{k+1}^\delta := x_k^\delta - \theta_k \lambda_k A_0^* (A_0 x_k^\delta - y_0^\delta), \tag{3}$$

where $\theta_k \in (0,2)$ is a relaxation parameter and $\lambda_k \geq 0$ gives the exact orthogonal projection of x_k^δ onto $H_{x_k^\delta}$.

It corresponds to a Landweber iteration with stepsize defined by (relaxed) orthogonal projections onto the separating sets $H_{x_k^\delta}$.

Proposed approach: A projective SGD type method (2/3)

• The PLWK method was originally proposed in [A.L. Svaiter, 2016] for solving systems of nonlinear ill-posed equations as in (1), (2), $N > 1$.

• It consists in coupling the PLW method (3) with the Kaczmarz (cyclic) strategy and incorporating a bang-bang parameter, i.e.,

$$x_{k+1}^\delta := x_k^\delta - \theta_k \, \lambda_k \, \omega_k \, A_{[k]}^* \big(A_{[k]} x_k^\delta - y_{[k]}^\delta \big) . \tag{4a}$$

The parameters θ_k, λ_k have the same meaning as in (3), while

$$\omega_k = \omega_k(\delta_{[k]}, y_{[k]}^\delta) := \begin{cases} 1 & \|A_{[k]} x_k^\delta - y_{[k]}^\delta\| > \tau \delta_{[k]} \\ 0 & \text{otherwise} \end{cases} , \tag{4b}$$

— $\tau > 1$ is a positive constant.

— $[k] := (k \mod N) \in \{0, \dots, N-1\}$.

Proposed approach: A projective SGD type method (3/3)

Main goals:

• We propose and analyze a projective version of the SGD method, the projective stochastic-gradient (pSGD) method.

• We modify the SGD, to obtain an efficient method for computing stable approximate solutions to large scale systems of ill-posed operator equations (1), (2).

• Differently from [A.L. Svaiter, 2016] we propose here a stochastic (noncyclic) method based on the iteration (4), which uses an *a priori* stopping rule in the case of noisy data.

Outline of the talk:

Main assumptions

Given:
- some guess $x_0 \in X$ for the solution of (2).
- a sequence sequence $(\theta_k) \in \mathbb{R}$ of relaxation parameters.
- a positive constant γ.

Assumptions:

(A1) There exists $x^\star \in X$ s.t. $A_i\, x^\star = y_i$, $i = 0, \ldots, N-1$;
here $y_i \in R(A_i)$ are exact data;

(A2) $A_i : X \to Y$ are linear, bounded and ill-posed operators, i.e.,
even if the operator $A_i^{-1} : R(A_i) \to X$ exists, it is not continuous;

(A3) The sequence (θ_k) satisfies $0 < \inf_k \theta_k$ and $\sup_k \theta_k < 2$;

(A4) We choose $\gamma > C := \max_i \|A_i\|$;

(A5) The stopping index $k_\delta^* = k^*(\delta) \in \mathbb{N}$, satisfies

$$\lim_{\delta \to 0^+} k_\delta^* = \infty\,, \quad \lim_{\delta \to 0^+} \delta\, k_\delta^* = 0.$$

Description of the method pSGD

Given x_0, (θ_k) and γ as above, the sequence $(x_k^\delta) \in X$ is genarated by the iteration formula

$$x_{k+1}^\delta = x_k^\delta - \theta_k \, \lambda_{I_k} \, A_{I_k}^* (A_{I_k} x_k^\delta - y_{I_k}^\delta), \ k = 0, 1, \dots \quad (5a)$$

where the stepsize $\lambda_{I_k} := \lambda_{I_k}(x_k^\delta)$ is given by

$$\lambda_{I_k}(x_k^\delta) := \left\{ \begin{array}{ll} \dfrac{\|A_{I_k} x_k^\delta - y_{I_k}^\delta\| (\|A_{I_k} x_k^\delta - y_{I_k}^\delta\| - \delta_{I_k})}{\|A_{I_k}^* (A_{I_k} x_k^\delta - y_{I_k}^\delta)\|^2} & , \text{if} \|A_{I_k}^* (A_{I_k} x_k^\delta - y_{I_k}^\delta)\| > \gamma \delta_{I_k} \\ 0 & , \text{otherwise.} \end{array} \right.$$

$$\quad (5b)$$

Remarks:

• In a fixed probability space $(\Omega, \mathcal{F}, \mathbb{P})$, (I_k) is an independent and identically distributed sequence of indexes taking values in $\{0, \dots, N-1\}$.

• $p_i = \mathbb{P}(I_k = i), \quad 0 < p_i < 1, \quad \sum_i p_i = 1.$

• Additionally to depending on I_k, $\lambda_{I_k} = \lambda_{I_k}(x_k^\delta)$ also depends on the realization of $I_0, \dots, I_{k-1}$ through the random variable x_k^δ.

Remarks 1/5:

Remark (Exact projections)

The **sPLWK method with exact projections** *is obtained by taking* $\theta_k = 1$ *in* (5a).

This amounts to define x_{k+1}^δ *as the orthogonal projection of* x_k^δ *onto* H_{I_k, x_k^δ}*, where*

$$H_{i,x} := \left\{ z \in X \mid \langle z - x, \ A_i^*(y_i^\delta - A_i x) \rangle \geq \left\| y_i^\delta - A_i x \right\| \left(\left\| y_i^\delta - A_i x \right\| - \delta_i \right) \right\}$$

A relaxed variant of the sPLWK method uses $\theta_k \in (0, 2)$*, so that* x_{k+1}^δ *can be interpreted as a relaxed projection of* x_k^δ *onto* H_{I_k, x_k^δ}*.*

Remark (Separation property)

The solution set $A_i^{-1}(y_i)$ of the ith-equation is contained in $H_{i,x}$ for $i = 0, \ldots, N-1$ and all $x \in X$.

• *Indeed, for each $x^* \in A_i^{-1}(y_i)$ we have*

$$\langle x^* - x, A_i^*(y_i^\delta - A_i x) \rangle = \langle y_i - y_i^\delta + y_i^\delta - A_i x, y_i^\delta - A_i x \rangle$$
$$\geq \|y_i^\delta - A_i x\| \left(\|y_i^\delta - A_i x\| - \delta_i \right).$$

• *Moreover, from the definition of $H_{i,x}$ follows that $x \in H_{i,x}$ if and only if $\|y_i^\delta - A_i x\| \leq \delta_i$.*

*These two facts allow us to conclude that the convex set $H_{i,x}$ **separates** $A_i^{-1}(y_i)$ from $x \in X$ whenever $\|y_i^\delta - A_i x\| > \delta_i$.*

Remark (Exact data case)

Notice that $A_i^*(A_i x - y_i) = 0$ iff $A_i x = y_i$.[1] Thus, (5b) can be written as

$$
\lambda_{I_k} := \begin{cases}
\dfrac{\left\|A_{I_k} x_k - y_{I_k}\right\|^2}{\left\|A_{I_k}^*(A_{I_k} x_k - y_{I_k})\right\|^2} & , \text{ if } \left\|A_{I_k} x_k - y_{I_k}\right\| > 0 \\
0 & , \text{ otherwise}
\end{cases}
$$

• If $A_{I_k} x_k \neq y_{I_k}$, then x_{k+1} is given by (5a) with
$\lambda_{I_k} = \left\|A_{I_k} x_k - y_{I_k}\right\|^2 \left\|A_{I_k}^*(A_{I_k} x_k - y_{I_k})\right\|^{-2}$;

• If $A_{I_k} x_k = y_{I_k}$, then $x_{k+1} = x_k$ and $\lambda_{I_k} = 0$.

Conclusions:
— $\left\|y_{I_k} - A_{I_k} x_k\right\| > 0$ is sufficient to guarantee that the convex set H_{I_k, x_k} separates $A_{I_k}^{-1}(y_{I_k})$ from x_k (see Remark 2).
— For any $\theta_k \in (0, 2)$, x_{k+1} given by (5a) is closer to the solution set $A_{I_k}^{-1}(y_{I_k})$ than x_k.

---

[1] Indeed, $A_i x - y_i \in R(A_i)$ & $[A_i^*(A_i x - y_i) = 0] \Rightarrow [A_i x - y_i \in N(A_i^*) = R(A_i)^\perp]$.

Remarks: 4/5

Remark (Noisy data case)

The sPLWK method in (5) can be interpreted as follows:

- *If $\left\| A_{I_k}^*(A_{I_k} x_k^\delta - y_{I_k}^\delta) \right\| > \gamma \delta_{I_k}$, $\quad x_{k+1}^\delta$ is given by (5a) with λ_{I_k} as in (5b);*

- *If $\left\| A_{I_k}^*(A_{I_k} x_k^\delta - y_{I_k}^\delta) \right\| \leq \gamma \delta_{I_k}$, $\quad x_{k+1}^\delta = x_k^\delta$ and $\lambda_{I_k} = 0$.*

Conclusions:

— *Due to (A4), inequality $\left\| A_{I_k}^*(A_{I_k} x_k^\delta - y_{I_k}^\delta) \right\| > \gamma \delta_{I_k}$ in (5b) implies $\left\| y_{I_k}^\delta - A_{I_k} x_k^\delta \right\| > C^{-1} \gamma \delta_{I_k} > \delta_{I_k}$.*

— *From Remark 2/5 follows: $\quad H_{I_k, x_k^\delta}$ separates $A_{I_k}^{-1}(y_{I_k})$ from x_k^δ.*

— *Thus, for any $\theta_k \in (0, 2)$, x_{k+1}^δ given by (5a) is closer to the solution set $A_{I_k}^{-1}(y_{I_k})$ than x_k^δ.*

Remarks: 5/5

Remark (Lower bound for the stepsizes λ_{I_k})

• In the exact data case, (A2) implies $\lambda_{I_k} \geq C^{-2}$ whenever $\|A_{I_k} x_k - y_{I_k}\| > 0$.

In other words, C^{-2} is a natural lower bound for the stepsizes defined in (5b), whenever x_k is not a solution of $A_{I_k} x = y_{I_k}$.

• In the noisy data case, (A2) and (A4) imply $\lambda_{I_k} \geq (\gamma - C)(\gamma C^2)^{-1}$, whenever $\|A_{I_k}^*(A_{I_k} x_k^\delta - y_{I_k}^\delta)\| > \gamma \delta_{I_k}$.

Outline of the talk:

Average "gain" $\mathbb{E}[\|x^* - x_{k+1}\|^2] - \mathbb{E}[\|x^* - x_k\|^2]$

Lemma (Average gain)

Let (A1), (A2) hold true and (x_k) be a sequence generated by the sPLWK method (5). Then, for any x^* solution of (2) we have

$$\mathbb{E}[\|x^* - x_{k+1}\|^2] - \mathbb{E}[\|x^* - x_k\|^2] = \theta_k(\theta_k - 2)\,\mathbb{E}[\lambda_I \|A_I x_k - y_I\|^2], \ k \geq 0 \tag{6}$$

Moreover,
$$\mathbb{E}[\|x^* - x_{k+1}\|^2] - \mathbb{E}[\|x^* - x_k\|^2] \leq \frac{\theta_k(\theta_k - 2)}{C^2}\,\mathbb{E}[\|A_I x_k - y_I\|^2], \ k \geq 0$$

Proposition (Monotonicity)

Let the assumptions of Lemma 1 hold. Moreover, let (A3) hold. Then, for any x^* solution of (2) we have

$$\mathbb{E}[\|x^* - x_{k+1}\|^2] \ \leq \ \mathbb{E}[\|x^* - x_k\|^2], \ k = 0, 1, \dots \tag{7}$$

Convergence result

Proposition

Let the assumptions of Lemma 1 hold. Moreover, let Assumption (A3) hold. Then, the following series are summable:

$$\sum_{k=0}^{\infty} \theta_k(2-\theta_k)\mathbb{E}[\lambda_I\|A_I x_k - y_I\|^2], \ \sum_{k=0}^{\infty} \theta_k \, \mathbb{E}[\lambda_I\|A_I x_k - y_I\|^2], \ \sum_{k=0}^{\infty} \mathbb{E}[\|A_I x_k - y_I\|^2].$$

Theorem (Convergence for exact data)

Let assumptions (A1), (A2) ,(A3) hold true. Then, any sequence (x_k) generated by the sPLWK method (5) converges in mean square to $x^\dagger$, the x_0-minimal norm solution of (2): $\mathbb{E}[\|x^\dagger - x_k\|^2] \to 0$ as $k \to \infty$.

Outline of the talk:

Average "gain" $\mathbb{E}[\|x^* - x_{k+1}^\delta\|^2] - \mathbb{E}[\|x^* - x_k^\delta\|^2]$

Lemma (Average gain)

Let assumptions (A1), (A2) hold true and (x_k^δ) be a sequence generated by the sPLWK method (5). Then, for any x^* solution of (2) it holds

$$\mathbb{E}[\|x^* - x_{k+1}^\delta\|^2] - \mathbb{E}[\|x^* - x_k^\delta\|^2] \leq$$
$$\leq \theta_k(\theta_k - 2)\mathbb{E}[\lambda_I \|A_I x_k^\delta - y_I^\delta\|(\|A_I x_k^\delta - y_I^\delta\| - \delta_I)], \ k = 0, \ldots, k_\delta^*. \ (8)$$

Proposition (Monotonicity)

Let the assumptions of Lemma 5 hold. Moreover, let (A3), (A4) hold. Then
$$\mathbb{E}[\|x^* - x_{k+1}^\delta\|^2] \leq \mathbb{E}[\|x^* - x_k^\delta\|^2], \ k = 0, \ldots, k_\delta^*$$
for any x^* solution of (2).

A stability result

Theorem (Stability)

Let (A1), (A2), (A3) hold. Let $(\delta^j) = (\delta_0^j, \ldots, \delta_{N-1}^j) \in (\mathbb{R}^+)^N$ be a sequence with $\|\delta^j\| \to 0$ as $j \to \infty$, and $(y^{\delta^j}) = (y_0^{\delta^j}, \ldots, y_{N-1}^{\delta^j}) \in Y^N$ be a corresponding sequence of noisy data satisfying (1).

Moreover, let $(x_l)_{l \in \mathbb{N}}$ and $(x_l^{\delta^j})_{l=0}^{k_\delta^*}$ be the sequences generated by the sPLWK method in the case of exact and noisy data respectively; all sequences are generated using the same $(l_0, \ldots, l_k, \ldots)$.

Then, for each $k \in \mathbb{N}$ it holds

$$\lim_{j \to \infty} \mathbb{E}\big[\big\|x_k^{\delta^j} - x_k\big\|^2\big] = 0. \tag{9}$$

A semi-convergence result

Theorem (Semi-convergence)

Let assumptions (A1), ..., (A5) hold. Let $(\delta^j) = (\delta_0^j, \ldots, \delta_{N-1}^j) \in \mathbb{R}^N$ be a zero sequence, $(\mathbf{y}^{\delta^j}) = (y_0^{\delta^j}, \ldots, y_{N-1}^{\delta^j}) \in Y^N$ a corresponding sequence of noisy data satisfying (1).

Moreover, for each $j \in \mathbb{N}$, let $(x_k^{\delta^j})_{k=0}^{k^*(\delta^j)}$ be the corresponding sequence generated by the sPLWK method (these sequences are generated using the same $(I_0, \ldots, I_k, \ldots)$).

Then, $(x_{k^*(\delta^j)}^{\delta^j})_j$ converges in mean square to $x^\dagger$,[2] i.e,

$$\lim_{j \to \infty} \mathbb{E}\left[\left\| x_{k^*(\delta^{jn})}^{\delta^j} - x^\dagger \right\|^2\right] = 0. \tag{10}$$

---

[2] Here $x^\dagger$ is the x_0-minimal norm solution of (2).

Outline of the talk:

(5.1) Prediction of CO-concentration in a gas sensor array

Description of the chemical experiment:

• We use a data set collected in a gas delivery platform facility at the ChemoSignals Laboratory in the BioCircuits Institute at University of California, San Diego.

• This **measurement system platform** provides versatility for obtaining the **desired concentrations of the chemical substances** of interest with high accuracy.

• The **data set** contains the **readings of 16 chemical sensors** (Figaro Inc., US) of 4 different types:
TGS-2600, TGS-2602, TGS-2610, TGS-2620 (4 units of each type).

• These sensors were exposed to the **mixture of Ethylene and CO** at **varying concentrations** in air Ethylene concentration ranges 0–20ppm, CO concentration ranges 0–600 ppm

• For this gas mixture, the **measurement** was constructed by the **continuous acquisition of the 16-sensor array signals** for a duration of approximately **12 hours** without interruption.

(5.1) Prediction of CO-concentration in a gas sensor array

• Concentration transitions were set at random times, in the interval 80s–120s, and to random concentration levels.

Figure: Gas sensor array with 16 sensors.

• Data set was constructed such that all possible transitions are present: increasing, decreasing, setting to zero the concentration of one volatile while the concentration of the other volatile is kept constant.

• At the beginning, ending, and approximately every 10^4s, additional predefined concentration patterns with pure gas mixtures were inserted.

• The concentration ranges for Ethylene and CO were selected such that the induced magnitudes of the sensor responses were similar.
(for gas mixtures, lower concentration levels were favored)

• Experimental data available at the UC Irvine Machine Learning Repository:
https://archive.ics.uci.edu/ml/index.php
Data-set *"Gas sensor array under dynamic gas mixtures"*
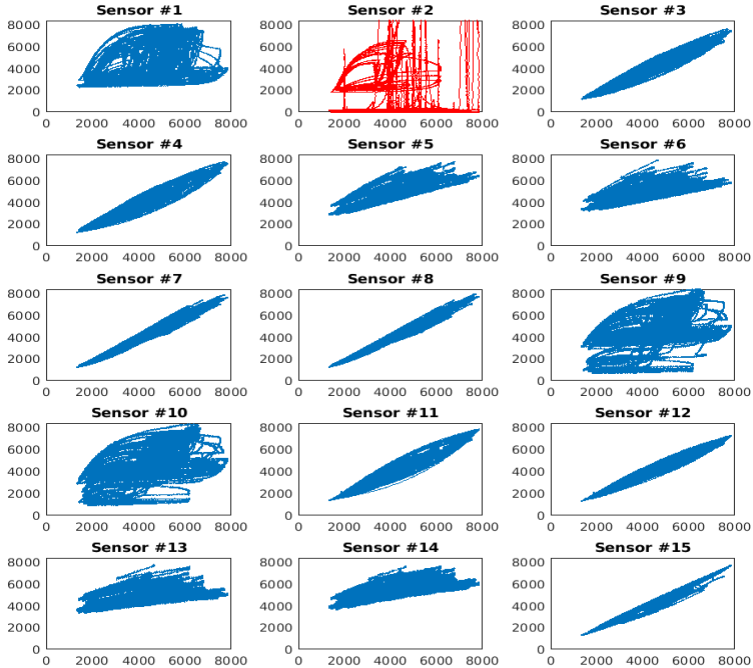
Figure: Scatter plots of Sensor #i data against Sensor #16 data

(5.1) Prediction of CO-concentration in a gas sensor array

Prediction of Gas-Concentration: Problem description

• We consider the problem of predicting the reading from sensor #16 based on the readings from the previous 14 sensors.
(readings from Sensor #2 are disregarded due to strong lack of accuracy)

• We use here a NN that inputs the readings from the first sensors and outputs a scalar value, which predicts the reading of the last sensor.

• Each sensor data consists of $4,188,262$ scalar measurements.
(Figure: scatter plots of sensor #i readings against sensor #16 readings)

• The structure of proposed NN reads:

— **Input:** $z \in \mathbb{R}^{14}$, readings of the first 14 sensors;

— **Output:** $NN(z; W, b) = \sigma(Wz + b) \in \mathbb{R}$,
   $W \in \mathbb{R}^{1 \times 14}$ matrix of weights;
   $b \in \mathbb{R}$ scalar bias;
   $\sigma : \mathbb{R} \to \mathbb{R}$ nonlinear activation function.

• This is a very simple NN with only one layer (the output layer).
  The dimention of the corresponding parameter space is 15.
  (this is the size of (W,b))

(5.1) Prediction of CO-concentration in a gas sensor array

• The inverse problem under consideration is a **NN training problem**. one aims to find an approximate solution (a pair of parameters (W,b)) to

$$F_i(W, b) = y_i^\delta, \; i = 0, \ldots, N_t - 1. \tag{11}$$

— $N_t < N$ is the size of the training set.

— y_i^δ readings of sensor #16 at instant $0 < i < N_t$.

— $F_i(W, b) := NN(z_i; W, b) = \sigma(W z_i + b)$.

— $z_i \in \mathbb{R}^{14}$ readings of the first 14 sensors at instant $0 < i < N_t$.

• Once the parameters (W, b) are determined, the performance $\mathcal{P}$ of the corresponding neural network $NN(\cdot; W, b)$ is defined by

$$\mathcal{P}\Big(NN(\cdot; W, b)\Big) := 1 - \frac{1}{N_T} \sum_{i=N_t}^{N_t+N_T-1} \frac{\left\| NN(z_i; W, b) - y_i^\delta \right\|}{\left\| y_i^\delta \right\|}. \tag{12}$$
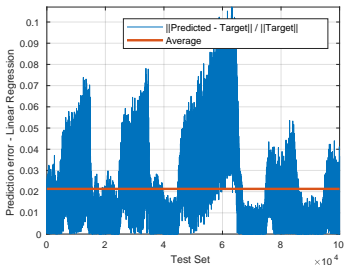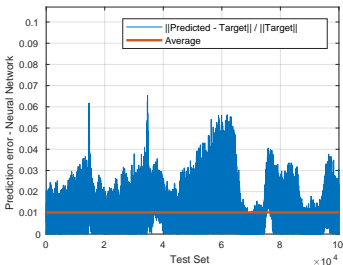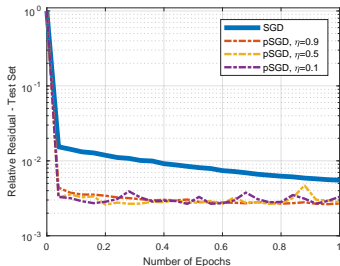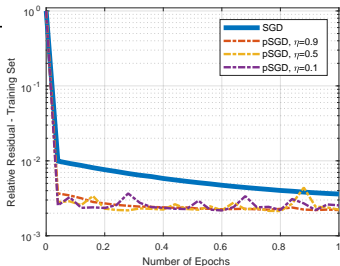
Figure: Prediction of CO-concentration in a gas-sensor array.
(TOP) Evolution of residual: Training set (Left) and Test set (Right);
(BOTTOM) Prediction Accuracy: Neural network(Left) and Linear regr.(Right)

(5.2) Classification problem for the MNIST database

The *Modified National Institute of Standards and Technology* (MNIST) database consists of images of handwritten digits.

Each image is accompanied by a corresponding label indicating the digit it represents.

This dataset is commonly used in the field of machine learning for developing neural network architectures, and for testing training algorithms for neural networks.

Figure: Sample images from the MNIST database (source Wikipedia).

- The MNIST database contains 60,000 training images (along with 10,000 testing images) of the ten digits.

- Each image consists of a 28×28 pixel array of grayscale levels.

- The corresponding data-files are accessible from http://yann.lecun.com/exdb/mnist/.

(5.2) Classification problem for the MNIST database

• The **mathematical model:** we use here a NN that inputs a 28×28 pixel array of grayscale levels (a vector in $\mathbb{R}^{784}$ with coordinates ranging from zero (black) to 255 (white)) and outputs a vector in $\mathbb{R}^{10}$.

• The **classification of the handwritten digit** depicted in the image is given by the coordinate of the output vector with maximal absolute value

• The architechture of the NN used in our experiments:

— **Input:** $z \in \mathbb{R}^{784}$, pixel array from the MNIST database;

— **Hidden layer:** $\tilde{z} := \sigma_1(W_1 z + b_1) \in \mathbb{R}^{64}$, where $W_1 \in \mathbb{R}^{64,784}$ and $b_1 \in \mathbb{R}^{64}$;

— **Output:** $NN(z; W_1, b_1, W_2, b_2) := \sigma_2(W_2 \tilde{z} + b_2) \in \mathbb{R}^{10}$, where $W_2 \in \mathbb{R}^{10,64}$ and $b_2 \in \mathbb{R}^{10}$.

W_1, W_2 are weight matrices;

b_1, b_2 are biases vectors;

$\sigma_1 : \mathbb{R}^{64} \to \mathbb{R}^{64}$ and $\sigma_2 : \mathbb{R}^{10} \to \mathbb{R}^{10}$ are nonlinear activation functions.

(5.2) Classification problem for the MNIST database

- The **classification of the input image** z is given by the scalar value $j \in \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$ defined by

$$j := arg \max_{0 \leq i \leq 9} \left| NN_i(z; W_1, b_1, W_2, b_2) \right|.$$

(notation: $NN(\cdot) = [NN_i(\cdot)]_{i=0}^{9} \in \mathbb{R}^{10}$)

- This NN has only 2-layers 784–64–10
  (one hidden layer and the output layer)

- the dimension of the corresponding parameter space is

$$50,890 \ = \ 64(784 + 1) + 10(64 + 1)$$

(i.e. the dimension of the set of parameters (W_1,W_2,b_1,b_2))

- Typically, much larger NN's are used for atacking the MNIST problem:
— The Deep NN in [Ciresan Et al. 2010] has 6–layers
$$784\text{-}2500\text{-}2000\text{-}1500\text{-}1000\text{-}500\text{-}10$$
— This NN achieves an accuracy rate of 99.65%.

(5.2) Classification problem for the MNIST database

• The inverse problem under consideration is a **NN training problem.**

• One aims to find an approximate solution (W_1, b_1, W_2, b_2) to the nonlinear system

$$F_i(W_1, b_1, W_2, b_2) \;=\; y_i, \;\; i = 0, \ldots, N_t - 1. \tag{13}$$

— $N_t = 60,000$ is the size of the training set

— $F_i(W_1, b_1, W_2, b_2) := NN(z_i; W_1, b_1, W_2, b_2)$
$$= \sigma_2\big(W_2\, \sigma_1(W_1 z_i + b_1) + b_2\big),$$

— z_i is the ith-image of the MNIST database for $i = 0, \ldots, N_t - 1$.

— The r.h.s. in (13) is a vector $y_i = (0, \ldots, 0, 1, 0, \ldots, 0) \in \mathbb{R}^{10}$
(the index of the coordinate "1" indicates the digit depicted in image z_i)

(note that the data in (13) is exact, i.e. the noise level is $\delta = 0$)

(5.2) Classification problem for the MNIST database

- The activation functions σ_1, $\sigma_2 : \mathbb{R} \to \mathbb{R}$ used in the above NN are variations of the *sigmoid function*, namely:

$$\sigma_1(t) = \tfrac{1}{2}\tanh(t/10) \quad and \quad \sigma_2(t) = 2\tanh(t/10).$$

- pSGD method is implemented for solving the NN training problem (13)
— Initial guess $(W_1^0, b_1^0, W_2^0, b_2^0)$ consists of random matrices/vectors.
— Three different runs of the pSGD method are presented in Figure 31.
— In each one of them the iteration $(W_1^k, b_1^k, W_2^k, b_2^k)$ is computed for 20 epochs, i.e. for $k = 1, \ldots, 20N_t$.
— In the first 2 runs we choose the sequence $\theta_k \equiv 1$ in (A4), while in the last run a random sequence $\theta_k \in (0,2)$ is choosen.
— The numerical results plotted in Figure 31 show:

(TOP) Evolution of relative residual on the training set:

$$\sum_{i=0}^{N_t-1} \frac{\left\| NN(z_i; W_1^k, b_1^k, W_2^k, b_2^k) - y_i \right\|}{\left\| NN(z_i; W_1^0, b_1^0, W_2^0, b_2^0) - y_i \right\|}.$$

(BOTTOM) Evolution of relative residual on the test set:

$$\sum_{i=N_t}^{N_t+N_T-1} \frac{\left\| NN(z_i; W_1^k, b_1^k, W_2^k, b_2^k) - y_i \right\|}{\left\| NN(z_i; W_1^0, b_1^0, W_2^0, b_2^0) - y_i \right\|}.$$

(5.2) Classification problem for the MNIST database
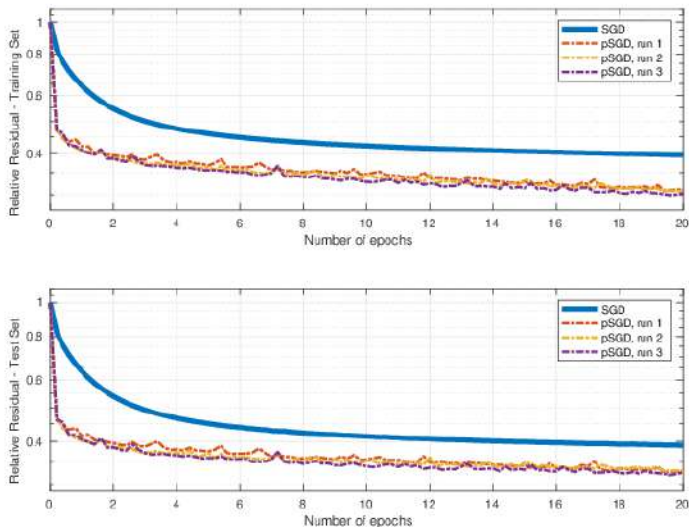
Figure: MNIST classification problem. (TOP) Evolution of the relative residual for training set; (BOTTOM) Evolution of the relative residual for test set.
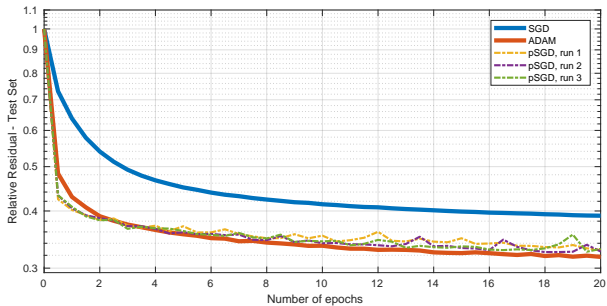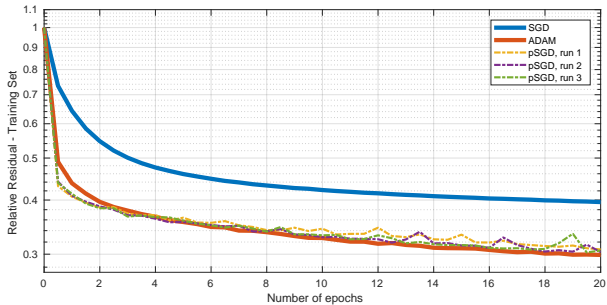
(5.2) Classification problem for the MNIST database

- $N_T = 10,000$ is the number of images in the MNIST database test set.

- For comparison, the SGD method was also implemented for solving (13) (the evolution of the corresponding residuals is also plotted)

- Following every $\frac{1}{10}N_t$ steps, the average relative residual is computed on the test set; the index $0 \leq k^* \leq 20N_t$ is chosen s.t.
  $(\mathbf{W_1^{k^*}}, \mathbf{b_1^{k^*}}, \mathbf{W_2^{k^*}}, \mathbf{b_2^{k^*}})$ **exhibits the smallest relative residual.**

- After selecting $(W_1^{k^*}, b_1^{k^*}, W_2^{k^*}, b_2^{k^*})$, the accuracy rate of the corresponding neural network $NN(\cdot; W_1^{k^*}, b_1^{k^*}, W_2^{k^*}, b_2^{k^*})$ is calculated using the test set.

- In this experiment, $k^* = 19.4N_t$ is obtained from the first run of the pSGD method.

- **The accuracy rate of $\mathbf{NN}(\cdot; \mathbf{W_1^{k^*}}, \mathbf{b_1^{k^*}}, \mathbf{W_2^{k^*}}, \mathbf{b_2^{k^*}})$ is 95.96%.**

(5.2) Classification problem for the MNIST database

Accuracy rate & Confusion matrix:

• The accuracy rate of $NN(\cdot; W_1^{k^*}, b_1^{k^*}, W_2^{k^*}, b_2^{k^*})$ is given by the trace of the confusion matrix divided by N_T.

• The confusion matrix is a table that is used to evaluate the performance of a classification model.

• It provides a summary of how well the model has classified the different classes in a dataset.

• It is typically used for problems like the MNIST classification, where the output of the model can belong to multiple classes.

• It displays the actual class labels of the data against the predicted class labels generated by the model.

• The main diagonal in Figure 3 represents the correctly classified instances, while the off-diagonal elements represent misclassifications.

• The final entry in a row/column represents the cumulative sum of all preceding elements in that particular row/column.

(5.2) Classification problem for the MNIST database
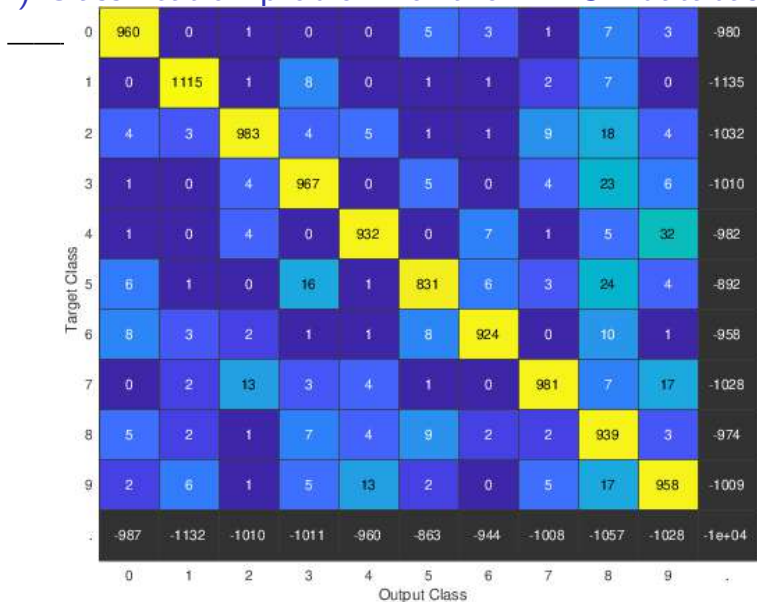
Figure: MNIST classification. Confusion matrix for $NN(\cdot; W_1^{k^*}, b_1^{k^*}, W_2^{k^*}, b_2^{k^*})$

Thank you for your time and your interest ! ! !

A level-set approach for problem of crack detection from electrical measurements

Antonio Leitão
acgleitao@gmail.com

Department of Mathematics
Federal University of St. Catarina, Brazil

Outline

1. Introduction

2. Modelling the parameter space

3. Numerical experiments

4. Tikhonov regularization

Collaborators (current project):
E. Hafemann (UFSC), A. De Cezaro (Rio Grande), A. Osses (Santiago)

📄 A. DeCezaro, E. Haffeman, A. Osses, A.L., *A regularization method based on level-sets for the problem of crack detection from electrical measurements,* Inv. Probl., submitted

Previous Collaborators:
O. Scherzer(2004/05), X.-C. Tai(2009/13), U. Ascher(2010), O. Dorn(2003)

Outline

Introduction	Modelling the parameter space	Numerical experiments	Tikhonov regularization	Bibliography
○●○○○○○○○	○○○○○○○○	○○○○○○○○○	○○○○○○○○○	○○

Abstract

- Inverse problem: to determine the position and shape of a crack in a bounded domain $\Omega \subset \mathbb{R}^2$ from electrical measurements on the boundary $\partial\Omega$.
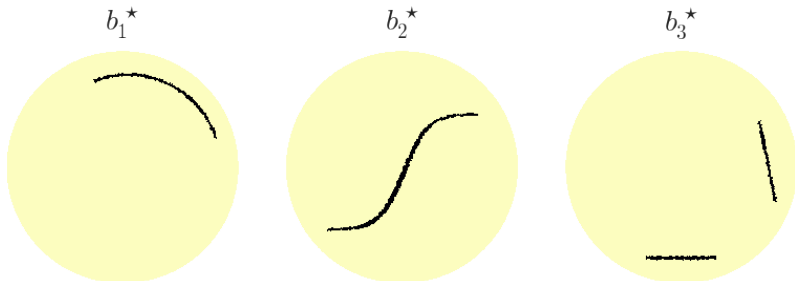
Figure: Tipical crack scenarios $b_k^\star$, $k = 1, 2, 3$.

• Based on the level-set approach in [2] and on the regularization strategy in [3], we propose a Tikhonov type method for stabilizing the inverse problem.

📄 D. Alvarez, O. Dorn, N. Irishina, M. Moscoso, *Crack reconstruction using a level-set strategy,* Journal of Computational Physics 228 (2009)

📄 A. De Cezaro, A. Leitão, X.-C. Tai, *On multiple level-set regularization methods for inverse problems,* Inverse Problems 25 (2009)

• An iterative method of multiple level-set type is derived from the optimality conditions for the Tikhonov functional, and a relation between this method and the iterated Tikhonov method is establised.

- We assume that the domain Ω has Lipschitz boundary and represents the specimen under investigation.
- A set of currents profiles $\{\eta_j\}_{j=1}^N$ are applied at $\partial\Omega$, for with, we have access to measurements of the corresponding potentials $\{u_j\}_{j=1}^N$ on $\partial\Omega$.
- The corresponding electric potential u_j satisfies

$$\nabla \cdot (b(x)\nabla u_j(x)) = 0, \, x \in \Omega, \qquad b(x)(u_j(x))_\nu = \eta_j(x), \, x \in \partial\Omega. \quad (1)$$

with $\int_{\partial\Omega} \eta_j = 0$, for $j = 1, \cdots, N$
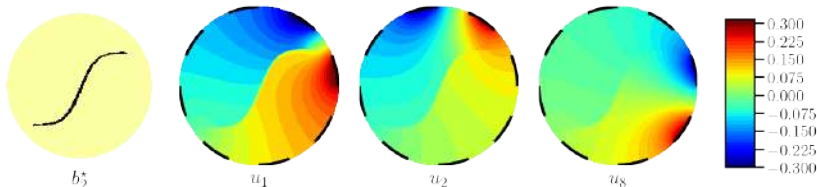
Figure: Typical NtD experiment. (LEFT) Crack $b_2^\star$. (RIGHT) Solutions u_1, u_2 and u_8 of (1) for the Neumann data η_1, η_2 and η_8 respectively.

Introduction	Modelling the parameter space	Numerical experiments	Tikhonov regularization	Bibliography
○○○○●○○○○	○○○○○○○○	○○○○○○○○○	○○○○○○○○○	○○

The Model problem

• Insulating cracks with finite conductivity contrast between the interior and the exterior of the crack are modeled by the conductivity coefficient $b(x)$.

— $b(x) = b_i$, if x is inside the crack;

— $b(x) = b_e$, if x is outside the crack;

— $b_e >> b_i$;

— $\beta > 0$ fixed crack thickness.

• This corresponds to the assumption that cracks can be modeled as a thin structure with small thickness along a curve contained in Ω.

• The inverse problem we are concerned with consists in identifying the coefficient function $b(x)$ from a finite number N of experiments, where the current profiles η_j, $j = 1, \ldots, N$, are chosen in an appropriate way and the corresponding measurements $\gamma_j := u_j|_{\partial\Omega}$ are available.

Introduction	Modelling the parameter space	Numerical experiments	Tikhonov regularization	Bibliography
○○○○○●○○	○○○○○○○○	○○○○○○○○○	○○○○○○○○○	○○

The Model problem

- Notation:

— $D_F := \{b \in L^\infty(\Omega); \overline{b} \geq b(x) \geq \underline{b} > 0, \text{ a.e. in } \Omega\}$

— $X := \{b \in L^\infty(\Omega); b(x) \geq \underline{b} > 0, \text{ a.e. in } \Omega\}$

— $Y := H^{1/2}(\partial\Omega)$

- If $\eta_j \in H^{-1/2}(\partial\Omega)$, the Neumann BVP in (1) has a unique solution

$$u_j \in H^1_* := \{u \in H^1(\Omega); \int_{\partial\Omega} u_j = 0\}.$$

- The crack detection problem can be written in terms of the system of nonlinear operator equations

$$
\begin{aligned}
F_j : D_F \subset X &\rightarrow Y \\
b &\mapsto F_j(b) = u_j|_{\partial\Omega} =: \gamma_j
\end{aligned}
\tag{2}
$$

- Literature overview (far from being complete):

— A. Friedman, M. Vogelius, Determining cracks by boundary measurements, Indiana University Mathematics Journal 38 (1989)

— M. Bruhl, M. Hanke, M. Pidcock, Crack detection using electrostatic measurements, ESAIM: Mathematical Modelling and Num. Anal. 35 (2001)

— Y. Boukari, H. Haddar, The factorization method applied to cracks with impedance boundary conditions, Inv. Probl. & Imaging 7 (2013)

— J. Guo, X. Zhu, The factorization method for cracks in EIT, Comp. Appl. Math. 40 (2021)

— A. Hauptmann, M. Ikehata, H. Itou, S. Siltanen, Revealing cracks inside conductive bodies by electric surface measurements, Inverse Problems 35 (2018)

— W-K. Won-Kwang Park, Performance analysis of multi-frequency topological derivative for reconstructing perfectly conducting cracks, Journal of Comput. Physics 335 (2017)

| Introduction | Modelling the parameter space | Numerical experiments | Tikhonov regularization | Bibliography |
| 00000000● | 00000000 | 000000000 | 000000000 | 00 |

Main results

- Main results:

  i) **Modelling the parameter space:**
  The parameters $b \in D_F$ are represented using pairs of level-set functions $(\varphi, \psi) \in H^1(\Omega)^2$, i.e., $b = P(\varphi, \psi)$ where P is a discontinuous operator;

  ii) **Tikhonov regularization approach:**
  The multiple level-set approach in *i*) is used to define a Tikhonov functional based on TV-H^1 regularization;

  iii) **Iterative method:**
  The optimality conditions for this Tikhonov functional allow the derivation of an iterative multiple level-set type method for solving the crack identification problem.

Introduction
00000000

Modelling the parameter space
●0000000

Numerical experiments
000000000

Tikhonov regularization
000000000

Bibliography
00

Outline

1. Introduction

2. Modelling the parameter space

3. Numerical experiments

4. Tikhonov regularization

| Introduction | Modelling the parameter space | Numerical experiments | Tikhonov regularization | Bibliography |
| 0000000 | 0●00000000 | 000000000 | 000000000 | 00 |

Motivation: Multiple leve-set methods for inverse problems

- Solve the abstract operator equation

$$F(u) = y, \quad \|y^\delta - y\|_Y \leq \delta,$$

$F : \mathbb{D} \subset X \to Y$ is a Fréchet diff. mapping X Banach, Y Hilbert space.

- Assumption: the solution u of the Inv.Probl. above is a simple function defined on a bounded domain $\Omega \subset \mathbb{R}^d$, $d = 2, 3$, and assuming at most N different values,

- Ansatz: A solution u can be represented in the form

$$u = c_1 H(\phi^1) H(\phi^2) + c_2 H(\phi^1)(1 - H(\phi^2)) +$$
$$c_3(1 - H(\phi^1)) H(\phi^2) + c_4(1 - H(\phi^1))(1 - H(\phi^2)) =: P(\phi^1, \phi^2),$$

("color level-set" or "multiple level-sel"; [Tai/Chan'04], [Chan/Vese'02] + Tai, Dorn, Ascher, van den Doel, A.L.).

- Rewrite the inverse problem in the form:

$$F(P(\phi^1, \phi^2)) = y,$$

and solve it in terms of (ϕ^1, ϕ^2), the level-set functions.

Introduction	Modelling the parameter space	Numerical experiments	Tikhonov regularization	Bibliography
0000000	0000000	000000000	000000000	00

Motivation: Multiple leve-set methods for inverse problems

- Choose a Tikhonov functional:
- The least square approach leads to the Santosa model

$$\mathcal{F}_\alpha(\phi^1, \phi^2) := \|F(P(\phi^1, \phi^2)) - y^\delta\|_Y^2,$$

- The ROF approach leads to the Chan-Vese model

$$\mathcal{F}_\alpha(\phi^1, \phi^2) := \|F(P(\phi^1, \phi^2)) - y^\delta\|_Y^2 + \alpha \sum_{j=1}^{2} |H(\phi^j)|_{\mathrm{BV}},$$

(in general one cannot guarantee the existence of a minimizers)

- The BV-H^1 approach leads to the models in [Scherzer/AL'05] and [Tai/AL'09]

$$\mathcal{G}_\alpha(\phi^1, \phi^2) := \|F(P(\phi^1, \phi^2)) - y^\delta\|_Y^2 + \alpha \sum_{j=1}^{2} \left\{ \beta |H(\phi^j)|_{\mathrm{BV}} + \|\phi^j - \phi_0^j\|_{H^1(\Omega)}^2 \right\}.$$

- (PLAY VIDEO) Example: Inverse Potential Problem in 2D.

| Introduction | Modelling the parameter space | Numerical experiments | Tikhonov regularization | Bibliography |
| 00000000 | 0000●0000 | 000000000 | 000000000 | 00 |

Level-set representation for cracks

Multiple level-set representation in [2] for cracks:

• A level-set function $\varphi : \Omega \to \mathbb{R}$ is chosen such that its zero level-set $\Gamma_\varphi := \{x \in \Omega;\ \varphi(x) = 0\}$ defines a connected curve within Ω; the cracks are located 'along' Γ_φ.

• Another level-set function $\psi : \Omega \to \mathbb{R}$ is chosen such the cracks are contained in the set $B := \{x \in \Omega;\ \psi(x) < 0\}$.

— The intersections of the level-set curve $\Gamma_\psi := \{x \in \Omega;\ \psi(x) = 0\}$ with Γ_φ coincide with the endpoints of the cracks.

— The position of the cracks corresponds to the set

$$S = S(\varphi, \psi) := \Gamma_\varphi \cap B.$$

Introduction	Modelling the parameter space	Numerical experiments	Tikhonov regularization	Bibliography
00000000	00000●000	000000000	000000000	00

Level-set representation for cracks

- We consider the cracks to have small fixed thickness $\beta > 0$ and conductivity $b_i > 0$ much smaller than the background value $b_e > 0$ (the three constants are known).

- The position of the cracks is represented by the set

$$S_\beta = S_\beta(\varphi, \psi) := \{x \in \Omega;\, 0 < \varphi(x) < \beta\} \cap \{x \in \Omega;\, \psi(x) < 0\}.$$
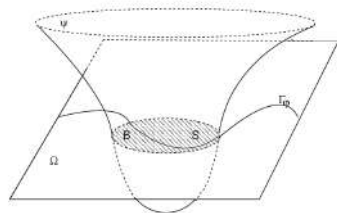

Figure: Multiple level-set representation for cracks.

| Introduction | Modelling the parameter space | Numerical experiments | Tikhonov regularization | Bibliography |
| 00000000 | 000000000 | 000000000 | 000000000 | 00 |
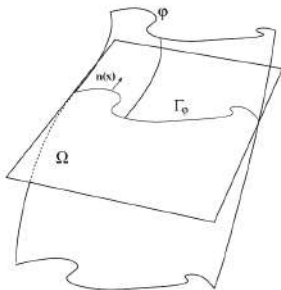
Level-set representation for cracks

- The conductivity distribution $b(x)$ in (1) is modeled by

$$b(x) = b_e + (b_i - b_e)\, \chi_{S_\beta}(x), \qquad (3)$$

where χ_{S_β} is the indicator function of the set S_β.

(multiple level-set representation of the parameter b)

- Following [3] we introduce the Heaviside projector

$$(H(\phi))(x) := \begin{cases} 1, & \text{if } \phi(x) > 0 \\ 0, & \text{if } \phi(x) \le 0 \end{cases},$$

and the translation $(H_\beta(\phi))(x) := H(\phi(x) - \beta)$.

- The conductivity distribution $b(x)$ can be written in the form

$$b = (b_i - b_e)\,[H(\phi) - H_\beta(\phi)]\, H(\psi) + b_e =: P(\phi, \psi). \qquad (4)$$

Introduction
00000000

Modelling the parameter space
0000000●0

Numerical experiments
000000000

Tikhonov regularization
000000000

Bibliography
00

Level-set representation for cracks

- As already observed in [3], the operator H maps $H^1(\Omega)$ into the space

$$\mathcal{V}_{0,1} := \{w \in L^\infty(\Omega) \mid w = \chi_S, \ S \subset \Omega \text{ measurable}, \ \mathcal{H}^1(\partial S) < \infty\}, \quad (5)$$

($\mathcal{H}^1(S)$ denotes the one-dimensional Hausdorff-measure of the set S)

- The operator P in (4) maps $H^1(\Omega) \times H^1(\Omega)$ into the admissible class

$$\mathcal{V} := \{w \in L^\infty(\Omega) \mid w = b_e + (b_i - b_e)\chi_S, \ S \subset \Omega \text{ measurable}, \ \mathcal{H}^1(\partial S) < \infty\},$$

- Within this framework, the inverse problem (2) can be written in the form of the system of operator equations

$$F_j(P(\varphi, \psi)) = \gamma_j^\delta, \ j = 1, \ldots, N. \tag{6}$$

(once a solution (φ, ψ) of (6) is obtained, a corresponding solution of (2) is given by $b = P(\varphi, \psi)$)

Introduction	Modelling the parameter space	Numerical experiments	Tikhonov regularization	Bibliography
00000000	0000000●	000000000	000000000	00

Towards stable approximate solutions

A Thikhonov approach:

• We follow [3] and introduce the Tikhonov functional

$$
\mathcal{G}_\alpha(\varphi, \psi) := \sum_{j=1}^{N} \|F_j(P(\varphi, \psi)) - \gamma_j^\delta\|_Y^2 + \alpha\big\{|H(\varphi)|_{\mathrm{BV}} + |H_\beta(\varphi)|_{\mathrm{BV}} + |H(\psi)|_{\mathrm{BV}}
$$
$$
+ \|\varphi - \varphi_0\|_{H^1(\Omega)}^2 + \|\psi - \psi_0\|_{H^1(\Omega)}^2\big\}, \quad (7)
$$

based on TV–H^1 penalization.

• The H^1–terms act simultaneously as a control on the size of the norm and as a regularization on the space $H^1(\Omega)$.

• The BV-seminorm terms are well know for penalizing the length of the Hausdorff measure of the boundary of the sets $\{x : \varphi(x) > 0\}$, $\{x : \varphi(x) > \beta\}$ and $\{x : \psi(x) > 0\}$.

Outline

1. Introduction
2. Modelling the parameter space
3. Numerical experiments
4. Tikhonov regularization

Introduction
0000000

Modelling the parameter space
00000000

Numerical experiments
0●0000000

Tikhonov regularization
000000000

Bibliography
00
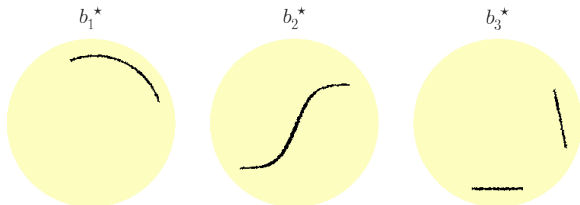
Exact solutions / Initial guess for the level-set method

Figure: Exact cracks $b_k^\star$, $k = 1, 2, 3$ used in the numerical experiments.
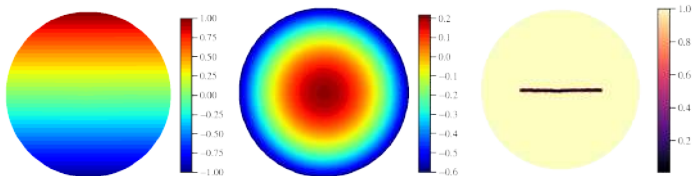

Figure: Initial guess for the level-set method: (LEFT) φ_0, (CENTER) ψ_0, (RIGHT) corresponding crack $P_\varepsilon(\varphi_0, \psi_0)$.
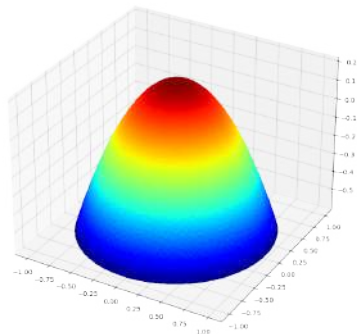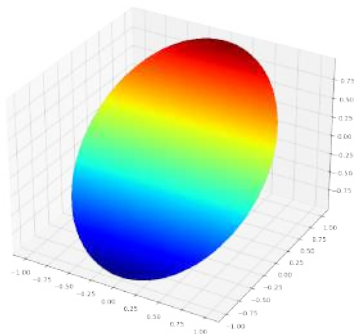
Introduction
00000000

Modelling the parameter space
00000000

Numerical experiments
000●00000

Tikhonov regularization
000000000

Bibliography
00

Initial guess for the level-set method

Figure: Initial guess for the level-set method: (LEFT) φ_0, (RIGHT) ψ_0.

Figure: Crack scenario $b_3^\star$ with $\delta = 1\%$. Evolution of $b_k = P_\varepsilon(\varphi_k, \psi_k)$ for $0 \le k \le 1500$.
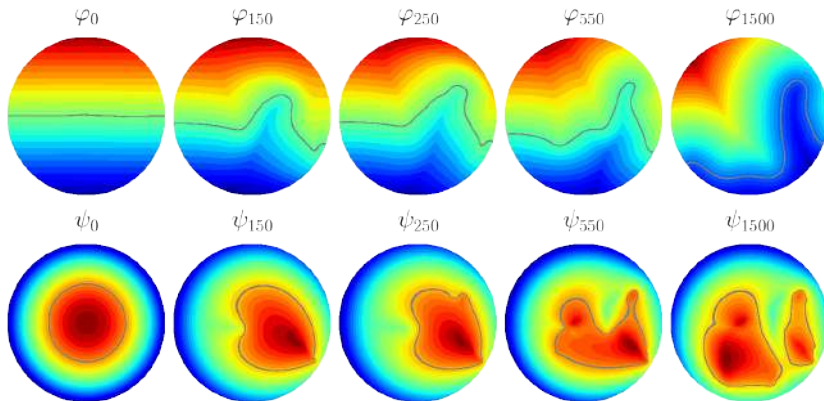
Figure: Crack scenario $b_3^\star$ with $\delta = 1\%$. Evolution of the level-set functions φ_k and ψ_k for $0 \leq k \leq 1500$.

Introduction
○○○○○○○○

Modelling the parameter space
○○○○○○○○

Numerical experiments
○○○○○●○○○

Tikhonov regularization
○○○○○○○○○

Bibliography
○○

Crack scenarios $b_1^\star, b_2^\star, b_3^\star$ with $\delta = 1\%$ & $\delta = 20\%$
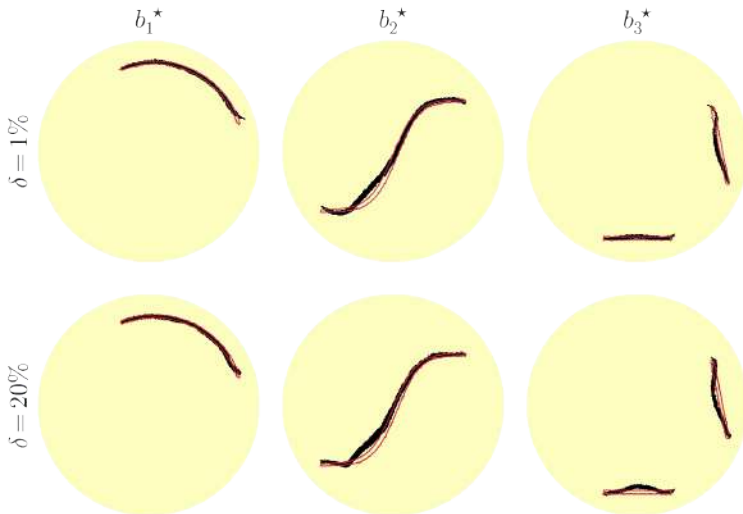
Figure: Reconstructions: Crack scenarios $b_1^\star$, $b_2^\star$, $b_3^\star$ divided by columns.
(TOP ROW) noise level $\delta = 1\%$.   (BOTOM ROW) noise level $\delta = 20\%$.

Introduction	Modelling the parameter space	Numerical experiments	Tikhonov regularization	Bibliography
000000000	000000000	00000000●00	000000000	00

Crack scenarios $b_1^\star, b_2^\star, b_3^\star$ with $\delta = 1\%$ & $\delta = 20\%$
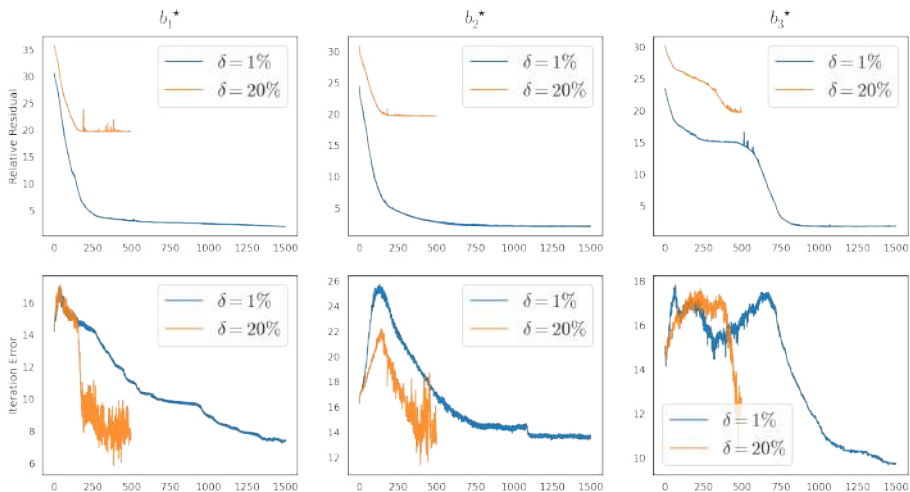
Figure: Residual/Error: Crack scenarios $b_1^\star$, $b_2^\star$, $b_3^\star$ divided by columns.
(TOP ROW) Relative residual.   (BOTTOM ROW) Relative iteration error.

Introduction
00000000

Modelling the parameter space
00000000

Numerical experiments
000000●0

Tikhonov regularization
000000000

Bibliography
00

Crack scenario $b_2^\star$ revisited / $N = 4, 8, 16$
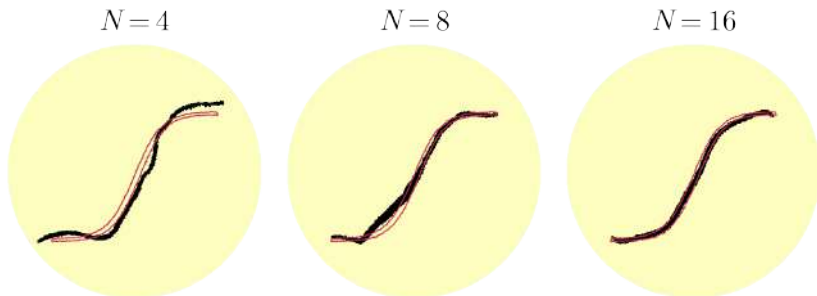
Figure: Crack scenario $b_2^\star$ revisited: Noise level $\delta = 20\%$.
Reconstruction results for distinct values of N.

- Current patterns:

(1 0 0 0 -1 0 0 0)   (1 -1 0 0 0 0 0 0)   (1 0 -1 0 0 0 0 0)

Figure: Crack scenario $b_1^\star$ revisited: $\delta = 20\%$ and $N = 16$.
Reconstruction results for distinct current patterns.

Introduction
○○○○○○○○

Modelling the parameter space
○○○○○○○○

Numerical experiments
○○○○○○○○○

Tikhonov regularization
●○○○○○○○○

Bibliography
○○

Outline

1. Introduction

2. Modelling the parameter space

3. Numerical experiments

4. Tikhonov regularization

| Introduction | Modelling the parameter space | Numerical experiments | Tikhonov regularization | Bibliography |
| 00000000 | 00000000 | 000000000 | 0●0000000 | 00 |

Main assumptions

(A1) $\Omega \subseteq \mathbb{R}^2$ is bounded with piecewise C^1 boundary $\partial\Omega$.

(A2) System (6) has a solution, i.e. there exists $b^* \in \mathcal{U}$ s.t. $F_j(b^*) = \gamma_j$, $j = 1, \ldots, N$.

There exist functions $\varphi^*, \psi^* \in H^1(\Omega)$ satisfying $P(\varphi^*, \psi^*) = b^*$, with $|\nabla\varphi^*| \neq 0$, $|\nabla\psi^*| \neq 0$ in a neighborhood of $\{\varphi^* \in [-\beta/2, \beta/2]\}$, $\{\psi^* = 0\}$ respectively.

It holds $H(\varphi^* + \beta/2) = z^1$, $H(\varphi^* - \beta/2) = z^2$, $H(\psi^*) = z^3$, for some $z^1, z^2, z^3 \in \mathcal{V}_{0,1}$.

- Continuous approximations to the operators P and H. Given $\varepsilon > 0$, define

$$P_\varepsilon(\varphi, \psi) := (b_i - b_e)[H_\varepsilon(\varphi) - H_{\beta,\varepsilon}(\varphi)] H_\varepsilon(\psi) + b_e , \qquad (8)$$

and

$$H_\varepsilon(\phi) := \left\{ \begin{array}{cl} 0, & \text{if } \phi < -\varepsilon \\ 1 + \frac{\phi}{\varepsilon}, & \text{if } \phi \in [-\varepsilon, 0] \\ 1, & \text{if } \phi > 0. \end{array} \right.$$

(the operators $H_{\beta,\varepsilon}$ are defined analogously)

Introduction	Modelling the parameter space	Numerical experiments	Tikhonov regularization	Bibliography
00000000	00000000	000000000	000●00000	00

Generalized minimizers

Definition (Generalized minimizers)

a) A tuple of functions $(z^1, z^2, z^3, \varphi, \psi) \in (L^\infty(\Omega))^3 \times (H^1(\Omega))^2$ is called **admissible** if there exist sequences $\{\varphi_k\}_{k \in \mathbb{N}}$, $\{\psi_k\}_{k \in \mathbb{N}}$ in $H^1(\Omega)$, and a sequence $\{\varepsilon_k\}_{k \in \mathbb{N}}$ of positive numbers converging to zero such that

$$\lim_{k \to \infty} \|\varphi_k - \varphi\|_{L^2(\Omega)} = 0, \quad \lim_{k \to \infty} \|\psi_k - \psi\|_{L^2(\Omega)} = 0,$$

$$\lim_{k \to \infty} \|H_{\varepsilon_k}(\varphi_k) - z^1\|_{L^1} = \lim_{k \to \infty} \|H_{\beta,\varepsilon_k}(\varphi_k) - z^2\|_{L^1} = \lim_{k \to \infty} \|H_{\varepsilon_k}(\psi_k) - z^3\|_{L^1} = 0.$$

b) A minimizer of $\hat{\mathcal{G}}_\alpha$ is considered to be any admissible tuple of the form $(z^1, z^2, z^3, \varphi, \psi)$ minimizing

$$\hat{\mathcal{G}}_\alpha(z^1, z^2, z^3, \varphi, \psi) := \sum_{j=1}^N \left\| F_j(q(z^1, z^2, z^3)) - \gamma_j^\delta \right\|_Y^2 + \alpha \rho(z^1, z^2, z^3, \varphi, \psi) \tag{9}$$

over all admissible tuples.

Introduction	Modelling the parameter space	Numerical experiments	Tikhonov regularization	Bibliography
00000000	00000000	000000000	000●00000	00

Generalized minimizers

Definition (continuation)

Here the functional ρ is defined by

$$
\rho(z^1, z^2, z^3, \varphi, \psi) := \inf \left\{ \liminf_{k \to \infty} \left(\mu_1 |H_{\varepsilon_k}(\varphi_k + \tfrac{\beta}{2})|_{\mathrm{BV}} + \mu_2 |H_{\varepsilon_k}(\varphi_k - \tfrac{\beta}{2})|_{\mathrm{BV}} \right. \right.
$$
$$
\left. \left. + \mu_3 |H_{\varepsilon_k}(\psi_k)|_{\mathrm{BV}} + \mu_4 \|\varphi_k - \varphi_0\|_{H^1}^2 + \mu_5 \|\psi_k - \psi_0\|_{H^1}^2 \right) \right\}, \quad (10)
$$

where the infimum is taken with respect to all sequences $\{\varepsilon_k\}$ and $\{(\varphi_k, \psi_k)\}$ satisfying (a).

c) A generalized minimizer of $\mathcal{G}_\alpha(\varphi, \psi)$ is a minimizer of $\hat{\mathcal{G}}_\alpha(z^1, z^2, z^3, \varphi, \psi)$ on the set of admissible tuples.

Lemma (Closedness of the set of admissible tuples)

Let $(z_k^1, z_k^2, z_k^3, \varphi_k, \psi_k)$ be a sequence of admissible tuples converging in $(L^1(\Omega))^3 \times (L^2(\Omega))^2$ to some $(z^1, z^2, z^3, \varphi, \psi) \in (L^\infty(\Omega))^3 \times (H^1(\Omega))^2$. Then $(z^1, z^2, z^3, \varphi, \psi)$ is an admissible tuple.

Introduction
00000000

Modelling the parameter space
00000000

Numerical experiments
000000000

Tikhonov regularization
0000●0000

Bibliography
00

Generalized minimizers

Lemma (coercivity and l.s.c. of ρ on the set of admissible tuples)

For each admissible quintuple $(z^1, z^2, z^3, \varphi, \psi)$, we have

$$\sum_{i=1}^{3} \mu_i |z^i|_{\mathrm{BV}} + \mu_4 \|\varphi - \varphi_0\|_{H^1}^2 + \mu_5 \|\psi - \psi_0\|_{H^1}^2 \leq \rho(z^1, z^2, z^3, \varphi, \psi). \quad (11)$$

Moreover, given a sequence $\{(z_k^1, z_k^2, z_k^3, \varphi_k, \psi_k)\}_{k \in \mathbb{N}}$ of admissible tuples such that $z_k^i \to z^i$ in $L^1(\Omega)$, $\varphi_k \rightharpoonup \varphi$ in $H^1(\Omega)$, $\psi_k \rightharpoonup \psi$ in $H^1(\Omega)$, where $(z^1, z^2, z^3, \varphi, \psi)$ is some admissible tuple, then

$$\rho(z^1, z^2, z^3, \varphi, \psi) \leq \liminf_{k \in \mathbb{N}} \rho(z_k^1, z_k^2, z_k^3, \varphi_k, \psi_k)$$

(i.e., ρ is weak-lower semi-continuous)

Introduction Modelling the parameter space Numerical experiments **Tikhonov regularization** Bibliography
0000000   00000000   000000000   00000●000   00

Well-Posedness of $\mathcal{G}_\alpha$

Proposition (Regularity property of the operators F_j)

Let the boundary data in the BVP (1) *satisfy* $\eta_j \in (W^{1-1/q,q}(\partial\Omega))'$, *for*
$q = p/(p-1)$, *for any* $p \in (2, p_0)$.
Then, the operators $F_j : D(F) \subset L^1(\Omega) \to Y$ *are continuous on* $D(F)$ *with*
respect to the $L^1(\Omega)$*-topology.*

Theorem (Well-Posedness of the functionals $\mathcal{G}_\alpha$)

The functional $\mathcal{G}_\alpha$ *in* (7) *attains generalized minimizers on the set of*
admissible tuples.

Sketch of the proof. First, notice that the set of admissible tuples is not empty.
Given a minimizing sequence of admissible quintuples for $\hat{\mathcal{G}}_\alpha$, it follows from
the coercivity of ρ, the Sobolev compact embedding (of H^1 in L^2) and the
compact embedding of BV into L^1, that this minimizing sequence converges
to some tuple which is admissible (due to Lemma 1).
From the weak lower semi-continuity of ρ together with the cont. of F_j and the
cont. of q, we conclude that the limit tuple is a minimizer of $\hat{\mathcal{G}}_\alpha$. $\square$

Introduction	Modelling the parameter space	Numerical experiments	Tikhonov regularization	Bibliography
0000000	00000000	000000000	000000●00	00

Convergence for exact and noisy data

- Main convergence and stability results.

Theorem (Convergence for exact data**)**

Assume that we have exact data, i.e. $\gamma_j^\delta = \gamma_j$. For every $\alpha > 0$, let $(z_\alpha^1, z_\alpha^2, z_\alpha^3, \varphi_\alpha, \psi_\alpha)$ denote a minimizer of $\hat{\mathcal{G}}_\alpha$ on the set of admissible tuples. Then, for every sequence of positive numbers $\{\alpha_k\}_{k \in \mathbb{N}}$ converging to zero there exists a subsequence, denoted again by $\{\alpha_k\}_{l \in \mathbb{N}}$, such that $(z_{\alpha_k}^1, z_{\alpha_k}^2, z_{\alpha_k}^3, \phi_{\alpha_k}^1, \phi_{\alpha_k}^2)$ is strongly convergent in $(L^1(\Omega))^3 \times (L^2(\Omega))^2$. Moreover, the limit is a solution of (6).

Theorem (Convergence for noisy data**)**

Let $\alpha = \alpha(\delta)$ be a function satisfying $\lim_{\delta \to 0} \alpha(\delta) = 0$ and $\lim_{\delta \to 0} \delta^2 \alpha(\delta)^{-1} = 0$. Moreover, let $\{\delta_k\}_{k \in \mathbb{N}}$ be a sequence of positive numbers converging to zero and $\gamma^{\delta_k} \in Y$ be corresponding noisy data satisfying (**??**). Then, there exist a subsequence (denoted again by $\{\delta_k\}$) and a sequence $\{\alpha_k := \alpha(\delta_k)\}$ such that $(z_{\alpha_k}^1, z_{\alpha_k}^2, z_{\alpha_k}^3, \varphi_{\alpha_k}, \psi_{\alpha_k})$ converges in $(L^1(\Omega))^3 \times (L^2(\Omega))^2$ to solution of (6).

| Introduction | Modelling the parameter space | Numerical experiments | Tikhonov regularization | Bibliography |
| 00000000 | 00000000 | 000000000 | 000000●0 | 00 |

A smoothed functional

- Consider the smoothed Tikhonov functional

$$\mathcal{G}_{\varepsilon,\alpha}(\varphi,\psi) := \sum_{j=1}^{N} \|F_j(P_\varepsilon(\varphi,\psi)) - \gamma_j^\delta\|_Y^2 + \alpha\big\{|H_\varepsilon(\varphi)|_{\mathrm{BV}} + |H_{\beta,\varepsilon}(\varphi)|_{\mathrm{BV}} + |H_\varepsilon(\psi)|_{\mathrm{BV}}$$
$$+ \|\varphi - \varphi_0\|_{H^1(\Omega)}^2 + \|\psi - \psi_0\|_{H^1(\Omega)}^2\big\}, \quad (12)$$

where $P_\varepsilon(\varphi,\psi) := q(H_\varepsilon(\varphi), H_{\beta,\varepsilon}(\varphi), H_\varepsilon(\psi))$.

Lemma (Well posedness of $\mathcal{G}_{\varepsilon,\alpha}$)

Given α, $\varepsilon > 0$ and $\varphi_0, \psi_0 \in H^1(\Omega)$, the functional $\mathcal{G}_{\varepsilon,\alpha}$ in (12) attains a minimizer on $(H^1(\Omega))^2$.

Theorem (Relation between minimizers of $\mathcal{G}_\alpha$ and $\mathcal{G}_{\varepsilon,\alpha}$)

Let $\alpha > 0$ be given. For each $\varepsilon > 0$ denote by $(\varphi_{\varepsilon_k,\alpha}, \psi_{\varepsilon_k,\alpha})$ a minimizer of $\mathcal{G}_{\varepsilon,\alpha}$. There exists a sequence of positive numbers $\{\varepsilon_k\}$ converging to zero such that $(H_{\varepsilon_k}(\varphi_{\varepsilon_k,\alpha}), H_{\varepsilon_k}(\psi\varepsilon_k,\alpha), H_{\beta,\varepsilon_k}(\psi\varepsilon_k,\alpha), \varphi_{\varepsilon_k,\alpha}, \psi_{\varepsilon_k,\alpha})$ converges strongly in $(L^1(\Omega))^3 \times (L^2(\Omega))^2$ and the limit minimizes $\hat{\mathcal{G}}_\alpha$ in the set of admissible 4-tuples.

Introduction | Modelling the parameter space | Numerical experiments | **Tikhonov regularization** | Bibliography
0000000 | 00000000 | 000000000 | 00000000● | 00

Step of the iterative multiple level-set method

- Conditions of optimality for $\mathcal{G}_{\varepsilon,\alpha}$

$$\alpha(\Delta - I)(\varphi_{k+1} - \varphi_k) = R^1_{\varepsilon,\alpha}(\varphi_k, \psi_k), \text{ in } \Omega \tag{13a}$$

$$(\varphi_{k+1} - \varphi_k)_\nu = 0, \text{ at } \partial\Omega \tag{13b}$$

$$\alpha(\Delta - I)(\psi_{k+1} - \psi_k) = R^2_{\varepsilon,\alpha}(\varphi_k, \psi_k), \text{ in } \Omega \tag{13c}$$

$$(\psi_{k+1} - \psi_k)_\nu = 0, \text{ at } \partial\Omega \tag{13d}$$

where

$$
\begin{aligned}
R^1_{\varepsilon,\alpha}(\varphi,\psi) &= \Theta^1_\varepsilon \, F'_j(P_\varepsilon(\varphi,\psi))^* \, (F_j(P_\varepsilon(\varphi,\psi)) - \gamma^\delta_j) + \\
&\quad + \textit{terms related to } |\nabla H_\varepsilon(\varphi)|_{\mathrm{BV}}, \ |\nabla H_{\beta,\varepsilon}(\varphi)|_{\mathrm{BV}}, \tag{14a}
\end{aligned}
$$

$$
\begin{aligned}
R^2_{\varepsilon,\alpha}(\varphi,\psi) &= \Theta^2_\varepsilon \, F'_j(P_\varepsilon(\varphi,\psi))^* \, (F_j(P_\varepsilon(\varphi,\psi)) - \gamma^\delta_j) + \\
&\quad + \textit{terms related to } |\nabla H_\varepsilon(\psi)|_{\mathrm{BV}}, \tag{14b}
\end{aligned}
$$

and

$$\Theta^1_\varepsilon(\varphi,\psi) = (b_i - b_e) \, H_\varepsilon(\psi) \, [H'_\varepsilon(\varphi) - H'_{\beta,\varepsilon}(\varphi)], \tag{15a}$$

$$\Theta^2_\varepsilon(\varphi,\psi) = (b_i - b_e) \, [H_\varepsilon(\varphi) - H_{\beta,\varepsilon}(\varphi)] \, H'_\varepsilon(\psi). \tag{15b}$$

Bibliography

📑 **A. De Cezaro, E. Haffeman, A. Osses, A.L.,** *A regularization method based on level-sets for the problem of crack detection from electrical measurements,* **Inverse Problems (2022), submitted**

📑 D. Alvarez, O. Dorn, N. Irishina, M. Moscoso, *Crack reconstruction using a level-set strategy,* Journal of Computational Physics 228 (2009)

📑 A. DeCezaro, X.-C. Tai, A.L., *On multiple level-set regularization methods for inverse problems,* Inverse Problems 25 (2009)

📑 F. Frühauf, O. Scherzer, A.L., *Analysis of regularization methods for the solution of ill-posed problems involving discontinuous operators,* SIAM Journal of Numerical Analysis 43 (2005)

📑 O. Scherzer, A.L., *On the relation between constraint regularization, level sets, and shape optimization,* Inverse Problems 19 (2003)

📑 F. Santosa, *A level-set approach for inverse problems involving obstacles,* ESAIM Control Optim. Calc. Var. 1 (1995/96)

📑 X.-C. Tai, T.F. Chan, *A survey on multiple level set methods with appl. for identifying piecewise constant functions,* Int.J.Num.Anal.Model.1 (2004)

Introduction
00000000

Modelling the parameter space
00000000

Numerical experiments
000000000

Tikhonov regularization
000000000

Bibliography
○●

The end

Thank You !!!