Seminario presso il Dipartimento di Matematica “Guido Castelnuovo”
Università degli Studi di Roma “La Sapienza”

6 novembre 2018

Regularization methods in variable exponent Lebesgue spaces

Claudio Estatico

Dipartimento di Matematica, Università degli Studi di Genova

Related joint works:

- Image restoration and preconditioning: Paola Brianzi, Fabio Di Benedetto; Pietro Dell'Acqua; Marco Donatelli
  Dep. of Mathematics (DIMA), Univ. Genova, Italy; DISAT, Univ. of Aquila; Univ. Insubria, Como, Italy.
- Microwave inverse scattering and Remote Sensing: Federico Benvenuto, Michele Piana and Alberto Sorrentino; Alessandro Fedeli, Matteo Pastorino and Andrea Randazzo; Flavia Lenti, Serge Gratton; David Titley-Peloquin; Matteo Alparone, Maurizio Migliaccio and Ferdinando Nunziata
  DIMA, Univ. Genova; Dep. of Engineering (DITEN), Univ. Genova, Italy ; CERFACS and ENSEEIHT, Université de Toulouse; France, Dep. of Bioresource Engineering, McGill University, Canada; Dep. of Engeneering, Univ. Napoli Parthenope, Italy.
- Subsurface prospecting: Patricia Diaz De Alba, Caterina Fenu, Giuseppe Rodriguez; Gian P. Deidda
  Dep. of Math. and Comp. Sci.; Dep. of Eng., Univ. Cagliari, Italy.

Introduction:

linear equations and inverse problems

in Hilbert vs Banach space settings

Inverse Problem

By the knowledge of some "observed" data y (i.e., the effect),
find an approximation of some model parameters x (i.e., the cause).

Given the data $y \in Y$, find (an approximation of) the unknown $x \in X$ such that

$$Ax = y$$

where $A : \mathcal{D} \subseteq X \longrightarrow Y$ is a known linear operator,

between two functional (**Hilbert** or **Banach**) spaces X and Y.

Inverse problems are ill-posed, they need regularization techniques.

(Recall.) Well-posed problem: (1) a solution always exists, AND (2) the solution is unique, AND (3) it depends continuously on the data.

(A classical) Example in Inverse Problems: Image Deblurring

Forward operator (blurring model by convolution):

A blurred version $y \in L^2(\mathbb{R}^2)$ of a true image $x \in L^2(\mathbb{R}^2)$ is given by

$$y(r) = \int_{\mathbb{R}^2} K(r-s)\, x(s)\, ds$$

where $r, s \in \mathbb{R}^2$, and K is the (known) impulse response of the imaging system, i.e., the point spread function (PSF). We write $y = Ax$.

Inverse problem (image deblurring):

Given (a noisy version of) y, find (an approximation of) x, by solving the linear equation

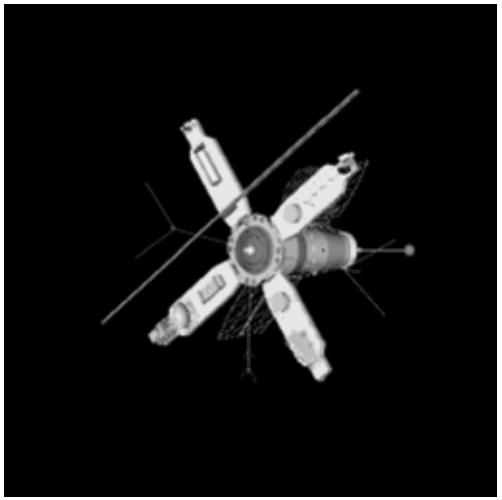
$$Ax = y\,.$$

(A classical) Example in Inverse Problems: Image Deblurring

Inverse problem (image deblurring):

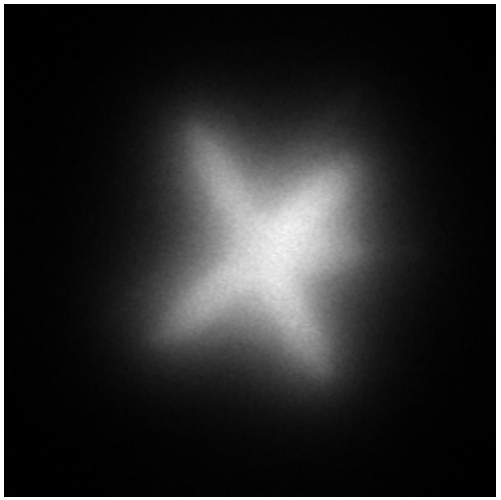
Given (a noisy version of) y, find (an approximation of) x, by solving

$$Ax = y\,.$$



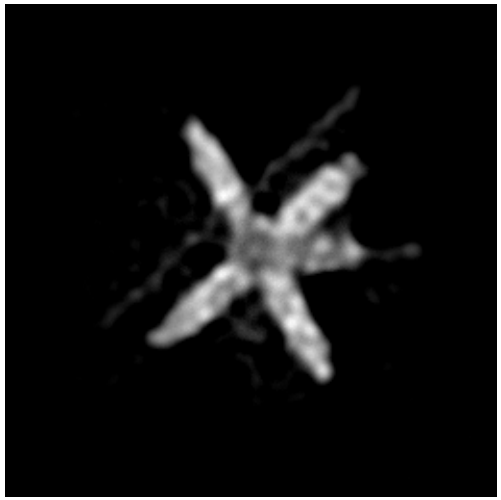
True Image

x



Blurred and noisy image

y



Restored Image

x_{rest}

Inverse Problems: Why do we need regularization?

I) In inverse problems, generally the bounded linear operator A is ill-posed (such as any compact operator in infinite dimensional spaces), so that its inverse in unbounded (i.e., the matrix of its discretization is ill-conditioned).

II) The data y is corrupted by noise, that is if the true data is $\bar{y} := A\bar{x} \in R(A) \subset Y$, where $\bar{x} \in X$ is the true solution and $R(A)$ is the range of A, we dispose only of a noisy data $y = \bar{y} + \eta$, where $\eta \in Y$ is the noise

The (noisy) linear system $Ax = y$, gives rise to the solution

$$x := A^{-1}y = A^{-1}(\bar{y} + \eta) = A^{-1}\bar{y} + A^{-1}\eta = \bar{x} + A^{-1}\eta$$

Since A is ill-conditioned, even if the noise is small $\|\eta\| \ll 1$, the effect of the noise on the solution can be very large

$$\|A^{-1}\eta\| \gg 1$$

What has happened, in short: the small noise η on the data is amplified in the inversion process, and the solution x is far from the exact one $\bar{x}$.

The conventional calculus lives in Hilbert spaces

A Hilbert space is a complete vector space endowed with a scalar product that allows (lengths and) angles to be defined and computed.

In any Hilbert spaces, what happens in our Euclidean world (that is, in our part of universe ...) always holds:

- Pythagorean theorem: if $u \perp v$, then $\|u+v\|^2 = \|u\|^2 + \|v\|^2$
- Parallelogram identity: $\|u+v\|^2 + \|u-v\|^2 = 2(\|u\|^2 + \|v\|^2)$
- Gradient of (the square of) the norm: $\nabla \frac{1}{2}\|u\|^2 = u$

The scalar product allows also for a natural definition of:

- orthogonal projection (best approximation) of u on v $P(u) = \langle u, v\rangle v / \|v\|^2$
- SVD decomposition (or eigenvalues/eigenvectors dec.) of linear operators (for separable Hilbert spaces, as any finite dimensional Hilbert space...)

Regularization: the "classical" framework in Hilbert spaces

In general, all the regularization methods for ill-posed functional equations have been deeply investigated in the context of **Hilbert spaces**.

Benefits:

Any linear (or linearized) operator in Hilbert spaces can be decomposed into a set of eigenfunctions by using the conventional spectral theory. This way, convergence and regularization properties of any solving method can be analyzed by considering the behavior of any single eigenfunction (i.e., we can use the Singular Value Decomposition - SVD ...).

Drawback:

Regularization methods in Hilbert spaces usually give rise to smooth (and over-smooth) solutions. In image deblurring, regularization methods in Hilbert spaces do not allow to obtain a good localization of the edges.

Regularization: the "novel" framework in Banach spaces

More recently, some regularization methods have been introduced and investigated in **Banach spaces**. In any Banach space, only distances between its elements can be defined and measured, but no scalar product (thus no "angle") is defined.

Benefits:

Due to the geometrical properties of Banach spaces, these regularization methods allow us to obtain solutions endowed with lower over-smoothness, which result, as instance, in a better localization and restoration of the discontinuities in imaging applications. Another useful property of the regularization in Banach space is that solutions are more sparse, that is, in general they can be represented by very few components.

Drawback:

The "Mathematics" is much more involving (the spectral theory cannot be used...). Convex analysis is required.

Regularization: Hilbert VS Banach spaces

	Hilbert spaces	Banach spaces
	$L^2(\mathbb{R}^n), H^s = W^{s,2}$	$L^p(\mathbb{R}^n), W^{s,p}, (p \geq 1); BV$
Benefits	Easier computation (Spectral theory, eigencomponents)	Better restoration of the discontinuities; Sparse solutions
Drawbacks	Over-smoothness (bad localization of edges)	Theoretically tricky (Convex analysis required)

Recap:

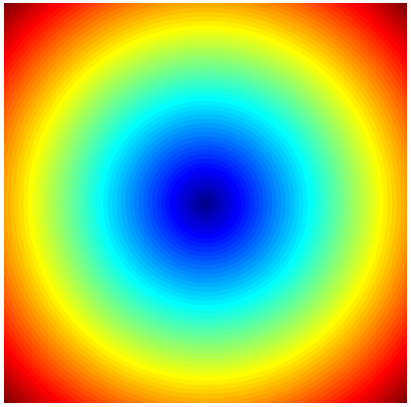
In Banach spaces we have no scalar product (so, no orthogonal projection), no Pythagorean theorem, no SVD...

The (sub-)gradient of (the square of) the norm is not linear, so that the least square solution (of a linear problem) is no more a linear problem...

Anyway, links between the two approaches are interesting and can help to derive new algorithms or improve old ones.

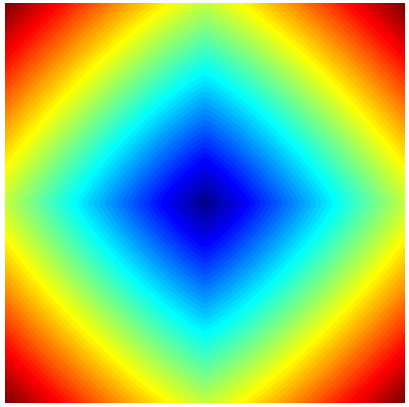
$L^2(R^2)$

$1/2 \quad \|x\|_2$



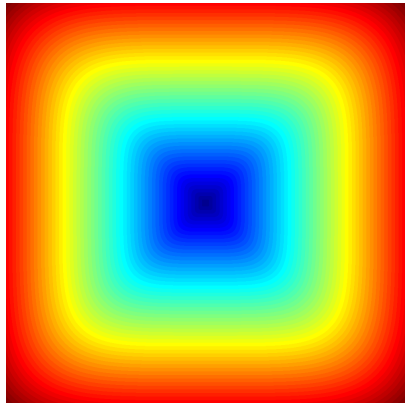
$L^{1.2}(R^2)$

$1/1.2 \quad \|x\|_{1.2}$

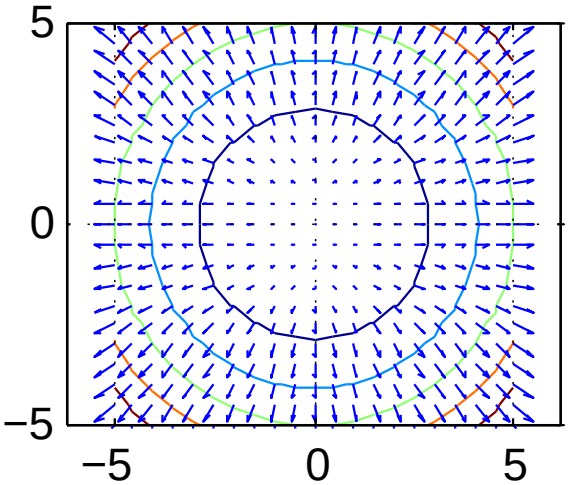


$L^5(R^2)$

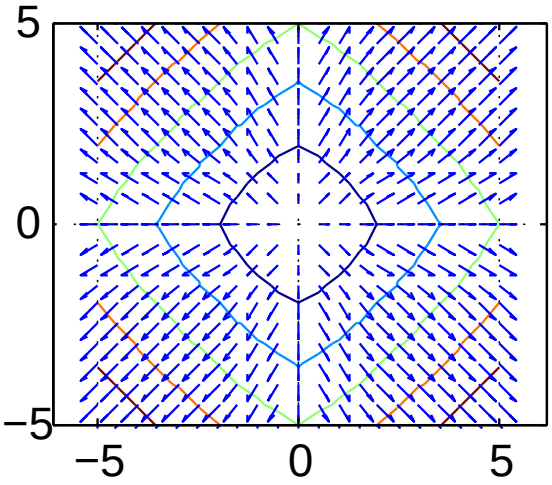
$1/5 \quad \|x\|_5$



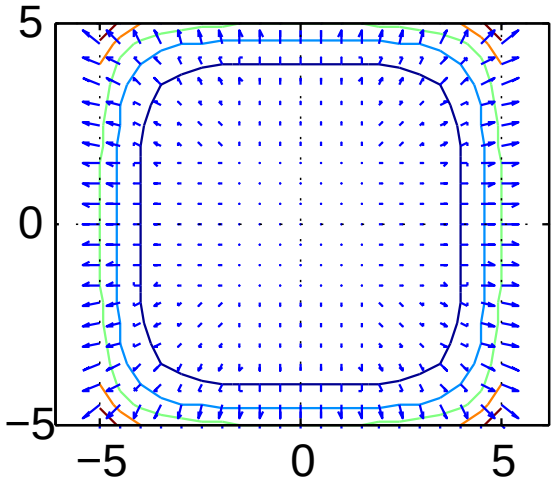
$\nabla (1/2 \quad \|x\|_2^2)$



$\nabla (1/1.2 \quad \|x\|_{1.2}^{1.2})$



$\nabla (1/5 \quad \|x\|_5^5)$



I

Iterative regularization algorithms:

From Hilbert to Banach space setting

Variational approach for inverse problems (i.e., optimization)

Instead of directly solving the operator equation $Ax = y$, we minimize the functional $H : X \longrightarrow [0, +\infty)$

$$H(x) = \frac{1}{r}\|Ax - y\|_Y^r \,,$$

or the Tikhonov-type variational regularization functional

$$H_\alpha(x) = \frac{1}{r}\|Ax - y\|_Y^r + \alpha\mathcal{R}(x) \,,$$

where $1 < r < +\infty$ and $\mathcal{R} : X \longrightarrow [0, +\infty)$ is a convex functional, and $\alpha > 0$ is the regularization parameter.

The "data-fitting" term $\frac{1}{r}\|Ax - y\|_Y^r$ is called as residual (usually in mathematics) or cost function (usually in engineering).

The "penalty" term $\mathcal{R}(x)$ is often $\frac{1}{r^*}\|x\|_X^{r^*}$, or $\frac{1}{r^*}\|\nabla x\|_X^{r^*}$ or $\frac{1}{r^*}\|Lx\|_X^{r^*}$, for a differential operator L which measures the non-regularity of x, where r^* is the Hölder conjugate of r, that is, $\frac{1}{r} + \frac{1}{r^*} = 1$.

Iterative regularization algorithms by gradient methods

In the variational approach, instead of solving straightly the operator equation $Ax = y$, we minimize the **residual functional** $H : X \longrightarrow [0, +\infty)$

$$H(x) = \frac{1}{r}\|Ax - y\|_Y^r \,.$$

Basic iterative scheme (set x_0 arbitrarily):

$$x_{k+1} = x_k - \lambda_k \Phi_A(x_k, y)$$

where the operator $\Phi_A : X \times Y \longrightarrow X$ returns a value

$$\Phi_A(x_k, y) \approx \nabla H(x_k) = \nabla \left(\frac{1}{r}\|Ax_k - y\|_Y^r \right),$$

that is, (an approximation of) the "gradient" of the functional $\frac{1}{r}\|Ax - y\|_Y^r$ in the point x_k and $\lambda_k > 0$ is the step length.

This way, the iterative schemes are all different generalizations of the basic gradient descent method.

The Landweber algorithm in Hilbert spaces

In the conventional case (i.e., both X and Y are Hilbert spaces) we consider the least square functional

$$H_2(x) = \frac{1}{2}\|Ax - y\|_Y^2 \,.$$

Direct computation of ∇H_2 by chaining rule for derivatives (in $\mathbb{R}^n$)

$$\begin{aligned}\nabla H_2(x) &= \left(\left(\left(\nabla \tfrac{1}{2}\|u\|^2\right)\big|_{u=Ax-y}\right)^* \mathcal{J}(Ax-y)\right)^* \\ &= ((Ax-y)^* A)^* = A^*(Ax-y)\end{aligned}$$

leads to the "simplest" iterative method: **the Landweber algorithm**

$$x_{k+1} = x_k - \lambda A^*(Ax_k - y)$$

Since H_2 is convex, there exists a non-empty set of stationary points, i.e. $\nabla H_2(x) = 0$, which are all minimum points of H_2.

The positive and constant step size $\lambda \in (0, 2/\|A^*A\|)$, yields $H_2(x_{k+1}) < H_2(x_k)$ and also guarantees the convergence of the iterations.

The Landweber algorithm in Hilbert spaces

$$x_{k+1} = x_k - \lambda A^*(Ax_k - y)$$

The Landweber algorithm as iterative regularization of $(A^*A)^{-1}A^*$

From

$$x_{k+1} = x_k - \lambda A^*(Ax_k - y)$$

by induction, it is simple to verify that

$$x_{k+1} = \lambda \sum_{i=0}^{k} (I - \lambda A^*A)^i A^*y + (I - \lambda A^*A)^{k+1} x_0 \,,$$

By considering the polynomial of degree k

$P_k(s) = \sum_{i=0}^{k} (1-s)^i = \frac{1-(1-s)^{k+1}}{s} \longrightarrow s^{-1}$ $(k \longrightarrow +\infty)$

if $0 < s < 2$, and $P_k(0) = k+1$, we have that, formally, if $\lambda \in (0, 2/\|A^*A\|_2)$,

$$x_k = \lambda P_k(\lambda A^*A)A^*y + (I - \lambda A^*A)^k x_0 \longrightarrow x = (A^*A)^{-1}A^*y \,,$$

which is a least square solution of $Ax = y$ (provided that $(A^*A)^{-1}$ exists).

From Hilbert to Banach spaces regularization (I)

$$x_{k+1} = x_k - \lambda A^*(Ax_k - y)$$

Formally, A^* is the dual operator of A, that is, the operator

$$A^* : Y^* \longrightarrow X^* \qquad \text{such that}$$

$$y^*(Ax) = (A^*y^*)(x)\,, \quad \forall x \in X \;\text{ and }\; \forall y^* \in Y^*\,,$$

where X^* and Y^* are the dual spaces of X and Y.

If X and Y are also Hilbert spaces, then X is isometrically isomorph to X^* and Y is isometrically isomorphic to Y^* (by virtue of Riesz representation Theorem), and the operator A^* can be identified with $A^* : Y \longrightarrow X$, where the isomorphisms $\mathcal{I}_Y : Y \longrightarrow Y^*$ and $\mathcal{I}_{X^*} : X^* \longrightarrow X$ have been implicitly applied.

$$A^* \leftrightarrow \mathcal{I}_{X^*} A^* \mathcal{I}_Y$$

This way, the Landweber method is well defined in Hilbert spaces (...only!)

From Hilbert to Banach spaces regularization (II)

$$x_{k+1} = x_k - \lambda A^*(Ax_k - y)$$

In Banach spaces, the term $A^*(Ax_k - y)$ is senseless since $A^* : Y^* \longrightarrow X^*$ cannot be applied to $Ax_k - y \in Y$. The same occurs for the sum of $x_k \in X$ and $-\lambda A^* w \in X^*$, whatever $w \in Y^*$ could be.

This shows how the situation is now much more challenging!

The key point: To generalize from Hilbert to Banach spaces we consider the so-called duality mappings.

A duality map is a special function which allows us to associate an element of a Banach space B with an element of its dual B^*.

Notation: For $b \in B$ and $b^* \in B^*$, let $b^*(b) \in \mathbb{R}$ be denoted by the pairing notation $b^*(b) =< b^*, b >=< b, b^* >$.

From Hilbert to Banach spaces regularization (III)

Formally, for $r > 1$ the duality map J_r^B of B is defined as the multi-valued operator $J_r^B : B \longrightarrow 2^{B^*}$, where 2^{B^*} denoted the power set of B^*, such that

$$J_r^B(b) = \{b^* \in B^* : < b^*, b >= \|b\|_B \|b^*\|_{B^*},\ \|b^*\|_{B^*} = \|b\|_B^{r-1}\}\,, \quad \forall b \in B.$$

The value of $r > 1$ has not essential meaning, since it acts on all the duality maps only as a weight, or scaling factor, by virtue of the identity

$$J_r^B(b) = \|b\|_B^{r-2} J_2^B(b) \qquad \forall b \in B\,, \ \ b \neq 0.$$

Moreover, if the Banach space is also a Hilbert space, then J_2^B is the isometric isomorphism between B^* and B, and it can be identified with the identity operator by virtue of the Riesz Theorem, that is, $J_2^B(b) = b \ \ \forall b \in B$, whenever B is a Hilbert space.

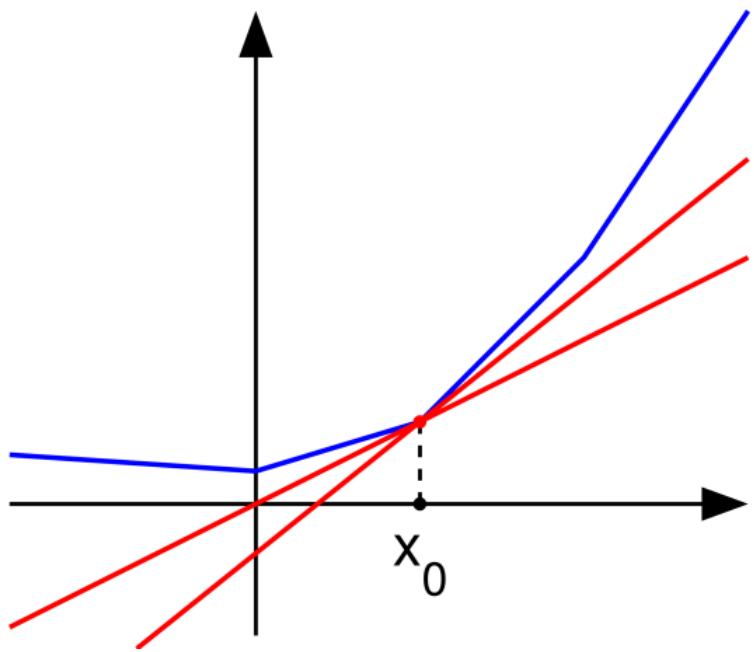
From Hilbert to Banach spaces regularization (IV)

The duality map has a more illustrative meaning in the context of minimization of convex functional.

Let f be a convex functional $f : B \longrightarrow \mathbb{R}$.

The subdifferential of f is the multi-valued operator $\partial f : B \longrightarrow 2^{B^*}$ s.t.

$$\bar{b}^* \in \partial f(b) \Longleftrightarrow f(c) \geq f(b) + < \bar{b}^*, c - b > \qquad \forall c \in B\,.$$



From Hilbert to Banach spaces regularization (V)

Theorem (Asplund)
Let B be a Banach space and let $r > 1$. A duality map J_r^B is the subdifferential of the convex functional $f : B \longrightarrow \mathbb{R}$ defined as $f(b) = \frac{1}{r}\|b\|_B^r$, that is

$$J_r^B = \partial f = \partial \left(\frac{1}{r} \| \cdot \|_B^r \right) .$$

The Asplund Theorem represents the key to extend the Landweber method to Banach spaces: The Landweber method corresponds now to a minimization procedure on the convex functional $H_r : X \longrightarrow \mathbb{R}$

$$H_r(x) = \frac{1}{r}\|Ax - y\|_Y^r .$$

Now $\partial H_r : Y \longrightarrow 2^{Y^*}$ can be computed by using the duality map J_r^Y

$$\partial H_r(x) = \partial \left(\frac{1}{r}\|Ax - y\|_Y^r \right) = A^* J_r^Y (Ax - y) .$$

The Landweber method in Banach spaces

Let $r > 1$ a fixed weight value. Let $x_0 \in X$ an initial guess (the null vector $x_0 = 0 \in X$ is often used in the applications), and set $x_0^* = J_r^X(x_0) \in X^*$.

$$\text{For } \; k = 0, 1, 2, \ldots$$

$$\begin{aligned} x_{k+1}^* &= x_k^* - \lambda_k A^* J_r^Y (A x_k - y)\,, \\ x_{k+1} &= J_{r^*}^{X^*}(x_{k+1}^*)\,, \end{aligned}$$

where r^* is the Hölder conjugate of r, that is, $\frac{1}{r} + \frac{1}{r^*} = 1$, and the step sizes λ_k are suitably chosen.

Here the duality map $J_r^X : X \longrightarrow 2^{X^*}$ acts on the iterates $x_k \in X$, and the duality map $J_{r^*}^{X^*} : X^* \longrightarrow 2^{X^{**}}$ acts on the iterates $x_k^* \in X^*$: In order to be well defined, it is only required that the space X is reflexive, that is X^{**} is isometrically isomorph to X, so that $J_{r^*}^{X^*} \subseteq X$.

Landweber iterative method in Hilbert spaces

$A : X \longrightarrow Y \qquad A^* : Y \longrightarrow X \qquad H_2(x) = \frac{1}{2}\|Ax - y\|_Y^2$

$$x_{k+1} = x_k - \lambda A^*(Ax_k - y)$$

Landweber iterative method in Banach spaces

$A : X \longrightarrow Y \qquad A^* : Y^* \longrightarrow X^* \qquad H_r(x) = \frac{1}{r}\|Ax - y\|_Y^r$

$$x_{k+1} = J_{r^*}^{X^*}(J_r^X x_k - \lambda_k A^* J_r^Y (Ax_k - y))$$

Some remarks:

(i) Any duality map is in general nonlinear (and multi-valued...), so that, differing from the Hilbert space case, the Landweber method is not linear.

(ii) In the Banach space L^p, $p \in (1, +\infty)$, by direct computation we have

$$J_r^{L^p}(x) = \|x\|_p^{r-p}|x|^{p-1}\mathrm{sgn}(x)$$

$J_r^{L^p}$ is a non-linear, single-valued, diagonal operator, which cost $O(n)$. $J_r^{L^p}$ does not increase the global numerical complexity $O(n \log n)$ of linear problems with (FFT-based) structured matrices. Note that $J_2^{L^2} = I$.

A convergence result for the Landweber method in Banach spaces

Proposition (Schöpfer,Louis,Schuster, *Inv. Prob.*, 2006)
Let X be a reflexive Banach space, and Y a (arbitrary) Banach space. Let $y \in \mathcal{R}(A)$ and let $x^\dagger$ the minimum norm pseudo-solution of $Ax = y$.

If $\lambda_k > 0$ is suitably (...) chosen for all k, then the sequence of the iterations (x_k) converges strongly to $x^\dagger$, that is,

$$\lim_{k \longrightarrow +\infty} \|x_k - x^\dagger\|_X = 0$$

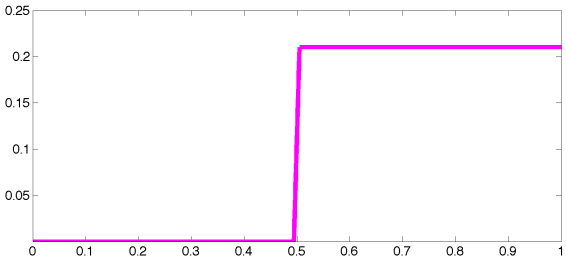
If the data y is noisy, an early stop of the iterations (by discrepancy principle) gives rise to an iterative regularization method.

In Hilbert setting, the iterations are defined in the (primal) space X.

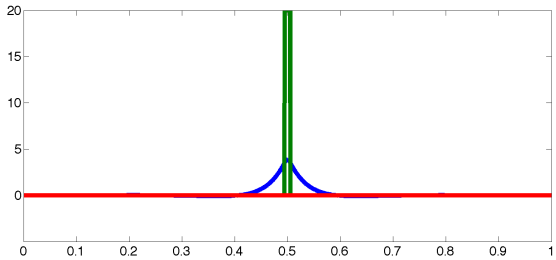
In Banach setting, the iterations are defined in the dual space X^* and are linked to the ("wide"...) Banach fixed point theory.

A simple 1d example

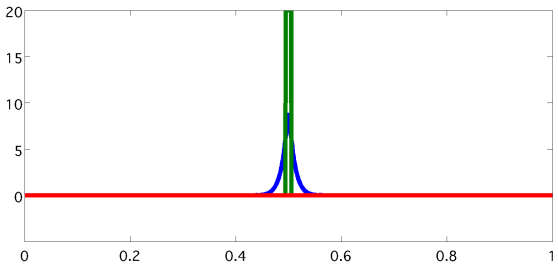
$$(F(x))(s) = \int_0^s x(t)\,dt$$



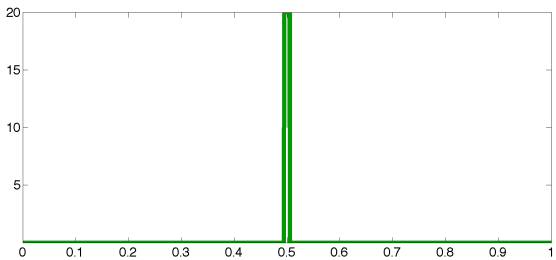
Data y



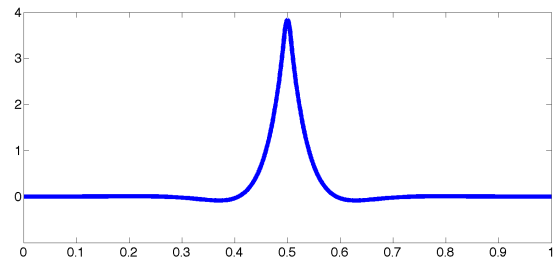
Hilbert Restoration ($p = 2$)



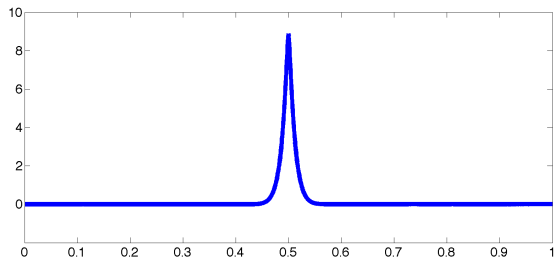
Banach Restoration ($p = 1.2$)



True solution x



Hilbert Restoration ZOOM ($p = 2$)

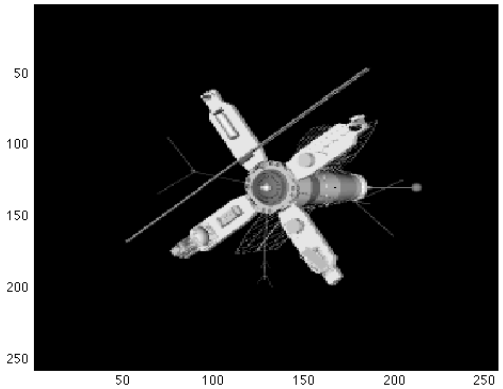


Banach Restoration ZOOM ($p = 1.2$)

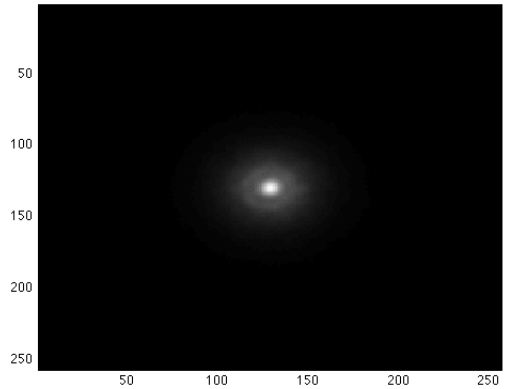
500 iterations, noiseless data. These numerical results in collaboration with Emanuele Frandi, University of Insubria, Italy.

A basic application: the (linear) Image Deblurring

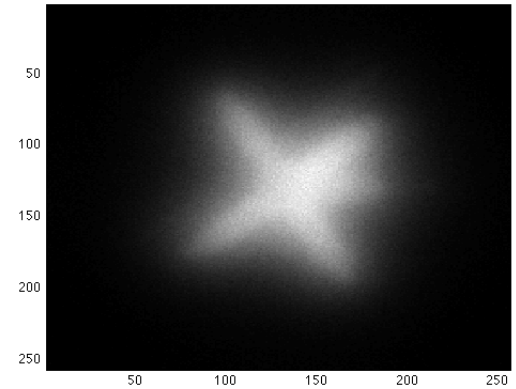
Landweber method (200 iterations)



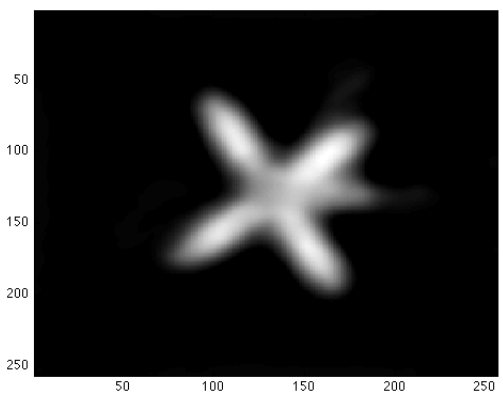
True image x



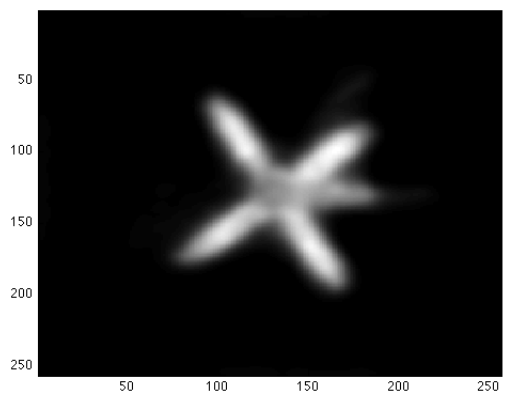
PSF (A)



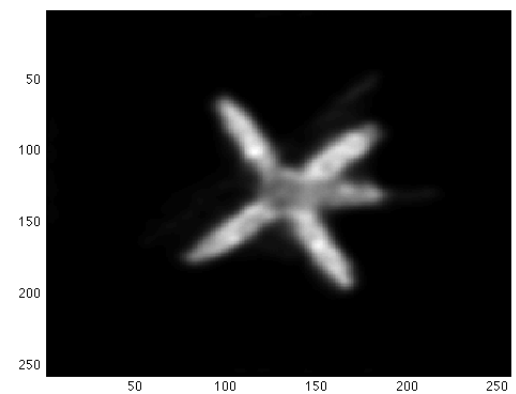
Blurred and noisy image y



Hilbert Restoration ($p = 2$)



Banach Restoration ($p = 1.5$)

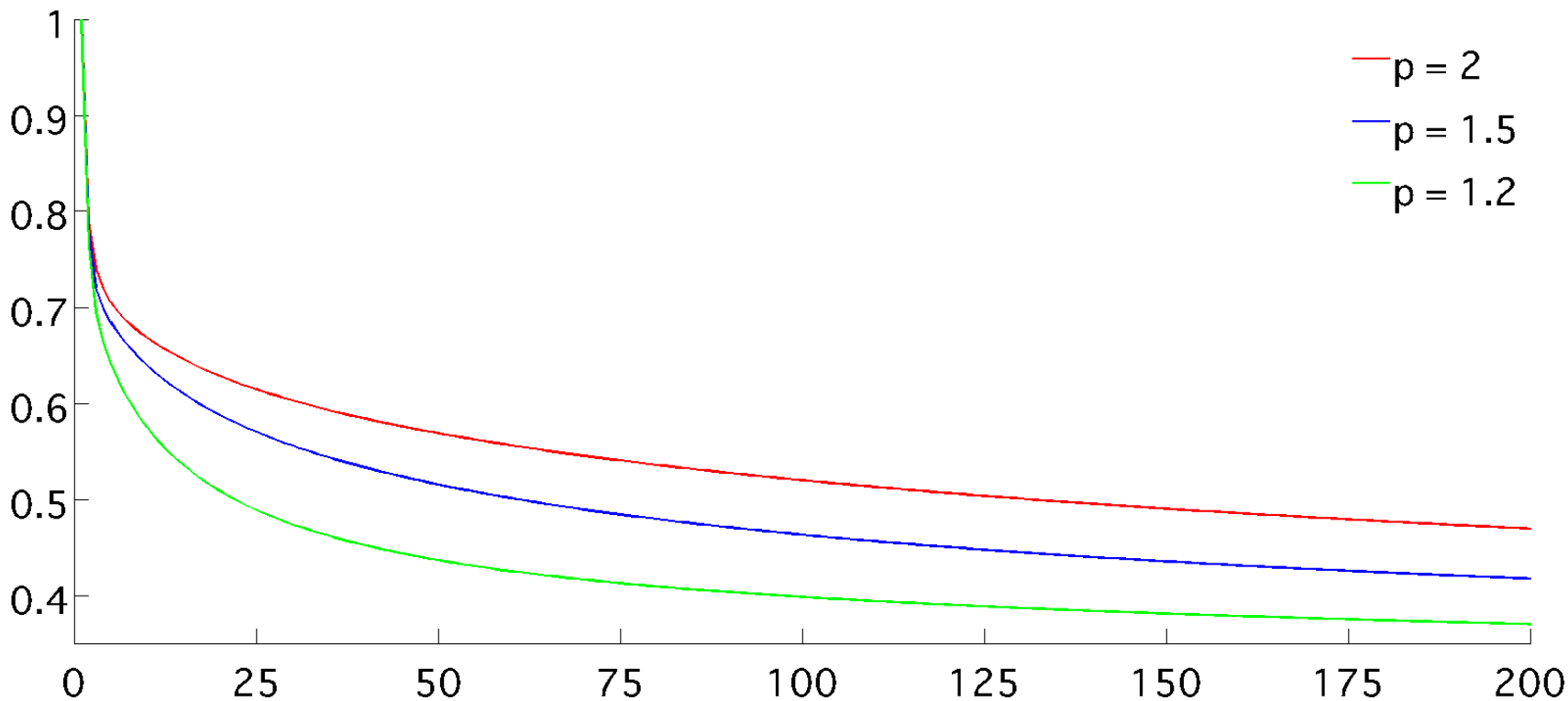


Banach Restoration ($p = 1.2$)

A classical image restoration example

Convergence history:

Relative restoration errors $\frac{\|x-x_k\|_2^2}{\|x\|_2^2}$ versus iteration index k



II

Extension of Conjugate Gradient Method:

From Hilbert to Banach space setting

The Conjugate Gradient in L^p Banach spaces

The Landweber algorithm in Banach spaces setting gives better restorations but it is still slow. Instead of $-\nabla(H(x_k))$, we consider a different anti-gradient descent direction, based on the same idea of the classical CG.

The CG(NR) method for min. of $H(x) = \frac{1}{2}\|Ax - y\|_2^2$, in Hilbert spaces:

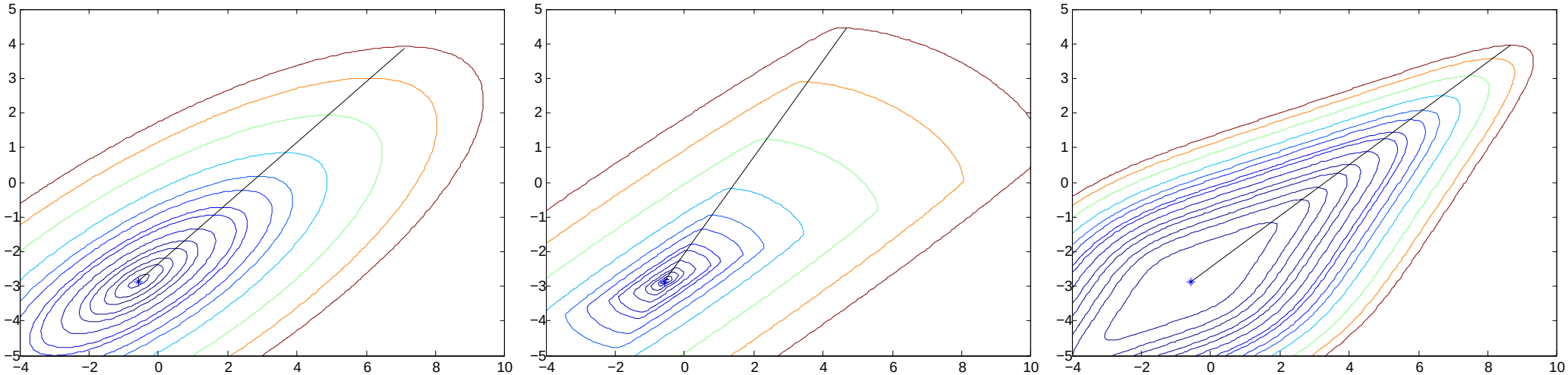
$$\begin{cases} x_{k+1} = x_k + \alpha_k p_k \\ p_{k+1} = -\nabla H(x_{k+1}) + \beta_k p_k = -A^*(Ax_{k+1} - y) + \beta_k p_k \end{cases}$$

where $p_0 = -\nabla H(x_0) = -A^*(Ax_0 - y)$, and

$$\alpha_k = \frac{(Ap_k)^T (y - Ax_k)}{(Ap_k)^T (Ap_k)}, \qquad \beta_k = \frac{\nabla H(x_{k+1})^T \nabla H(x_{k+1})}{\nabla H(x_k)^T \nabla H(x_k)}.$$

The step size α_k is called optimal, since $\alpha := \alpha_k$ solves the linear equation

$$\frac{d}{d\alpha} H(x_k + \alpha p_k) = 0\,.$$



$L^2(\mathbb{R}^2)$ $L^{1.2}(\mathbb{R}^2)$ $L^5(\mathbb{R}^2)$

The Conjugate Gradient method in L^p Banach spaces

The CG(NR) method for min. of $H(x) = \frac{1}{p}\|Ax - y\|_Y^p$, in $Y = L^p$:

$$\begin{cases} x_{k+1}^* = & x_k^* + \alpha_k p_k^* \\ x_{k+1} = & J_{p^*}^{X^*}(x_{k+1}^*) \\ p_{k+1}^* = & -\nabla H(x_{k+1}) + \beta_k p_k^* = -A^* J_p^Y (Ax_{k+1} - y) + \beta_k p_k^* \end{cases}$$

where $p_0^* = -\nabla H(x_0) = -A^* J_p^Y (Ax_0 - y)$,

the step size α_k solves (with approximation) the nonlinear equation

$$\frac{d}{d\alpha} H\Big(J_{p^*}^{X^*}(x_k^* + \alpha p_k^*) \Big) = 0\,,$$

and the coefficient β_k satisfies a "Fletcher-Reeves"-like formula

$$\beta_k = \gamma \frac{||Ax_{k+1} - y||_p^p}{||Ax_k - y||_p^p} \qquad \text{with } \gamma < 1/2$$

Convergence of CG in L^p Banach spaces

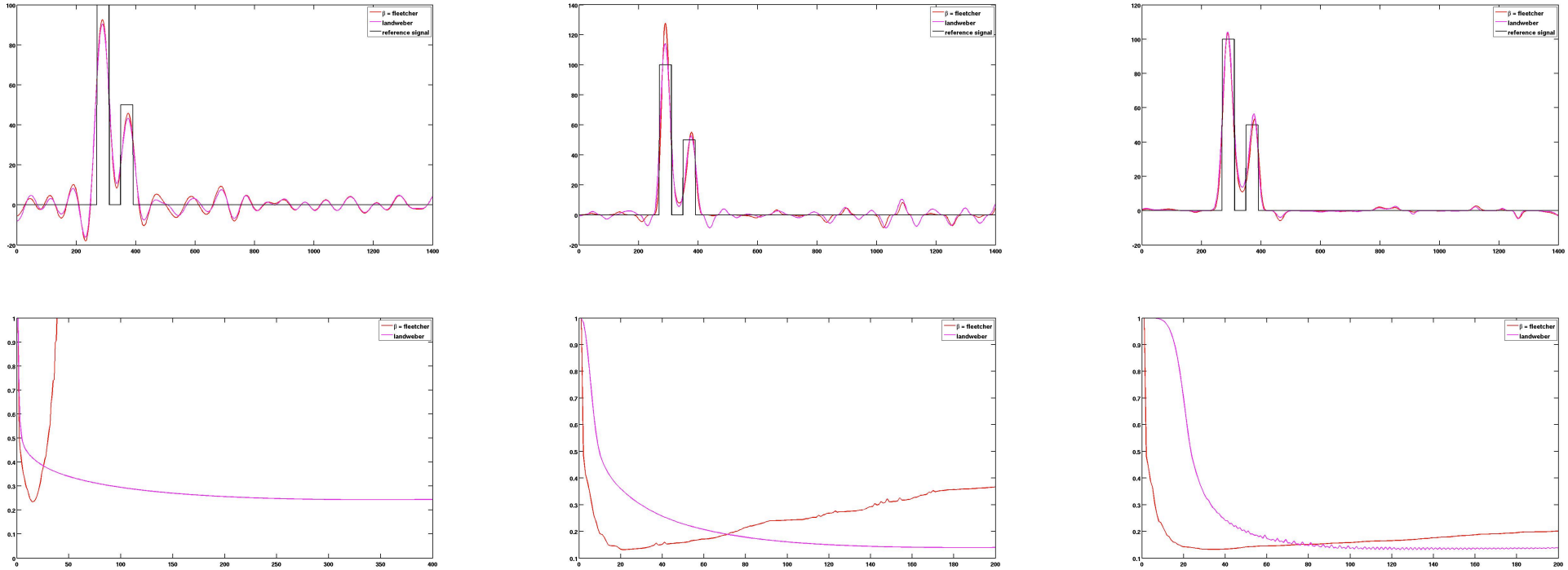
Theorem (E., Gratton, Lenti, Titley-Peloquin; *Num. Math.*, 2017)

(...) the sequence $\{x_k\}_k$ of the CG method in L^p converges strongly to the minimum norm pseudo-solution $x^\dagger$ of $Ax = y$, i.e.

$$\|x_k - x^\dagger\|_p \longrightarrow 0 \qquad (k \longrightarrow +\infty)\,.$$

If the data y is noisy, an early stop of the iterations (by discrepancy principle) gives rise to an iterative regularization method.

Remark: Differing from the conventional CG in Hilbert spaces, NO finite (n steps) convergence in $\mathbb{R}^n$.
However, Banach space finite steps convergence still holds in a simplified setting [Herzog R., Wollner W.; J. inverse ill-posed prob., 2016].



L^2 $L^{1.5}$ $L^{1.2}$

(Relative Errors vs iterations - Red: CG; Magenta: Landweber)

Some comments about the results:

For $p = 2$ the CG method is too fast (i.e., not enough regularization), and does not provide the same good quality of the slowest Landweber.

For the smaller $p = 1.5, 1.2$ the CG decelerates, and it gives the same quality of restoration give by the Landweber method, in much less iterations.

III

Iterative projection algorithms:

formal similarity between Hilbert and Banach space settings

Iterated projections (Row Action Methods)

The points $\tilde{x}$ satisfying the i-th row equation $\langle a_i, \tilde{x} \rangle = b_i$ of the $m \times n$ linear system $Ax = b$ define an hyperplane

$$\mathcal{Q}_i = \{\tilde{x} \in \mathbb{R}^n : \langle a_i, \tilde{x} \rangle = b_i\}$$

The solution x of the linear systems belongs to all the hyperplanes $\mathcal{Q}_i$, for $i = 1, \ldots, n$.

In Hilbert spaces, BECAUSE

- the orthogonal projection $P_i(z)$ of a point z onto one hyperplane $\mathcal{Q}_i$ is easy to perform

$$P_i(z) = z + \frac{b_i - \langle a_i, z \rangle}{\|a_i\|_2^2} a_i^T,$$

- and the projection $P_i(z)$ is closer to the solution x than z itself,

THEN

iterative projecting $x_{k+1} = P_i(x_k)$ onto different hyperplanes $\mathcal{Q}_i$ gives rise to a "low-cost" sequence (x_k) converging to the solution x.

Some families of projection methods (in Hilbert spaces)

- **ART** (Algebraic Reconstruction Techniques) or Kaczmarz's methods

$$x_{k+1} = x_k + \lambda_k \frac{b_i - \langle a_i, x_k \rangle}{\|a_i\|^2} a_i^T$$

where the row index i depends on the iteration index k (in many ways, for instance $i = k \mod m$, where m is the number of rows of A, or even randomly (!)) and $\lambda_k \in \mathbb{R}$ is a relaxation parameter.

- **SIRT** (Simultaneous Iterative Reconstruction Techniques)

The same "fixed" iteration x_k is used (i.e., simultaneously) for a "complete" set of m different projections, that is, for a full set of indexes $i = 1, 2, \ldots, m$.

In this case, a matrix form iteration holds

$$x_{k+1} = x_k + \lambda_k S A^* M (b - A x_k)$$

SIRT: Cimmino method, DROP (Diagonally Relaxed Orthogonal Projection), CAV (Component Averaging), BIP (Block-Iterative Projections), ...

Cimmino and DROP methods

In the SIRT family $x_{k+1} = x_k + \lambda_k S A^* M (b - A x_k)$, we have:

- the **Cimmino method** (1938), where the new iteration is the average of a complete set of m projections

$$x_{k+1} = \frac{1}{m} \sum_{i=1}^{m} P_i(x_k) = x_k + \frac{1}{m} \sum_{i=1}^{m} \frac{b_i - \langle a_i, x_k \rangle}{\|a_i\|^2} a_i^T$$

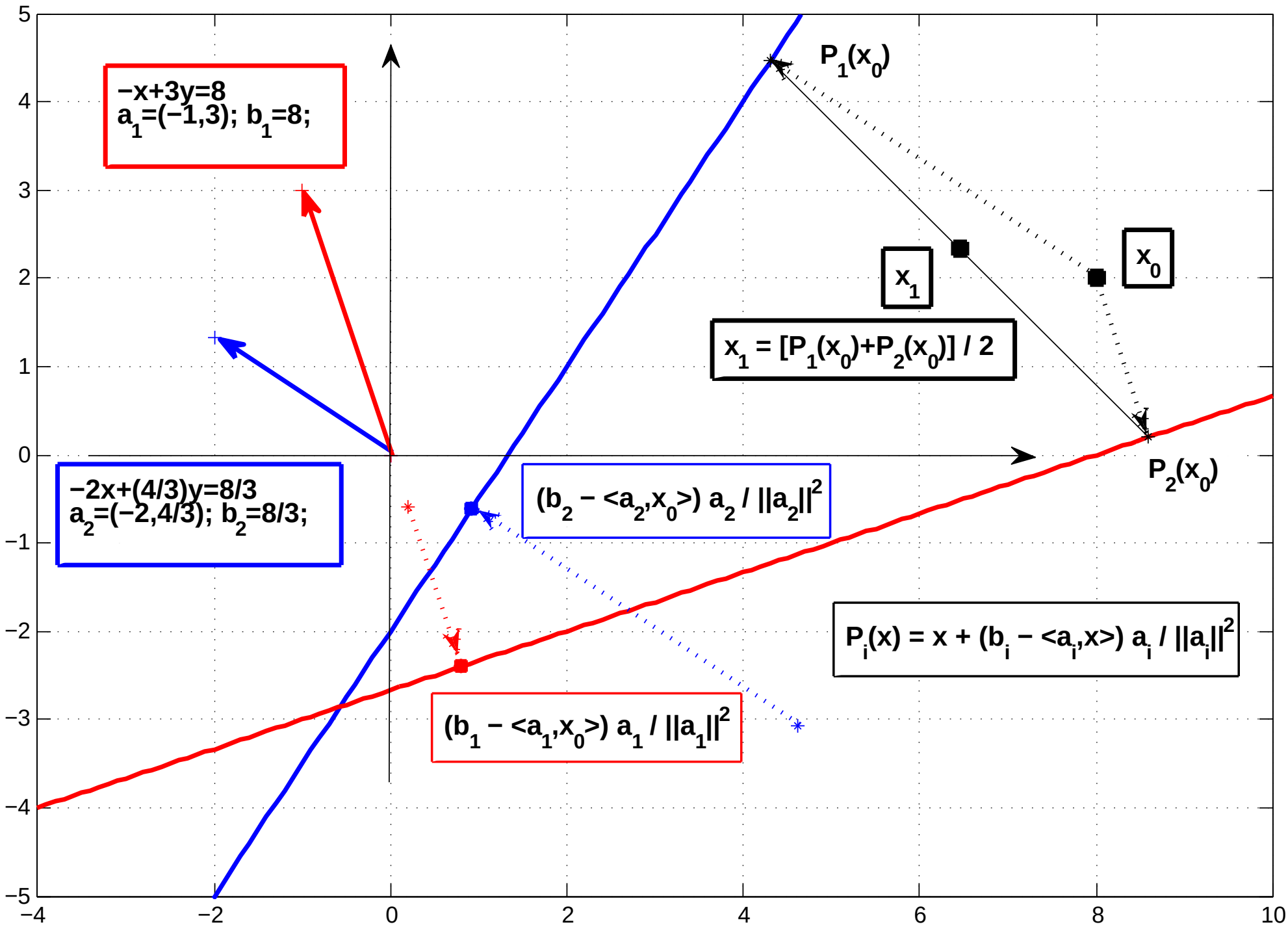
so that

$$\begin{aligned} S &= I, \\ M &= \frac{1}{m} D = \frac{1}{m} \mathrm{diag}(1/\|a_1\|^2, 1/\|a_2\|^2, \ldots, 1/\|a_m\|^2). \end{aligned}$$

- the **DROP method** (Diagonally Relaxed Orthogonal Projection), where

$$\begin{aligned} S &= \mathrm{diag}(1/r_1, 1/r_2, \ldots, 1/r_n)\,, \\ M &= D = \mathrm{diag}(1/\|a_1\|^2, 1/\|a_2\|^2, \ldots, 1/\|a_m\|^2)\,, \end{aligned}$$

being r_j the number of non-zeroes elements of the j-th column of A.


−x+3y=8
a1=(−1,3); b1=8;
−2x+(4/3)y=8/3
a2=(−2,4/3); b2=8/3;
P1(x0)
P2(x0)
x0
x1
x1 = [P1(x0)+P2(x0)] / 2
(b2 − <a2,x0>) a2 / ||a2||2
(b1 − <a1,x0>) a1 / ||a1||2
Pi(x) = x + (bi − <ai,x>) ai / ||ai||2
5
4
3
2
1
0
−1
−2
−3
−4
−5
−4
−2
0
2
4
6
8
10

DROP method (in Hilbert spaces)

$A : X \longrightarrow Y \qquad A^* : Y \longrightarrow X \qquad H_M(x) = \frac{1}{2}\|Ax - y\|_M^2$

$$x_{k+1} = S(S^{-1}x_k - \lambda_k A^* M(Ax_k - y))$$

Landweber iterative method in Banach spaces

$A : X \longrightarrow Y \qquad A^* : Y^* \longrightarrow X^* \qquad H_r(x) = \frac{1}{r}\|Ax - y\|_Y^r$

$$x_{k+1} = J_{r^*}^{X^*}(J_r^X x_k - \lambda_k A^* J_r^Y (Ax_k - y))$$

Some remarks about

J_r^Y, $J_{r^*}^{X^*}$ (in L^p Banach spaces) VS M, S (in L^2 Hilbert spaces):

- J_r^Y is non-linear and M is linear; $J_{r^*}^{X^*}$ is non-linear and S is linear;
- all are diagonal and positive operators;
- all cost $O(n)$ operations;
- the action of the matrix M in DROP is "similar" to the one of the duality map $J_r^Y = J_2^{L^p}$ with $1 < p < 2$ in Landweber-Banach.

DROP method (in Hilbert spaces)

$$A : X \longrightarrow Y \qquad A^* : Y \longrightarrow X \qquad H_M(x) = \tfrac{1}{2}\|Ax - y\|_M^2$$

$$x_{k+1} = S(S^{-1}x_k - \lambda_k A^* M(Ax_k - y))$$

Landweber iterative method in Banach spaces

$$A : X \longrightarrow Y \qquad A^* : Y^* \longrightarrow X^* \qquad H_r(x) = \tfrac{1}{r}\|Ax - y\|_Y^r$$

$$x_{k+1} = J_{r^*}^{X^*}(J_r^X x_k - \lambda_k A^* J_r^Y (Ax_k - y))$$

In summary, the Landweber iterative method in Banach spaces can also be viewed as a non-linear generalization of well-known projection algorithms for linear systems.

The oblique geometry of the M-induced norm $\|\cdot\|_M$ in Hilbert space is replaced by the L^p-norm $\|\cdot\|_{L^p}$ in L^p Banach space setting.

IV

Preconditioning:

from Hilbert space (primal) preconditioning to Banach space primal or dual preconditioning

Preconditioning in Banach Space

Preconditioned system in Hilbert space, with (invertible) preconditioner D

$$A^*Ax = A^*y \quad \Longleftrightarrow \quad DA^*Ax = DA^*y$$

Preconditioned Landweber method in Hilbert spaces

$$x_{k+1} = x_k - \lambda DA^*(Ax_k - y)$$

where the preconditioner D is a regularized (polynomial, rational, circulant) "low-cost" approximation (a structured and/or sparse matrix) of the inverse of A^*A, i.e.,

$$D \approx (A^*A)^{-1}.$$

Now in Banach space A^*A is not well defined (. . . it cannot even be written!)

We generalize preconditioning schemes to Banach space in two ways:

- **Primal preconditioning**
- **Dual preconditioning**

Primal Preconditioning in Banach Space

Remember, in Hilbert space:

$$x_{k+1} = x_k - \lambda D A^*(Ax_k - y) \qquad \text{with } \; D \approx (A^*A)^{-1}$$

Primal preconditioner : $\quad D : X \longrightarrow X$

The Primal preconditioner is a regularized approximation of $(J_{r^*}^{X^*} A^* J_r^Y A)^{-1}$

From Hilbert space:

$$x_{k+1} = D(D^{-1}x_k - \lambda A^*(Ax_k - y))$$

To Banach space:

$$x_{k+1} = D J_{r^*}^{X^*}(J_r^X D^{-1} x_k - \lambda A^* J_r^Y (Ax_k - y))$$

Recalling that $A : X \longrightarrow Y$, $A^* : Y^* \longrightarrow X^*$, $J_r^X : X \longrightarrow X^*$, $J_{r^*}^{X^*} : X^* \longrightarrow X$, $J_r^Y : Y \longrightarrow Y^*$, the iterations are well-defined.

Dual Preconditioning in Banach Space

Remember, in Hilbert space:

$$x_{k+1} = x_k - \lambda D A^*(Ax_k - y) \qquad \text{with } \; D \approx (A^*A)^{-1}\,.$$

Dual preconditioner : $\quad D = X^* \longrightarrow X^*$

The Dual preconditioner is a regularized approximation of $(A^* J_r^Y A J_{r^*}^{X^*})^{-1}$.

From Hilbert space:

$$x_{k+1} = x_k - \lambda D A^*(Ax_k - y)$$

To Banach space:

$$x_{k+1} = J_{r^*}^{X^*}(J_r^X x_k - \lambda D A^* J_r^Y (Ax_k - y))$$

Recalling that $A : X \longrightarrow Y$, $A^* : Y^* \longrightarrow X^*$, $J_r^X : X \longrightarrow X^*$, $J_{r^*}^{X^*} : X^* \longrightarrow X$, $J_r^Y : Y \longrightarrow Y^*$, the iterations are well-defined.

Convergence analysis of preconditioning in L^p Banach Space

These preconditioned algorithms can be read as fixed-point iterations in the dual space X^*, that is,

$$x^*_{k+1} = g(x^*_k)\,,$$

where $g : X^* \longrightarrow X^*$ in the dual preconditioning is

$$g(t) = t - \lambda D A^* J_r^Y (A J_{r^*}^{X^*}(t) - y)\,.$$

By simple direct computation, its Jacobian $\mathcal{J}_g$ in L^p Banach spaces is

$$\mathcal{J}_g(x^*) = I - \lambda D A^* A + E\,,$$

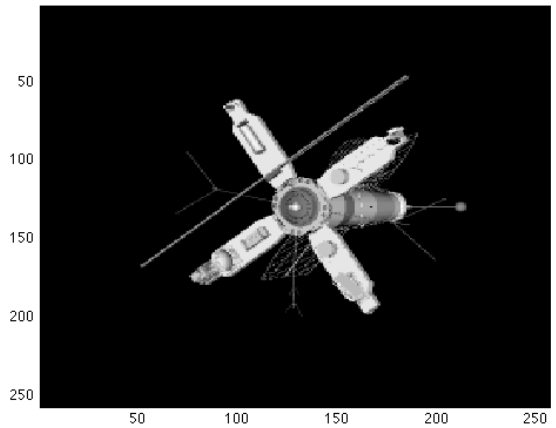
where E is a matrix of small norm for p close to (the Hilbert value) $p = 2$.

This way, by a Bauer-Fike argument, the eigenvalues of $\mathcal{J}_g(x^*)$ are "perturbations" of those of the iteration matrix $I - \lambda D A^* A$ of the conventional Hilbert case. It follows that, for $p \approx 2$, the spectral radius of $\mathcal{J}_g(x^*)$ is essentially the same as that of the Hilbert case.

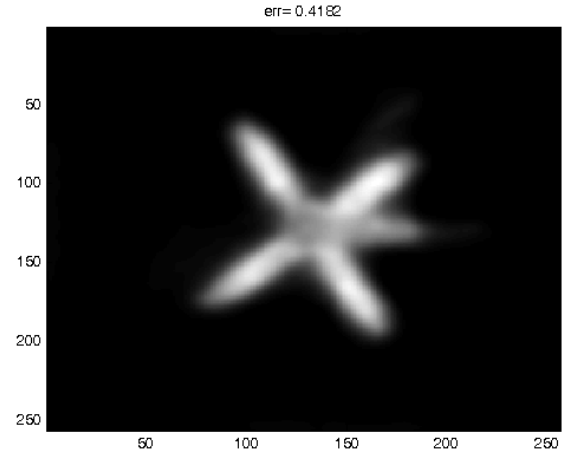
In particular, choosing D as an optimal approximation of $(A^*A)^\dagger$ will bring to an acceleration of the convergence, at least for $p \approx 2$.

(Regularized) Preconditioning for image restoration

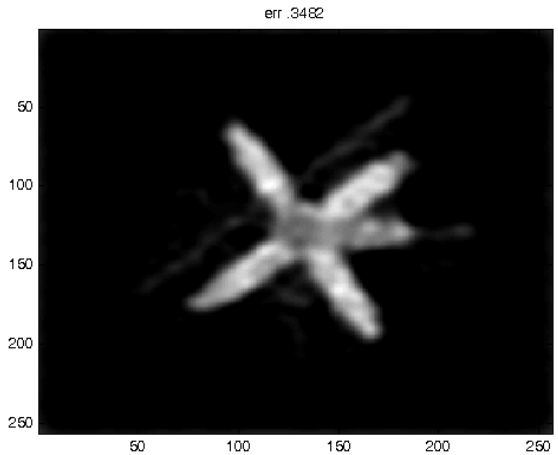
Preconditioner: Thikonov-filtered T. Chan optimal circulant preconditioner.



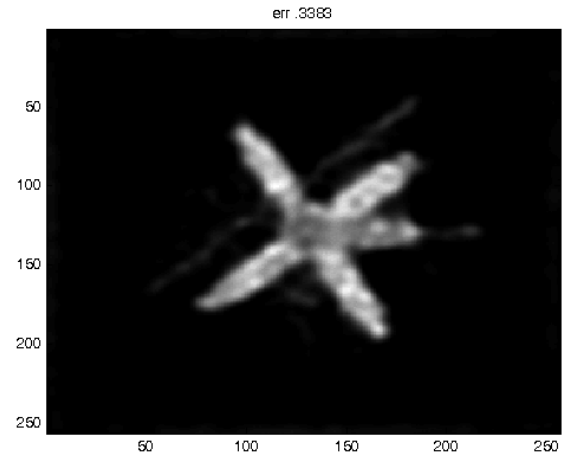
True image x



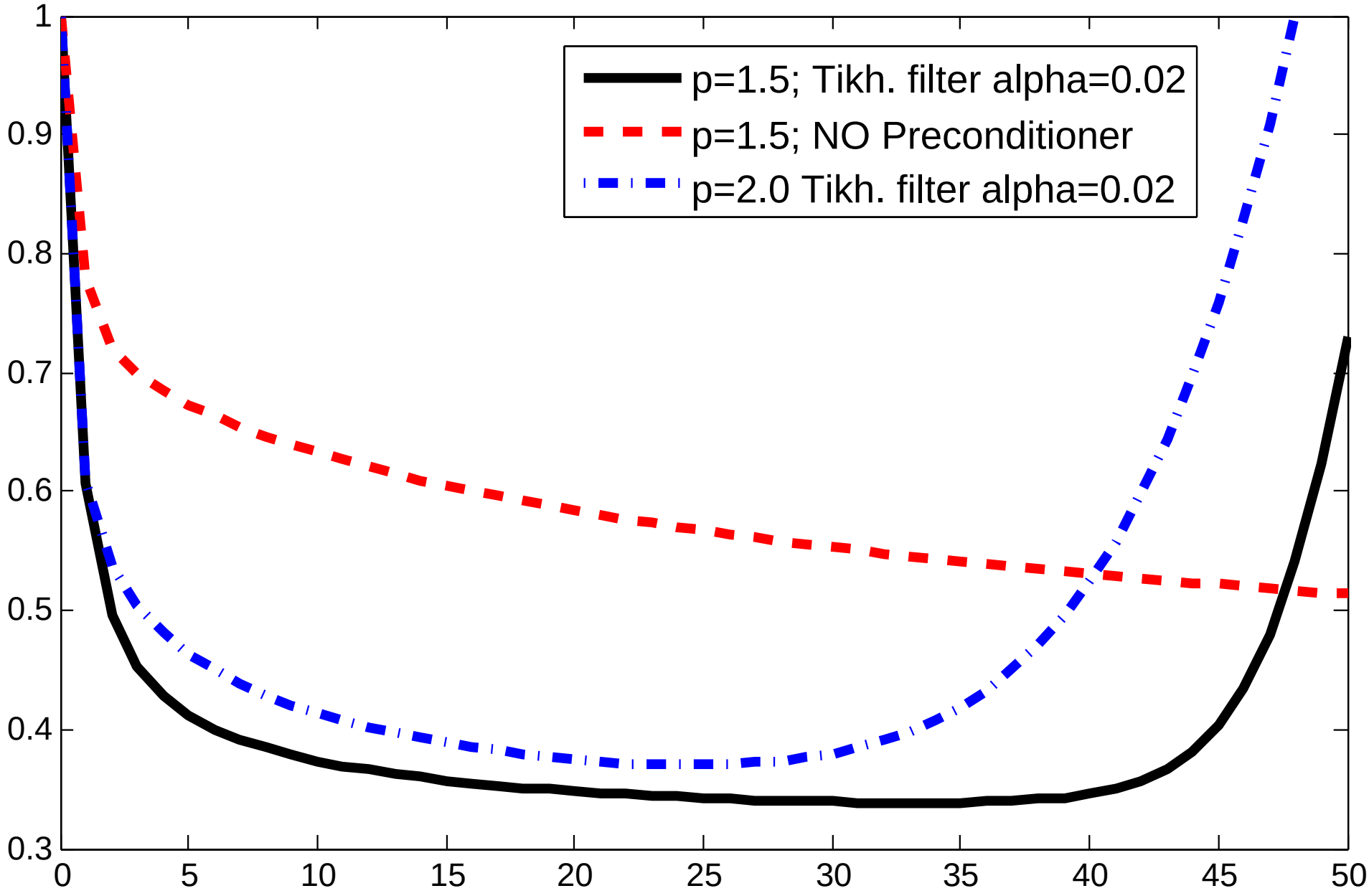
NO Prec. 200 Iter. ($p = 1.5$)
200 Iter. Rel. Err 0.4182



Tikhonov Prec. ($p = 1.8$)
93 Iter. Rel. Err 0.3482



Tikhonov Prec. 33 Iter.($p = 1.5$)
33 Iter. Rel. Err 0.3383


p=1.5; Tikh. filter alpha=0.02
p=1.5; NO Preconditioner
p=2.0 Tikh. filter alpha=0.02
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0
5
10
15
20
25
30
35
40
45
50

V

Extension to a non conventional Banach space:

The variable exponent Lebesgue spaces $L^{p(\cdot)}$

A "new" framework: variable exponent Lebesgue spaces $L^{p(\cdot)}$

In image restoration, often different regions of the image require different "amount of regularization".

Setting different levels of regularization is useful because background, low intensity, and high intensity values require different filtering levels (see Nagy, Pauca, Plemmons, Torgersen, J Opt Soc Am A, 1997).

The idea: the ill-posed functional equation $Ax = y$ is solved in $L^{p(\cdot)}$ Banach spaces, namely, the variable exponent Lebesgue spaces, a special case of the so-called Musielak–Orlicz functional spaces (first proposed in two seminal papers in 1931 and 1959, but intensively studied just in the last 10 years).

In a variable exponent Lebesgue space, to measure a function f, instead of a constant exponent p all over the domain, we have a pointwise variable (i.e., a function) exponent $1 \leq p(\cdot) \leq +\infty$.

This way, different regularization levels on different regions of the image to restore can be automatically and adaptively assigned.

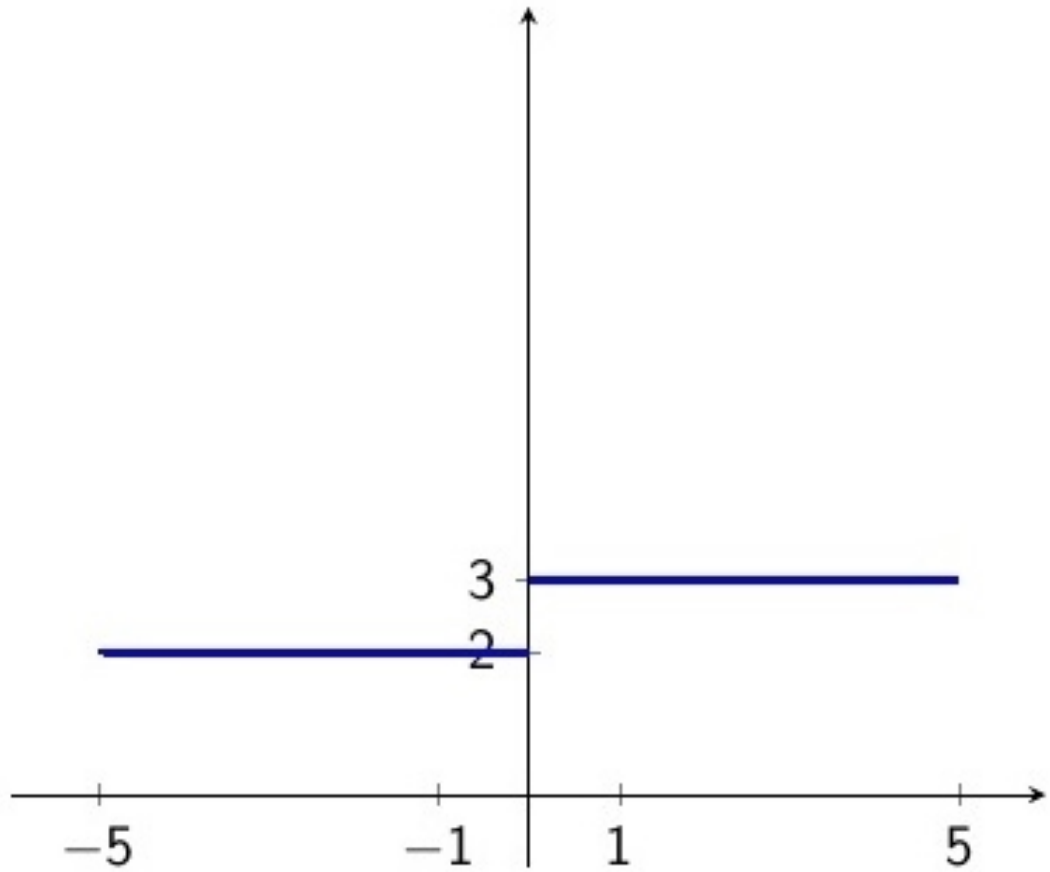
A sketch on variable exponent Lebesgue spaces $L^{p(\cdot)}$

$L^p(\Omega)$	$L^{p(x)}(\Omega)$
$1 \le p \le \infty$ p is constant	$p(x) : \Omega \to [1, \infty]$ $p(x)$ is a measurable function
$\|f\|_p = \left(\int_\Omega \lvert f(x)\rvert^p \, dx \right)^{1/p}$ $\|f\|_\infty = \operatorname{ess\,sup} \lvert f(x)\rvert$	$\|f\|_{p(\cdot)} = \left(\int_\Omega \lvert f(x)\rvert^{p(x)} \, dx \right)^{1/???}$ $\dots$
$f \in L^p(\Omega) \iff \int_\Omega \lvert f(x)\rvert^p \, dx < \infty$	$f \in L^{p(\cdot)}(\Omega) \iff ???$

In the following, $\Omega_\infty = \{x \in \Omega : p(x) = \infty\}$ has zero measure.

$$\Omega = [-5, 5]$$

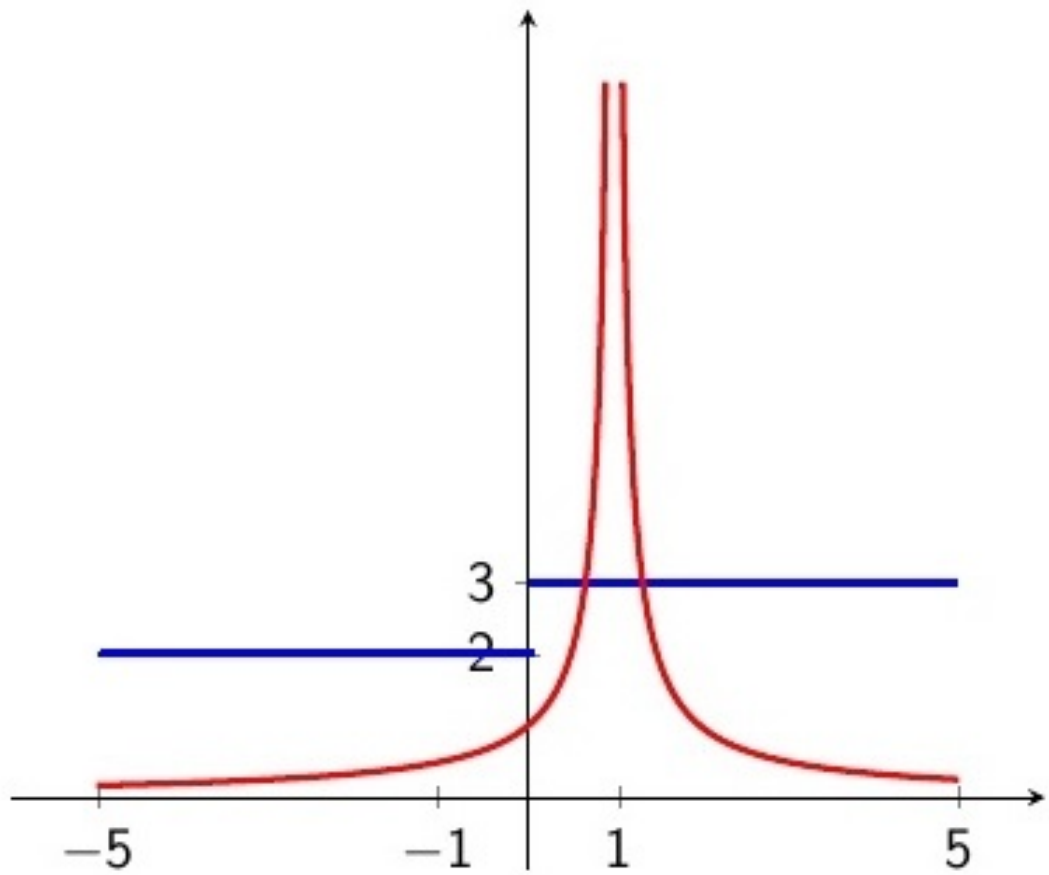
$$p(x) = \begin{cases} 2 & \text{se} \quad -5 \leq x \leq 0 \\ 3 & \text{se} \quad 0 < x \leq 5 \end{cases}$$



$$\Omega = [-5, 5]$$

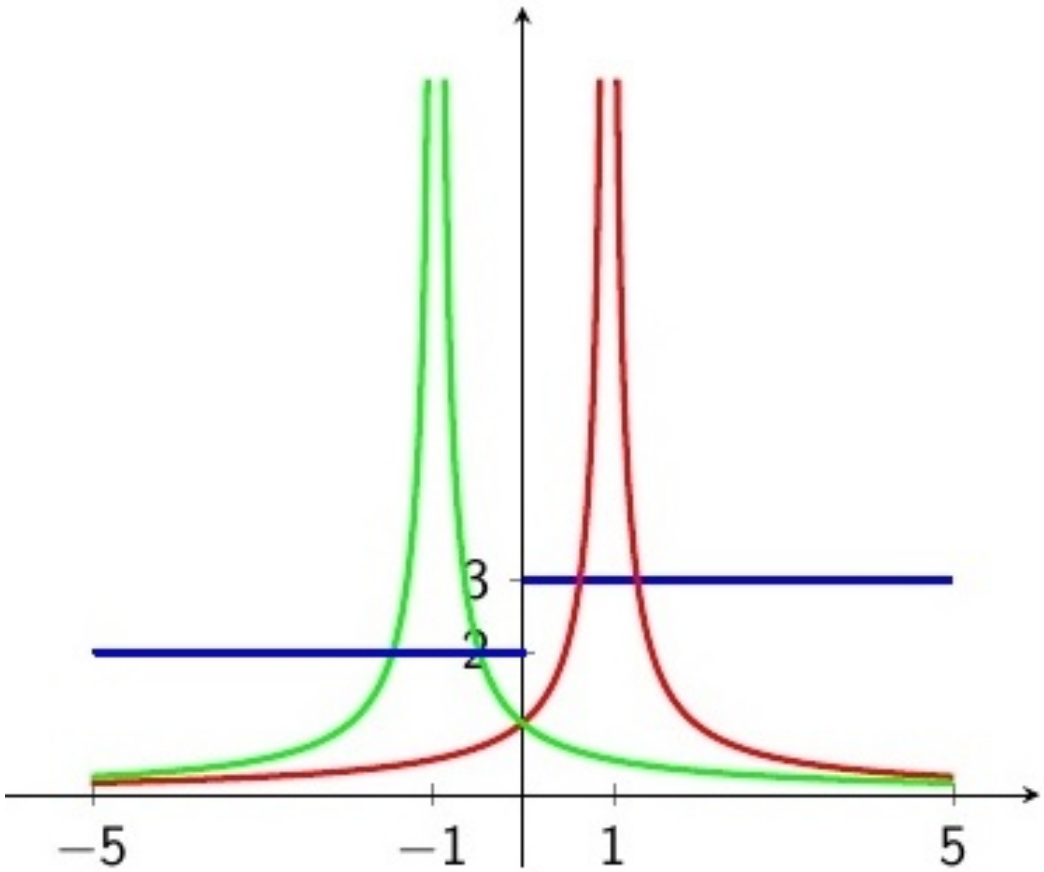
$$p(x) = \begin{cases} 2 & \text{se} \quad -5 \leq x \leq 0 \\ 3 & \text{se} \quad 0 < x \leq 5 \end{cases}$$

$$f(x) = \frac{1}{|x-1|^{1/3}} \notin L^{p(\cdot)}([-5, 5])$$



$$\Omega = [-5, 5] \qquad p(x) = \begin{cases} 2 & \text{se} \quad -5 \leq x \leq 0 \\ 3 & \text{se} \quad 0 < x \leq 5 \end{cases}$$

$$f(x) = \frac{1}{|x+1|^{1/3}} \in L^{p(\cdot)}([-5, 5]) \qquad f(x) = \frac{1}{|x-1|^{1/3}} \notin L^{p(\cdot)}([-5, 5])$$



The norm of variable exponent Lebesgue spaces

In the conventional case L^p, the norm is $\|f\|_{L^p} = \Big(\int_\Omega |f(x)|^p dx \Big)^{1/p}$.

In $L^{p(\cdot)}$ Lebesgue spaces, the definition and computation of the norm is not straightforward, since we have not a constant value for computing the ("mandatory") radical.

$$\|f\|_{L^{p(\cdot)}} = \Big(\int_\Omega |f(x)|^{p(x)} dx \Big)^{1/???}.$$

The solution: compute first the modular (for $1 \le p(\cdot) < +\infty$)

$$\varrho(f) = \int_\Omega |f(x)|^{p(x)} dx \,,$$

and then obtain the (so called Luxemburg [1955]) norm by solving a 1D minimization problem

$$\|f\|_{L^{p(\cdot)}} = \inf \Big\{ \lambda > 0 : \varrho\Big(\frac{f}{\lambda}\Big) \le 1 \Big\}.$$

The elements of a variable exponent Lebesgue space

$$\varrho(f) = \int_\Omega |f(x)|^{p(x)} dx\,,$$

$$\|f\|_{L^{p(\cdot)}} = \inf\left\{\lambda > 0 : \varrho\Big(\frac{f}{\lambda}\Big) \le 1\right\}.$$

The Lebesgue space

$$L^{p(\cdot)}(\Omega) = \left\{f : \Omega \to \mathbb{R} \ \mid \ \|f\|_{L^{p(\cdot)}} < \infty\right\}$$

is a Banach space.

In the case of a constant function exponent $p(x) = p$, this norm is exactly the classical one $\|f\|_p$, indeed

$$\varrho\Big(\frac{f}{\lambda}\Big) = \int_\Omega \Big(\frac{|f(x)|}{\lambda}\Big)^p dx = \frac{1}{\lambda^p}\int_\Omega |f(x)|^p dx = \frac{1}{\lambda^p}\|f\|_p^p$$

and

$$\inf\left\{\lambda > 0 : \frac{1}{\lambda^p}\|f\|_p^p \le 1\right\} = \|f\|_p$$

Modulus VS Norm

In (classical) L^p, norm and modulus are "the same" apart from a p-root:

$$\|f\|_p < \infty \qquad \Longleftrightarrow \qquad \int_\Omega |f(x)|^p \, dx < \infty$$

In $L^{p(\cdot)}$, norm and modulus are really different:

$$\|f\|_{p(\cdot)} < \infty \qquad \Leftarrow \not\Rightarrow \qquad \varrho(f) < \infty$$

Indeed, the following holds

$$\|f\|_{p(\cdot)} < \infty \iff \text{there exist a } \lambda > 0 \text{ s.t. } \varrho\left(\frac{f}{\lambda}\right) < \infty$$

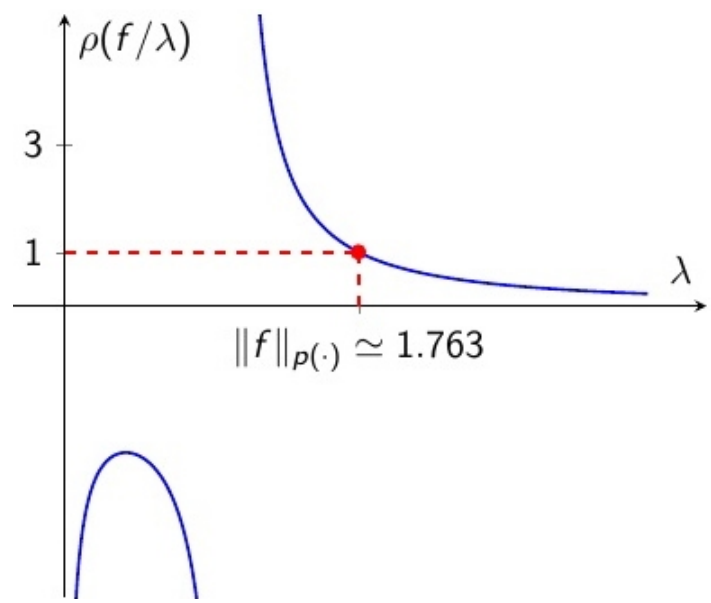
(and notice that λ can be chosen large enough ...).

A simple example of a strange behavior

$$\Omega = [1, \infty) \qquad f(x) \equiv 1 \qquad p(x) = x$$

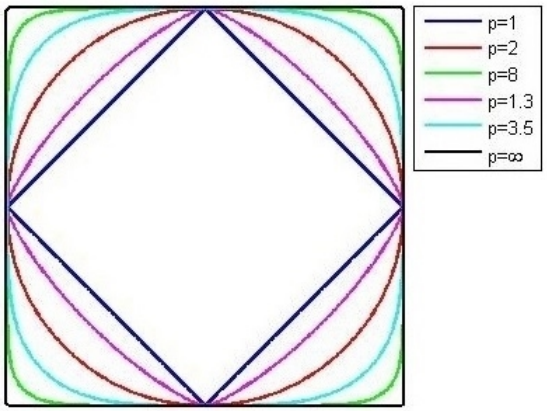
$$\varrho(f) = \int_1^\infty 1^x \, dx = \infty \qquad \text{BUT} \qquad \|f\|_{p(\cdot)} \simeq 1.763$$

Indeed $\quad \varrho(f/\lambda) = \displaystyle\int_1^\infty \left(\frac{1}{\lambda}\right)^x dx = \frac{1}{\lambda \log \lambda} < \infty, \quad$ for any $\lambda > 1$

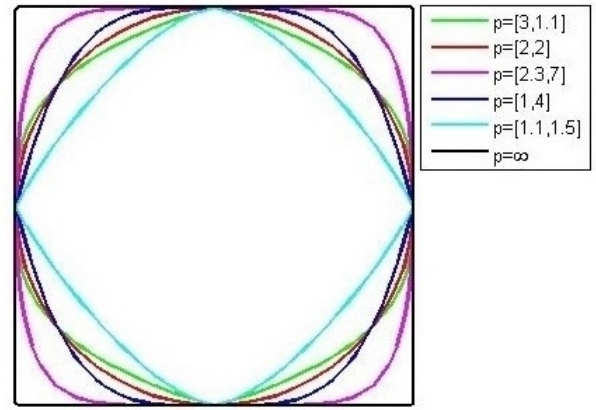


The vector case: the Lebesgue spaces of sequences $l^{p(\cdot)} = (l_n^{p_n})$

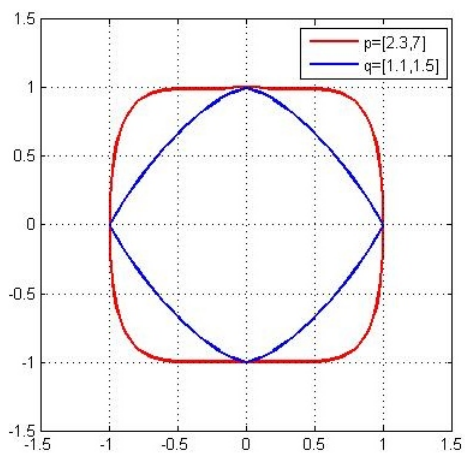
The unit circle of $x = (x_1; x_2)$ in $\mathbb{R}^2$ with variable exponents $p = (p_1; p_2)$.



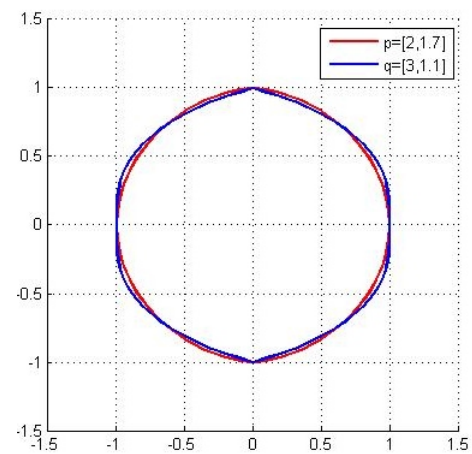
$\|x\|_p$



$\|x\|_{p(\cdot)}$



Inclusion if $(p_1, p_2) \geq (q_1, q_2)$ (as classical)



No inclusion in general

Properties of variable exponent Lebesgue spaces $L^{p(\cdot)}$

$$\text{Let} \quad p_- = \operatorname{ess\,inf}_{\Omega} |p(x)| \ , \quad \text{and} \quad p_+ = \operatorname{ess\,sup}_{\Omega} |p(x)| \ .$$

If $p_+ = \infty$, then $L^{p(\cdot)}(\Omega)$ is a "bad" (although very interesting) Banach space, with poor geometric properties (i.e., not useful for our regularization schemes).

If $1 < p_- \leq p_+ < \infty$, then $L^{p(\cdot)}(\Omega)$ is a "good" Banach space, since many properties of classical Lebesgue spaces L^p still hold.

This is the natural framework for our iterative methods in Banach spaces, because:

- $L^{p(\cdot)}$ is uniformly smooth, uniformly convex, and reflexive,
- its dual space is well defined, $\left(L^{p(\cdot)}\right)^* \simeq L^{q(\cdot)}$, where $\frac{1}{p(x)} + \frac{1}{q(x)} = 1$,
- we can define its duality map.

The duality map of the variable exponent Lebesgue space

By extending the duality maps, we can define into $L^{p(\cdot)}$ all the iterative methods developed in L^p (Landweber, Steepest descent, CG, Mann iter.).

For any constant $1 < r < +\infty$, we recall that the duality map, that is, the (sub-)differential of the functional $\frac{1}{r}\|f\|_{L^p}^r$, in the classical Banach space L^p, with constant $1 < p < +\infty$, is defined as follows

$$\Big(J_{L^p}(f)\Big)(x) = \frac{|f(x)|^{p-1}\,\mathrm{sgn}(f(x))}{\|f\|_p^{p-r}}\,.$$

By generalizing a result of P. Matei [2012], we have that the corresponding duality map in variable exponent Lebesgue space is defined as follows

$$\Big(J_{L^{p(\cdot)}}(f)\Big)(x) = \frac{1}{\int_\Omega \frac{p(x)\,|f(x)|^{p(x)}}{\|f\|_{p(\cdot)}^{p(x)}}\,dx}\,\frac{p(x)\,|f(x)|^{p(x)-1}\,\mathrm{sgn}(f(x))}{\|f\|_{p(\cdot)}^{p(x)-r}}\,,$$

where any product and any ratio have to be considered as pointwise.

The adaptive algorithm in variable exponent Lebesgue spaces

It is a numerical evidence that, in L^p image deblurring,

- dealing with small $1 \approx p << 2$ improves sparsity and allows a better restoration of the edges of the images and of the zero-background,
- dealing with $p \approx 2$ (even $p > 2$), allows a better restoration of the regions of pixels with the highest intensities.

The idea: to use a scaled into $[1, 2]$ version of the (re-)blurred data y as function of the exponent $p(\cdot)$ for the variable exponent Lebesgue spaces $L^{p(\cdot)}$ where computing the solution. Example (linear interpolation):

$$p(\cdot) = 1 + [A^*y(\cdot) - \min(A^*y)]/[\max(A^*y) - \min(A^*y)]$$

The Landweber (i.e., fixed point) iterative scheme in this $L^{p(\cdot)}$ Banach space can be modified as adaptive iteration algorithm, by recomputing, after each fixed number of iterations, the exponent function $p_k(\cdot)$ by means of the k-th restored image x_k (instead of the first re-blurred data Ay), that is

$$p_k(\cdot) = 1 + [x_k(\cdot) - \min(x_k)]/[\max(x_k) - \min(x_k)]$$

The conjugate gradient method in $L^{p(\cdot)}$ for image restoration

Let $p(\cdot) = 1 + [A^*y(\cdot) - \min(A^*y)]/[\max(A^*y) - \min(A^*y)]$

$$\varrho_0^* = -A^* J_r^{L^{p(\cdot)}}(Ax_0 - y)$$

For $k = 1, 2, 3, \ldots$

$$\alpha_k = \arg\min_\alpha \tfrac{1}{r}\|A(x_k + \alpha\varrho_k) - y\|^r_{L^{p(\cdot)}}$$

$$x^*_{k+1} = x^*_k + \alpha_k \varrho^*_k \qquad x_{k+1} = J_{s'}^{(L^{p(\cdot)})^*}(x^*_{k+1})$$

$$\beta_{k+1} = \gamma \frac{\|Ax_{k+1} - y\|^r_{L^{p(\cdot)}}}{\|Ax_k - y\|^r_{L^{p(\cdot)}}} \qquad \text{with } \gamma < 1/2$$

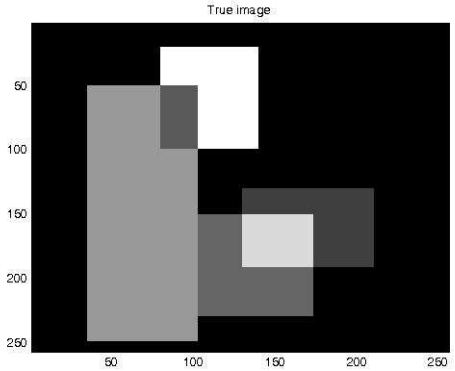
$$\varrho^*_{k+1} = -A^* J_r^{L^{p(\cdot)}}(Ax_{k+1} - y) + \beta_{k+1}\varrho^*_k$$

(and recompute $p(\cdot) = 1 + [x_k(\cdot) - \min(x_k)]/[\max(x_k) - \min(x_k)]$ each m iterations by using the last iteration x_k).

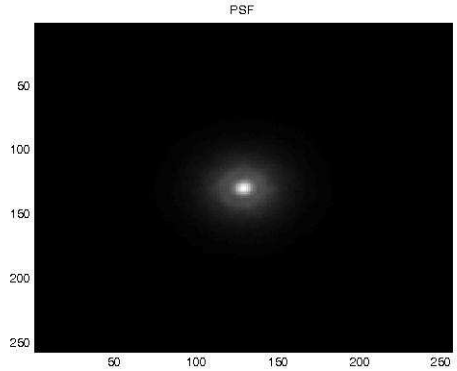
Numerical results

for the variable exponent Lebesgue spaces $L^{p(\cdot)}$

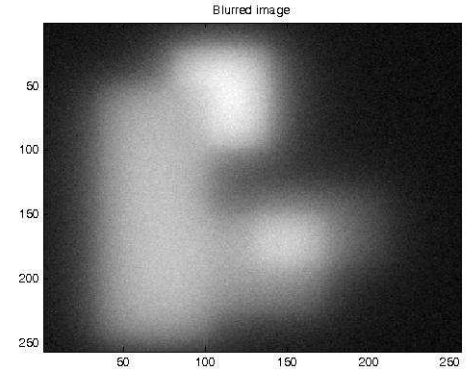
Special thanks to Dr. Brigida Bonino.



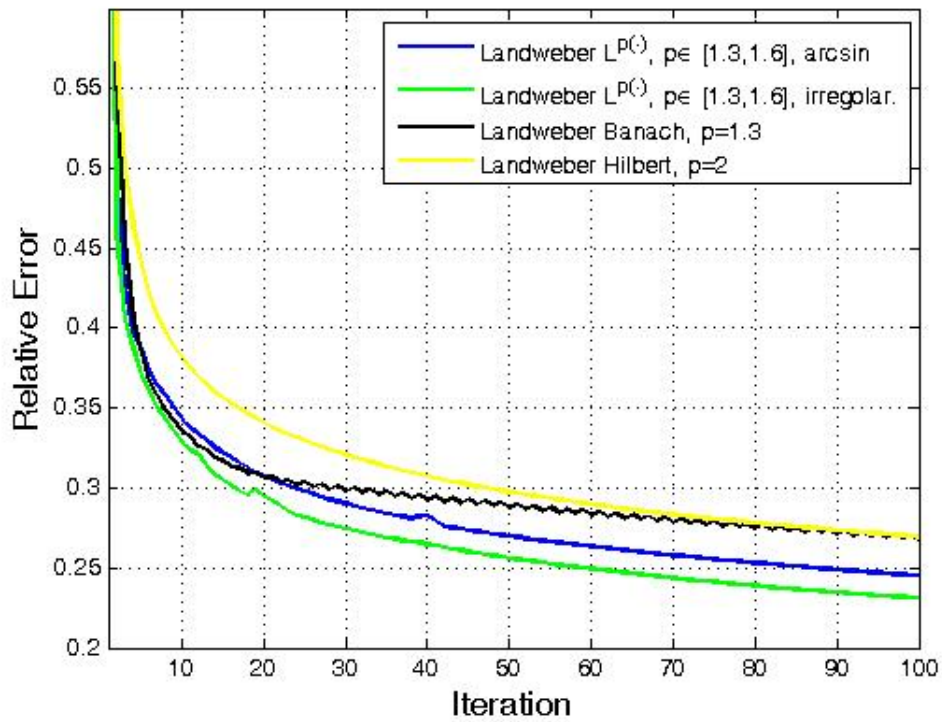
True image

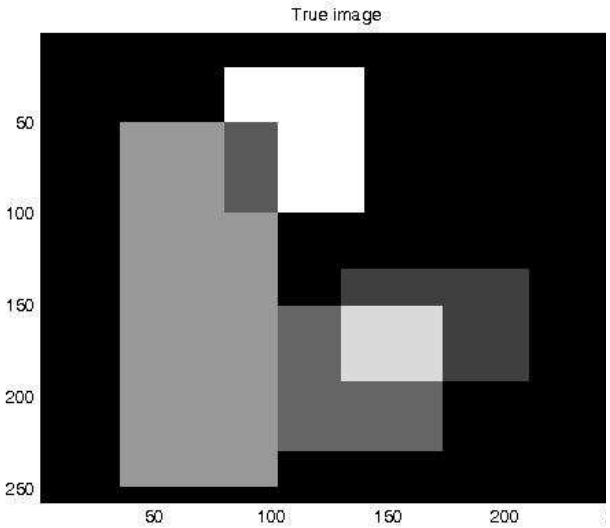


Point Spread Function

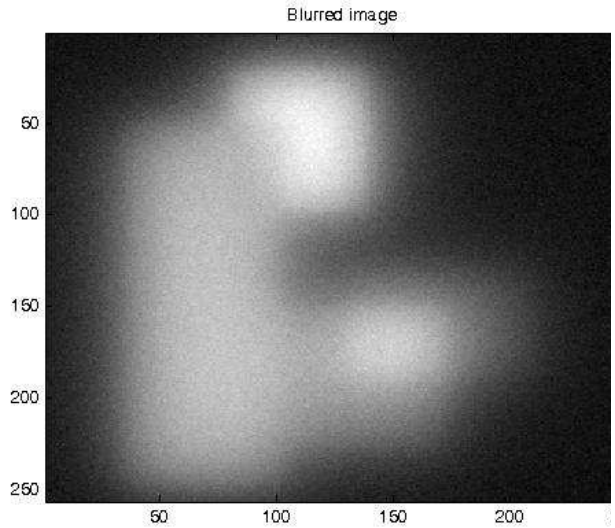


Blurred image (noise = 4.7%)

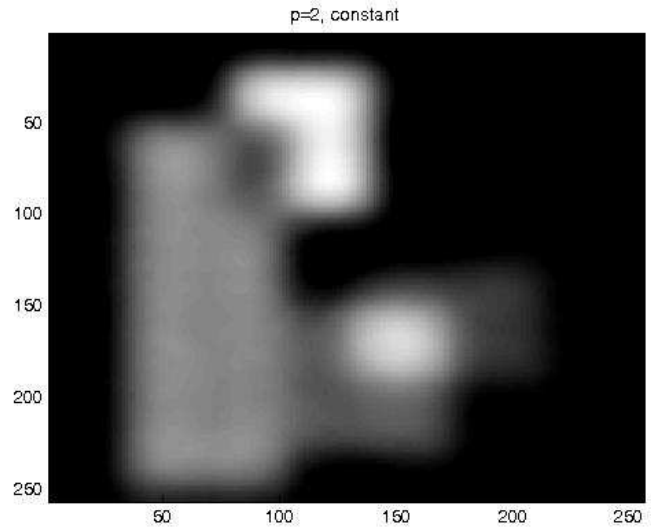




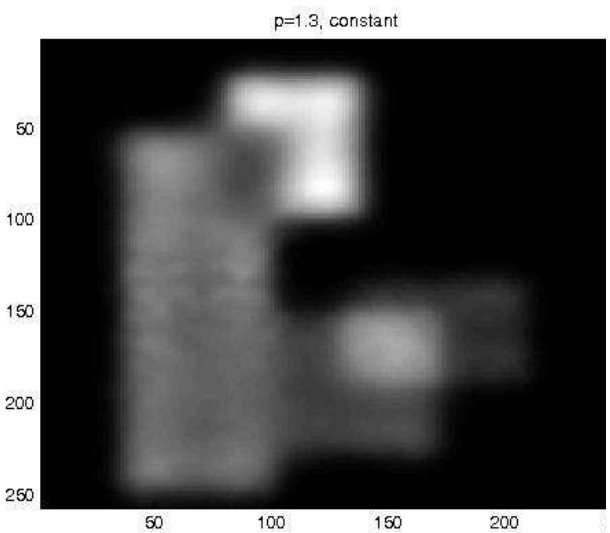
True image



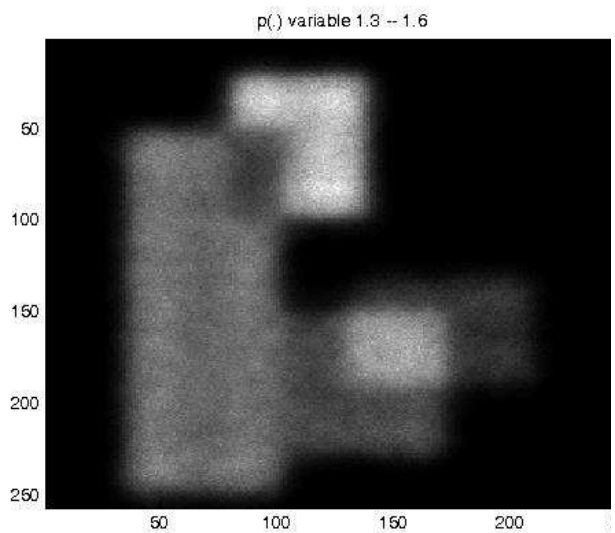
Blurred (noise = 4.7%)



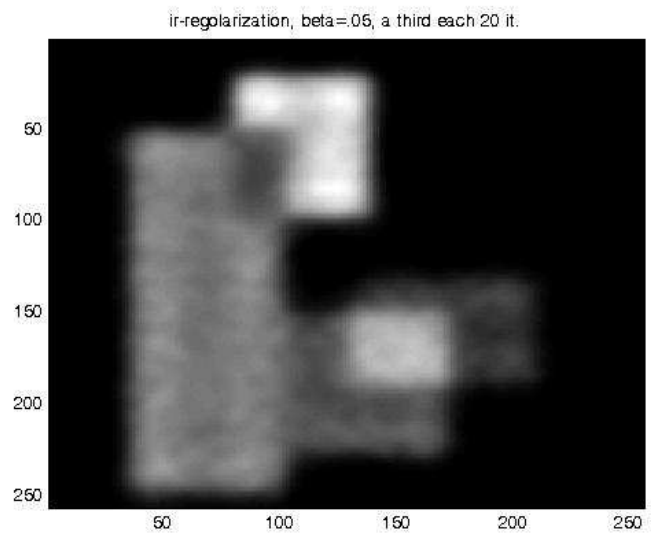
$p = 2 \quad (0.2692)$



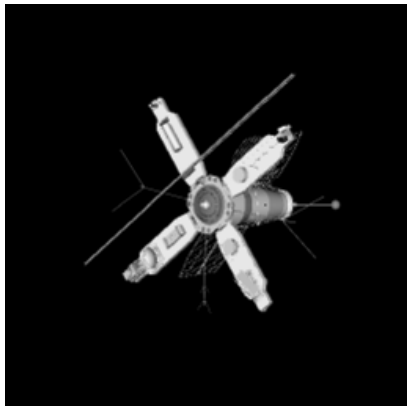
$p = 1.3 \quad (0.2681)$



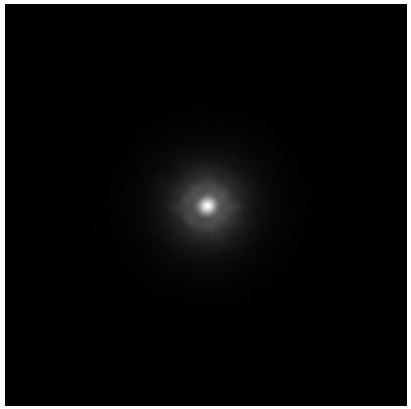
$p = 1.3 - 1.6 \quad (0.2473)$



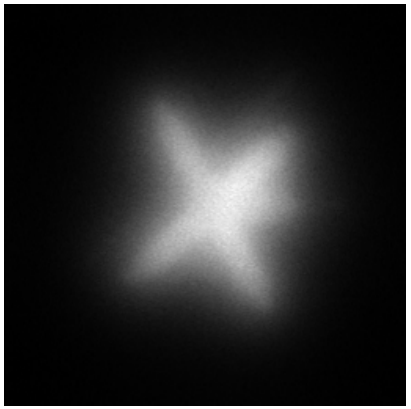
$1.3 - 1.6$ and irreg. (0.2307)



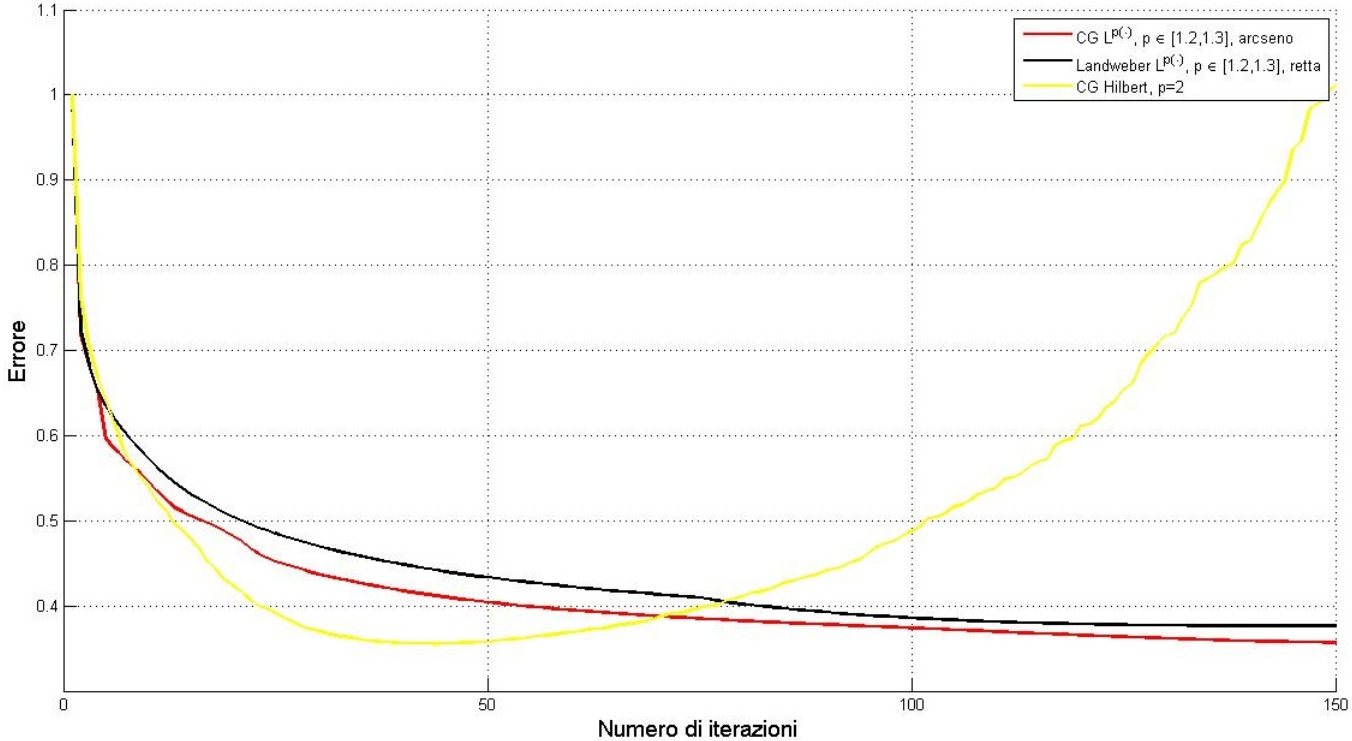
True image

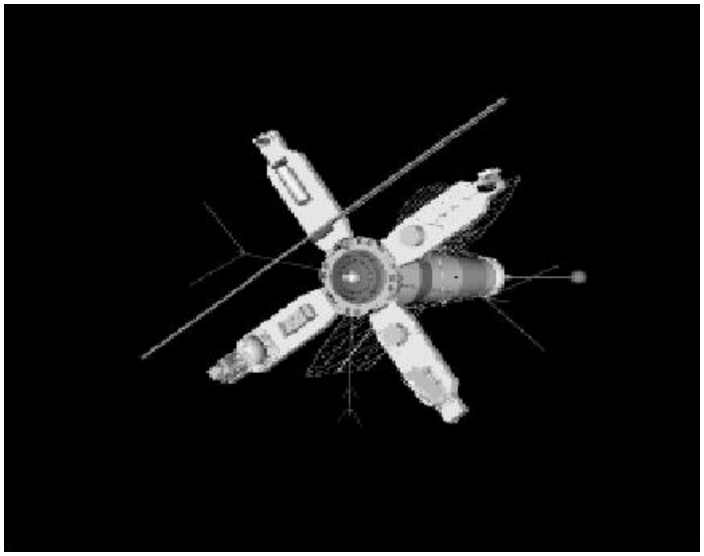


Point Spread Function

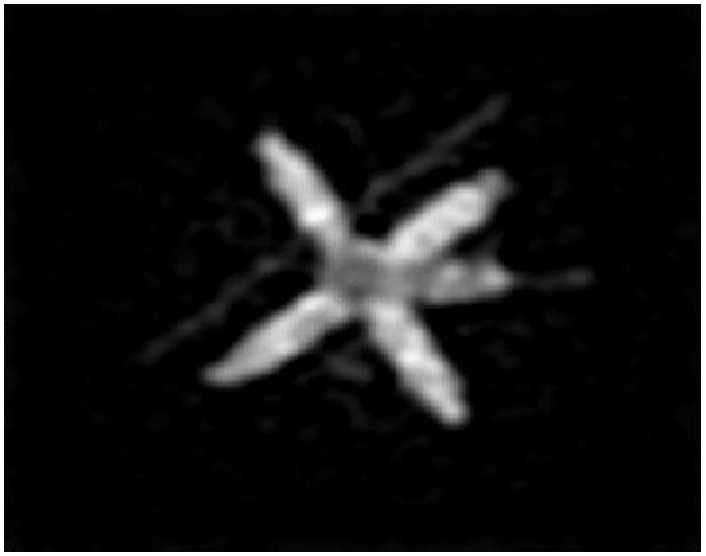


Blurred image (noise = 4.7%)

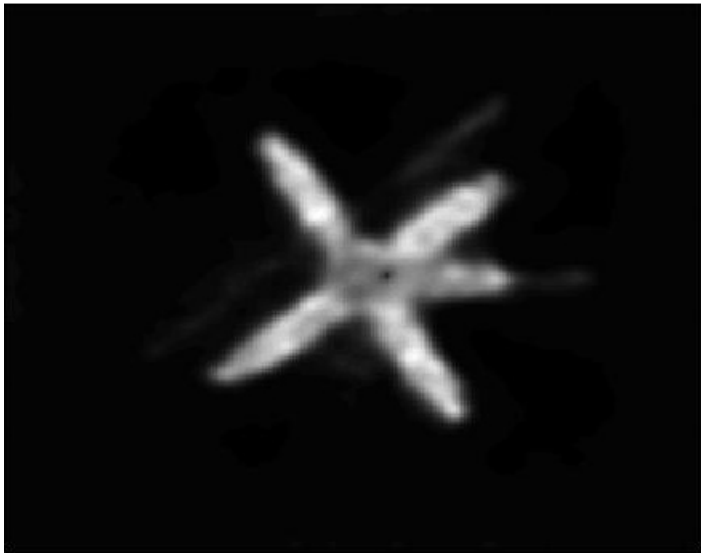




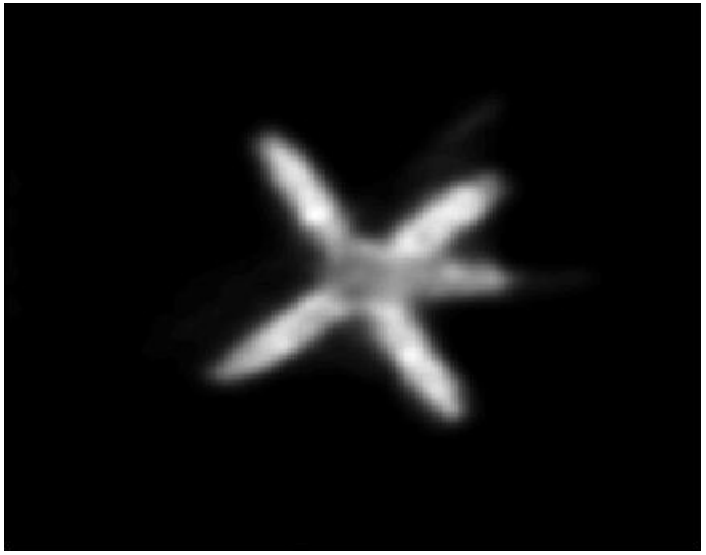
True image



CG $p(\cdot)$ (it. 150; rre: 0.3569)



CG $p = 2$ (it. 45; rre: 0.3557)



Landw. $p(\cdot)$ (it. 150; rre: 0.3766)

VI

Biomedical, civil engineering and geoscience applications

Inverse scattering for brain stroke detection

Inverse scattering for nondestructive testing

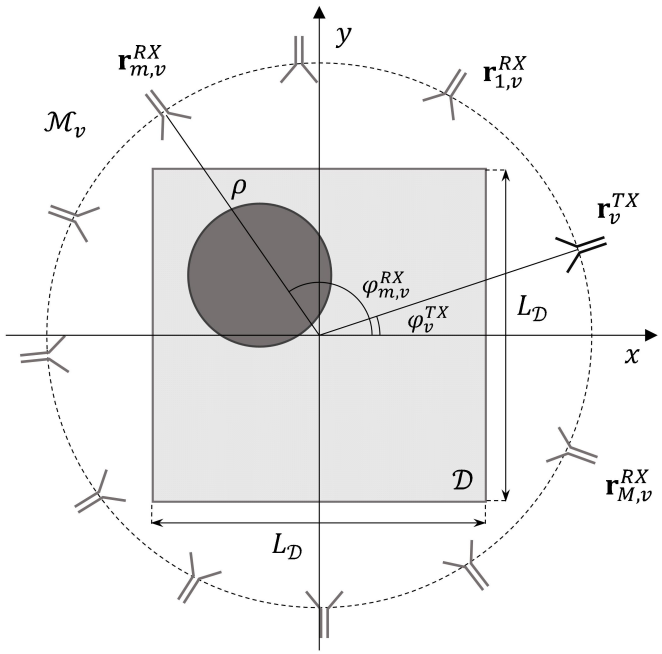
Inverse scattering for remote sensing

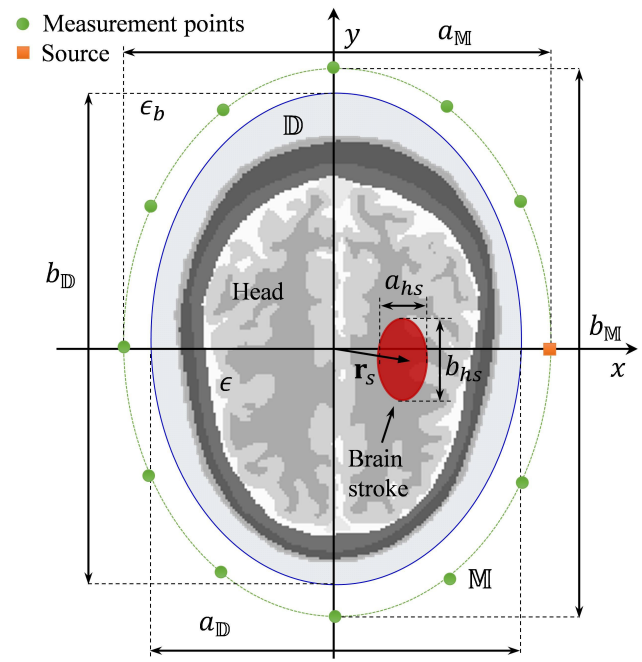
Biomedical application

(Nonlinear) inverse scattering for brain stroke detection

An application:

The Microwave Inverse Scattering (nonlinear imaging)





Input: scattered electromagnetic (EM) field on Ω_{obs} (observation domain)
Output: dielectric properties, i.e. the object, in Ω_{inv} (investigation dom.)
The model leads to a nonlinear integral equation -particles interaction-.

Features: very low degree of invasivity; provide information about the dielectric properties (instead of density); microwave cheap and easy to generate.
Applications: medical imaging (microwave tomography), nondestructive evaluations of materials (civil eng., cultural heritage), subsurface prospecting,...

Linear Imaging vs Nonlinear Imaging

Inverse problem in Imaging: to reconstruct the true image x from the knowledge of the acquired image $A(x)$, where A is an integral operator.

Linear imaging (Image Deblurring Problem)

$$[A\,x](r) = \int_{\Omega_{inv}} G(r-s)\,x(s)\,ds$$

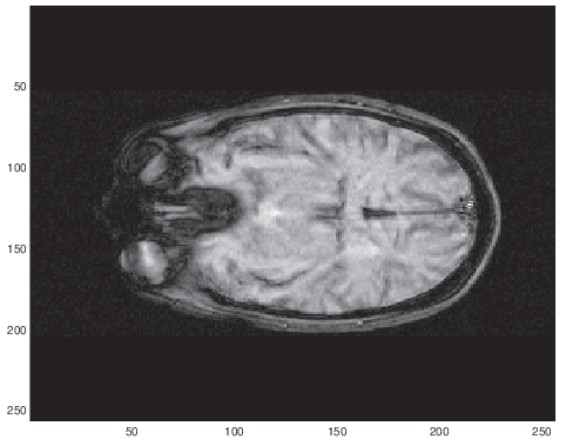
$\forall r \in \Omega_{inv}$, i.e. the 2D (or 3D) investigation domain.

Nonlinear imaging (Inverse Scattering Problem)

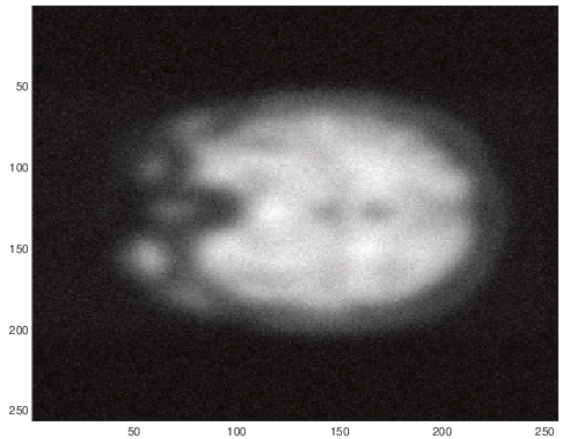
$$[A(x)](r) = \int_{\Omega_{inv}} G(r-s)\left[N(x, A(x))\right](s)\,ds$$

$\forall r \in \Omega_{obs}$, i.e. the 2D (or 3D) observation domain, where N is a nonlinear functional (sometimes not completely known).

Linear Imaging (image deblurring):

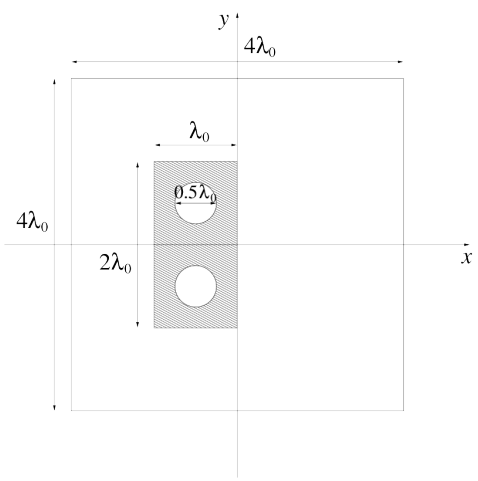


True object to restore

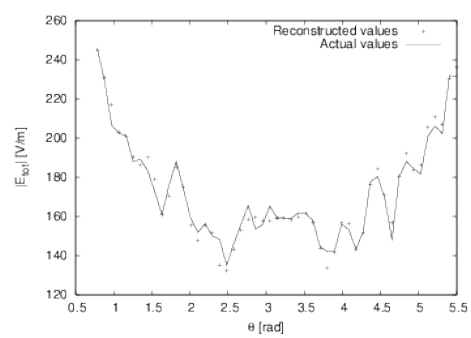


Input data for restoration

Nonlinear imaging (microwave inverse scattering):



True object to restore
(dielectric properties)



Input data for restoration
(EM scattered field)

Microwave Inverse Scattering

Forward non-linear integral operator $A : X \longrightarrow Y$, where X is the space of the dielectric permittivity, and Y of the scattered EM field,

$$[A(x)](r) = \int_{\Omega_{inv}} G(r-s) \Big[[A(x)](s) + y_{inc}(s) \Big] x(s)\, ds \qquad \forall r \in \Omega_{obs}.$$

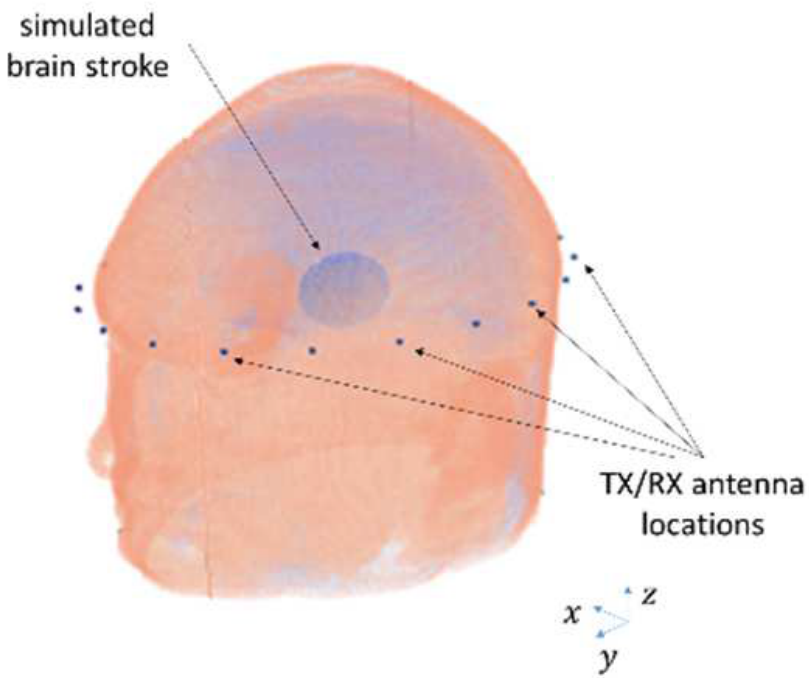
$x \in X$ is the scatterer (i.e., the unknown to retrieve),
$A(x) \in Y$ is the EM scattered field, unknown in Ω_{inv},
$y_{inc} \in Y$ is the EM incident field, known everywhere,
G is the known integral kernel (i.e., the Hankel function $H_0^{(2)}$ for the 2D case).

Inverse Scattering Problem

INPUT: the scattered field $y_{scat} := A(x)$ on the observation domain Ω_{obs}
OUTPUT: the scattering potential x in the investigation domain Ω_{inv}

by solving of the nonlinear equation: $A(x) = y_{scat}$

Brain Stroke Microwave Imaging



Three-dimensional head model

Experimental setup

In this application of microwave imaging, the unknown target is the dielectric permittivity of brain tissues and hemorrhagic stroke.

The apparatus is rotated (multiple views) and different incident waves are used (multiple illuminations), in order to provide different acquisitions of the scattered field y_{scat} onto Ω_{obs}.

Weak scatterers: Born approximations

Together to the actual unknown x to retrieve, inside the investigation domain Ω_{inv} the total electric field y_{tot} is unknown too !

(First order) Born Approximation

If $y_{scat} \ll y_{inc}$ (weak scatterers), then $y_{tot} = y_{inc} + y_{scat} \approx y_{inc}$, which is known everywhere, leading to

$$\int_{\Omega_{inv}} G(r-s)\, y_{B1}(s)\, x(s)\, ds = y_{scat}(r)\,,$$

which is a **linear** equation with $y_{B1} = y_{inc}$.

Second order Born Approximation

We consider the **quadratic (i.e., nonlinear)** functional equation

$$\int_{\Omega_{inv}} G(r-s)\, y_{B2}(s)\, x(s)\, ds = y_{scat}(r)\,,$$

where

$$y_{B2}(s) = y_{inc}(s) + \int_{\Omega_{inv}} G(s-s')\, y_{B1}(s')\, x(s')\, ds'$$

The computation of the Fréchet Derivative

The Fréchet Derivative of the Second order Born operator

$$A(x) = \int_{\Omega_{inv}} G(r-s)\, y_{B2}(s)\, x(s)\, ds$$

at the point x is the linear operator $A'_x : X \longrightarrow Y$ such that

$$A(x+h) = A(x) + A'_x h + O(\|h\|^2)\,.$$

By simple algebraic computations

$$\begin{aligned}[A'_x h](r) &= (A_1 h)(r) + \\ &+ \int_{\Omega_{inv}} G(r-s) x(s) [A_1 h](s) ds \\ &+ \int_{\Omega_{inv}} G(r-s) h(s) [A_1 x](s) ds\end{aligned}$$

where $A_1 : X \longrightarrow Y$ is the classical first order Born approximation

$$[A_1 v](r) = \int_{\Omega_{inv}} G(r-s)\, y_{inc}(s)\, v(s)\, ds$$

Strong scatterers: full formulation (without Born approx.)

If $y_{scat} \not\ll y_{inc}$ (strong scatterers), in addition to the actual unknown x to retrieve, we consider y_{tot} as unknown inside the investigation domain Ω_{inv}, and we obtain two coupled integral equations.

In the observation domain Ω_{obs} (i.e., measured data):

$$\int_{\Omega_{inv}} G(r-s)\, y_{tot}(s)\, x(s)\, ds = y_{scat}(r)\,,$$

$$\forall r \in \Omega_{obs}.$$

In the investigation domain Ω_{inv}:

$$y_{tot}(r) - \int_{\Omega_{inv}} G(r-s)\, y_{tot}(s)\, x(s)\, ds = y_{inc}(r)\,,$$

$$\forall r \in \Omega_{inv}.$$

Remark We rely on different acquisitions of $y_{tot}(r)$ on Ω_{obs} (that is, both multiple views and multiple illuminations) by an index $e = 1, \ldots, E$ related to the particular view and illumination.

Full Formulation for both scatterer and scattered field

By introducing the nonlinear operator A defined as

$$[A(y^1_{tot}, \dots, y^E_{tot}, x)](r) = \begin{pmatrix} \int_{\Omega_{inv}} G(r-s)\, y^1_{tot}(s)\, x(s)\, ds \\ \vdots \\ \int_{\Omega_{inv}} G(r-s)\, y^E_{tot}(s)\, x(s)\, ds \\ y^1_{tot}(r) - \int_{\Omega_{inv}} G(r-s)\, y^1_{tot}(s)\, x(s)\, ds \\ \vdots \\ y^E_{tot}(r) - \int_{\Omega_{inv}} G(r-s)\, y^E_{tot}(s)\, x(s)\, ds \end{pmatrix}$$

and the data vector $\quad y = \left(y^1_{scat}, \dots, y^E_{scat}, y^1_{inc}, \dots, y^E_{inc}\right)^T,$

the **inverse scattering problem** becomes:

find $x \in L^p(\Omega_{inv})$ and $y^e_{tot} \in L^p(\Omega_{inv})$, $e = 1, \dots, E$, such that

$$A(y^1_{tot}, \dots, y^E_{tot}, x) = y$$

The computation of the Fréchet Derivative for linearization

The Fréchet Derivative of the operator A at the point x is the linear operator $A'_x : X \longrightarrow Y$ such that $A(x+h) = A(x) + A'_x h + O(\|h\|^2)$.

The linearization gives rise to the sparse and structured matrix:

$$A'_x = \begin{pmatrix} a_{x,1} & 0 & \dots & 0 & A_{u,1} \\ 0 & A_{x,2} & \ddots & \vdots & A_{u,2} \\ \vdots & \ddots & \ddots & 0 & \vdots \\ 0 & \ddots & \ddots & A_{x,E} & A_{u,E} \\ I - A_x & \ddots & \ddots & 0 & -A_{u,1} \\ 0 & I - A_x & \ddots & \vdots & -A_{u,2} \\ \vdots & \ddots & \ddots & 0 & \vdots \\ 0 & \dots & 0 & I - A_x & -A_{u,E} \end{pmatrix}$$

where the **linear** operators $\{A_{u,e}\}_{e=1,\dots,E}$, $\{A_{x,e}\}_{e=1,\dots,E}$, and A_x are

$$[A_{x,e}h](r) = \int_{\Omega_{inv}} G(r-s)\, x(s)\, h(s)\, ds \qquad r \in \Omega^e_{obs}$$

$$[A_{u,e}h](r) = \int_{\Omega_{inv}} G(r-s)\, h(s)\, y^e_{tot}(s)\, ds \qquad r \in \Omega^e_{obs}$$

$$[A_x h](r) = \int_{\Omega_{inv}} G(r-s)\, x(s)\, h(s)\, ds \qquad r \in \Omega_{inv}$$

Two levels iterative regularization algorithm in Banach space

I- Outer Iterations: Gauss-Newton Method

Let $j = 0$ and x_0 be the initial guess. Compute the Fréchet derivative A'_{x_j}, and solve w.r.t. h_j the linear system

$$A'_{x_j} h_j = y - A(x_j)$$

by means of the inner iterations II.
Then update $x_{j+1} = x_j + h_j$, until a stopping rule (discrepancy principle, GCV, ...) holds true.

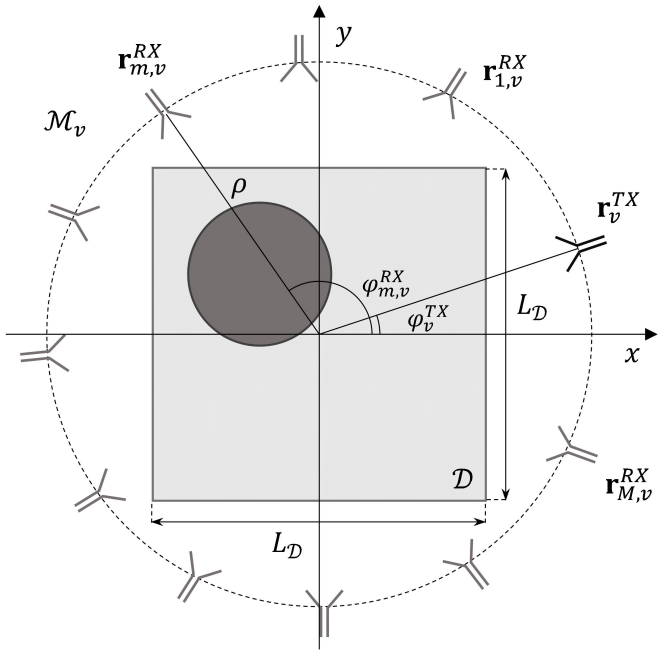
II- Inner Iterations: CG (Landweber) method in Banach space

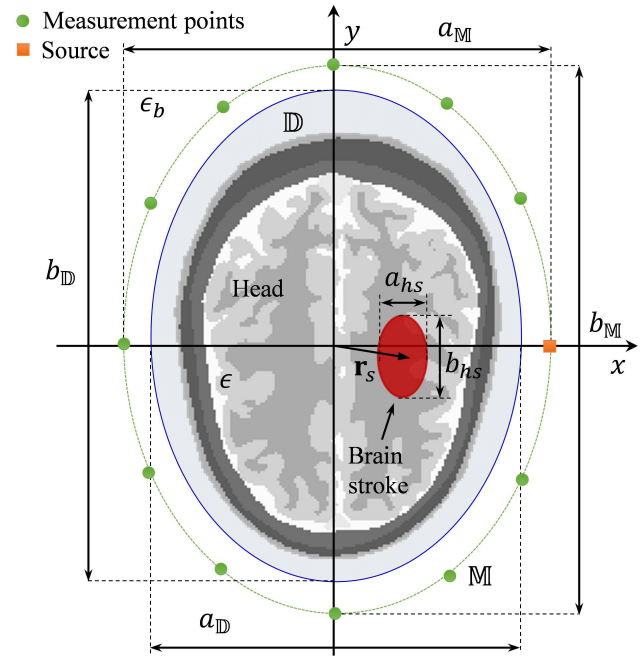
Compute a regularized solution of the Newton linear system by means of the CG (Landweber) algorithm in Banach space

$$h_{j,k+1} = J_{r^*}^{X^*}\left(J_r^X h_{j,k} - \lambda_k A'^{*}_{x_j} J_r^Y \left(A'_{x_j} h_{j,k+1} - (y - A(x_j))\right)\right)$$

and set $\quad h_j := h_{j,k_{\max}}$

Numerical results (synthetic 2D case)





Conjugate Gradient Method in L^p Banach spaces

Zubal Head Model, $a_{\mathbb{D}} = 16$ cm $b_{\mathbb{D}} = 20$cm and $a_{\mathbb{M}} = 18$cm $b_{\mathbb{M}} = 22$ cm

Cole-Cole model for brain tissues and stroke $a_{hs} = 1 \div 3$cm $b_{hs} = 2a_{hs}$

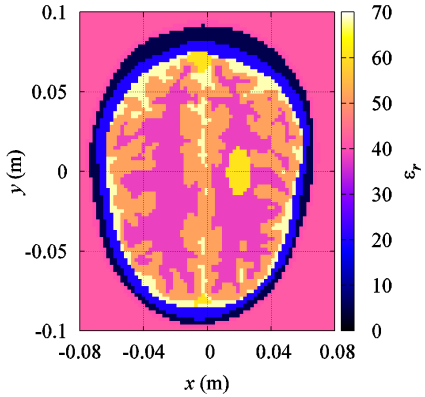
<u>Frequency</u>: $F = 400-600-800-1200$ MHz; <u>Measurement points</u>: $M = 30$

<u>Direct solver</u>: 5199 square subdomains of side length 2.2 mm.

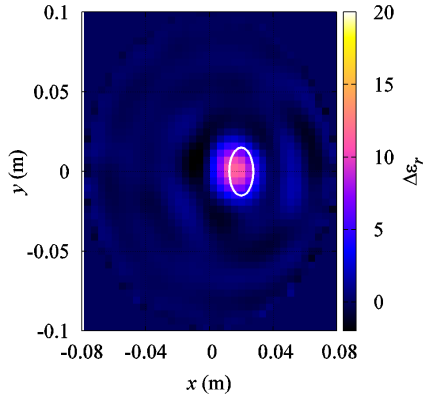
<u>Inversion</u>: 1300 cells of side length 4.4 mm.

<u>Noise</u>: Gaussian with 0 mean, SNR=25dB (Relative noise=6%).

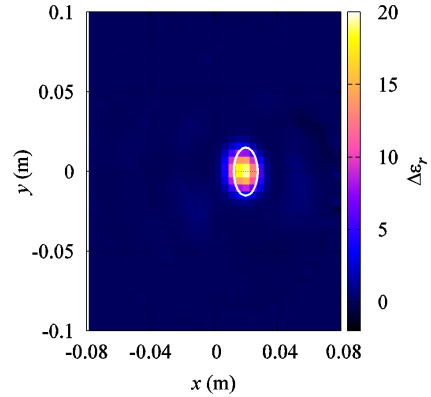
Numerical results (2D case)



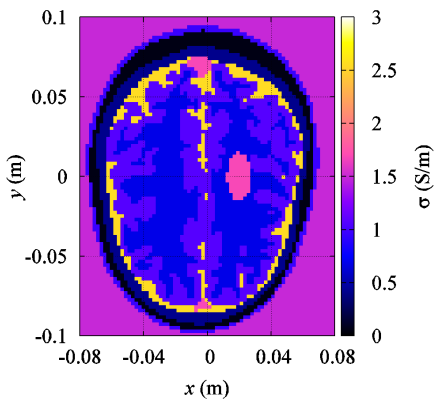
Actual relative dielectric permittivity;



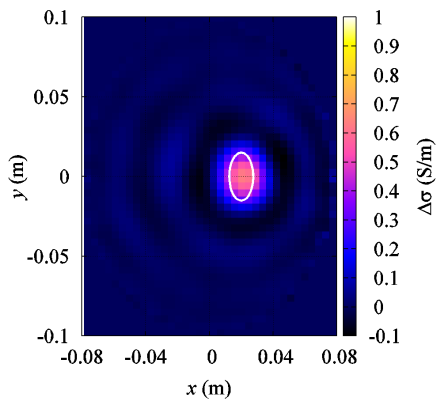
Reconstructed (differential): in Hilbert space;



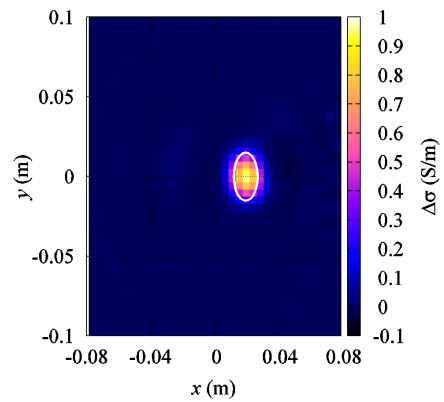
in Banach space (p=1.4)



Actual electric conductivity;



Reconstructed (differential): in Hilbert space;



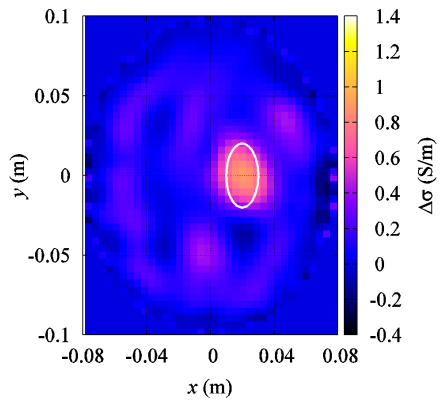
in Banach space (p=1.4)

Head model with hemorrhagic stroke of minor axis $a_{hs} = 1.5$cm.

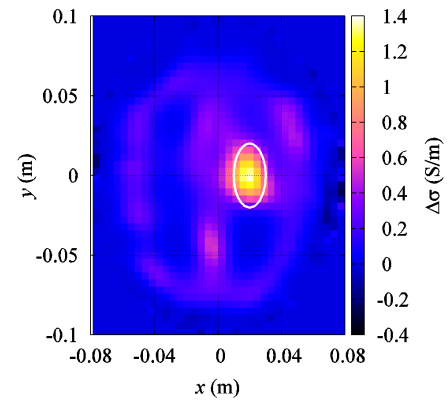
Initial guess: healthy head, supposed to be known (follow-up monitoring, or previous MRI/CT scan of the patients head). F=600 MHz.

Numerical results (2D case)

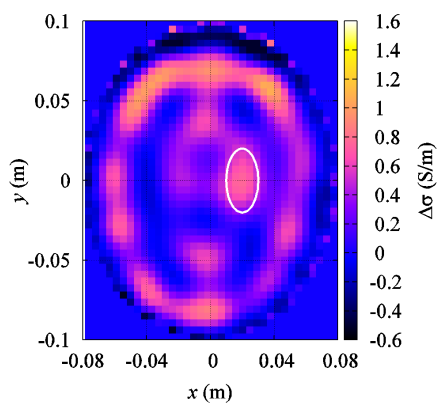
Reconstructed (differential) electric conductivity ($a_{hs} = 2$cm)



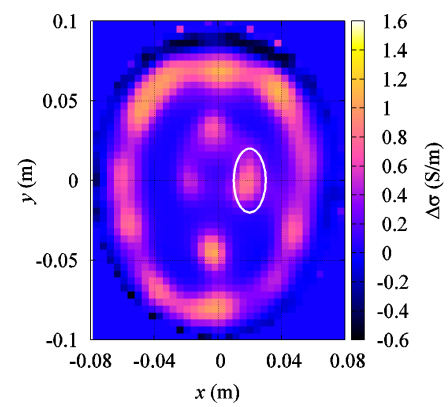
in Hilbert space



in Banach space (p=1.3)



in Hilbert space



in Banach space (p=1.3)

<u>Initial guess</u>: TOP partially known reference model;
. BOTTOM unknown reference model. F=600 MHz.

Numerical results (2D case): Landweber vs CG methods

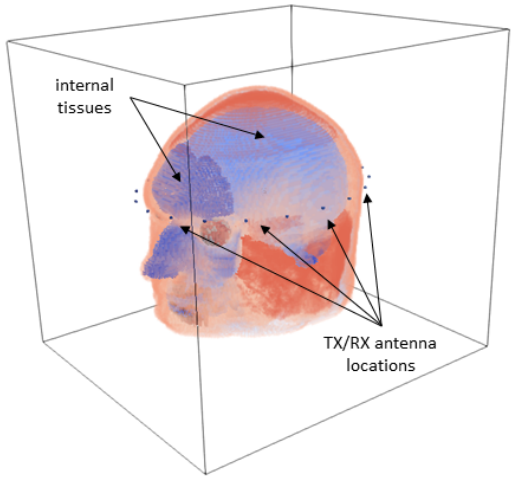
Initial guess: Healty head

Inner Solver	Relative Error			p optimal	Inner It.	Outer It.
	Total	Background	Scatterer			
Lw	0.116	0.007	0.226	1.3	64	3
CG	0.098	0.011	0.161	1.4	22	4

Initial guess: Homogeneous (gray matter)

Inner Solver	Relative Error			p optimal	Inner It.	Outer It.
	Total	Background	Scatterer			
Lw	0.745	0.128	1.105	1.5	28	4
CG	0.713	0.155	0.986	1.4	12	5

Numerical results (synthetic and laboratory 3D case)



Experimental setup

Conjugate Gradient Method in L^p Banach spaces

AustinMan 3D phantom, ($gprMax$ open source): $3.6 \cdot 10^6$ cubic cells.
Cole-Cole model for the hemorrhagic stroke ellipsoid of $1-3$cm of size

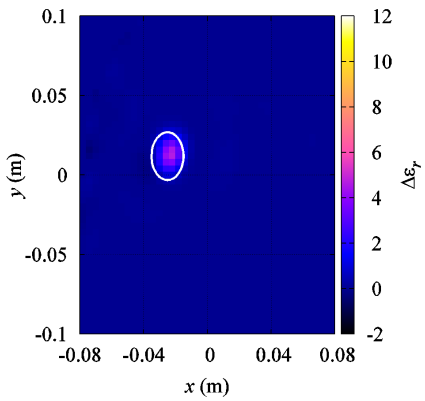
Frequency: $F = 800$ MHz; Measurement points: $M = 20$

Direct solver (gprMax open source): 1741 cells.

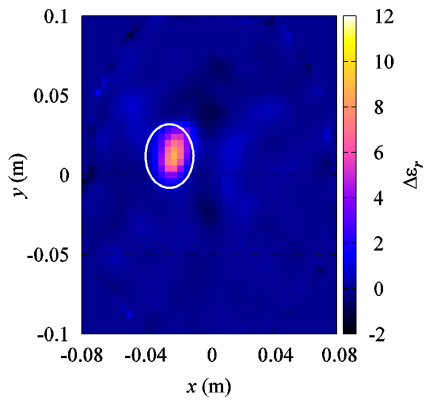
Inversion: 1300 cells .

Noise: Gaussian with 0 mean, SNR=25dB (Relative noise=6%).

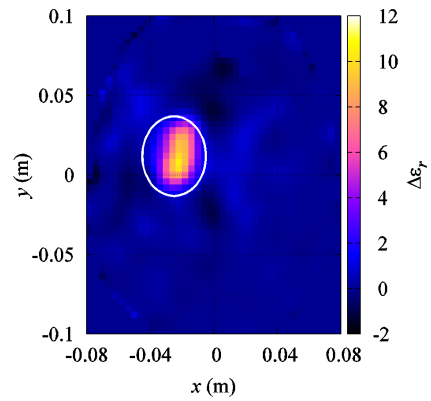
Numerical results (synthetic 3D case)



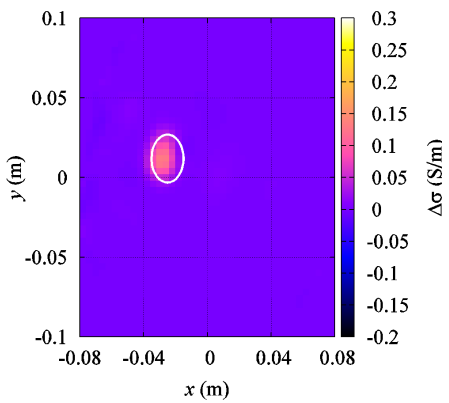
Small stroke 1cm;



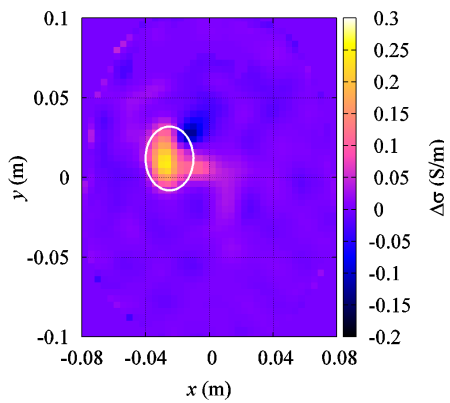
Medium stroke 1.5cm;



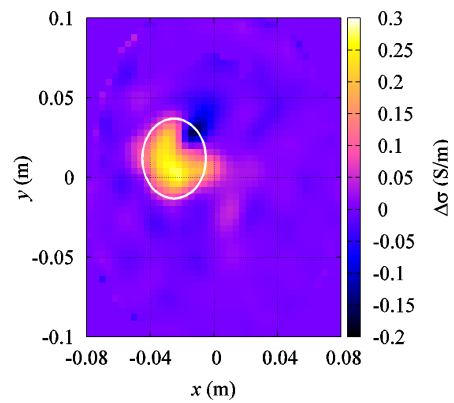
Large stroke 2cm



Small stroke 1cm;



Medium stroke 1.5cm;



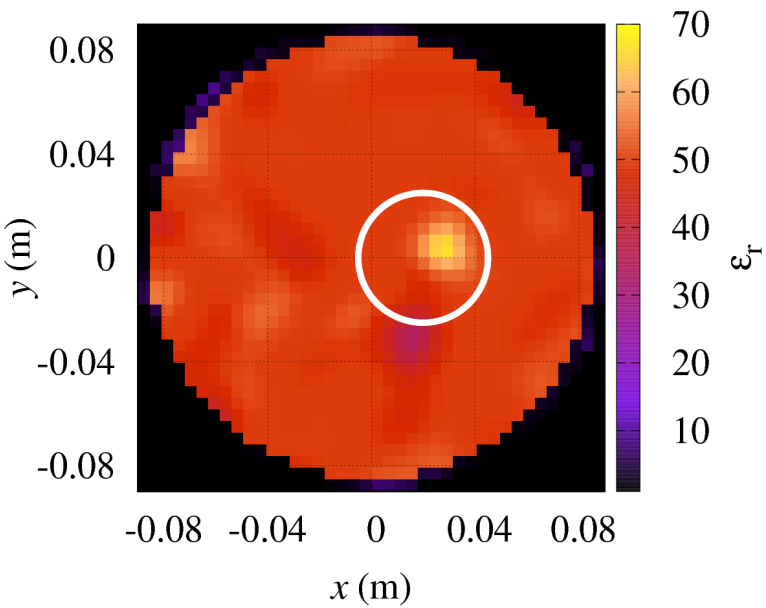
Large stroke 2cm

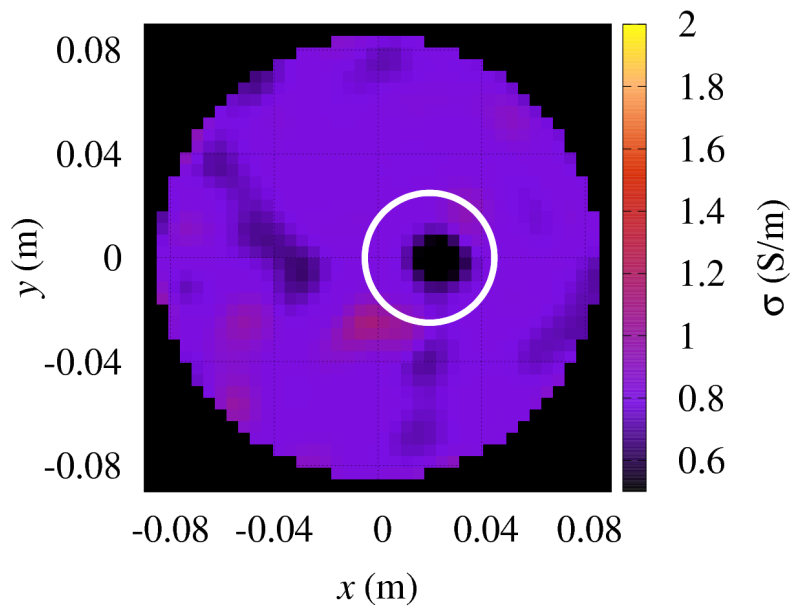
Banach space L^p, with $p = 1.4$

TOP: Reconstructed (differential) relative dielectric permittivity

BOTTOM: Reconstructed (differential) relative electric conductivity

Numerical results (laboratory testing 3D case)





Hilbert VS Banach restorations

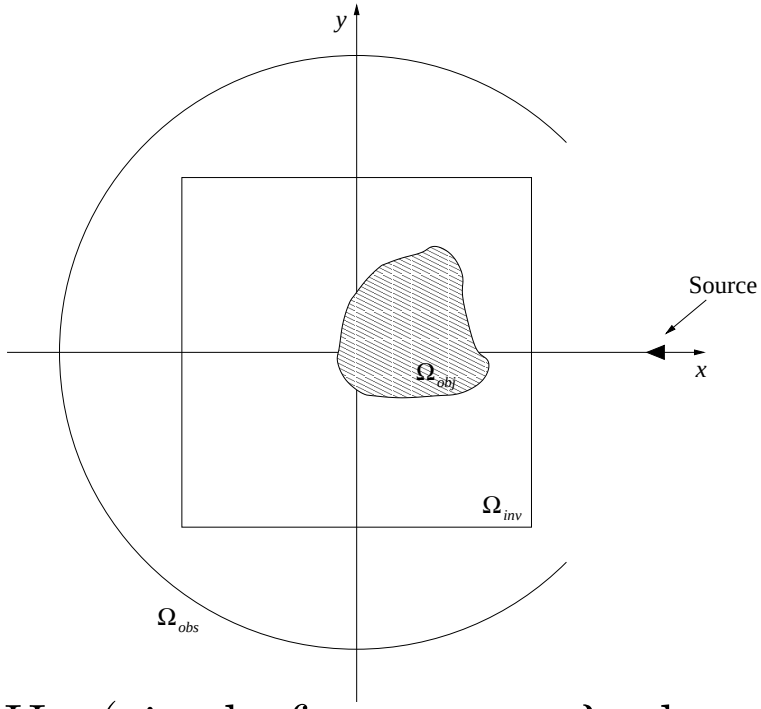
p	Small stroke		Medium stroke		Large stroke	
	Background	Scatterer	Background	Scatterer	Background	Scatterert
2	0.001	0.274	0.015	0.271	0.021	0.264
1.4	0.017	0.277	0.048	0.269	0.093	0.263

F=800MHz

Civil engineering application

(Nonlinear) inverse scattering for nondestructive testing

Numerical results



Frequency: $F = 1,2$ GHz (single frequency, λ about 0.5 m.).

Views: $V = 8$ (the apparatus is 8 times rotated by $2\pi/8$).

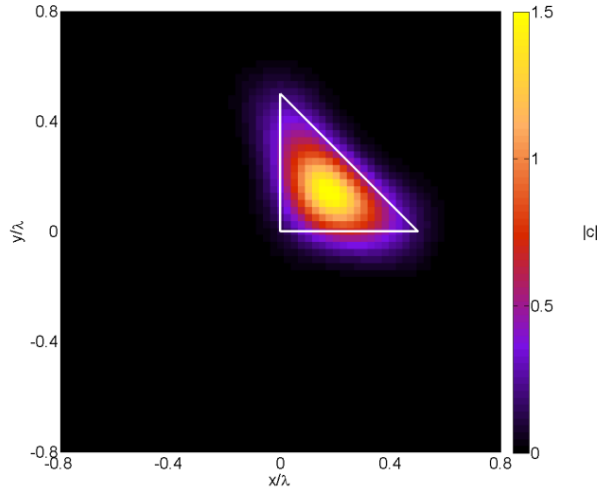
Measurement points: $M = 241$ points equispaced on Ω_{inv}, radius 1.67 m.

Investigation domain: square with side of about 1 m., discretization 63×63.

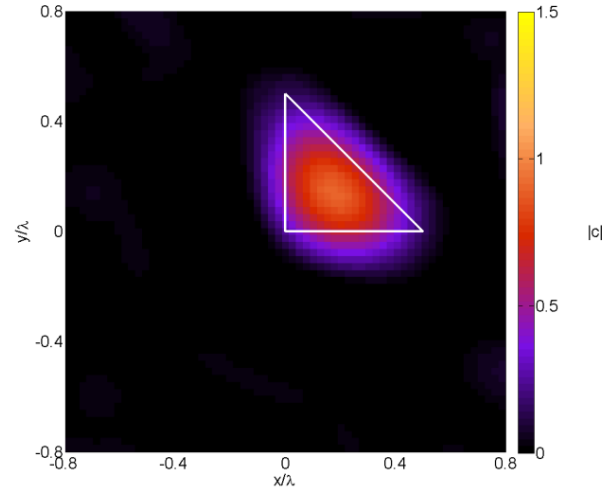
Noise: Gaussian with 0 mean, SNR=20dB, Relative noise=10%.

Initial guess: empty scene.

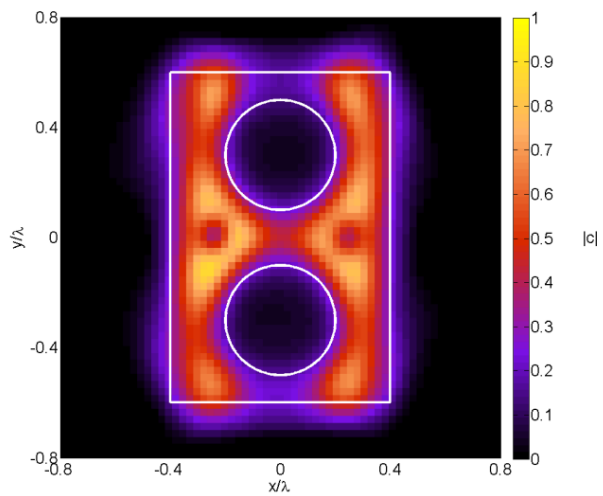
Gauss-Newton steps: 10.



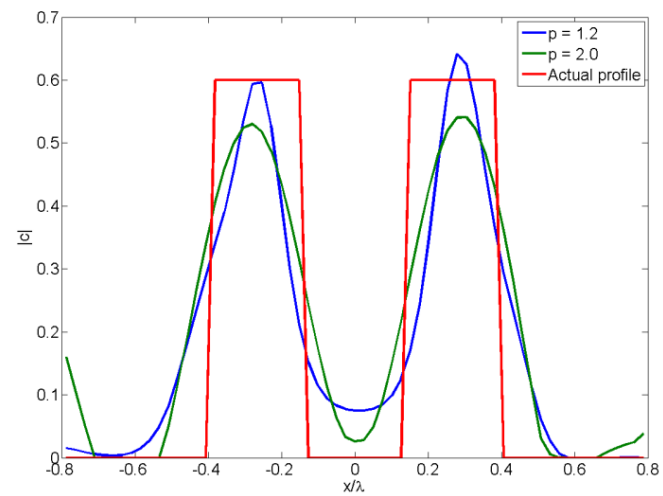
Banach Space regularization ($p = 1.2$)



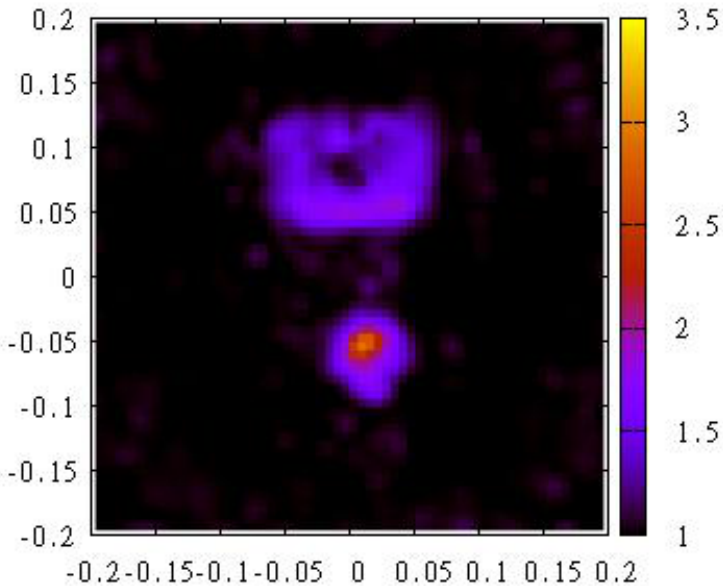
Hilbert Space regularization ($p = 2$)

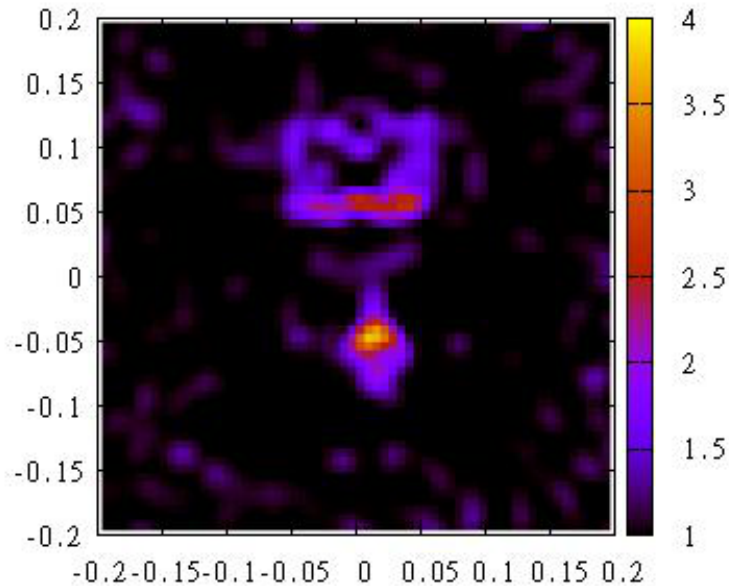


Banach Space regularization ($p = 1.2$)



Banach vs Hilbert (horizontal cut $y = -0.4$)



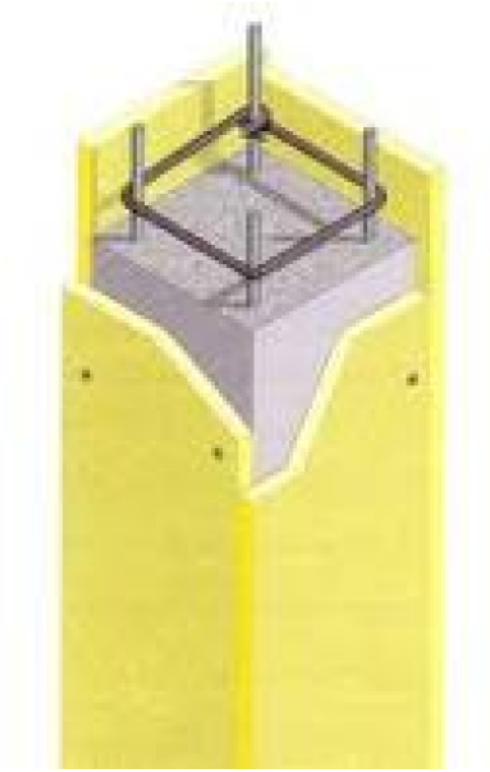


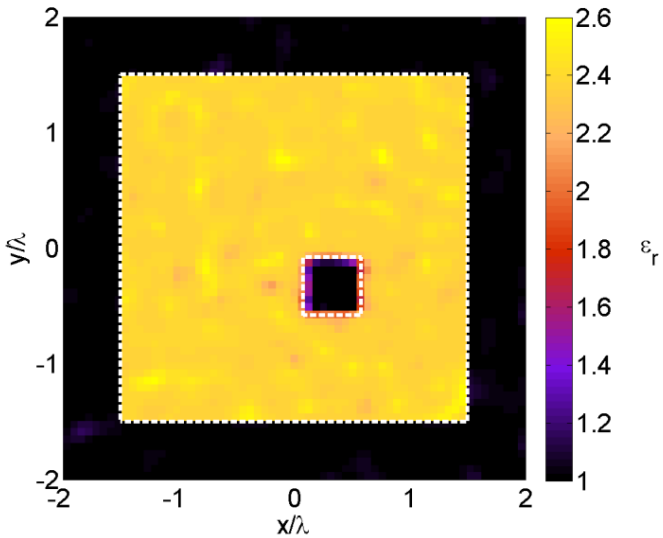
Real data from Institut Fresnel, Marseille
J.M. Geffrin, P. Sabouroux, C. Eyrau, *Inv. Probl.*, 2005

Banach space regularization ($p = 1.2$)

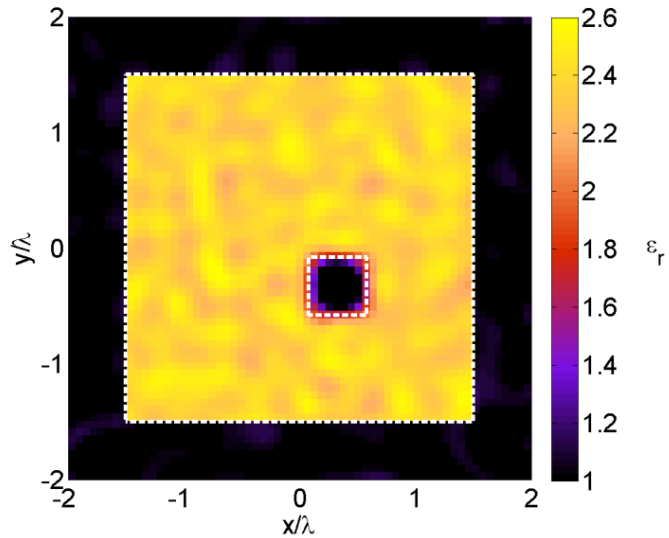
Hilbert space ($p = 2$)

An application in civil engineering : cement pillar non-destructive control

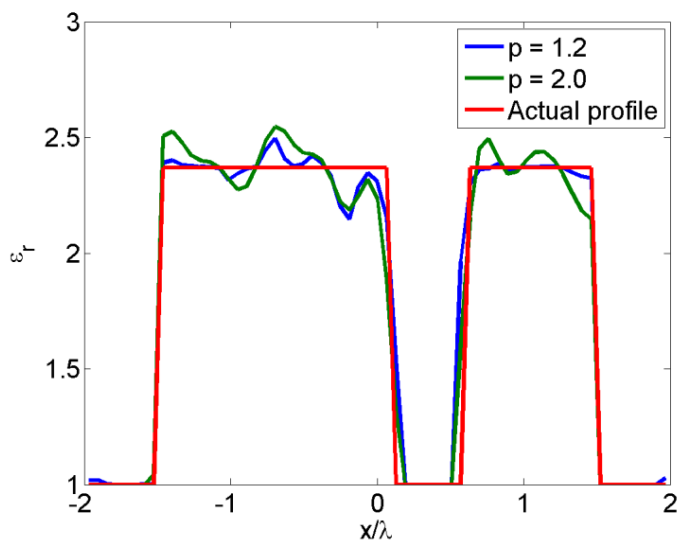




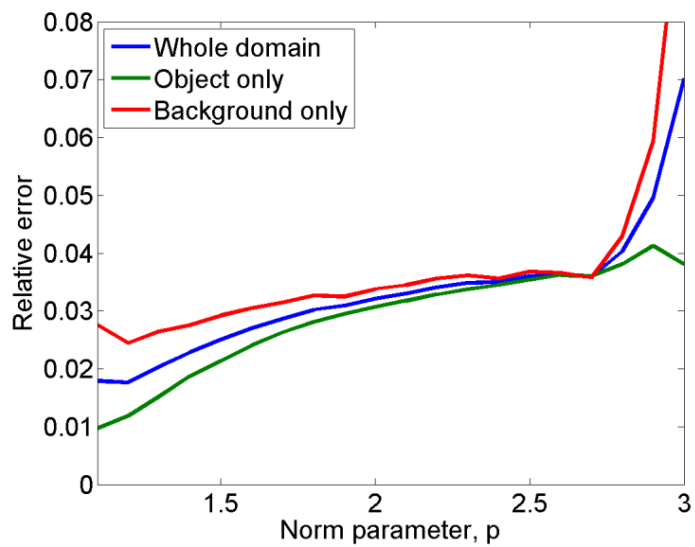
Banach Space regularization ($p = 1.2$)



Hilbert Space regularization ($p = 2$)



Banach vs Hilbert (horizontal cut $y = -0.4$)



Banach vs Hilbert (convergence w.r.t.p)

Banach vs Hilbert regularization:
Computational time parameter VERSUS norm parameter p.

p	Comput. time [s]	Outer iter.
1.2	526	5
1.6	912	11
2.0	1033	13
2.4	1099	14

An application in goescience

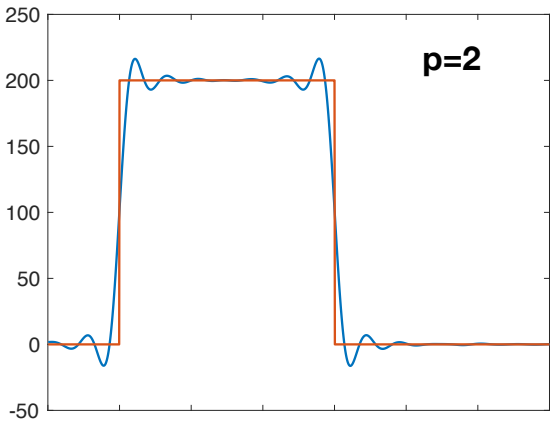
Inverse scattering for remote sensing

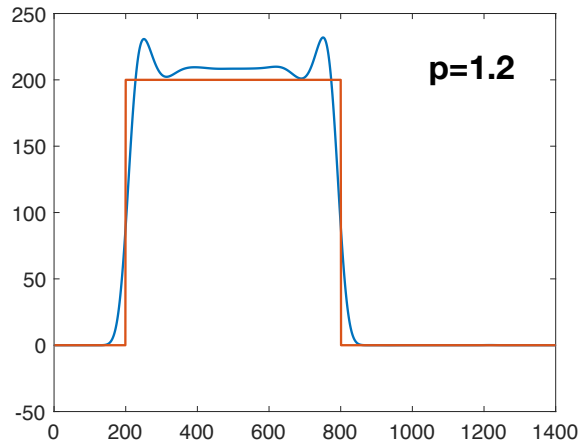
A remote sensing application: spatial resolution enhancement of microwave radiometer data

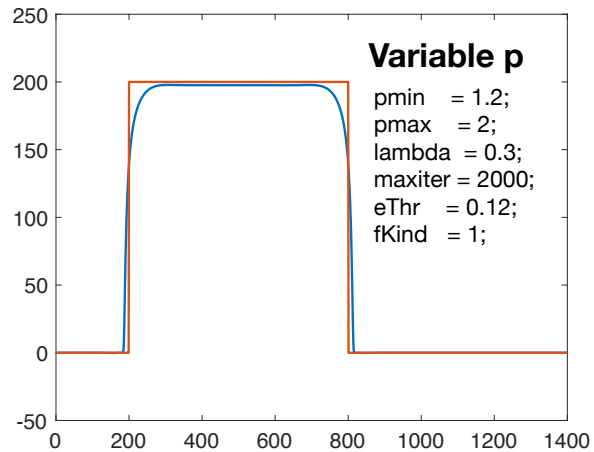
It is an under-determined problem (with a special structured matrix), since we want to reconstruct the High frequency components reduced by the Low Pass filter of the radiometer.

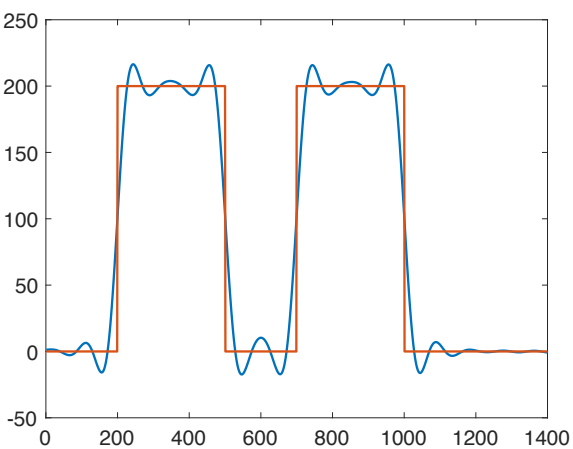
Courtesy: monde-geospatial.com

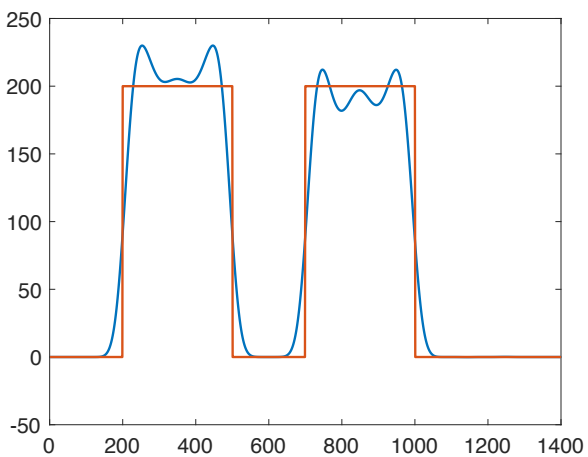
RECT E DOUBLE RECT

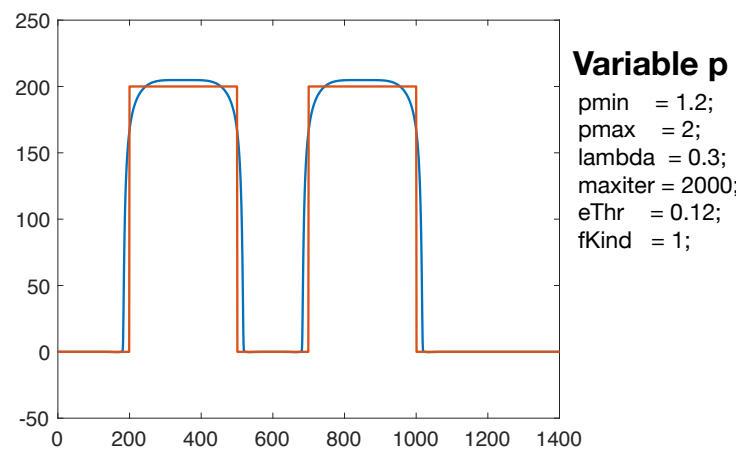


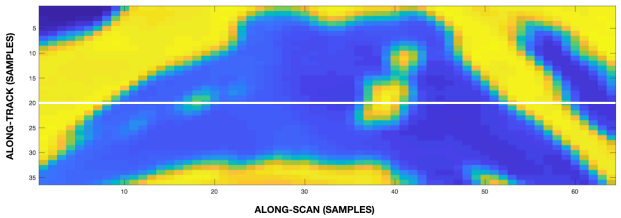


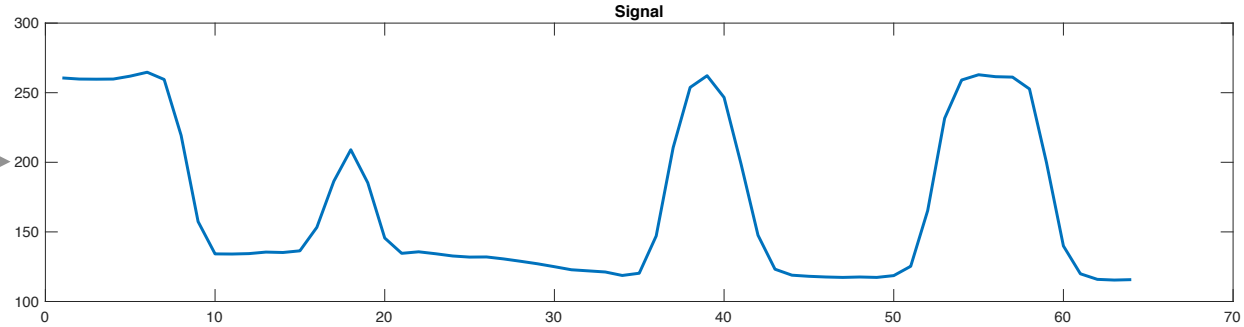


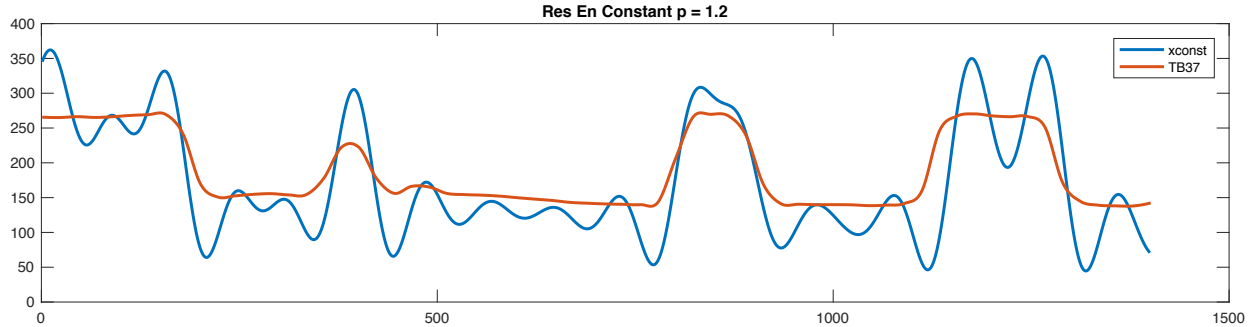


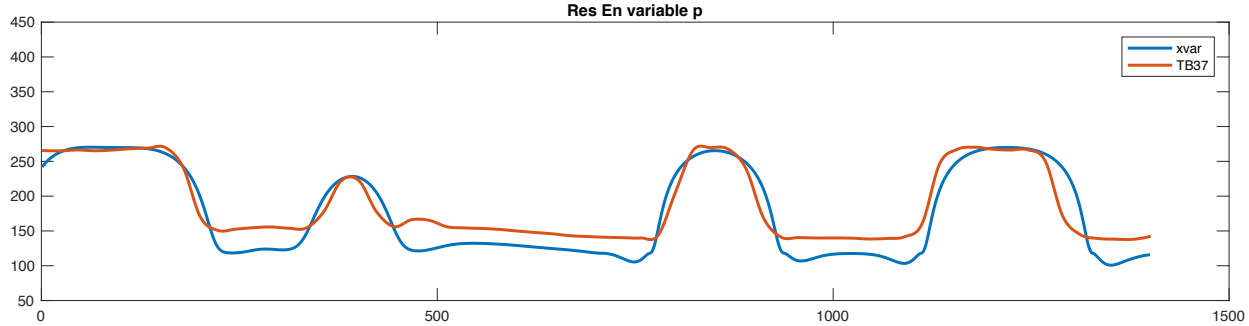





ALONG-TRACK (SAMPLES)
ALONG-SCAN (SAMPLES)


Signal


Res En Constant p = 1.2
xconst
TB37


Res En variable p
xvar
TB37

Thank you for your attention

References

[1] F. Schöpfer, A.K. Louis, T. Schuster, Nonlinear iterative methods for linear ill-posed problems in Banach spaces, *Inverse Problems*, vol. 22, pp. 311–329, 2006.

[2] M. O. Scherzer, M. Grasmair, H. Grossauer, M. Haltmeier, F. Lenzen, *Variational Methods in Imaging*, Springer, Berlin, 2008.

[3] Schuster, T., Kaltenbacher, B., Hofmann, B., and Kazimierski, K. S., *Regularization Methods in Banach Spaces.* Radon Series on Computational and Applied Mathematics, vol. 10, De Gruyter, 2012.

[4] C. Estatico, M. Pastorino, and A. Randazzo, A novel microwave imaging approach based on regularization in L^p Banach spaces, *IEEE Transactions on Antennas and Propagation*, 60, 3373–3381, 2012.

[5] P. Brianzi, F. Di Benedetto, C. Estatico, Preconditioned iterative regularization in Banach spaces, *Computational Optimization and Applications*, vol. 54, pp. 263–282, 2013.

[6] F. Lenti, F. Nunziata, C. Estatico, M. Migliaccio, On the resolution enhancement of radiometer data in Banach spaces, *IEEE Trans. Geoscience and Remote Sensing*, vol. 52, pp. 1834–1842, 2014.

[7] P. Dell'Acqua, C. Estatico, Acceleration of multiplicative iterative algorithms for image deblurring by duality maps in Banach spaces, *Applied Numerical Mathematics*, vol. 99, pp. 121–136, 2016.

[8] C. Estatico, S. Gratton, F. Lenti, D. Titley-Peloquin, A conjugate gradient like method for p-norm minimization in functional spaces, *Numerische Mathematik*, vol. 137, pp. 895–922, 2017.

[9] C. Estatico, A. Fedeli, M. Pastorino, A. Randazzo, Quantitative Microwave Imaging Method in Lebesgue Spaces with Non-Constant Exponents, *IEEE Transactions on Antennas and Prop.*, in press, 2018.

[10] M. Alparone, F. Nunziata, C. Estatico, F. Lenti, M. Migliaccio, An adaptive Lp-penalization method to enhance the spatial resolution of microwave radiometer measurements, *IEEE TGRS*, major rev., 2018.