Progetto di Ricerca GNCS 2016

PING – Problemi Inversi in Geofisica

Firenze, 6 aprile 2016

Regularized nonconvex minimization for image restoration

Claudio Estatico

Joint work with: Fabio Di Benedetto, Flavia Lenti

Dipartimento di Matematica, Università di Genova, Italy

IRIT – Institut de Recherche en Informatique de Toulouse, France

Outline

I - Inverse Problems, Image restoration, and Tikhonov-type variational approaches for solution by optimization.

II - Minimization of the residual by gradient-type iterative methods in (Hilbert and) Banach spaces.

III - Acceleration of the convergence via operator-dependent penalty terms: the "ir-regularization".

IV - The vector space of the DC (difference of convex) functions, and some relations with linear algebra.

V - Numerical results in imaging and geoscience, for the accelerated method by "ir-regularization".

Inverse Problem

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

Given the (noisy) data $g \in G$, find (an approximation of) the unknown $f \in F$ such that

$$A(f) = g$$

where $A : \mathcal{D} \subseteq F \longrightarrow G$ is a known functional operator,
and F and G are two functional (here Hilbert or Banach) spaces.

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

(A classical) Example of Inverse Problem: Image Deblurring

Forward operator (blurring model):

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

$$g(x) = \int_{\mathbb{R}^2} h(x,y)\, f(y)\, dy$$

where $x, y \in \mathbb{R}^2$, and $h(\bullet, \bullet)$ is the known impulse response of the imaging system, i.e., the point spread function (PSF).

Inverse problem (image deblurring):

Given (a noisy version of) g, find (an approximation of) f, by solving the functional linear equation

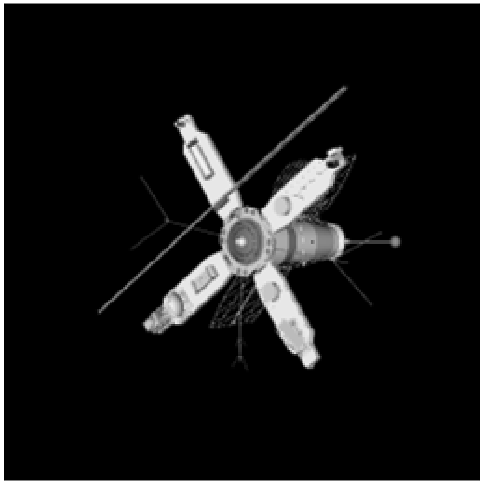
$$Af = g\,.$$

(A classical) Example of Inverse Problem: Image Deblurring

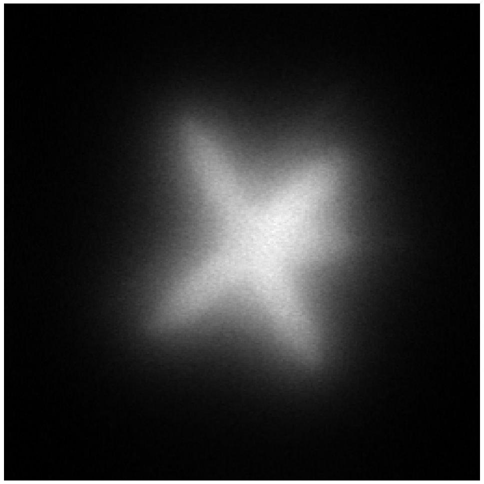
Inverse problem (image deblurring):

Given (a noisy version of) g, find (an approximation of) f, by solving

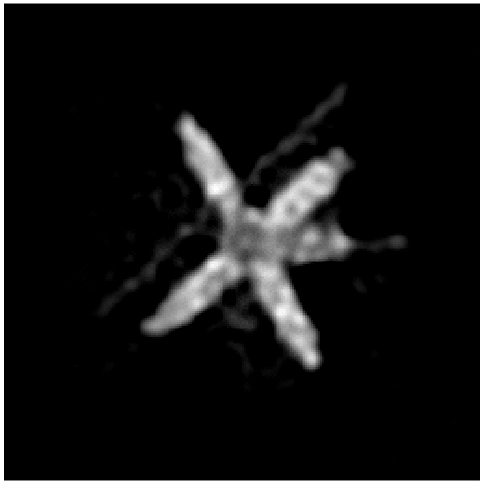
$$Af = g\,.$$



True Image



Blurred and noisy image



Restored Image

Solution of inverse problems by optimization

Optimization techniques (by variational approaches) are very useful to solve the functional equation $A(f) = g$.

These methods minimize the functional $\Phi : F \longrightarrow \mathbb{R}$

$$\Phi(f) = \|A(f) - g\|_G^p \,,$$

or the Tikhonov-type variational regularization functional Φ_α

$$\Phi_\alpha(f) = \|A(f) - g\|_G^p + \alpha\mathcal{R}(f)\,,$$

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

The "data-fitting" term $\|A(f) - g\|_Y^p$ is called residual (usually in mathematics) or cost function (usually in engineering).

The "penalty" term $\mathcal{R}(f)$ is often $\|f\|_F^p$, or $\|\nabla f\|_F^p$ or $\|Lf\|_F^p$, for a differential operator L which measures the "non-regularity" of f.

Some comments on regularization in Banach spaces.

Several regularization methods for ill-posed functional equations have been formulated as minimization problems, first in the context of Hilbert spaces (i.e., the classical approach) and later in Banach spaces (i.e., the more recent approach).

Convex optimization in Banach spaces (such as L^1 for sparse recovery or L^p, $p > 1$ for smoother regularization) helps to derive new algorithms.

	Hilbert spaces	Banach spaces
Benefits	Easier computation (Spectral theory, eigencomponents)	Better restoration of the discontinuities; Sparse solutions
Drawbacks	Over-smoothness (bad localization of edges)	Theoretical involving (Convex analysis required)

Minimization of the residual by gradient-type iterative methods

For the Tikhonov-type functional $\Phi_\alpha(f) = \|A(f) - g\|_G^p + \alpha \mathcal{R}(f)$, the basic minimization approach is the gradient-type, which reads as

$$f_{k+1} = f_k - \tau_k \Psi_A(f_k, g)$$

where

$$\Psi_A(f_k, g) \approx \partial \left(\|A(f) - g\|_G^p + \alpha \mathcal{R}(f) \right),$$

is an approximation of the (sub-)gradient of the minimization functional at point f_k, and $\tau_k > 0$ is the step length.

In Banach setting, these iterations are defined in the dual space and are linked to the (“wide”...) fixed point theory.

Schöpfer, Louis, Hein, Scherzer, Schuster, Kazimierski, Kaltenbacher, Q. Jin, Tautenhahn, Neubauer, Hofmann, Daubechies, De Mol, Fornasier, Tomba.

The Landweber iterative method in Hilbert spaces

The (modified) Landweber algorithm is the simplest method for the minimization of $\Phi_\alpha(f) = \frac{1}{2}\|Af - g\|_2^2 + \alpha\frac{1}{2}\|f\|_2^2$, when the operator A is linear.

Since the gradient of Φ_α is

$$\nabla\Phi_\alpha(f) = A^*(Af - g) + \alpha f \, ,$$

we have the following iterative scheme:

Let $f_0 \in F$ be an initial guess (the null vector $f_0 = 0 \in F$ is often used in the applications).

For $k = 0, 1, 2, \ldots$

$$f_{k+1} = f_k - \tau(A^*(Af_k - g) + \alpha f_k)$$

where $\tau \in (0, 2(\|A\|_2^2 + \alpha)^{-1})$ is a fixed step length.

The Landweber iterative method in Banach spaces

The Landweber algorithm in Hilbert spaces has been extended to Banach spaces setting. Again, it is the basic method for the minimization of $\Phi_\alpha(f) = \frac{1}{p}\|Af - g\|_G^p + \alpha\frac{1}{p}\|f\|_F^p$, where $p > 1$ is a fixed weight value.

Let $f_0 \in F$ be an initial guess (or simply the null vector $f_0 = 0$).

For $k = 0, 1, 2, \ldots$

$$f_{k+1} = J_{p^*}^{F^*}\Big(J_p^F f_k - \tau_k(A^* J_p^G(Af_k - g) + \alpha J_p^F f_k)\Big)$$

where p^* is the Hölder conjugate of p, that is, $\frac{1}{p} + \frac{1}{p^*} = 1$.

The duality map $J_p^F : F \longrightarrow 2^{F^*}$ acts on the iterates $f_k \in F$, and the duality map $J_{p^*}^{F^*} : F^* \longrightarrow 2^F$ acts on the iterates $f_k^* \in F^*$.

N.B. Any duality map allow to associate an element of any Banach space B with its dual in the dual space B^*. Here B is reflexive.

Landweber iterative method in Hilbert spaces

$A : F \longrightarrow G \qquad A^* : G \longrightarrow F \qquad H_2(f) = \frac{1}{2}\|Af - g\|_G^2 + \frac{1}{2}\|f\|_F^2$

$$f_{k+1} = f_k - \tau(A^*(Af_k - g) + \alpha f)$$

Landweber iterative method in Banach spaces

$A : F \longrightarrow G \qquad A^* : G^* \longrightarrow F^* \qquad H_p(f) = \frac{1}{p}\|Af - g\|_G^p + \frac{1}{p}\|f\|_F^p$

$$f_{k+1} = J_{p^*}^{F^*}\Big(J_p^F f_k - \tau_k(A^* J_p^G(Af_k - g) + \alpha J_p^F f_k)\Big)$$

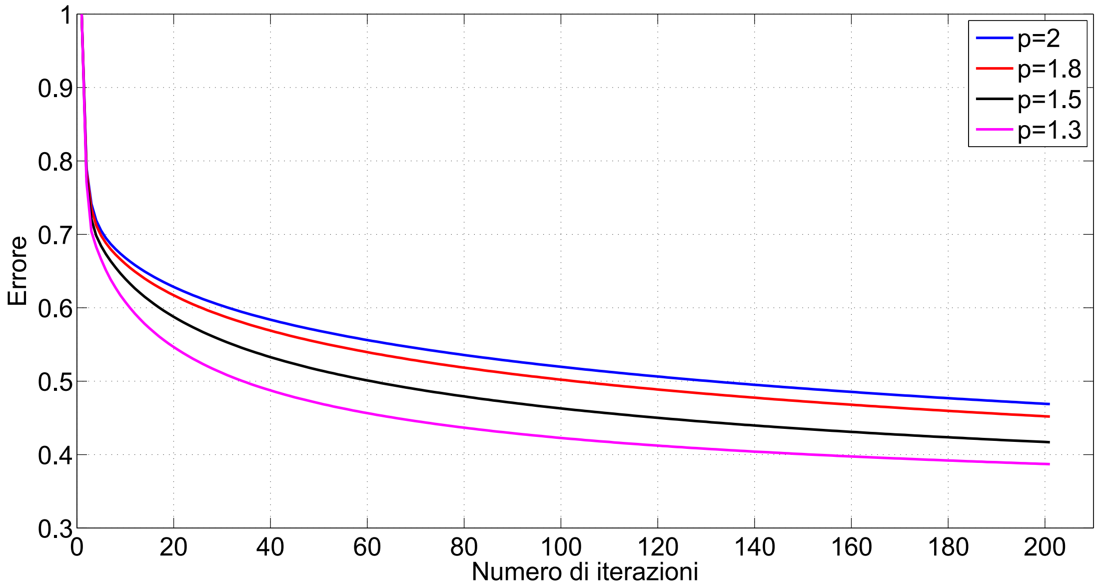
Some remarks.

In the Banach space L^p, by direct computation we have

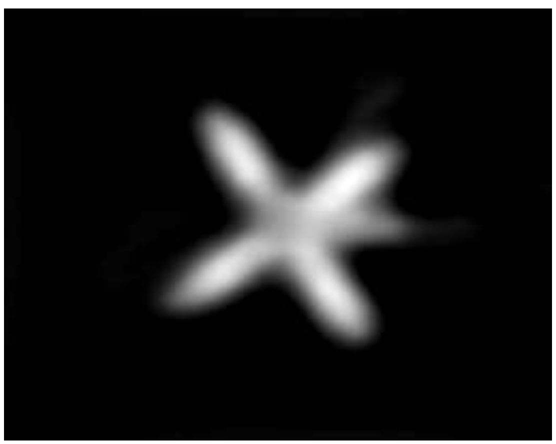
$$J_r^{L^p}(f) = \|f\|_p^{r-p}|f|^{p-1}\mathrm{sgn}(f)\,.$$

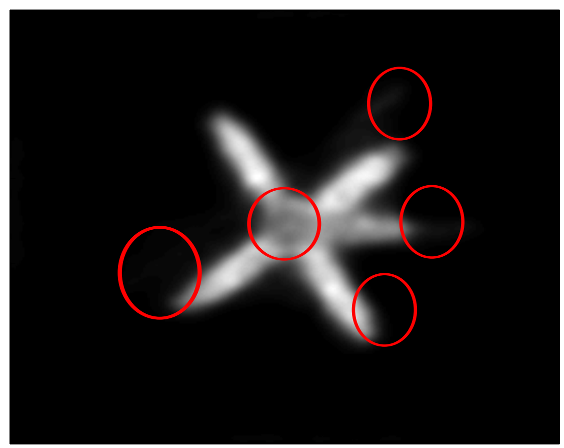
It is a non-linear, single-valued, diagonal operator, which cost $O(n)$ operations, and does not increase the global numerical complexity $O(n \log n)$ of shift-invariant image restoration problems solved by FFT.

A numerical evidence: Hilbert vs. Banach



Relative Restoration Errors $RRE(k) = \|f_k - f\|_2 / \|f\|_2$ vs. Iteration Number





Landweber in Hilbert spaces ($p = 2$) Landweber in Banach spaces ($p = 1.3$)

(200-th iteration, with $\alpha = 0$)

Improvement of regularization effects via operator-dependent penalty terms (I)

In the Tikhonov regularization functional $\Phi_\alpha(f) = \|Af - g\|_G^p + \alpha\mathcal{R}(f)$, widely used penalty terms $\mathcal{R}(f)$ include:

(i) $\|f\|^p$, or $\|f - f_0\|^p$, where f_0 is a priori guess for the true solution, whit L^p-norm, $1 < p < +\infty$, or the Sobolev spaces $W^{l,p}$-norm;

(ii) $\|f\|_S^2 = (Sf, f)$ in the Hilbertian case, where $S : F \longrightarrow F$ is a fixed linear positive definite (often is the Laplacian, $S = \triangle$).

(iii) $\int |\nabla f|$ for Total Variation regularization;

(iv) $\sum_i |(f, \phi_i)|$ or the L^1-norm $\int |f|$ for regularization with sparsity constrains;

In the blue case (ii) with the S-norm, the Landweber iteration becomes:

$$f_{k+1} = f_k - \tau(A^*(Af_k - g) + \alpha S f_k)$$

Improvement of regularization effects via operator-dependent penalty terms (II)

All of the classical penalty terms do not depend on the operator A of the functional equation $Af = g$, but only on f.

On the other hand, it is reasonable that the "regularity" of a solution depends on the properties of the operator A too.

Recalling that, in inverse problems:

Spectrum of A^*A	$\longleftrightarrow$	Subspace	Components
$\lambda(A^*A)$ small	$\longleftrightarrow$	Noise Space	High Frequencies
$\lambda(A^*A)$ large	$\longleftrightarrow$	Signal Space	Low Frequencies

The idea: [T. Huckle and M. Sedlacek, 2012]
The penalty term should measure "how much" the solution f is in the noise space, which depends on A.

Improvement of regularization effects via operator-dependent penalty terms (III)

In [HS12], the key proposal is based on the following operator S

$$S = \left(I - \frac{A^*A}{\|A\|^2} \right),$$

so that $\quad \|f\|_S^2 \approx \begin{cases} large & \text{if } f \text{ is heavily in the noise space of } A \\ small & \text{if } f \text{ is lightly in the noise space of } A \end{cases}$

The linear (semi-)definite operator S is a high pass-filter (please notice that S is NOT a regularization filter, which are all low pass...).
This way, the S-norm is able to measure the "non-regularity" w.r.t. the properties of the actual model-operator A.

In the previous literature, the Tikhonov regularization functional $\Phi_\alpha(f) = \|Af - g\|^2 + \alpha\|f\|_S^2$ is solved, only in Hilbert spaces, by direct methods via Euler-Lagrange normal equations. This way, the direct solver benefits of the very high regularization effects given by $\|f\|_S^2$.

The Landweber method in Banach space for the Tikhonov regularization functional with high-pass filter S

We apply the idea of the high-pass filter S to the iterative Landweber method in Banach spaces for the minimization of the Tikhonov regularization functional $\Phi_\alpha(f) = \|Af - g\|_G^p + \alpha\|f\|_S^p$.
The iteration becomes:

$$f_{k+1} = J_{p^*}^{F^*}\left(J_p^F f_k - \tau_k(A^* J_p^G(Af_k - g) + \alpha S J_p^F f_k)\right),$$

where

$$S = \left(I - \frac{A^* J_p^G A}{\|A\|\|A^*\|}\right).$$

Drawback: The high-regularization effects of the S-norm slow down the convergence of the iterations.

In general, the classical Landweber method is a slow iterative regularization algorithm. The S-norm further reduces the convergence speed. Hence, the S-norm is much more useful for direct solvers rather than iterative ones.

Acceleration of the Landweber method by S-norm

The operator S reduces the components in the signal space and keep the component in the noise space. The operator $-S$ does the opposite.

If we simply change the sign in the related term of the iteration, that is, **from**

$$f_{k+1} = J_{p^*}^{F^*}\Big(J_p^F f_k - \tau_k(A^* J_p^G(Af_k - g) + \alpha S J_p^F f_k)\Big)\,,$$

to

$$f_{k+1} = J_{p^*}^{F^*}\Big(J_p^F f_k - \tau_k(A^* J_p^G(Af_k - g) - \alpha S J_p^F f_k)\Big)\,,$$

then the convergence speed is improved w.r.t. the classical Landweber method.

The action of $-S$ is a *"ir-regularization"* [Di Stefano, 2012], which reduces the (over-smoothing) regularization effects of the iterative method.

The iteration is a descent method for the non-convex functional

$$\tilde{\Phi}_\alpha(f) = \|Af - g\|_G^p - \alpha\|f\|_S^p\,.$$

Minimization of difference of convex functions

The *ir-regularization* functional

$$\tilde{\Phi}_\alpha(f) = \|Af - g\|_G^p - \alpha \|f\|_S^p \, .$$

is not convex. However, it is composed by the difference of two convex functional, $\|Af - g\|_G^p$ and $\|f\|_S^p$.

These kind of functions, called DC -difference of convex- functions (or delta-convex functions), have been exhaustively analyzed since about 1950.

The class of DC functions is a remarkable subclass of locally Lipschitz functions that is of interest both in analysis and optimization.

It is naturally the smallest vector space containing all continuous convex functions on a given set. And it is surprisingly "large"!

If you know the DC decomposition (as in our case), there exist algorithms for the global minimum based on primality-duality.

The vector space of DC (difference of convex) functions

Let DC(X) be the vector space of scalar DC functions on the open convex set $X \subset \mathbb{R}^n$ (Euclidean case). Let

$$h = h_1 - h_2 \in \mathrm{DC}(X)$$

where both h_1 and h_2 are convex functions on X. Obviously h_1 and h_2 can be chosen non-negative.

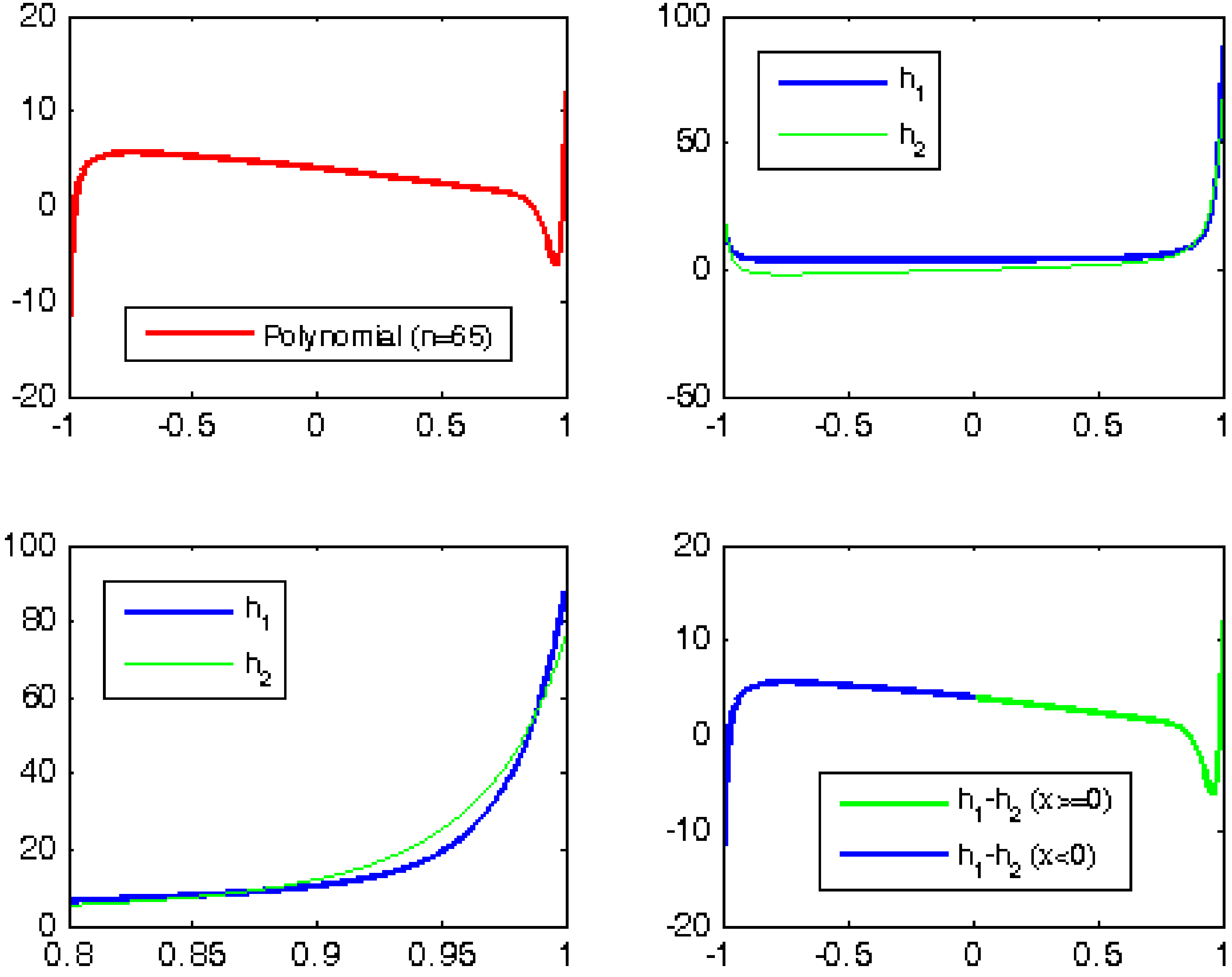
The class of DC functions is a subclass of locally Lipschitz functions.

(i) In the simplest case $n = 1$, $h(x)$ is a DC function if and only if it has left and right derivatives and these derivatives are of bounded variation on every closed bounded interval interior to X (that is, f' is a difference of two nondecreasing functions).

(ii) Thanks to (i), any polynomials on $\mathbb{R}^n$ is DC. Indeed, each polynomial p can be decomposed as $p = q - r$ where r, q are nonnegative convex functions. Easy proof: $x^{2n+1} = (x^+)^{2n+1} - (x^-)^{2n+1}$ and x^{2n} are DC

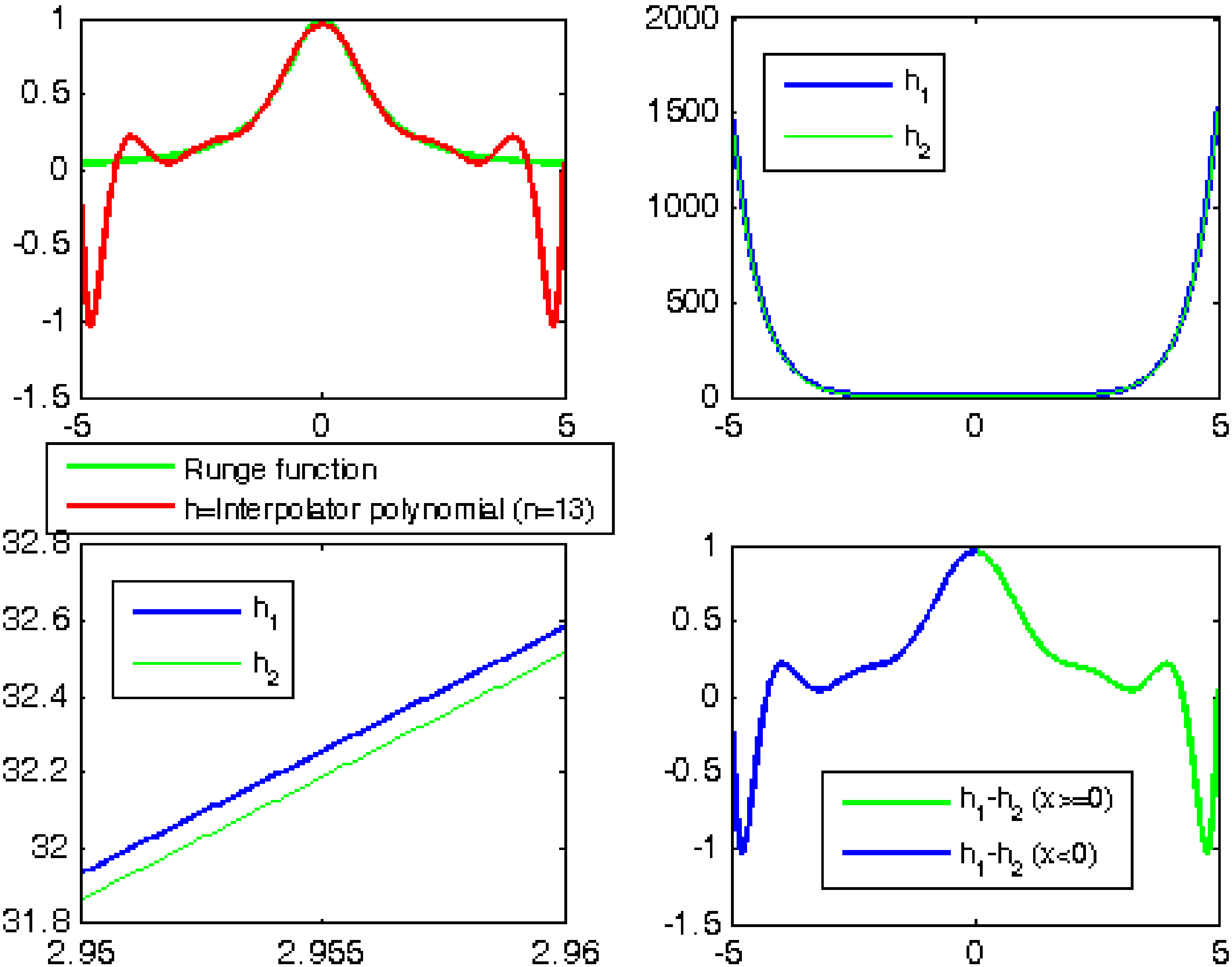
(iii) Thanks to (ii), DC functions are dense uniformly in $C(X)$ for compact X. Any continuous function can be well approximated by DC functions.

DC decomposition of a random polynomial



$$h = h_1 - h_2$$

DC decomp. of the interp. polynomial of Runge function



$$h = h_1 - h_2$$

A sketch on linear algebra: Eigenvalues and DC functions (I)

Let A be a $n \times n$ symmetric matrix with eigenvalues $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$.

(i) The quadratic form

$$R(x) = \frac{1}{2} x^t A x$$

is DC on $\mathbb{R}^n$.

Indeed, there are many decomposition with positive semi-definite A^+ and A^- such that $R(x) = \frac{1}{2} x^t A^+ x - \frac{1}{2} x^t A^- x$.

(ii) The kth-largest eigenvalue function

$$\lambda_k : A \longrightarrow \lambda_k(A)$$

is DC on the space of symmetric matrices.

Proof: $\lambda_k = \sum_{j=1}^{j=k} \lambda_j - \sum_{j=1}^{j=k-1} \lambda_j$, and the sum of the first largest eigenvalues is convex, i.e.,
$\sum_{j=1}^{j=k} \lambda_j (tA_1 + (1-t)A_2) \leq t \sum_{j=1}^{j=k} \lambda_j(A_1) + (1-t) \sum_{j=1}^{j=k} \lambda_j(A_2)$.

Eigenvalues and DC functions (II)

Let A be a $n \times n$ symmetric positive definite matrix with eigenvalues $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n > 0$.
There are some further more involving facts. An example:

From the Rayleigh quotient, we know that the largest eigenvalues is

$$\lambda_1 = \max\{x^t A x \; : \; \|x\| \leq 1\}\,.$$

Via dualization schemes for convex constrains, the optimization theory for DC function shows that

$$\lambda_1 = -\min\{x^t A^{-1} x - 2\|x\| \; : \; x \in \mathbb{R}^n\}\,, \qquad \text{as well as}$$

$$\lambda_1 = -\min\{\|x\|^2 - 2\sqrt{x^t A x} \; : \; x \in \mathbb{R}^n\}\,.$$

Eigenvalues and DC functions (III)

We verify the first one (the second one is similar). Recall that A is a $n \times n$ symmetric positive definite matrix with eigenvalues $\lambda_1 \geq \ldots \geq \lambda_n > 0$.

$$\lambda_1 = -\min\{x^t A^{-1} x - 2\|x\| \, : \, x \in \mathbb{R}^n\}$$

Indeed, let u be fixed, with $\|u\| = 1$, and consider $x = \varrho u$, with $\varrho \geq 0$.

For $x^t A^{-1} x - 2\|x\| = (u^t A^{-1} u)\varrho^2 - 2\varrho := s(\varrho)$, the minimum is attained for $s'(\tilde{\varrho}) = 2(u^t A^{-1} u)\tilde{\varrho} - 2 = 0$, that is, at $\tilde{\varrho} = (u^t A^{-1} u)^{-1}$.

This minimum is:
$s(\tilde{\varrho}) = (u^t A^{-1} u)(u^t A^{-1} u)^{-2} - 2(u^t A^{-1} u)^{-1} = -(u^t A^{-1} u)^{-1}$.

Searching for the minimum of all the minima of $s(\tilde{\varrho})$ (i.e, all over the directions u), we obtain:

$\min\{x^t A^{-1} x - 2\|x\| \, : \, x \in \mathbb{R}^n\} = \min_{\|u\|=1}\{-(u^t A^{-1} u)^{-1}\} =$
$= -\max_{\|u\|=1}\{(u^t A^{-1} u)^{-1}\} = -(\lambda_{\min}(A^{-1}))^{-1} = -\lambda_{\max}(A) = -\lambda_1$.

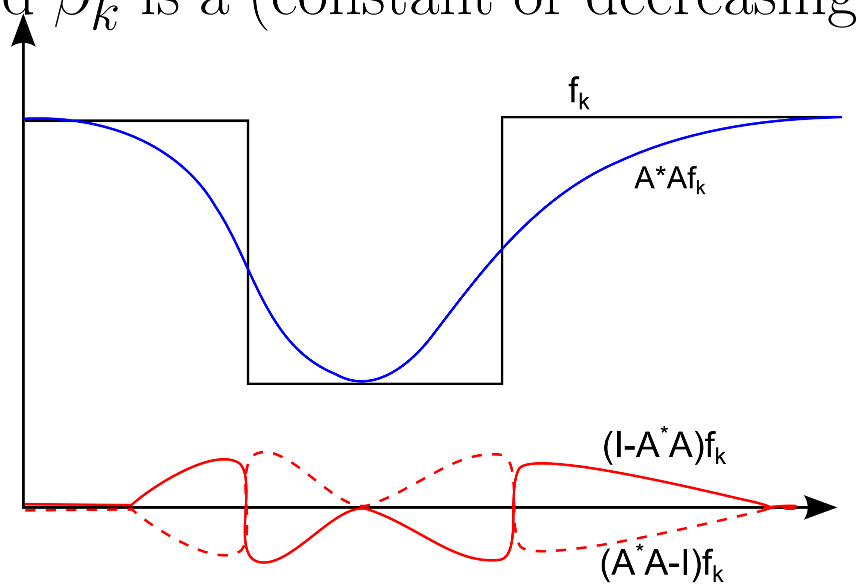
Numerical results (I): The Landweber method in Banach spaces with S-norm acceleration.

We consider the Landweber method in Banach spaces, with ir-regularization penalty term (i.e., minim. of a DC func.). We write the iteration as follows:

$$f_{k+1} = J_{p^*}^{F^*}\left(J_p^F f_k - \tau A^* J_p^G (Af_k - g) + \beta_k S J_p^F f_k\right),$$

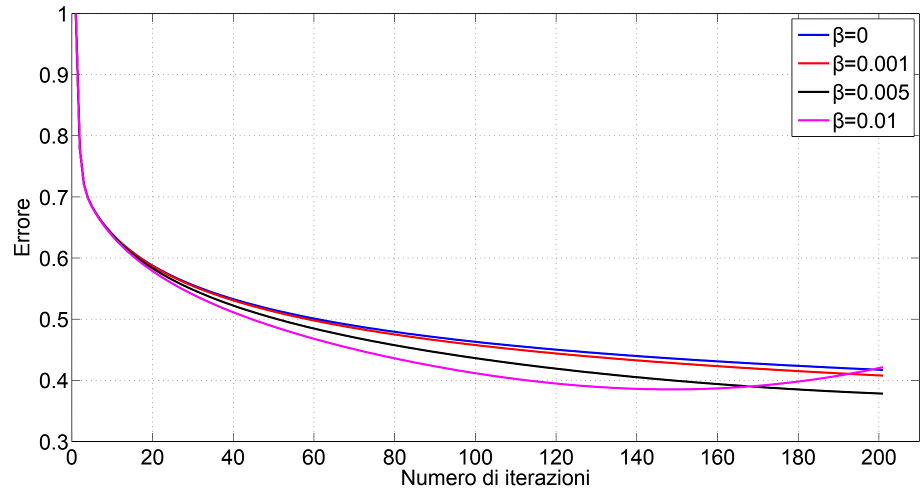
$$S = \left(I - \frac{A^* J_p^G A}{\|A\|\|A^*\|}\right),$$

$\tau = (\|A\|\dot{\|}A^*\|)^{-1}$ and β_k is a (constant or decreasing) sequence of weights.

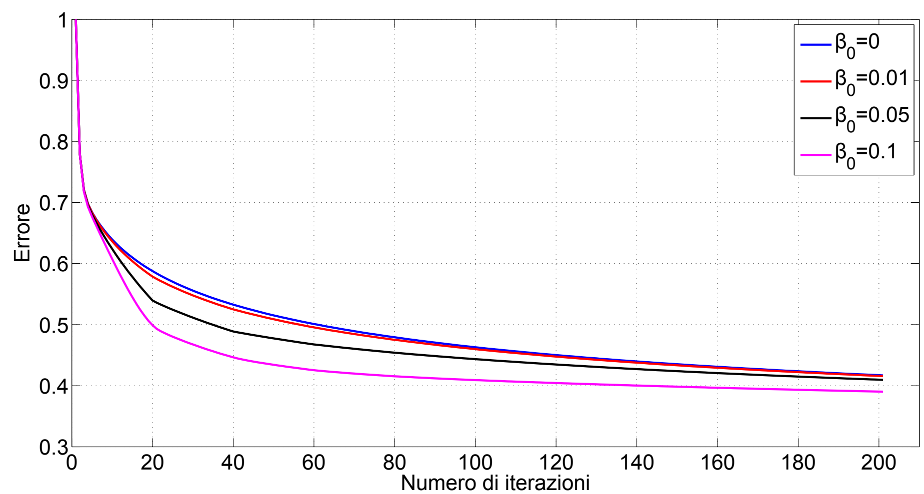


Numerical results (I)

Relative Restoration Errors $RRE(k) = \|f_k - f\|_2 / \|f\|_2$ vs. Iteration Number



$\beta_k = \beta$ (constant sequence); $F = G = L^p$ with $p = 1.5$.



$\beta_k = \frac{1}{5^{\lfloor k/20 \rfloor}} \beta_0$ (decreasing sequence); $F = G = L^p$ with $p = 1.5$.

Numerical results (II): The preconditioned version

Next, we consider the preconditioned version of the method, where the preconditioner D is a regularization preconditioner in the dual space, that is, $D : G^* \longrightarrow G^*$.

The dual-preconditioner D is built by means of (an extension to Banach spaces of) a filtering procedure of the T.Chan preconditioner in Hilbert space.

$$f_{k+1} = J_{p^*}^{F^*}\left(J_p^F f_k - \tau_k D A^* J_p^G(Af_k - g) + \beta_k S J_p^F f_k\right),$$

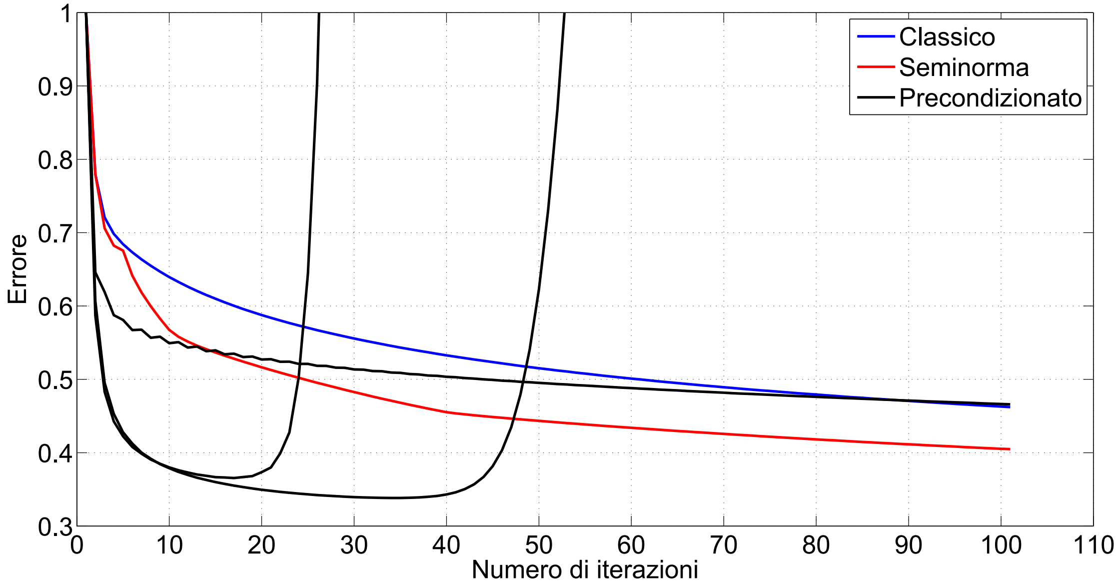
$$S = \left(I - \frac{A^* J_p^G A}{\|A\|\|A^*\|}\right).$$

The preconditioners allow to speed up the convergence, but usually they give rise to instability (that is, to a fast semi-convergence).

The classical preconditioned method is faster than the method with ir-regularization. However, surprisingly enough, the ir-regularization improves the stability of the preconditioned method.

Numerical results (II): preconditioner VS ir-regularization

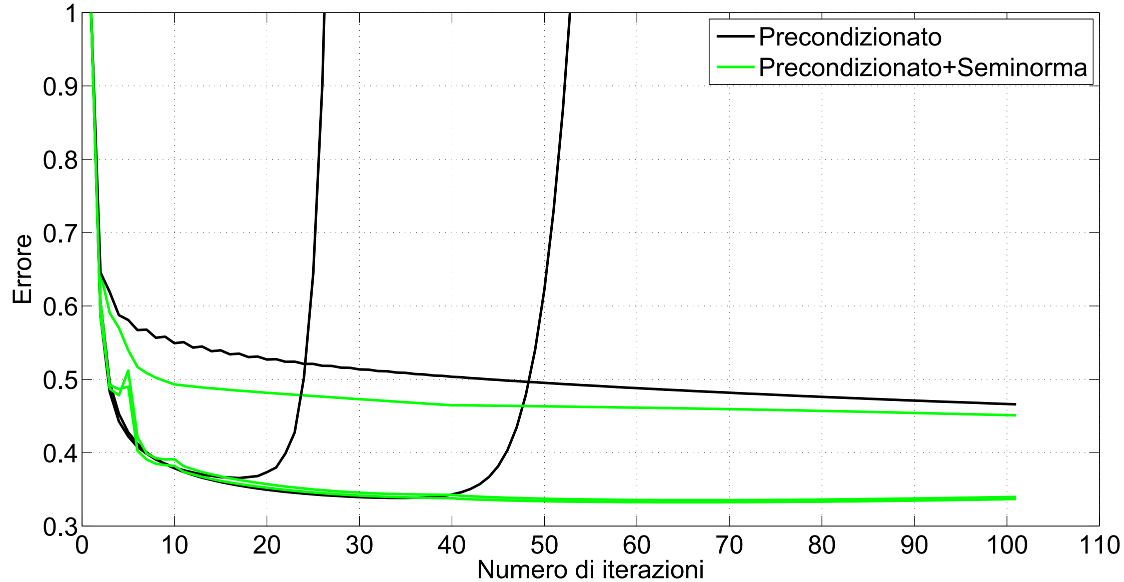
Relative Restoration Errors $RRE(k) = \|f_k - f\|_2 / \|f\|_2$ vs. Iteration Number



ir-regularization at 100-th iteration   preconditioner at 36-th iteration.

Numerical results (II): preconditioner AND ir-regularization

Relative Restoration Errors $RRE(k) = \|f_k - f\|_2 / \|f\|_2$ vs. Iteration Number

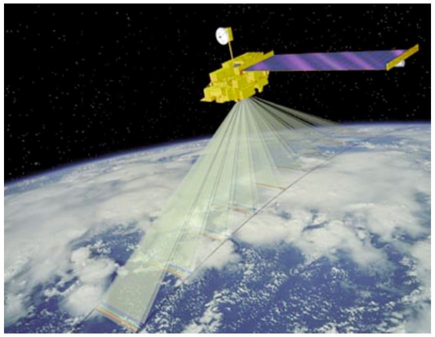


preconditioner, at 36-th iteration

prec. AND ir-regulariz. at 36-th iteration.

Numerical results (III): A geophysical application (Hilbert)

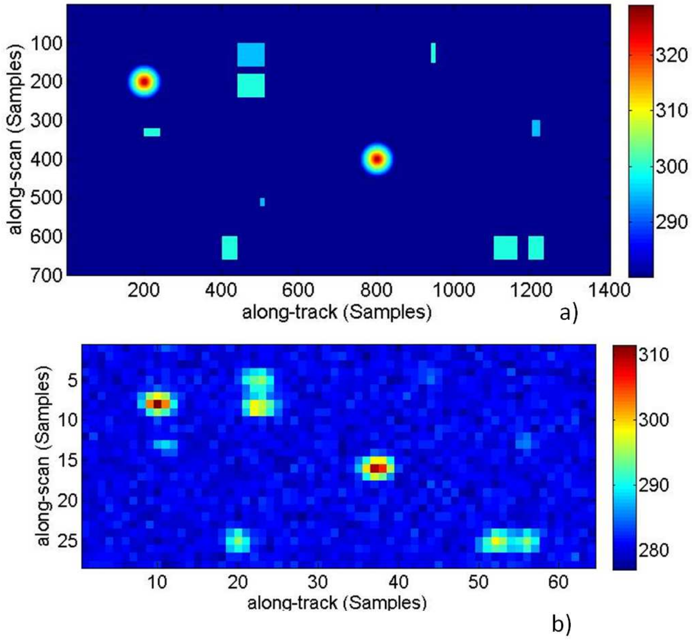
Aim: To enhance the spatial resolution of simulated Special Sensor Microwave/Imager (SSM/I) radiometer measurements (in Hilbert spaces).



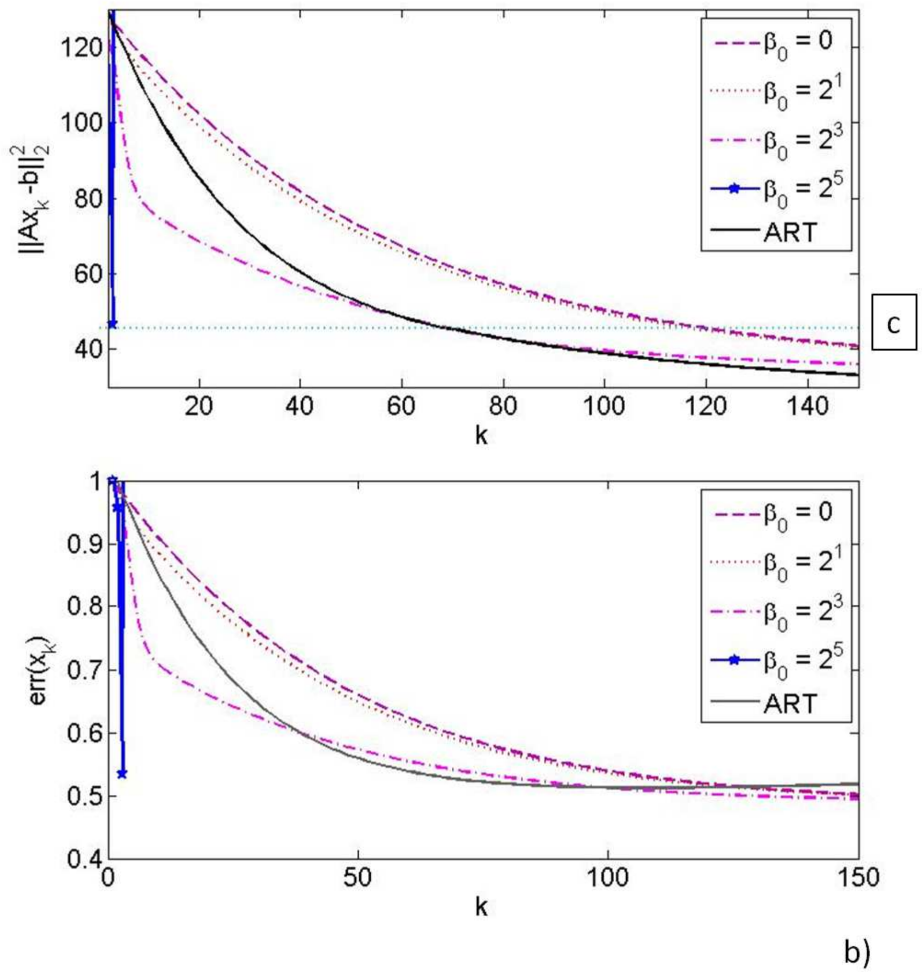
Unknown: brightness temperature on a 1400 700 km Earth's surface

Data: remotely sensed measurements via Fredholm integral operator

Noise: 10% Gaussian, zero mean



Numerical results (III): A geophysical application

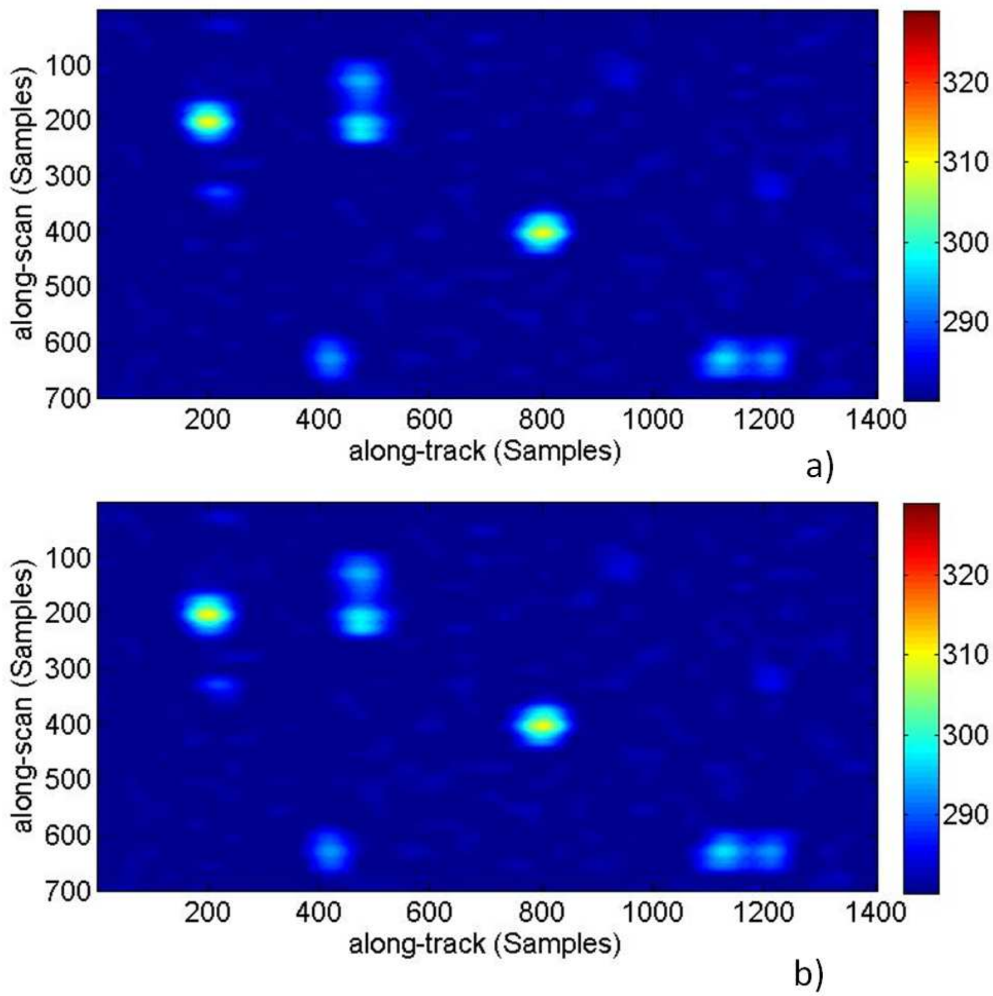


TOP: 2-norm of the residual versus the iteration index k.

BOTTOM: Relative error between the k-reconstructed and the reference field versus k. c is an upper bound related to the 2-norm of the noise.

ART = algebraic reconstruction technique (to compare with). $\beta_k = \beta_0/2^k$.

Numerical results (III): A geophysical application



TOP: Reconstructed field using the conventional Landweber method ($\beta_0 = 0\,,\ \ k = 175$).
BOTTOM: Reconstructed field using the improved Landweber method ($\beta_0 = 2^3\,,\ \ k = 120$).
Relative restoration error: 0.53, both.

Conclusions

- Regularization iterative methods can be accelerated by ir-regularization.

- Theory for global minimization of DC functions via dualization should be applied to obtain new iterative schemes.

- Extension to variable Lebesgue spaces will be analyzed.

Thank you for your attention.

References

[1] T. Huckle, M. Sedlacek,
Tykhonov-Phillips regularization with operator dependent seminorms,
Numer. Alg., vol. 60, pp. 339–353, 2012.

[2] A. Di Stefano,
Metodo di Landweber Modificato in spazi di Banach: accelerazione mediante precondizionamento e seminorme, Master Thesis, Department of Mathematics, University of Genova, 2012.

[3] F. Lenti, F. Nunziata, C. Estatico, M. Migliaccio,
Spatial resolution enhancement of Earth observation products using an acceleration technique for iterative methods, *IEEE Geosc. Rem. Sens. Lett.*, vol. 12, pp. 269-273, 2015.

[4] P. Brianzi, F. Di Benedetto, C. Estatico,
Preconditioned iterative regularization in Banach spaces,
Comp. Optim. Appl., vol. 54, pp. 263–282, 2013.

[5] P. Brianzi, F. Di Benedetto, C. Estatico, L. Surace
Irregularization accelerates iterative regularization,
in submission.

[6] 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.