Compressive algorithms: beyond adaptive wavelet methods in PDE's (II part)

Massimo Fornasier

Johann Radon Institute for Computational and Applied Mathematics
Austrian Academy of Sciences

Wolfgang Pauli Institute
Workshop "Adaptive numerical methods for PDE's"
January 26, 2008

massimo.fornasier@oeaw.ac.at
http://www.ricam.oeaw.ac.at/people/page/fornasier/

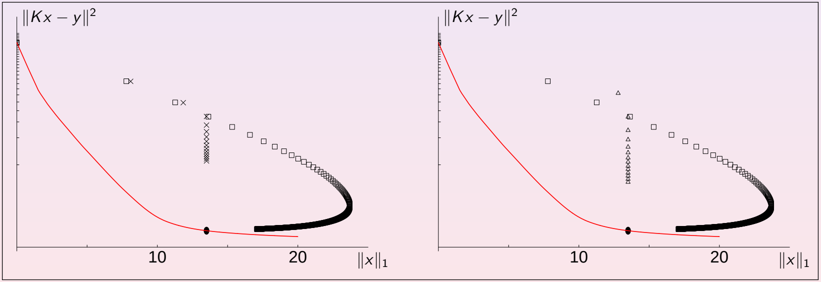
Projected gradient iterations

A first intuitive way to avoid this long "external" detour is to force the successive iterates to remain within the ball B_R. One method to achieve this is to substitute for the thresholding operations the projection $\mathbb{P}_R$ onto B_R. The algorithm

$$x^{(n+1)} = \mathbb{P}_R\left[x^{(n)} + \beta^{(n)}\Phi^*(y - \Phi x^{(n)})\right],$$

does lead, in numerical simulations, to promising, converging results.

Performances



Projected gradient iterations: the descent parameter choice

($\star$) The main issue is to determine how large we can choose the successive $\beta^{(n)}$, and still prove norm convergence.

($\star$) The projected steepest descent algorithm may allow in practice for large steps, "jumping from-face-to-face" of the reference ℓ_1-ball B_R.

($\star$) Typically, within the few initial iterations, the algorithm executes very large jumps to reach quickly the "right face/edge", and then it does not leave it anymore, denoting the capability to adapt the iteration in the vicinity of the solution.

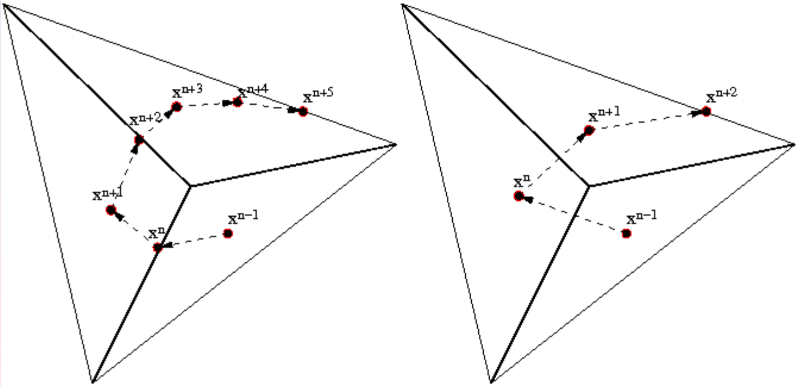
Projected gradient iterations: the descent parameter choice

- (⋆) The main issue is to determine how large we can choose the successive $\beta^{(n)}$, and still prove norm convergence.
- (⋆) The projected steepest descent algorithm may allow in practice for large steps, "jumping from-face-to-face" of the reference ℓ_1-ball B_R.
- (⋆) Typically, within the few initial iterations, the algorithm executes very large jumps to reach quickly the "right face/edge", and then it does not leave it anymore, denoting the capability to adapt the iteration in the vicinity of the solution.

RICAM JOHANN-RADON-INSTITUTE FOR COMPUTATIONAL AND APPLIED MATHEMATICS

Projected gradient iterations: the descent parameter choice

- ($\star$) The main issue is to determine how large we can choose the successive $\beta^{(n)}$, and still prove norm convergence.
- ($\star$) The projected steepest descent algorithm may allow in practice for large steps, “jumping from-face-to-face” of the reference ℓ_1-ball B_R.
- ($\star$) Typically, within the few initial iterations, the algorithm executes very large jumps to reach quickly the “right face/edge”, and then it does not leave it anymore, denoting the capability to adapt the iteration in the vicinity of the solution.


x^{n+3}
x^{n+4}
x^{n+5}
x^{n+2}
x^{n+1}
x^n
x^{n-1}
x^{n+2}
x^{n+1}
x^n
x^{n-1}

The two major ingredients

- $(\star)$ The descent parameter: We say that the sequence $\left(\beta^{(n)}\right)_{n\in\mathbb{N}}$ *satisfies Condition (B) with respect to the sequence* $\left(x^{(n)}\right)_{n\in\mathbb{N}}$ if there exists n_0 so that:

$$\begin{array}{ll} \textit{(B1)} & \bar{\beta} := \sup\{\, \beta^{(n)}\,;\, n \in \mathbb{N}\,\} < \infty \\ & \quad \text{and} \quad \inf\{\, \beta^{(n)}\,;\, n \in \mathbb{N}\,\} \geq 1 \\ \textit{(B2)} & \beta^{(n)}\|\Phi(x^{(n+1)} - x^{(n)})\|^2 \leq r\|x^{(n+1)} - x^{(n)}\|^2 \end{array}$$

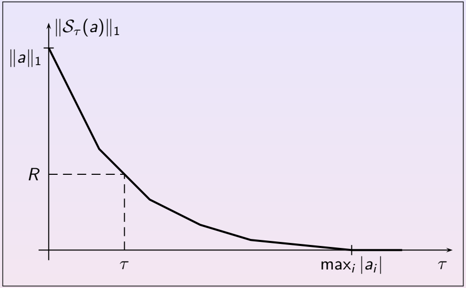
- $(\star)$ The projection $\mathbb{P}_R$ on the ℓ_1-ball B_R is implemented by $\mathbb{P}_R(x) = \mathbb{S}_{\tau(R;x)}(x)$, where $\tau(R;x) > 0$ is such that $\|\mathbb{S}_{\tau(R;x)}(x)\|_1 = R$. The complexity is $\mathcal{O}(m \log m)$.

The two major ingredients

(⋆) The descent parameter: We say that the sequence $\left(\beta^{(n)}\right)_{n\in\mathbb{N}}$ *satisfies Condition (B) with respect to the sequence* $\left(x^{(n)}\right)_{n\in\mathbb{N}}$ if there exists n_0 so that:

$$
\begin{aligned}
&(B1) && \bar{\beta} := \sup\{\beta^{(n)};\ n\in\mathbb{N}\} < \infty \\
& && \text{and} \quad \inf\{\beta^{(n)};\ n\in\mathbb{N}\} \geq 1 \\
&(B2) && \beta^{(n)}\|\Phi(x^{(n+1)}-x^{(n)})\|^2 \leq r\|x^{(n+1)}-x^{(n)}\|^2
\end{aligned}
$$

(⋆) The projection $\mathbb{P}_R$ on the ℓ_1-ball B_R is implemented by $\mathbb{P}_R(x) = \mathbb{S}_{\tau(R;x)}(x)$, where $\tau(R;x) > 0$ is such that $\|\mathbb{S}_{\tau(R;x)}(x)\|_1 = R$. The complexity is $\mathcal{O}(m\log m)$.



For a given vector $a \in \ell_2$, $\|\mathbb{S}_\tau(a)\|_1$ is a piecewise linear continuous and decreasing function of τ (strictly decreasing for $\tau < \max_i |a_i|$) . The knots are located at $\{|a_i|, i : 1 \dots m\}$ and 0. Finding τ such that $\|\mathbb{S}_\tau(a)\|_1 = R$ ultimately comes down to a linear interpolation.

Computation of the projection

In practice and more explicitely, the projection onto the ℓ_1-ball B_R is therefore computed as follows: Order the entries of a by magnitude such that $|a_{i_1}| \geq |a_{i_2}| \geq \ldots \geq |a_{i_m}|$. Let $n \in \{1, \ldots, m\}$ be the largest index satisfying

$$|a_{i_n}| \geq \frac{1}{n-1} \left(\sum_{k=1}^{n-1} |a_{i_k}| - R \right).$$

Then

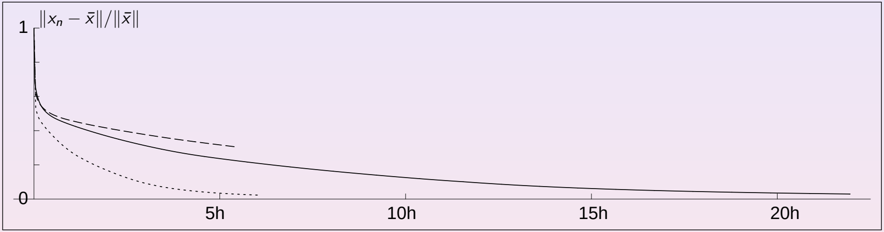
$$(\mathbb{P}_R(a))_{i_j} = a_{i_j} - \frac{\mathrm{sgn}(a_{i_j})}{n} \left(\sum_{k=1}^{n} |a_{i_k}| - R \right), \quad j = 1, \ldots, n,$$

$$(\mathbb{P}_R(a))_{i_j} = 0, \quad j = n+1, \ldots, m.$$

Theorem (Daubechies, F., Loris)

The sequence $\left(x^{(n)}\right)_{n\in\mathbb{N}}$, *where the* step-length sequence $\left(\beta^{(n)}\right)_{n\in\mathbb{N}}$ *satisfies Condition (B) with respect to the* $x^{(n)}$, *converges in norm to a minimizer of* $\mathcal{D}(x) = \|\Phi x - y\|_2^2$ *on* B_R.

Some results in a geophysics application



Convergence rate of the thresholded Landweber algorithm (solid line), the projected Landweber algorithm (dashed line) and the projected steepest descent algorithm (dotted line). The projected steepest descent algorithm converges about four times faster than the thresholded Landweber iteration.

A subspace correction method

- We consider the minimization of $\mathcal{J} = \mathcal{J}_\tau$, by alternating subspace corrections.
- We start by decomposing the "domain" of the sequences Λ into two disjoint sets Λ_1, Λ_2 so that $\Lambda = \Lambda_1 \cup \Lambda_2$.
- Associated to a decomposition $\mathcal{C} = \{\Lambda_1, \Lambda_2\}$ we define the *extension operators* $E_i : \ell_2(\Lambda_i) \to \ell_2(\Lambda)$, $(E_i v)_\lambda = v_\lambda$, if $\lambda \in \Lambda_i$, $(E_i v)_\lambda = 0$, otherwise, $i = 1, 2$. The adjoint operator, which we call the *restriction operator*, is denoted by $R_i := E_i^*$.

With these operators we define the functional $\mathcal{J}(x_1, x_2)$, $\mathcal{J} : \ell_2(\Lambda_1) \times \ell_2(\Lambda_2) \to \mathbb{R}$, given by

$$\mathcal{J}(x_1, x_2) := \mathcal{J}(E_1 x_1 + E_2 x_2).$$

In analogy to the Schwartz multiplicative algorithm, we analyze the following algorithm:

$$\begin{cases} x_1^{(n+1)} = \operatorname{argmin}_{v_1 \in \ell_2(\Lambda_1)} \mathcal{J}(v_1, x_2^{(n)}) \\ x_2^{(n+1)} = \operatorname{argmin}_{v_2 \in \ell_2(\Lambda_2)} \mathcal{J}(x_1^{(n+1)}, v_2) \\ x^{(n+1)} := E_1 x_1^{(n+1)} + E_2 x_2^{(n+1)}. \end{cases}$$

Let us observe that
$\|E_1 x_1 + E_2 x_2\|_{\ell_1(\Lambda)} := \|x_1\|_{\ell_1(\Lambda_1)} + \|x_2\|_{\ell_1(\Lambda_2)}$, hence

$$
\begin{aligned}
& \operatorname{argmin}_{v_1 \in \ell_2(\Lambda_1)} \mathcal{J}(v_1, x_2^{(n)}) \\
= \; & \operatorname{argmin}_{v_1 \in \ell_2(\Lambda_1)} \|(y - \Phi E_2 x_2^{(n)}) - \Phi E_1 v_1\|_2^2 + \tau \|v_1\|_1 .
\end{aligned}
$$

A similar formulation holds for $\operatorname{argmin}_{v_2 \in \ell_2(\Lambda_2)} \mathcal{J}(x_1^{(n+1)}, v_2)$. This means that the solution of the local problems on Λ_i is of the *same* kind as the original problem $\operatorname{argmin}_{x \in \ell_2(\Lambda)} \mathcal{J}(x)$, but the dimension for each has been reduced.

A sequential algorithm

This leads to the following sequential algorithm

$$
\begin{cases}
\begin{cases}
x_1^{(n+1,0)} = x_1^{(n,L)} \\
x_1^{(n+1,\ell+1)} = \mathbb{S}_\tau \left(x_1^{(n+1,\ell)} + R_1 \Phi^* ((y - \Phi E_2 x_2^{(n,M)}) - \Phi E_1 x_1^{(n+1,\ell)}) \right) \\
\ell = 0, \ldots, L-1
\end{cases} \\
\begin{cases}
x_2^{(n+1,0)} = x_2^{(n,M)} \\
x_2^{(n+1,\ell+1)} = \mathbb{S}_\tau \left(x_2^{(n+1,\ell)} + R_2 \Phi^* ((y - \Phi E_1 x_1^{(n+1,L)}) - \Phi E_2 x_2^{(n+1,\ell)}) \right) \\
\ell = 0, \ldots, M-1
\end{cases} \\
x^{(n+1)} := E_1 x_1^{(n+1,L)} + E_2 x_2^{(n+1,M)}.
\end{cases}
$$

This leads to the following sequential algorithm

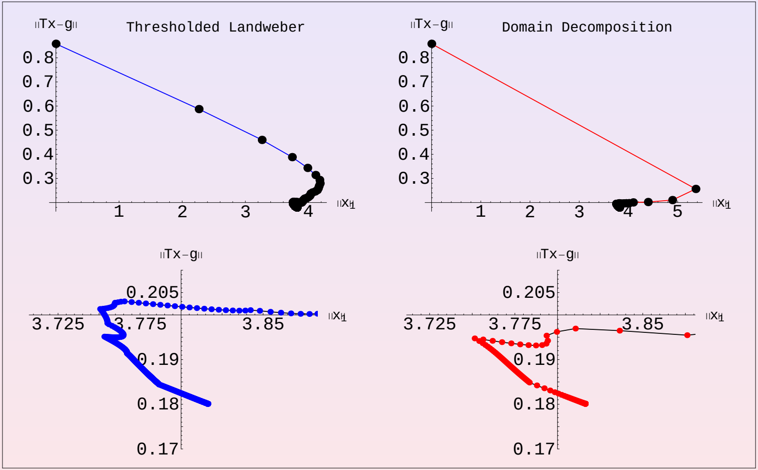
$$\left\{\begin{array}{l}\left\{\begin{array}{ll} x_1^{(n+1,0)} = x_1^{(n,L)} \\ x_1^{(n+1,\ell+1)} = \mathbb{S}_\tau\left(x_1^{(n+1,\ell)} + R_1\Phi^*((y - \Phi E_2 x_2^{(n,M)}) - \Phi E_1 x_1^{(n+1,\ell)})\right) \\ \ell = 0,\dots,L-1 \end{array}\right. \\ \left\{\begin{array}{ll} x_2^{(n+1,0)} = x_2^{(n,M)} \\ x_2^{(n+1,\ell+1)} = \mathbb{S}_\tau\left(x_2^{(n+1,\ell)} + R_2\Phi^*((y - \Phi E_1 x_1^{(n+1,L)}) - \Phi E_2 x_2^{(n+1,\ell)})\right) \\ \ell = 0,\dots,M-1 \end{array}\right. \\ x^{(n+1)} := E_1 x_1^{(n+1,L)} + E_2 x_2^{(n+1,M)}. \end{array}\right.$$

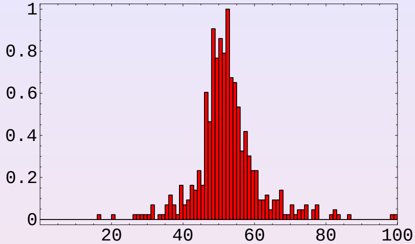
This leads to the following parallel algorithm

$$\begin{cases} \begin{cases} x_1^{(n+1,0)} = x_1^{(n,L)} \\ x_1^{(n+1,\ell+1)} = \mathbb{S}_\tau\left(x_1^{(n+1,\ell)} + R_1\Phi^*((y - \Phi E_2 R_2 x^{(n)}) - \Phi E_1 x_1^{(n+1,\ell)})\right) \\ \ell = 0,\dots,L-1 \end{cases} \\ \begin{cases} x_2^{(n+1,0)} = x_2^{(n,M)} \\ x_2^{(n+1,\ell+1)} = \mathbb{S}_\tau\left(x_2^{(n+1,\ell)} + R_2\Phi^*((y - \Phi E_1 R_1 x^{(n)}) - \Phi E_2 x_2^{(n+1,\ell)})\right) \\ \ell = 0,\dots,M-1 \end{cases} \\ x^{(n+1)} := \frac{E_1 x_1^{(n+1,L)} + E_2 x_2^{(n+1,M)} + x^{(n)}}{2}. \end{cases}$$

Theorem (Fornasier)

The (sequential and parallel) subspace correction algorithms produce a sequence $(x^{(n)})_{n\in\mathbb{N}}$ in $\ell_2(\Lambda)$ whose strong accumulation points are minimizers of the functional $\mathcal{J}$. In particular, the set of strong accumulation points is non-empty. If the minimizer is unique then the whole sequence $(x^{(n)})_{n\in\mathbb{N}}$ converges to it.


Thresholded Landweber
Domain Decomposition
‖Tx−g‖
‖x‖1
0.8
0.7
0.6
0.5
0.4
0.3
1
2
3
4
5
0.205
0.19
0.18
0.17
3.725
3.775
3.85

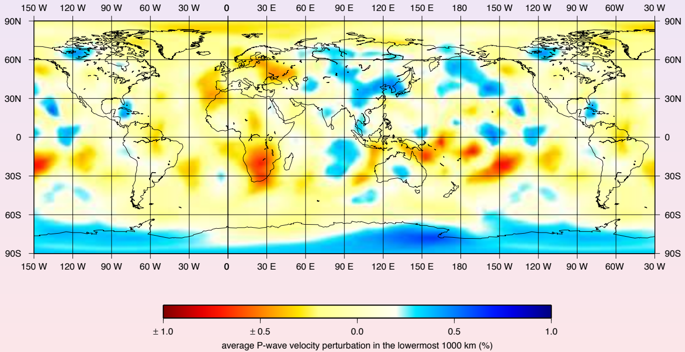


We assume $\mathcal{K} = \mathbb{R}^{40}$ and $\mathcal{H} = \mathbb{R}^{10}$, Φ is a 40×10 (scaled) random matrix with $\|\Phi\| < 1$, and $y \in \mathbb{R}^{10}$ is a random vector. We fix the regularization parameter $\tau = 0.1$. The figure shows the normalized frequency for multiple random trials of the percentage ratio between the number of operations required by the sequential domain decomposition method in order to achieve an accuracy of 10^{-15} and the one required by the thresholded Landweber iteration. Here $L = M = 8$.

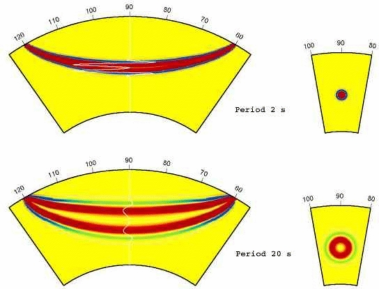
Some interesting applications

- Earth global tomography/Terrestrial seismic tomography;
- Magnetoencephalography (MEG);
- ... many more ...

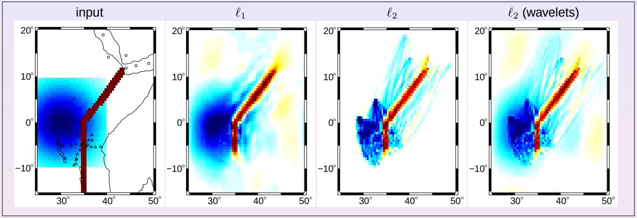
Global terrestrial seismic tomography (Dahlen, Nolet, Montelli)



Banana-Doughnut sensitivity kernels by Tony Dahlen

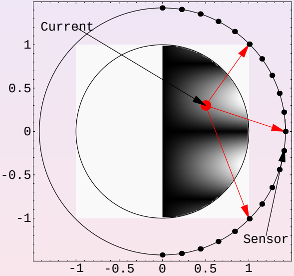


Global terrestrial seismic tomography via sparsity



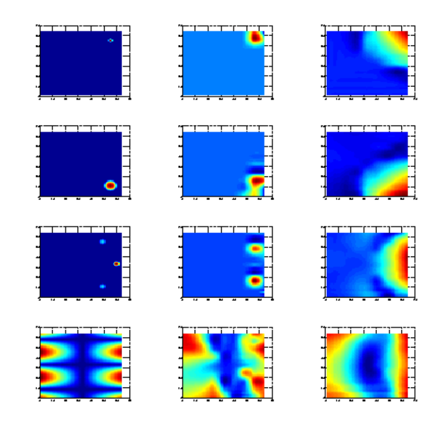
Toy 2D velocity model for the East African rift and adjacent continental craton, showing the seismic stations (triangles) and earthquake events (circles); reconstructed model using iterative thresholding vs Tikhonov reg. (joint work with F. A. Dahlen, I. Daubechies, I. Loris, and G. Nolet.)

Magnetoencephalography



Left: Machine for magnetoencephalography. **Right**: 2D model. Sensors are distributed on a semicircle. (joint work with F. Pitolli.)

Current density reconstruction



Conclusions

(1) Combinatorial problems can be approached effectively by ℓ_1-minimization;

(2) Fast algorithms are currently developed;

(3) Subspace corrections/domain decompositions help to further reduce the dimensionality and to speed-up computations;

(3) Many open problems to be understood concerning rates of convergence and complexity of these algorithms.

RICAM JOHANN-RADON-INSTITUTE FOR COMPUTATIONAL AND APPLIED MATHEMATICS

Conclusions

(1) Combinatorial problems can be approached effectively by ℓ_1-minimization;

(2) Fast algorithms are currently developed;

(3) Subspace corrections/domain decompositions help to further reduce the dimensionality and to speed-up computations;

(3) Many open problems to be understood concerning rates of convergence and complexity of these algorithms.

Conclusions

(1) Combinatorial problems can be approached effectively by ℓ_1-minimization;

(2) Fast algorithms are currently developed;

(3) Subspace corrections/domain decompositions help to further reduce the dimensionality and to speed-up computations;

(3) Many open problems to be understood concerning rates of convergence and complexity of these algorithms.

Conclusions

(1) Combinatorial problems can be approached effectively by ℓ_1-minimization;

(2) Fast algorithms are currently developed;

(3) Subspace corrections/domain decompositions help to further reduce the dimensionality and to speed-up computations;

(3) Many open problems to be understood concerning rates of convergence and complexity of these algorithms.

For Further Reading

- S. Dahlke, M. Fornasier, and T. Raasch, Adaptive frame methods for elliptic operator equations, Adv. Comput. Math., Vol. 27 No. 1, 2007, 27-63.
- S. Dahlke, M. Fornasier, T. Raasch, M. Primbs, and M. Werner, Nonlinear and adaptive frame approximation schemes for elliptic PDEs: theory and numerical experiments, Bericht Nr. 2007-7, Philipps-Universität Marburg, 2007.
- S. Dahlke, M. Fornasier, T. Raasch, R. Stevenson, and M. Werner, Adaptive frame methods for elliptic operator equations: the steepest descent approach, IMA J. Numer. Anal., Vol. 27 No. 4, 2007, 717-740.

- I. Daubechies, M. Fornasier, I. and Loris, Accelerated Projected Gradient Method for Linear Inverse Problems with Sparsity Constraints, to appear in J. Fourier Anal. Appl.
- M. Charina, C. Conti, and M. Fornasier, Adaptive frame methods for nonlinear variational problems, to appear in Numerische Mathematik.
- M. Fornasier, Domain decomposition methods for linear inverse problems with sparsity constraints, Inverse Problems, Vol. 23 No. 6, 2007, 2505â2526.

- M. Fornasier and H. Rauhut, Recovery algorithms for vector valued data with joint sparsity constraints, to appear in SIAM J. Numer. Anal.
- M. Fornasier and H. Rauhut, Iterative thresholding algorithms, to appear in Appl. Comput. Harmon. Anal.
- M. Fornasier and C.-B. Schönlieb, Subspace correction methods for total variation and ℓ_1-minimization, math.NA.arXiv:0712.2258.