Compressive algorithms: beyond adaptive wavelet methods in PDE's

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/

Outline

1. Compressive algorithms
   - Introduction
   - Things to be understood
2. Frames in Hilbert spaces: towards the sparsity concept
   - Definitions and main properties
   - Gabor and wavelet frames and audio signals
3. Adaptive frame methods for PDEs
   - Frame discretization of elliptic equations
   - Optimal algorithms
   - Conclusions
4. Sparse recovery: beyond wavelet approximation
   - Compression and ℓ_1-minimization
   - Compressed sensing and other applications
   - ℓ_1-minimization: re-weighted least square method
   - ℓ_1-minimization with noisy data

Outline

1. Compressive algorithms
   - Introduction
   - Things to be understood
2. Frames in Hilbert spaces: towards the sparsity concept
   - Definitions and main properties
   - Gabor and wavelet frames and audio signals
3. Adaptive frame methods for PDEs
   - Frame discretization of elliptic equations
   - Optimal algorithms
   - Conclusions
4. Sparse recovery: beyond wavelet approximation
   - Compression and ℓ_1-minimization
   - Compressed sensing and other applications
   - ℓ_1-minimization: re-weighted least square method
   - ℓ_1-minimization with noisy data

Outline

1. Compressive algorithms
   - Introduction
   - Things to be understood
2. Frames in Hilbert spaces: towards the sparsity concept
   - Definitions and main properties
   - Gabor and wavelet frames and audio signals
3. Adaptive frame methods for PDEs
   - Frame discretization of elliptic equations
   - Optimal algorithms
   - Conclusions
4. Sparse recovery: beyond wavelet approximation
   - Compression and ℓ_1-minimization
   - Compressed sensing and other applications
   - ℓ_1-minimization: re-weighted least square method
   - ℓ_1-minimization with noisy data

Outline

1. Compressive algorithms
   - Introduction
   - Things to be understood
2. Frames in Hilbert spaces: towards the sparsity concept
   - Definitions and main properties
   - Gabor and wavelet frames and audio signals
3. Adaptive frame methods for PDEs
   - Frame discretization of elliptic equations
   - Optimal algorithms
   - Conclusions
4. Sparse recovery: beyond wavelet approximation
   - Compression and ℓ_1-minimization
   - Compressed sensing and other applications
   - ℓ_1-minimization: re-weighted least square method
   - ℓ_1-minimization with noisy data

Outline

1. Compressive algorithms
   - Introduction
   - Things to be understood
2. Frames in Hilbert spaces: towards the sparsity concept
   - Definitions and main properties
   - Gabor and wavelet frames and audio signals
3. Adaptive frame methods for PDEs
   - Frame discretization of elliptic equations
   - Optimal algorithms
   - Conclusions
4. Sparse recovery: beyond wavelet approximation
   - Compression and ℓ_1-minimization
   - Compressed sensing and other applications
   - ℓ_1-minimization: re-weighted least square method
   - ℓ_1-minimization with noisy data

Compressive algorithm

Compressive Algorithms (CA) are an approach to efficient adaptive computing that take advantage of the property of solutions of certain PDE's and variational problems to be characterized by few major features, which are recovered by adaptive nonlinear iterations.

The approach to efficient computing via CA responds to the need of addressing large scale problems. CA tend to use the minimal number of degrees of freedom, and are very successfully applied in several problems. Their analysis is challenging.

Outline

1. Compressive algorithms
   - Introduction
   - Things to be understood
2. Frames in Hilbert spaces: towards the sparsity concept
   - Definitions and main properties
   - Gabor and wavelet frames and audio signals
3. Adaptive frame methods for PDEs
   - Frame discretization of elliptic equations
   - Optimal algorithms
   - Conclusions
4. Sparse recovery: beyond wavelet approximation
   - Compression and ℓ_1-minimization
   - Compressed sensing and other applications
   - ℓ_1-minimization: re-weighted least square method
   - ℓ_1-minimization with noisy data

(1) Let us stress from now that the "**relevant features**" might not be merely, e.g., large *wavelet coefficients*, but they can be expressed in terms of more sophisticated representations of the solution.

(2) To start, we will use *redundant frame expansions* (frames) as a prototype of compressible representation.

(3) However, we may consider also solutions of singular PDE's with discontinuities along curves, and these will be the interesting features to be recovered during the solution process.

RICAM JOHANN · RADON · INSTITUTE FOR COMPUTATIONAL AND APPLIED MATHEMATICS

(1) Let us stress from now that the "**relevant features**" might not be merely, e.g., large *wavelet coefficients*, but they can be expressed in terms of more sophisticated representations of the solution.

(2) To start, we will use *redundant frame expansions* (frames) as a prototype of compressible representation.

(3) However, we may consider also solutions of singular PDE's with discontinuities along curves, and these will be the interesting features to be recovered during the solution process.

(1) Let us stress from now that the "**relevant features**" might not be merely, e.g., large *wavelet coefficients*, but they can be expressed in terms of more sophisticated representations of the solution.

(2) To start, we will use *redundant frame expansions* (frames) as a prototype of compressible representation.

(3) However, we may consider also solutions of singular PDE's with discontinuities along curves, and these will be the interesting features to be recovered during the solution process.

Outline

1. Compressive algorithms
   - Introduction
   - Things to be understood
2. Frames in Hilbert spaces: towards the sparsity concept
   - Definitions and main properties
   - Gabor and wavelet frames and audio signals
3. Adaptive frame methods for PDEs
   - Frame discretization of elliptic equations
   - Optimal algorithms
   - Conclusions
4. Sparse recovery: beyond wavelet approximation
   - Compression and ℓ_1-minimization
   - Compressed sensing and other applications
   - ℓ_1-minimization: re-weighted least square method
   - ℓ_1-minimization with noisy data

Frames

Informally, a *frame* is a collection of “linear dependent” vectors $\mathcal{F} = \{f_n : n = 1, 2, 3, \ldots\}$. Hence,

$$f = \sum_n c_n f_n,$$

where the coefficients c_n are NOT unique.
A frame codify a signal f in a **redundant way**.

Frames in Hilbert spaces

Definition

A countable subset $\mathcal{F} = \{f_n : n \in \mathbb{N}\}$ of a separable Hilbert space $\mathcal{H}$ is a frame for $\mathcal{H}$ if there exists constants $A, B > 0$ such that the following pseudo-Parseval formula holds:

$$A\|f\|_{\mathcal{H}}^2 \leq \sum_{n \in \mathbb{N}} |\langle f, f_n \rangle|^2 \leq B\|f\|_{\mathcal{H}}^2,$$

for all $f \in \mathcal{H}$.

Operators of analysis and synthesis

Two operators are associated to a frame $\mathcal{F} = \{f_n : n \in \mathbb{N}\}$:

- **Analysis operator**: $F : \mathcal{H} \to \ell_2(\mathbb{N}) : f \mapsto \langle f, f_n \rangle$;
- **Synthesis operator**: $F^* : \ell_2(\mathbb{N}) \to \mathcal{H} : c \mapsto \sum_{n \in \mathbb{N}} c_n f_n$.

We define the so-called **frame operator** S by

$$S = F^* F, \quad Sf = \sum_n \langle f, f_n \rangle f_n.$$

The frame operator

The operator S is self-adjoint, positive, and boundedly invertible; We have the following reproducing formulas

$$f = S^{-1}Sf = \sum_n \langle f, f_n \rangle S^{-1} f_n = SS^{-1} f = \sum_n \langle f, S^{-1} f_n \rangle f_n.$$

The set $\tilde{\mathcal{F}} := \{\tilde{f}_n := S^{-1} f_n : n \in \mathbb{N}\}$ is again a frame, called **the canonical dual frame.**

The canonical dual

The canonical dual determines coefficients $c_n(f) = \langle f, \tilde{f}_n \rangle$ for the synthesis of f. However, since $\ker F^* \neq \{0\}$ in general, there exists an infinite collection of different coefficients $(d_n(f))_{n\in\mathbb{N}}$ such that $f = \sum_n c_n(f) f_n = \sum_n d_n(f) f_n, \quad c \neq d$.

Proposition

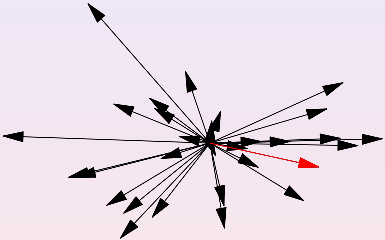
If $f = \sum_n d_n f_n$ for some scalars $(d_n)_n$, then

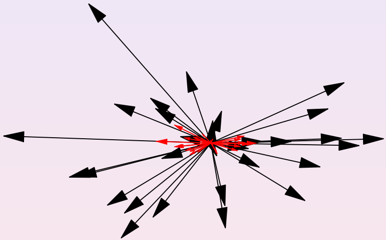
$$\sum_n |d_n|^2 = \sum_n |\langle f, S^{-1} f_n \rangle|^2 + \sum_n |\langle f, S^{-1} f_n \rangle - d_n|^2.$$

In particular, the sequence $(\langle f, S^{-1} f_n \rangle)_n$ has the minimal energy, i,e., $\ell_2 - norm$, among all such sequences.

The canonical dual

Frame elements, and one in red

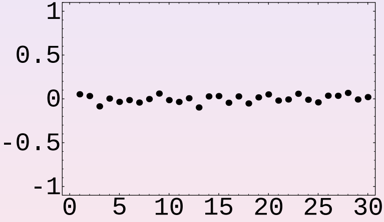




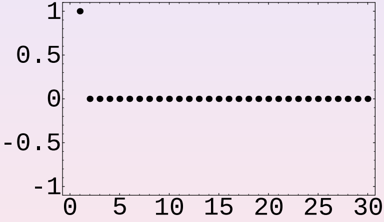
Frame and its canonical dual

The canonical dual

The canonical dual has not all the possible virtues.



Canonical dual coefficients



Sparsest coefficients

Compression and robustness

Since the coefficients are not unique, two fundamental positive features of frames are

- frames do allow for sparser representations (i.e., few coefficients are nonzero) of an element $f \in \mathcal{H}$; The slogan is: *"The larger is my dictionary, the shorter will be the phrases I can compose by using proper terminology"*;
- frames do improve robustness not only under perturbations

$$f = \sum_n c_n f_n \approx \sum_n \tilde{c}_n f_n, \quad c_n \approx \tilde{c}_n,$$

also under erasures, i.e., the loss of some coefficients, see the concept of "democratic expansion" (or Kashin's representation) introduced first by Calderbank-Daubechies.

Outline

1. Compressive algorithms
   - Introduction
   - Things to be understood
2. Frames in Hilbert spaces: towards the sparsity concept
   - Definitions and main properties
   - Gabor and wavelet frames and audio signals
3. Adaptive frame methods for PDEs
   - Frame discretization of elliptic equations
   - Optimal algorithms
   - Conclusions
4. Sparse recovery: beyond wavelet approximation
   - Compression and ℓ_1-minimization
   - Compressed sensing and other applications
   - ℓ_1-minimization: re-weighted least square method
   - ℓ_1-minimization with noisy data

Nontrivial examples

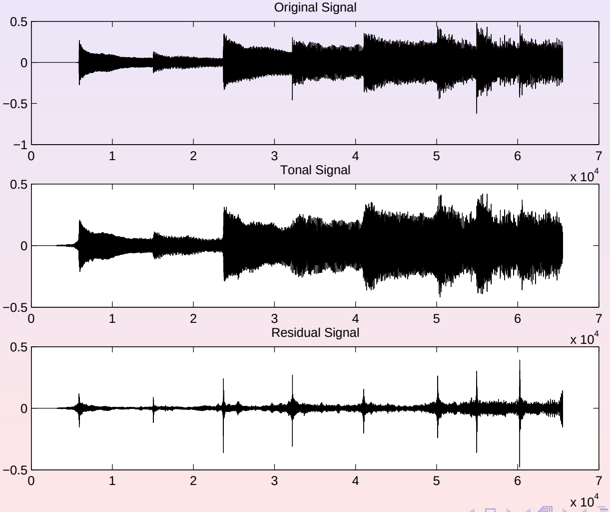
The *modulation*, *translation*, and *dilation* operators are resp.:

$$M_\omega f(t) = e^{2\pi i \omega t} f(t), \quad T_x f(t) = f(t-x), \quad D_a f(t) = |a|^{-\frac{d}{2}} f\left(\frac{t}{a}\right),$$

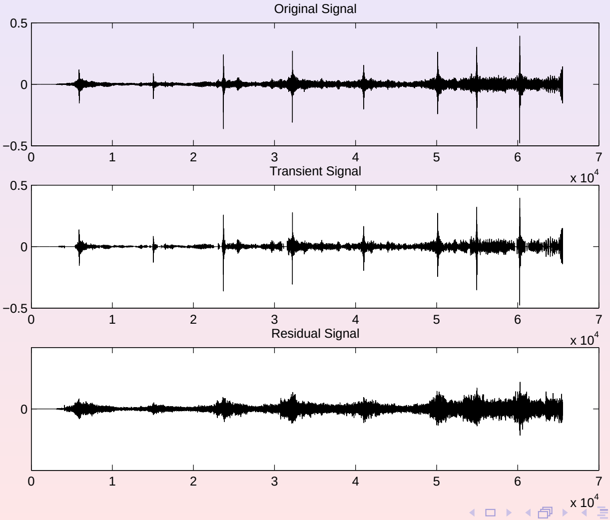
for $x, \omega \in \mathbb{R}^d$, and $a \in \mathbb{R}$, here d is the dimension.

- For $\Lambda \subset \mathbb{R}$, a family $\{T_x f : x \in \Lambda\}$ is a frame for its span if and only if is a Riesz basis;
- For $\Lambda \subset \mathbb{R} \times \mathbb{R}$, if a family of functions $\{M_\omega T_x f : (x, \omega) \in \Lambda\}$ is a frame, then it is called a *Gabor or Weyl-Heisenberg frame*;
- For $\mathcal{J} \subset \mathbb{R} \times \mathbb{R}$, if a family of functions $\{D_a T_x f : (x, a) \in \mathcal{J}\}$ is a frame, then it is called a *wavelet frame*.

Practical meaning of Gabor and wavelet frames



Practical meaning of Gabor and wavelet frames



The analysis of Gabor and wavelet frames

- *Few elements* of a Gabor frame can approximate very well functions constituted by strong local harmonic parts;
- *Few elements* of a wavelet frame/basis can approximate very well functions characterized by, e.g., strong transitions, discontinuities, impulsive parts.

Hence a frame constituted by the *hybrid* UNION of Gabor and wavelet frames is a suitable tool in order to COMPRESS, e.g., audio signals.

Outline

1. Compressive algorithms
   - Introduction
   - Things to be understood
2. Frames in Hilbert spaces: towards the sparsity concept
   - Definitions and main properties
   - Gabor and wavelet frames and audio signals
3. Adaptive frame methods for PDEs
   - Frame discretization of elliptic equations
   - Optimal algorithms
   - Conclusions
4. Sparse recovery: beyond wavelet approximation
   - Compression and ℓ_1-minimization
   - Compressed sensing and other applications
   - ℓ_1-minimization: re-weighted least square method
   - ℓ_1-minimization with noisy data

Elliptic equations: theoretical setting

- $\Omega \subset \mathbb{R}^d$ Lipschitz domain
- H Hilbert space, $H \subset L_2(\Omega) \subset H'$ Gelfand triple
- $\mathcal{L} : H \to H'$ linear, boundedly invertible

$$a(\cdot, \cdot) := \langle \mathcal{L}\cdot, \cdot \rangle : H \times H \to \mathbb{R}$$

  symmetric, continuous, H-elliptic: $a(v, v) \eqsim \|v\|_H^2$.
- task: for $f \in H'$, solve

$$\mathcal{L}u = f,$$

  i.e., find $u \in H$ s.t.

$$a(u, v) = \langle f, v \rangle, \quad v \in H.$$

Discretization via bases

Cohen, Dahmen, DeVore (1998):

- I choose wavelet basis $\Psi = \{\psi_\lambda\}_{\lambda \in \mathcal{J}} \subset L_2(\Omega)$ with

$$\|\mathbf{c}\|_{\ell_2} \approx \|\mathbf{c}^T D^{-1} \Psi\|_H.$$

- This implies

$$\mathcal{L}u = f \Rightarrow \mathbf{L}\mathbf{u} = \mathbf{f},$$

  where

$$\begin{aligned} \mathbf{L} &:= D^{-1} \langle \mathcal{L}\Psi, \Psi \rangle D^{-1}, \\ \mathbf{f} &:= D^{-1} \langle f, \Psi \rangle. \end{aligned}$$

- problem: construction of wavelet bases on general domains difficult/complicated!

Discretization via frames

Dahlke, F., Raasch, Stevenson (2003-2006)

- choose a frame $\Psi = \{\psi_\lambda\}_{\lambda \in \mathcal{J}} \subset L_2(\Omega)$ with

$$\|g\|_{L_2} \eqsim \|\langle g, \Psi \rangle\|_{\ell_2}.$$

- Ψ is a Gelfand frame for (H, L_2, H') if

$$F^* : \ell \to H : c \mapsto c^T \Psi, \quad \tilde{F} : H \to \ell : g \mapsto \langle g, \tilde{\Psi} \rangle$$

are bounded.

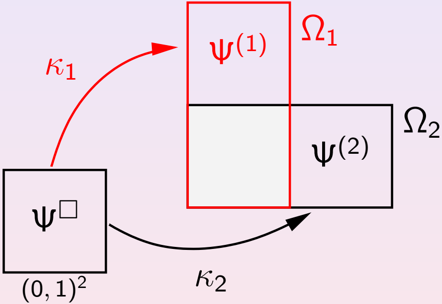
Mapping diagram

$$\begin{tikzcd}
\mathcal{H} \arrow[r] \arrow[rr, bend left, "\mathcal{L}"] \arrow[d, shift right, "\widetilde{F}"'] & L_2(\Omega) \arrow[r] & \mathcal{H}' \arrow[d, shift right, "F"'] \\
\ell \arrow[r, "\mathbf{D}"'] \arrow[u, shift right, "F^*"'] & \ell_2(\mathcal{J}) \arrow[r, "\mathbf{D}^*"'] & \ell' \arrow[u, shift right, "\widetilde{F}^*"']
\end{tikzcd}$$

Construction of wavelet Gelfand frames

- $H = H_0^t(\Omega)$;
- reference Riesz basis $\Psi^{\square} \subset H_0^t(\square)$, $\square := (0,1)^d$;
- overlapping decomposition $\Omega = \sum_{i=1}^n \Omega_i$;
- $\kappa_i : \square \to \Omega_i$, C^m-diffeomorphisms, $m \geq t$;
- appropriate lifting yields Gelfand frames $\Psi = \bigcup_{i=1}^n \Psi_i$.

Gelfand frame construction



Solution of the discrete problem

The operator $\mathbf{L} : \ell_2(\mathcal{J}) \to \ell_2(\mathcal{J})$ is going to be bounded, symmetric, positive, but has **nontrivial kernel.** Nevertheless we have

$$\mathbf{L} : ran(L) \to ran(L)$$

is boundedly invertible;

- (ideal) damped Richardson iteration:

$$(R) \quad \mathbf{u}^{(n+1)} = \mathbf{u}^{(n)} + \mathbf{r}^{(n)}, \quad \mathbf{r}^{(n)} = \mathbf{f} - \mathbf{L}\mathbf{u}^{(n)}$$

- (ideal) steepest descent method

$$(SD) \quad \mathbf{u}^{(n+1)} = \mathbf{u}^{(n)} + \frac{\langle \mathbf{r}^{(n)}, \mathbf{r}^{(n)} \rangle}{\langle \mathbf{L}\mathbf{r}^{(n)}, \mathbf{r}^{(n)} \rangle} \mathbf{r}^{(n)}.$$

Outline

1. Compressive algorithms
   - Introduction
   - Things to be understood
2. Frames in Hilbert spaces: towards the sparsity concept
   - Definitions and main properties
   - Gabor and wavelet frames and audio signals
3. Adaptive frame methods for PDEs
   - Frame discretization of elliptic equations
   - Optimal algorithms
   - Conclusions
4. Sparse recovery: beyond wavelet approximation
   - Compression and ℓ_1-minimization
   - Compressed sensing and other applications
   - ℓ_1-minimization: re-weighted least square method
   - ℓ_1-minimization with noisy data

Basic discrete procedures

Assume that we have the following procedures at our disposal:

- **RHS**$[\varepsilon, \mathbf{g}] \to \mathbf{g}_\varepsilon$: determines for $\mathbf{g} \in \ell_2(\mathcal{J})$ a finitely supported $\mathbf{g}_\varepsilon \in \ell_2(\mathcal{J})$ such that
$$\|\mathbf{g} - \mathbf{g}_\varepsilon\|_{\ell_2(\mathcal{J})} \leq \varepsilon;$$
- **APPLY**$[\varepsilon, \mathbf{N}, \mathbf{v}] \to \mathbf{w}_\varepsilon$: determines for $\mathbf{N} \in B(\ell_2(\mathcal{J}))$ and for a finitely supported $\mathbf{v} \in \ell_2(\mathcal{J})$ a finitely supported $\mathbf{w}_\varepsilon$ such that
$$\|\mathbf{N}\mathbf{v} - \mathbf{w}_\varepsilon\|_{\ell_2(\mathcal{J})} \leq \varepsilon;$$
- **COARSE**$[\varepsilon, \mathbf{v}] \to \mathbf{v}_\varepsilon$: determines for a finitely supported $\mathbf{v} \in \ell_2(\mathcal{J})$ a finitely supported $\mathbf{v}_\varepsilon \in \ell_2(\mathcal{J})$ with at most N significant coefficients, such that
$$\|\mathbf{v} - \mathbf{v}_\varepsilon\|_{\ell_2(\mathcal{J})} \leq \varepsilon.$$
Moreover, $N \lesssim N_{\min}$ holds, $N_{\min}$ being the minimal one.

The algorithm

SOLVE$[\varepsilon, \mathbf{L}, \mathbf{f}] \to \mathbf{u}_\varepsilon$:
Let $\theta < 1/3$ and $K \in \mathbb{N}$ be fixed such that $3\rho^K < \theta$.
$i := 0,\ \mathbf{v}^{(0)} := 0,\ \varepsilon_0 := \|\mathbf{L}^{-1}_{|\operatorname{ran}(\mathbf{L})}\| \|\mathbf{f}\|_{\ell_2(\mathcal{J})}$
While $\varepsilon_i > \varepsilon$ do
$i := i + 1$
$\varepsilon_i := 3\rho^K \varepsilon_{i-1}/\theta$
$\mathbf{f}^{(i)} := \mathbf{RHS}[\frac{\theta\varepsilon_i}{6\alpha K}, \mathbf{f}]$
$\mathbf{v}^{(i,0)} := \mathbf{v}^{(i-1)}$
For $j = 1, ..., K$ do
$\mathbf{v}^{(i,j)} := \mathbf{v}^{(i,j-1)} - \alpha(\mathbf{APPLY}[\frac{\theta\varepsilon_i}{6\alpha K}, \mathbf{L}, \mathbf{v}^{(i,j-1)}] - \mathbf{f}^{(i)})$
od
$\mathbf{v}^{(i)} := \mathbf{COARSE}[(1-\theta)\varepsilon_i, \mathbf{v}^{(i,K)}]$
od
$\mathbf{u}_\varepsilon := \mathbf{v}^{(i)}$.

The algorithm

$\mathbf{SOLVE}[\varepsilon, \mathbf{L}, \mathbf{f}] \to \mathbf{u}_\varepsilon$:
Let $\theta < 1/3$ and $K \in \mathbb{N}$ be fixed such that $3\rho^K < \theta$.
$i := 0,\ \mathbf{v}^{(0)} := 0,\ \varepsilon_0 := \|\mathbf{L}^{-1}_{|\operatorname{ran}(\mathbf{L})}\| \|\mathbf{f}\|_{\ell_2(\mathcal{J})}$
While $\varepsilon_i > \varepsilon$ do
$i := i + 1$
$\varepsilon_i := 3\rho^K \varepsilon_{i-1}/\theta$
$\mathbf{f}^{(i)} := \mathbf{RHS}[\frac{\theta\varepsilon_i}{6\alpha K}, \mathbf{f}]$
$\mathbf{v}^{(i,0)} := \mathbf{v}^{(i-1)}$
For $j = 1, ..., K$ do
$\mathbf{v}^{(i,j)} := \mathbf{v}^{(i,j-1)} - \alpha(\mathbf{APPLY}[\frac{\theta\varepsilon_i}{6\alpha K}, \mathbf{L}, \mathbf{v}^{(i,j-1)}] - \mathbf{f}^{(i)})$
od
$\mathbf{v}^{(i)} := \mathbf{COARSE}[(1-\theta)\varepsilon_i, \mathbf{v}^{(i,K)}]$
od
$\mathbf{u}_\varepsilon := \mathbf{v}^{(i)}$.

The proof of convergence and optimality

Convergence and optimal complexity:

- (R) Riesz basis case: Cohen, Dahmen, DeVore (2000)
- (R) Frame case: Stevenson (2003)
- (SD) Riesz basis case: Dahmen, Urban, Vorloeper (2002) Canuto, Urban (2003)
- (SD) Dahlke, F., Raasch, Stevenson, Werner (2005)

Theorem

Let $\mathbf{Q}$ *be the orthogonal projection onto* $\mathrm{ran}(\mathbf{L})$*. Assume that there exists a solution* $\mathbf{u} \in \mathrm{ran}(\mathbf{L}) \cap \ell_\tau^w$, $1/\tau = s + 1/2$, $s \in (0, s^*)$*. If* $\mathbf{Q}$ *is bounded on the weak-*ℓ_τ *space and* $K > 0$ *in the inner loop is large enough, then*

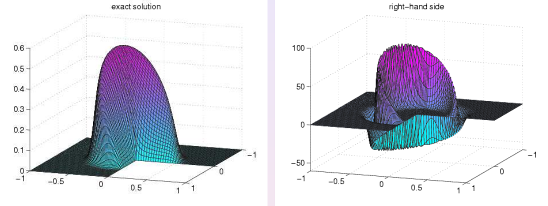
(1) $\|\mathbf{Q}(\mathbf{u} - \mathbf{u}_\varepsilon)\|_{\ell_2} \leq \varepsilon$;

(2) *the support size of* $\mathbf{u}_\varepsilon$ *and the number of algebraic equations to compute it are* $\mathcal{O}(\varepsilon^{-\frac{1}{s}} \|\mathbf{u}\|_{\ell_\tau^w}^{1/s})$

The presence of $\mathbf{Q}$ in (1) is harmless. Indeed, due to the fact that $\mathrm{ran}(\mathbf{L}) = \mathrm{ran}(D^{-1}F)$,

$$u_\varepsilon = \sum_i \sum_\lambda (\mathbf{Q}\mathbf{u}_\varepsilon)_{i,\lambda} \psi_{i,\lambda} = \sum_i \sum_\lambda (\mathbf{u}_\varepsilon)_{i,\lambda} \psi_{i,\lambda}$$

Numerical examples



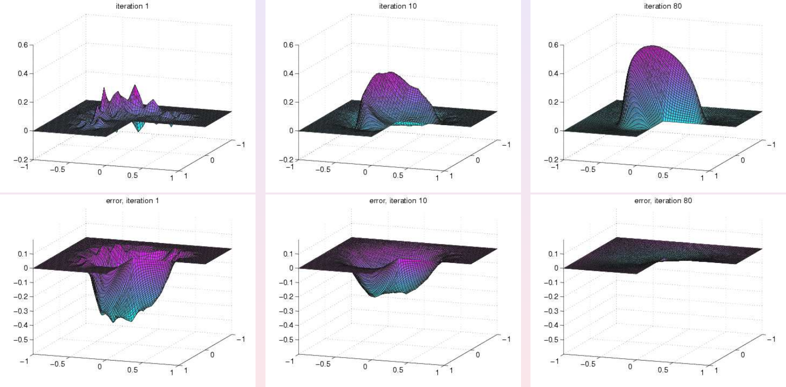
Let us consider the Poisson's equation on the L-shaped domain:

$$-\Delta u = f, \quad u|_{\partial\Omega} = 0,$$

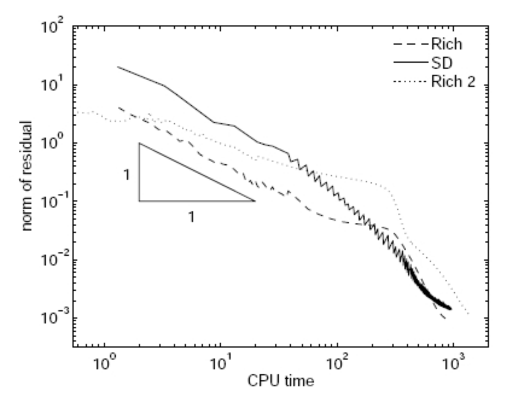
and choose f with singularity at the re-entrant corner.

Numerical examples

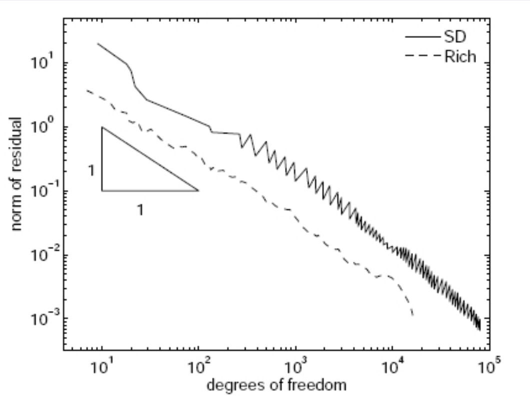
some (SD) iterations:



Numerical examples



Numerical examples



Besov regularity problems

L-shaped domain: $\Omega := (-1,1)^2 \setminus [0,1)^2$, decomposed into two overlapping subdomains $\Omega = \Omega_1 \cup \Omega_2$ with $\Omega_1 := (-1,0) \times (-1,1)$ and $\Omega_2 := (-1,1) \times (-1,0)$. $\phi : [0, \frac{3\pi}{2}] \to \mathbb{R}_{\geq 0}$ is a smooth function with $\phi(\theta) = 1$ for $\theta \leq \frac{\pi}{2}$ and $\phi(\theta) = 0$ for $\theta \geq \pi$. With $(r(x), \theta(x))$ being the polar coordinates of x with respect to the reentrant corner, the functions $\sigma_1 := \phi \circ \theta$ and $\sigma_2 := 1 - \sigma_1$ form a partition of unity subordinate to the patches Ω_i.

For sufficiently smooth wavelets, and $0 < s < (d-t)/n$, one has the relation

$$u \in B^{sn+t}_{\tau}(L_{\tau}(\Omega)) \text{ if and only if } D\tilde{F}u = \left(2^{|\lambda|t}\langle u, \tilde{\psi}_\lambda\rangle_{L_2(\Omega)}\right)_{\lambda\in\mathcal{I}} \in \ell_\tau(\mathcal{I}),$$

where $\tau = (s+\frac{1}{2})^{-1}$, and $D\tilde{F}u$ are exactly the unique expansion coefficients of u with respect to the Riesz basis $D^{-1}\Psi$ in $H^t_0(\Omega)$.
Unfortunately we cannot say much about the *canonical dual* frame coefficients of a Gelfand frame $\Psi := \Psi^{(1)} \cup \Psi^{(2)}$ for the L-shape domain.

However for the sequence of expansion coefficients

$$\mathbf{u} = (2^{t|\lambda|}\langle u, \sigma_i \tilde{\psi}_{i,\lambda}\rangle)_{\lambda \in \mathcal{I}^\square, i=1,2}$$

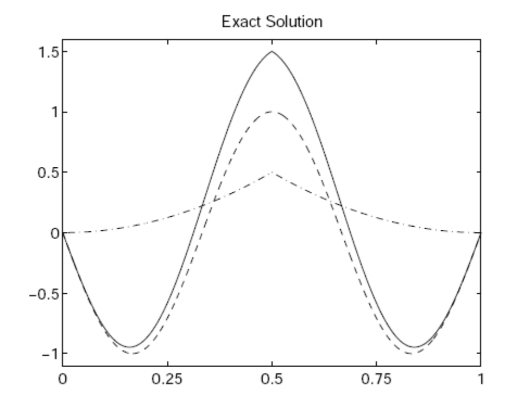
we have the following result.

Theorem (Dahlke, F., Primbs, Raasch, Werner)

Let u be the variational solution to the Poisson eq. on the L-shaped domain. Let the right-hand side f be contained in $H^\mu(\Omega)$ for a $\mu > 0$. Then, the sequence of frame coefficients $(2^{|\lambda|}\langle u, \sigma_i \tilde{\psi}_{i,\lambda}\rangle)_{\lambda \in \mathcal{I}^\square, i=1,2}$ belongs to the space $\ell_{\tau_0}(\mathcal{I})$, where $\frac{1}{\tau_0} = \frac{s-1}{2} + \frac{1}{2}$, for all $1 < s < \min\{d, \mu + 2\}$.

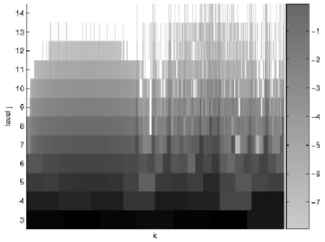
It is important to stress the fact that in the case of a wavelet basis the analogous statement holds under the only slightly milder requirement $f \in H^\mu(\Omega)$, $\mu > -1/2$.

Numerical examples



Numerical examples

$\Omega_1 = (0, 0.7)$ and $\Omega_2 = (0.3, 1)$



Domain decomposition methods

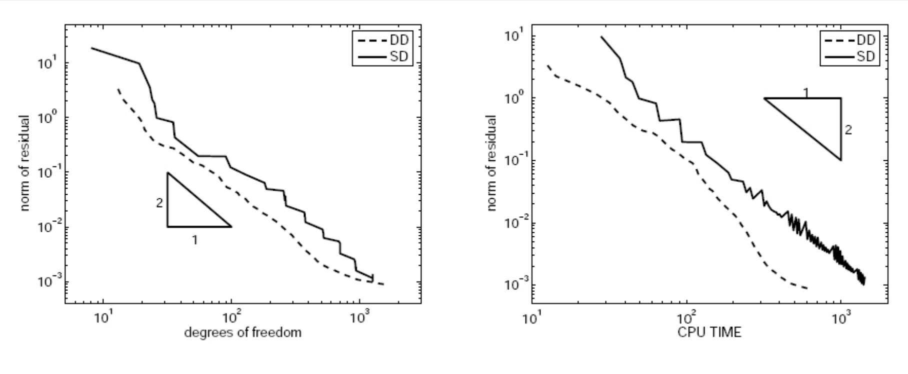
Algorithm 1. *Multiplicative Schwarz iteration*

for $k = 1, \ldots,$
$\mathbf{u}_0^{k-1} = \mathbf{u}^{k-1}$
for $i = 1, \ldots, n$
$\mathbf{u}_i^{k-1} = \mathbf{u}_{i-1}^{k-1} + Q_i^T \tilde{\mathbf{L}}_i^{-1} Q_i(\mathbf{f} - \mathbf{L}\mathbf{u}_{i-1}^{k-1})$
endfor
$\mathbf{u}^k = \mathbf{u}_n^{k-1}$
endfor

Algorithm 2. *Additive Schwarz iteration*

for $k = 1, \ldots,$
$\mathbf{u}^k = \mathbf{u}^{k-1} + \alpha \sum_{i=1}^n Q_i^T \tilde{\mathbf{L}}_i^{-1} Q_i(\mathbf{f} - \mathbf{L}\mathbf{u}^{k-1})$
endfor

Numerical examples (work in progress)



Outline

1. Compressive algorithms
   - Introduction
   - Things to be understood
2. Frames in Hilbert spaces: towards the sparsity concept
   - Definitions and main properties
   - Gabor and wavelet frames and audio signals
3. Adaptive frame methods for PDEs
   - Frame discretization of elliptic equations
   - Optimal algorithms
   - Conclusions
4. Sparse recovery: beyond wavelet approximation
   - Compression and ℓ_1-minimization
   - Compressed sensing and other applications
   - ℓ_1-minimization: re-weighted least square method
   - ℓ_1-minimization with noisy data

(1) The construction of wavelet frames on domains is a rather easy task;

(2) Optimality benchmarks can be established for wavelet frame approximations;

(3) Optimal algorithms can be realized for wavelet frame approximations;

(4) The redundancy effects can be diminished by subspace corrections/domain decompositions;

(5) Towards hybrid and highly redundant approximations.

(1) The construction of wavelet frames on domains is a rather easy task;

(2) Optimality benchmarks can be established for wavelet frame approximations;

(3) Optimal algorithms can be realized for wavelet frame approximations;

(4) The redundancy effects can be diminished by subspace corrections/domain decompositions;

(5) Towards hybrid and highly redundant approximations.

(1) The construction of wavelet frames on domains is a rather easy task;

(2) Optimality benchmarks can be established for wavelet frame approximations;

(3) Optimal algorithms can be realized for wavelet frame approximations;

(4) The redundancy effects can be diminished by subspace corrections/domain decompositions;

(5) Towards hybrid and highly redundant approximations.

(1) The construction of wavelet frames on domains is a rather easy task;

(2) Optimality benchmarks can be established for wavelet frame approximations;

(3) Optimal algorithms can be realized for wavelet frame approximations;

(4) The redundancy effects can be diminished by subspace corrections/domain decompositions;

(5) Towards hybrid and highly redundant approximations.

(1) The construction of wavelet frames on domains is a rather easy task;

(2) Optimality benchmarks can be established for wavelet frame approximations;

(3) Optimal algorithms can be realized for wavelet frame approximations;

(4) The redundancy effects can be diminished by subspace corrections/domain decompositions;

(5) Towards hybrid and highly redundant approximations.

Outline

1. Compressive algorithms
   - Introduction
   - Things to be understood
2. Frames in Hilbert spaces: towards the sparsity concept
   - Definitions and main properties
   - Gabor and wavelet frames and audio signals
3. Adaptive frame methods for PDEs
   - Frame discretization of elliptic equations
   - Optimal algorithms
   - Conclusions
4. Sparse recovery: beyond wavelet approximation
   - Compression and ℓ_1-minimization
   - Compressed sensing and other applications
   - ℓ_1-minimization: re-weighted least square method
   - ℓ_1-minimization with noisy data

Finite frames

To construct finite frames is easy:

Proposition

Every collection of vectors $\Phi = \{\psi_i \in \mathbb{R}^n : i = 1, ..., N\}$ is a frame for its span. We will always assume in the following that $n \leq N$.

The traditional Vs. new compression paradigms

(1) We showed that different classes of functions/signals can be represented with respect to certain *bases* Φ in a very parsimonious way, which leads to most of the modern compression methods (JPEG,MP3 etc.).

(2) This classical paradigm is based on the concept of best N-term approximation, i.e., re-arrange the entries by magnitude and keep the most important ones.

(3) Frames do allow for *sparser* representations BUT the computation of the best N-term approximation is a difficult combinatorial problem.

The traditional Vs. new compression paradigms

(1) We showed that different classes of functions/signals can be represented with respect to certain *bases* Φ in a very parsimonious way, which leads to most of the modern compression methods (JPEG,MP3 etc.).

(2) This classical paradigm is based on the concept of best N-term approximation, i.e., re-arrange the entries by magnitude and keep the most important ones.

(3) Frames do allow for *sparser* representations BUT the computation of the best N-term approximation is a difficult combinatorial problem.

RICAM
JOHANN-RADON-INSTITUTE
FOR COMPUTATIONAL AND APPLIED MATHEMATICS

The traditional Vs. new compression paradigms

(1) We showed that different classes of functions/signals can be represented with respect to certain *bases* Φ in a very parsimonious way, which leads to most of the modern compression methods (JPEG,MP3 etc.).

(2) This classical paradigm is based on the concept of best N-term approximation, i.e., re-arrange the entries by magnitude and keep the most important ones.

(3) Frames do allow for *sparser* representations BUT the computation of the best N-term approximation is a difficult combinatorial problem.

The problem

Given a frame $\Phi = \{\phi_i \in \mathbb{R}^n : i = 1, ..., N\}$ and a signal $y \in \mathbb{R}^n$ we would like to compute $x^* \in \mathbb{R}^N$ with $\|z^*\|_{\ell_0} := \# \operatorname{supp}(x) << n$ such that

$$\Phi x^* = y.$$

or

$$x^* = \operatorname{argmin}_{\Phi z = y} \|z\|_{\ell_0}.$$

The problem

How could we find such a x^*? One possibility is to consider any set T of $k << n$ column indices and find the least squares solution $x^T := \operatorname{argmin}_z \|\Phi_T z - y\|_{\ell_2}$. Finding x^T is numerically simple. After finding each x^T, we choose T^* which minimizes the residual. However, the method is numerically prohibitive, it requires solving $\binom{N}{k}$ least squares problems.

Possible practical solutions

The known practical solutions to the problem above assume special properties of Φ:

(1) Greedy algorithms, e.g., OMP (requires that Φ has incoherency properties, e.g., it is a union of incoherent bases – "sines and spikes");

(2) Convex relaxation, i.e., ℓ_1-minimization (basis pursuit)

$$x^* = \operatorname{argmin}_{\Phi z = y} \|z\|_{\ell_1}.$$

It needs special well conditioning of any small column subspace of Φ.

Possible practical solutions

The known practical solutions to the problem above assume special properties of Φ:

(1) Greedy algorithms, e.g., OMP (requires that Φ has incoherency properties, e.g., it is a union of incoherent bases – "sines and spikes");

(2) Convex relaxation, i.e., ℓ_1-minimization (basis pursuit)

$$x^* = \operatorname{argmin}_{\Phi z = y} \|z\|_{\ell_1}.$$

It needs special well conditioning of any small column subspace of Φ.

Why ℓ_1-minimization can work

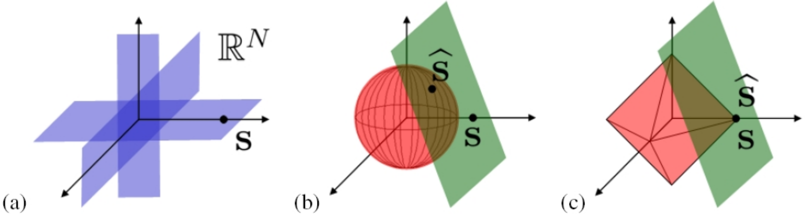


Figure 2: *(a) A sparse vector* s *lies on a* K*-dimensional hyperplane aligned with the coordinate axes in* $\mathbb{R}^N$ *and thus close to the axes. (b) Compressive sensing recovery via* ℓ_2 *minimization does not find the correct sparse solution* s *on the translated nullspace (green hyperplane) but rather the non-sparse vector* $\widehat{s}$. *(c) With enough measurements, recovery via* ℓ_1 *minimization does find the correct sparse solution* s.

Why ℓ_1-minimization can work

(1) ℓ_1-norm as a sparsity-promoting functional is first used in reflection seismology (1970-1980)

(2) Rigorous results began to appear in the late-1980's (Donoho, Logan, and Stark).

(3) Applications for ℓ_1-minimization in statistical estimation (mid-1990's, Tibshirani) with the LASSO algorithm (iterative soft-thresholding).

(4) In signal processing, Basis Pursuit (mid-1990's Donoho).

Why ℓ_1-minimization can work

(1) ℓ_1-norm as a sparsity-promoting functional is first used in reflection seismology (1970-1980)

(2) Rigorous results began to appear in the late-1980's (Donoho, Logan, and Stark).

(3) Applications for ℓ_1-minimization in statistical estimation (mid-1990's, Tibshirani) with the LASSO algorithm (iterative soft-thresholding).

(4) In signal processing, Basis Pursuit (mid-1990's Donoho).

Why ℓ_1-minimization can work

(1) ℓ_1-norm as a sparsity-promoting functional is first used in reflection seismology (1970-1980)

(2) Rigorous results began to appear in the late-1980's (Donoho, Logan, and Stark).

(3) Applications for ℓ_1-minimization in statistical estimation (mid-1990's, Tibshirani) with the LASSO algorithm (iterative soft-thresholding).

(4) In signal processing, Basis Pursuit (mid-1990's Donoho).

Why ℓ_1-minimization can work

(1) ℓ_1-norm as a sparsity-promoting functional is first used in reflection seismology (1970-1980)

(2) Rigorous results began to appear in the late-1980's (Donoho, Logan, and Stark).

(3) Applications for ℓ_1-minimization in statistical estimation (mid-1990's, Tibshirani) with the LASSO algorithm (iterative soft-thresholding).

(4) In signal processing, Basis Pursuit (mid-1990's Donoho).

Restricted Isometry Property (Candès, Tao)

A critical property needed in order to allow the best possible performances is the so-called *Restricted Isometry Property* (RIP).

Definition

We say that the matrix Φ has the RIP of order k if there exists a $0 < \delta_k < 1$ such that

$$(1-\delta_k)\|x_T\|_{\ell_2}^2 \leq \|\Phi_T x_T\|_2^2 \leq (1+\delta_k)\|x_T\|_{\ell_2}^2,$$

for all x and for *all* subset T with $\#T \leq k$.

A fundamental result

Theorem (Candès, Romberg, Tao - Cohen, Dahmen, DeVore)

Let Φ be a frame which satisfies the RIP of order $3k$ for $\delta_{3k} \leq \delta < 1$. Then there exists a decoder Δ such that

$$\|x - \Delta(\Phi x)\|_{\ell_2} \leq C \frac{\sigma_k(x)_{\ell_1}}{\sqrt{k}},$$

where $\sigma_k(x)_X = \|x - x_k\|_X$, and x_k is the best k-term approximation of x in norm X.

- The best possible (i.e. the largest) k for which this theorem can hold is $k \asymp \frac{n}{\log(N/n)+1}$;
- If $\#(\mathrm{supp}(x)) \leq k$ then $\sigma_k(x) = 0$ and we have the perfect reconstruction $x = \Delta(\Phi x)$.

The fundamental open problem

There exist frames Φ with RIP of optimal order $k \asymp \frac{n}{\log(N/n)+1}$?

- yes;
- ... but ALL the constructions of such frame are NOT deterministic[1].

[1] Refer to the Terence Tao's blog http://terrytao.wordpress.com/2007/07/02/open-question-deterministic-uup-matrices

Some probabilistic constructions

- The simplest examples are $n \times N$ matrices Φ whose entries $\phi_{i,j}$ are independent realizations of Gaussian random variables
$$\phi_{i,j} \sim \mathcal{N}(0, \frac{1}{n});$$
- One can also use matrices where the entries are independent realizations of $\pm$ Bernoulli random variables

$$\phi_{i,j} := \begin{cases} +\frac{1}{\sqrt{n}}, & \text{with probability } \frac{1}{2} \\ -\frac{1}{\sqrt{n}}, & \text{with probability } \frac{1}{2} \end{cases}$$

Fourier ensembles

The discrete Fourier transform $\hat{x} = \Psi x$ is defined by the matrix

$$\Psi_{\omega,t} = \frac{1}{\sqrt{n}} e^{-2\pi i \frac{\omega t}{n}}.$$

Our frame is going to be constructed extracting submatrix Φ by random choice of rows of Ψ. Therefore on the set $\{1, ..., N\}$ we apply some random selectors which are $\delta_1, \dots, \delta_N$ independent Bernoulli random variables taking value 1 with probability $\delta = n/N$. We define the rows by the set $\Omega = \{j \in \{1, ..., N\} : \delta_j = 1\}$. Clearly $\mathbb{E}|\Omega| = n$.

Encoding via random projections $\Rightarrow$ Compressed Sensing

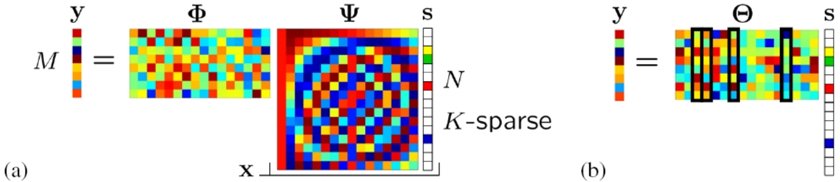


Figure 1: *(a) Compressive sensing measurement process with (random Gaussian) measurement matrix Φ and discrete cosine transform (DCT) matrix Ψ. The coefficient vector* s *is sparse with $K = 4$. (b) Measurement process in terms of the matrix product $\Theta = \Phi\Psi$ with the four columns corresponding to nonzero s_i highlighted. The measurement vector* y *is a linear combination of these four columns.*

ℓ_1-minimization is a practical decoder

Theorem (Donoho, Candes, Tao '05-'06)

Let Φ be frame which satisfies the RIP of order $2k$ for $\delta_{2k} \leq \delta < 1/3$. Then, for any $x \in \mathbb{R}^N$ with $\#(\mathrm{supp}(x)) \leq k$ and $y = \Phi x$, the decoder

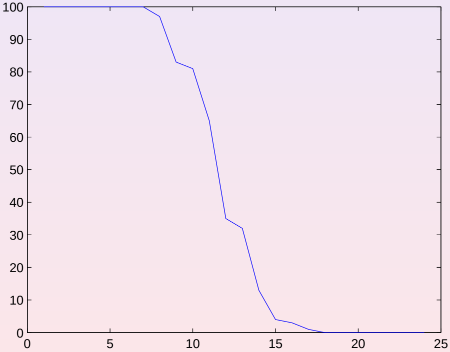
$$\Delta(y) := \mathrm{argmin}_{\Phi z = y} \|z\|_{\ell_1},$$

reconstructs $x = \Delta(y)$ exactly.

The variational problem can be simply reformulated as a Linear Program (LP) in standard form:

$$\mathrm{argmin}_{[\Phi|-\Phi]v=y} 1^T v, \quad v \geq 0, \quad v = [z_+ | z_-].$$

ℓ_1-minimization performances



Outline

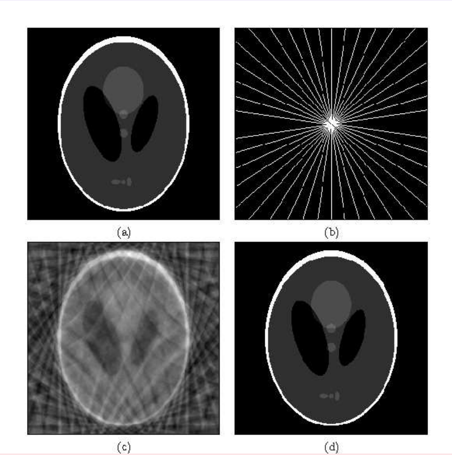
1. Compressive algorithms
   - Introduction
   - Things to be understood
2. Frames in Hilbert spaces: towards the sparsity concept
   - Definitions and main properties
   - Gabor and wavelet frames and audio signals
3. Adaptive frame methods for PDEs
   - Frame discretization of elliptic equations
   - Optimal algorithms
   - Conclusions
4. Sparse recovery: beyond wavelet approximation
   - Compression and ℓ_1-minimization
   - Compressed sensing and other applications
   - ℓ_1-minimization: re-weighted least square method
   - ℓ_1-minimization with noisy data

Two first applications that attracted the attention:

- CS for tomography (Candes, Romberg, Tao);
- CS in digital image acquisition (Wakin, Laska, Duarte, Baron, Sarvotham, Takhar, Kelly, and Baraniuk);

Nevertheless more and more applications are continuously found.

Tomography and Fourier ensembles



Tomography and Fourier ensembles

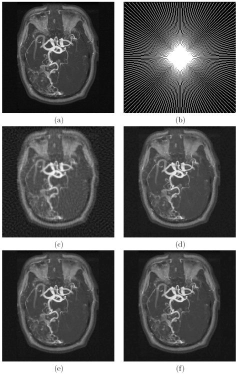
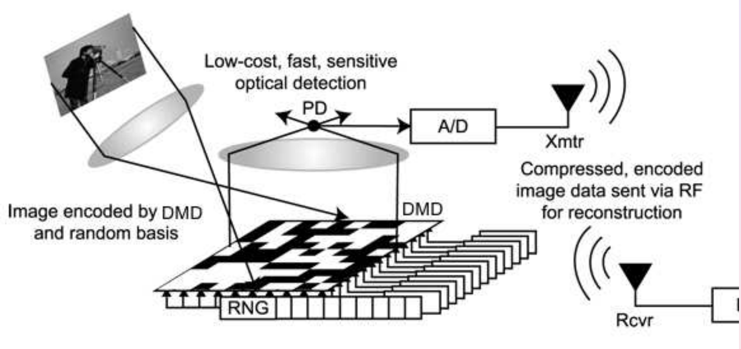
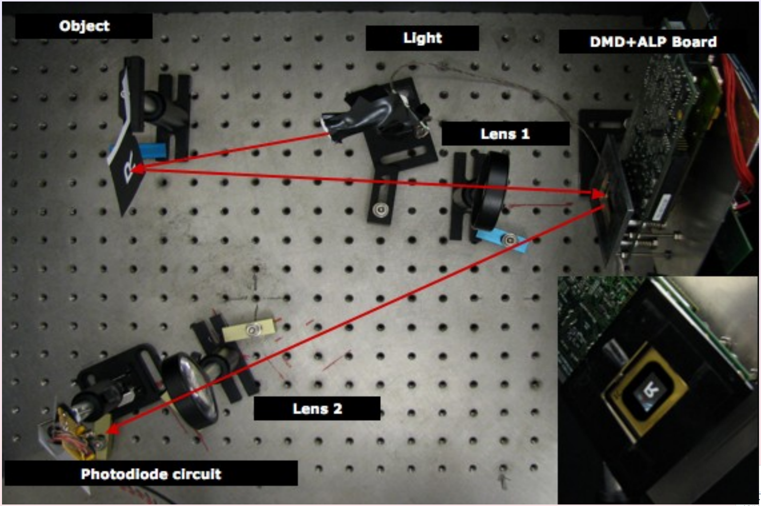
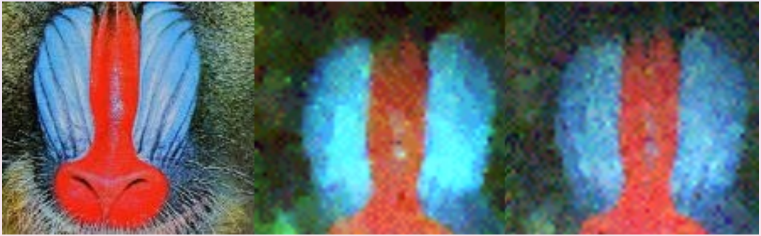


Figure 12: *(a) Original MR angiogram. (b) Fourier sampling pattern. (c) Backprojection, PSNR = 29.00dB. (d) Minimum TV reconstruction, PSNR = 34.23dB. (e) ℓ_1 analysis reconstruction, PSNR = 34.37dB. (f) Reweighted ℓ_1 analysis reconstruction, PSNR = 34.78dB.*

Rice camera and pseudo-random ensembles




Object
Light
DMD+ALP Board
Lens 1
Lens 2
Photodiode circuit



Left: Original mandrill 4096 pixels;

Center: reconstruction with 20% rand. meas.;

Right: reconstruction with 40% rand. meas.

Few recent new applications

- designing sparse interconnect wiring,
- designing sparse control system feedback gains,
- sparse trades in portfolio optimization with fixed transaction cost,
- design well connected sparse graphs,
- sparse gene network systems identification
- ...
- any combinatorial problem with prescribed linear or nonlinear equation constraints

A lapidary statement

The use of ℓ_1 regularization has become so widespread that it could arguably be considered the "modern least squares". From this lapidary statement it follows the clear need for efficient algorithms for the minimization problem.

Outline

1. Compressive algorithms
   - Introduction
   - Things to be understood
2. Frames in Hilbert spaces: towards the sparsity concept
   - Definitions and main properties
   - Gabor and wavelet frames and audio signals
3. Adaptive frame methods for PDEs
   - Frame discretization of elliptic equations
   - Optimal algorithms
   - Conclusions
4. Sparse recovery: beyond wavelet approximation
   - Compression and ℓ_1-minimization
   - Compressed sensing and other applications
   - ℓ_1-minimization: re-weighted least square method
   - ℓ_1-minimization with noisy data

Fast iterative methods

- Let Φ be an $n \times N$ (compressed sensing) CS matrix.
- We are interested in decoding $y = \Phi z$ with $z \in \mathbb{R}^N$ and $y \in \mathbb{R}^n$.
- The set of all solutions to $y = \Phi z$ will be denoted by $\mathcal{F}(y)$, which can also be described by the affine space $z + \mathcal{N}$, where $\mathcal{N}$ is the null space of Φ.
- We denote by x the minimal ℓ_1-norm solution

$$x := \arg \min_{z \in \mathcal{F}(y)} \|z\|_1$$

We seek for faster alternatives to using linear programming (i.e., interior point methods) to find x.

CS matrices (NSP)

The Null Space Property (NSP) of order k for $\gamma > 0$ says that

$$\|\eta_T\|_1 \leq \gamma \|\eta_{T^c}\|_1,$$

for all sets T of cardinality less than k and all $\eta \in \mathcal{N}$.
If Φ satisfies RIP of order $(b+1)k$ for $\delta > 0$, where $b \geq 1$ is an integer, then Φ satisfies the NSP of order k for $\gamma := \frac{1+\delta}{\sqrt{b}(1-\delta)}$.

Re-weighted Iterative Least Square (Osborne '85)

For $w \in \mathbb{R}_+^N$ and $\varepsilon > 0$, define

$$\mathcal{J}(z, w, \varepsilon) = \frac{1}{2}\left[\sum_{j=1}^N z_j^2 w_j + \sum_{j=1}^N (\varepsilon^2 w_j + w_j^{-1})\right].$$

We initialize by $w^{(0)} = (1, \dots, 1)$ and $\varepsilon_0 = 1$. Then, recursively

$$\begin{aligned} x^{(n+1)} &:= \arg\min_{z \in \mathcal{F}(y)} \mathcal{J}(z, w^{(n)}, \varepsilon_n), \\ \varepsilon_{n+1} &:= \min\{\varepsilon_n, \frac{r(x^{(n+1)})_K}{N}\} \\ w^{(n+1)} &:= \arg\min_{w>0} \mathcal{J}(x^{(n+1)}, w, \varepsilon_{n+1}) \\ &= \left([(x_j^{(n+1)})^2 + \varepsilon_{n+1}^2]^{-1/2}\right)_{j=1}^N \end{aligned}$$

Re-weighted least squares

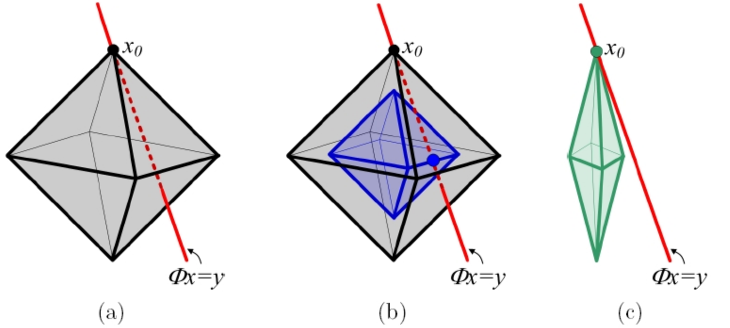


Figure 1: *Weighting ℓ_1 minimization to improve sparse signal recovery. (a) Sparse signal x_0, feasible set $\Phi x = y$, and ℓ_1 ball of radius $\|x_0\|_{\ell_1}$. (b) There exists an $x \neq x_0$ for which $\|x\|_{\ell_1} < \|x_0\|_{\ell_1}$. (c) Weighted ℓ_1 ball. There exists no $x \neq x_0$ for which $\|Wx\|_{\ell_1} \leq \|Wx_0\|_{\ell_1}$.*

Theorem (Daubechies, DeVore, F., Güntürk)

Let $k \geq 1$ and define $K = k + 6$. We assume that Φ satisfies the NSP of order $3K$ for $\gamma \leq 1/2$. Let x be the unique minimum ℓ_1-norm point in $\mathcal{F}(y)$. Then, for each $y \in \mathbb{R}^m$, the Algorithm above converges and its limit $\bar{x}$ satisfies

$$\|x - \bar{x}\|_1 \leq C_1 \sigma_k(x)_{\ell_1}, \qquad C_1 := \frac{5(1+\gamma)}{1-\gamma}$$

In particular if x is k-sparse then $x^{(n)}$ converges to x.

Theorem (Daubechies, DeVore, F., Güntürk)

For a given $0 < \rho < 1$, *assume that* Φ *satisfies NSP of order* $3K$ *with constant* $\gamma > 0$ *such that*

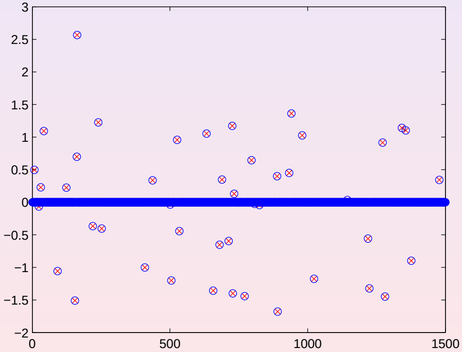
$$\mu := \frac{\gamma}{1-\rho}\left(1 + \frac{1}{K-k}\right) < 1.$$

Moreover, assume that $|\operatorname{supp}(x)| \leq k$, *i.e.,* x *is* k*-sparse. Here we denote* $E_n := \|x^{(n)} - x\|_1$. *Let* $n_0 \in \mathbb{N}$ *be such that*

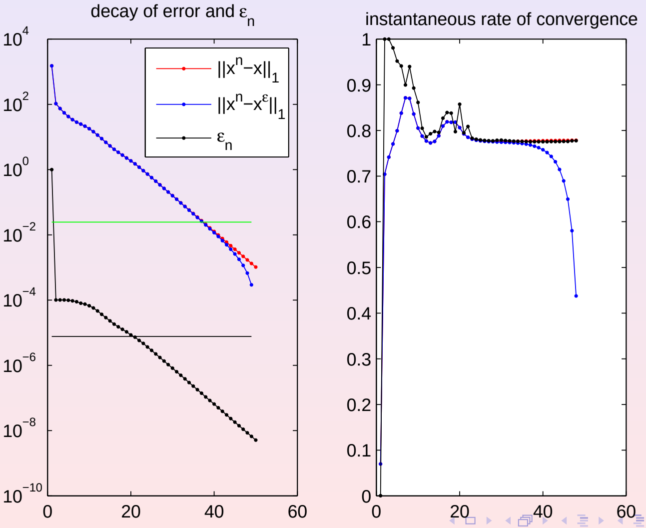
$$E_{n_0} \leq R^* := \rho \min_{j \in T} |x_j| = \rho r(x)_k.$$

Then for all $n \geq n_0$, *we have* $E_{n+1} \leq \mu E_n$. *Consequently* $x^{(n)}$ *converges to* x *exponentially.*

Exact reconstruction



Performances

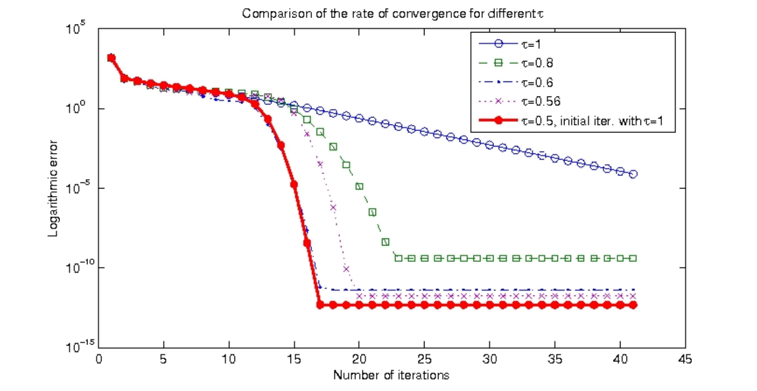


Re-weighted Iterative Least Square: ℓ_τ minimization $\tau < 1$

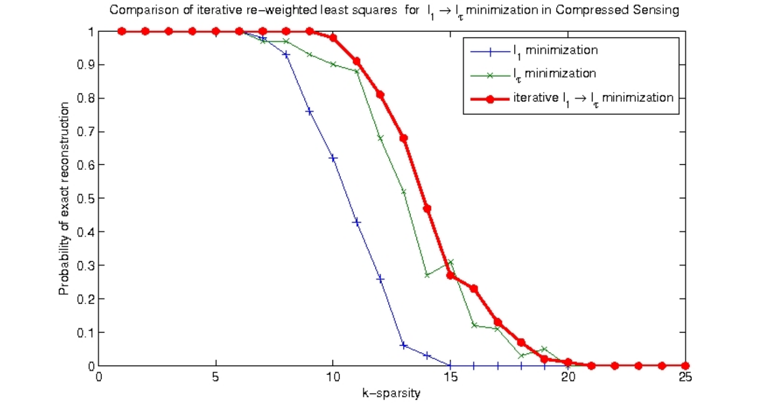
We initialize by $w^{(0)} = (1, \dots, 1)$ and $\varepsilon_0 = 1$. Then, recursively

$$
\begin{aligned}
x^{(n+1)} &:= \arg \min_{z \in \mathcal{F}(y)} \mathcal{J}(z, w^{(n)}, \varepsilon_n), \\
\varepsilon_{n+1} &:= \min\{\varepsilon_n, \frac{r(x^{(n+1)})_K}{N}\} \\
w^{(n+1)} &:= \arg\min_{w>0} \mathcal{J}(x^{(n+1)}, w, \varepsilon_{n+1}) \\
&= \left([(x_j^{(n+1)})^2 + \varepsilon_{n+1}^2]^{\frac{\tau}{2}-1} \right)_{j=1}^N
\end{aligned}
$$

Re-weighted Iterative Least Square: ℓ_τ minimization $\tau < 1$



Re-weighted Iterative Least Square: ℓ_τ minimization $\tau < 1$



Outline

1. Compressive algorithms
   - Introduction
   - Things to be understood
2. Frames in Hilbert spaces: towards the sparsity concept
   - Definitions and main properties
   - Gabor and wavelet frames and audio signals
3. Adaptive frame methods for PDEs
   - Frame discretization of elliptic equations
   - Optimal algorithms
   - Conclusions
4. Sparse recovery: beyond wavelet approximation
   - Compression and ℓ_1-minimization
   - Compressed sensing and other applications
   - ℓ_1-minimization: re-weighted least square method
   - ℓ_1-minimization with noisy data

ℓ_1 minimization with noisy data

- $(\star)$ In realistic cases the data y are not provided noise-free, thus we have to assume $\Phi x = y + e$. Here Φ can be the matrix of a bounded operator with respect to basis/frame coordinates.
- $(\star)$ To deal with the reconstruction of x a *regularization* mechanism is required.
- $(\star)$ Regularization techniques try to take advantage of prior knowledge one may have about the nature of x. We assume that x is sparse (or compressible).
- $(\star)$ The recovery is realized by minimizing

$$\mathcal{J}_\tau(x) := \|\Phi x - y\|_2^2 + 2\tau\|x\|_1.$$

ℓ_1 minimization with noisy data

- ($\star$) In realistic cases the data y are not provided noise-free, thus we have to assume $\Phi x = y + e$. Here Φ can be the matrix of a bounded operator with respect to basis/frame coordinates.
- ($\star$) To deal with the reconstruction of x a *regularization* mechanism is required.
- ($\star$) Regularization techniques try to take advantage of prior knowledge one may have about the nature of x. We assume that x is sparse (or compressible).
- ($\star$) The recovery is realized by minimizing

$$\mathcal{J}_\tau(x) := \|\Phi x - y\|_2^2 + 2\tau \|x\|_1.$$

ℓ_1 minimization with noisy data

- (⋆) In realistic cases the data y are not provided noise-free, thus we have to assume $\Phi x = y + e$. Here Φ can be the matrix of a bounded operator with respect to basis/frame coordinates.
- (⋆) To deal with the reconstruction of x a *regularization* mechanism is required.
- (⋆) Regularization techniques try to take advantage of prior knowledge one may have about the nature of x. We assume that x is sparse (or compressible).
- (⋆) The recovery is realized by minimizing

$$\mathcal{J}_\tau(x) := \|\Phi x - y\|_2^2 + 2\tau\|x\|_1.$$

ℓ_1 minimization with noisy data

- ($\star$) In realistic cases the data y are not provided noise-free, thus we have to assume $\Phi x = y + e$. Here Φ can be the matrix of a bounded operator with respect to basis/frame coordinates.
- ($\star$) To deal with the reconstruction of x a *regularization* mechanism is required.
- ($\star$) Regularization techniques try to take advantage of prior knowledge one may have about the nature of x. We assume that x is sparse (or compressible).
- ($\star$) The recovery is realized by minimizing

$$\mathcal{J}_\tau(x) := \|\Phi x - y\|_2^2 + 2\tau\|x\|_1.$$

ℓ_1 minimization promotes sparsity

($\star$) As observed above, for certain classes of matrices Φ, the ℓ_1-minimization does compute the sparsest solution.

($\star$) Even for Φ outside this classes, ℓ_1-minimization seems to lead to very good approximations to the sparsest solutions.

ℓ_1 minimization promotes sparsity

($\star$) As observed above, for certain classes of matrices Φ, the ℓ_1-minimization does compute the sparsest solution.

($\star$) Even for Φ outside this classes, ℓ_1-minimization seems to lead to very good approximations to the sparsest solutions.

A minimizing algorithm

Several authors (e.g., Donoho, Starck, Tibshirani) proposed independently an iterative soft-thresholding algorithm to approximate the solution $\bar{x}(\tau)$. More precisely, $\bar{x}(\tau)$ is the limit of the sequence $x^{(n)}$ defined by

$$x^{(n+1)} = \mathbb{S}_\tau \left[x^{(n)} + \Phi^* y - \Phi^* \Phi x^{(n)} \right] ,$$

starting from an arbitrary $x^{(0)}$, where $\mathbb{S}_\tau$ is the soft-thresholding operation defined by $\mathbb{S}_\tau(x)_\lambda = S_\tau(x_\lambda)$ with

$$S_\tau(x) = \begin{cases} x - \tau & x > \tau \\ 0 & |x| \leq \tau \\ x + \tau & x < -\tau \end{cases} .$$

Performances

The algorithm converges initially relatively fast, then it overshoots the value $\|\bar{x}(\tau)\|_1$ (where $\bar{x}(\tau) := \lim_{n\to\infty} x^{(n)}$), and it takes very long (forever!) to re-correct back. In other words, starting from $x^{(0)} = 0$, the algorithm generates a path $\{x^{(n)};\ n \in \mathbb{N}\}$ that is initially fully contained in the ℓ_1-ball $B_R := \{x \in \ell_2(\Lambda) : \|x\|_1 \leq R\}$, with $R := \|\bar{x}(\tau)\|_1$. Then it gets out of the ball to slowly inch back to it in the limit.

Performances