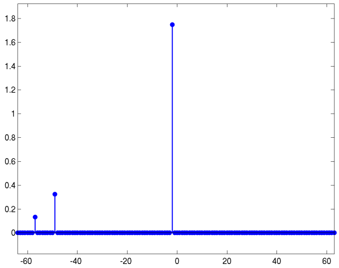
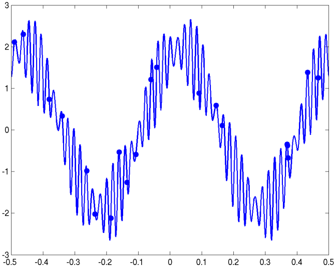
Computing sparse solutions of underdetermined structured systems by greedy methods

Stefan Kunis
(joint work with Holger Rauhut)

Chemnitz University of Technology
http://www.tu-chemnitz.de/∼skunis





Outline

Sparse reconstruction

- compressed sensing, compressive sampling, sparse reconstruction, ...
- a sparse reconstruction problem, $M \ll N$, $\mathbf{A} \in \mathbb{C}^{M \times N}$, $\mathbf{b} \in \mathbb{C}^{M}$ given, find

$$\min_{\mathbf{c} \in \mathbb{C}^N} |\mathrm{supp}(\mathbf{c})| \quad \text{s.t.} \quad \mathbf{A}\mathbf{c} = \mathbf{b} \tag{1}$$

- denote sparsity $S = |\mathrm{supp}(\mathbf{c})|$, interesting case

$$S \approx M \ll N$$

- structured matrices $\mathbf{A}$ with fast matrix-vector arithemtic

Sparse reconstruction

- basis pursuit principle, (Donoho, Stark [1989]; Candes, Romberg, Tao [2004-]; Donoho, Tanner [2004-]; Rauhut [2005-]; Rudelson, Vershynin [2006-])

$$\min_{\mathbf{c}\in\mathbb{C}^N} \|\mathbf{c}\|_1 \quad \text{s.t.} \quad \mathbf{A}\mathbf{c} = \mathbf{b} \tag{2}$$

- smallest number δ_S such that for all $\mathbf{c} \in \mathbb{C}^N$, $\text{supp}(\mathbf{c}) \leq S$:

$$1 - \delta_S \leq \frac{\mathbf{c}^{\mathsf{H}}\mathbf{A}^{\mathsf{H}}\mathbf{A}\mathbf{c}}{\mathbf{c}^{\mathsf{H}}\mathbf{c}} \leq 1 + \delta_S \tag{3}$$

Theorem. [Candes, Tao]

- if $\mathbf{A}$ satisfies $\delta_S + \delta_{2S} + \delta_{3S} < 1$ then (2) solves (1)
- condition (3) is satisfied with probabilty $1 - \epsilon$ for random matrices with Gaussian entries for

$$M \geq C_\delta \cdot S \cdot \log(N/\epsilon)$$

Sparse reconstruction

- general purpose schemes that solve basis pursuit (2) are slow
- theoretical computer science, computational time sublinear in N, (Mansour [1995]; Daubechies, Gilbert, Muthukrishnan, Strauss, Zou [2005-])
- greedy methods for random matrices with Gaussian or Bernoulli entries, (Tropp, Gilbert [2005-])
- connection to approximation theory, (Temlyakov [2003-]; Cohen, Dahmen, DeVore [2006])

Nonequispaced FFT

- trigonometric polynomials $f : \mathbb{T} \to \mathbb{C}$,

$$f(x) = \sum_{k=-N/2}^{N/2-1} \hat{f}_k \mathrm{e}^{-2\pi \mathrm{i} k x}$$

- discrete Fourier transform - DFT ($\mathcal{O}(N^2)$ or by FFT $\mathcal{O}(N \log N)$)

$$f_j = \sum_{k=-N/2}^{N/2-1} \hat{f}_k \mathrm{e}^{-2\pi \mathrm{i} k j / N}, \quad j = -N/2, \ldots, N/2-1$$

- for a finite sampling set $\mathcal{X} \subset \mathbb{T}$ - nonequispaced DFT ($\mathcal{O}(MN)$)

$$\mathbf{f} = \mathbf{A}\hat{\mathbf{f}}, \qquad (a_{j,k}) = \left(\mathrm{e}^{-2\pi \mathrm{i} k x_j}\right) \in \mathbb{C}^{M \times N}$$

Nonequispaced FFT

- fast Fourier tranform - FFT (Cooley, Tukey 1965; Frigo, Johnson 1997-)

$$\mathcal{O}\left(N^d \log N\right)$$

- nonequispaced FFT (Dutt, Rokhlin 1993; Beylkin 1995-; Potts, Steidl, Tasche 1997-; Greengard, Lee 2004; Potts, K. 2002-)

$$\mathcal{O}\left(N^d \log N + |\log \varepsilon|^d M\right)$$

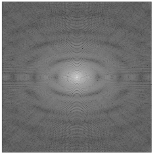
- software NFFT3.0 (Keiner, Potts, K.)

http://www.tu-chemnitz.de/∼potts/nfft

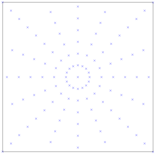
Nonequispaced FFT

- building block for
  - computation with curvelets, ridgelets
  - sparse FFTs (Gilbert, Muthukrishnan, Strauss 2005)
  - computation with RBF kernels (fast Gauss transform)
  - reconstruction schemes in computerised tomography, magnetic resonance imaging
- typical example from CT

phantom



Fourier transform (log)



polar grid

Reconstruction from samples

- reconstruction of signals from their Fourier transform, common practice (Shannon sampling theorem):

  "sampling rate" $\sim$ "bandwidth"   (Nyquist criteria)

- $\mathbb{T} = [-\frac{1}{2}, \frac{1}{2})$, $N \in 2\mathbb{N}$, $\hat{f}_k \in \mathbb{C}$; consider the trigonometric polynomial

$$f : \mathbb{T} \to \mathbb{C}, \quad f(x) = \sum_{k=-N/2}^{N/2-1} \hat{f}_k \mathrm{e}^{-2\pi \mathrm{i} k x}$$

- f can be reconstructed from $M \geq N$ samples $y_j = f(x_j)$
- conditions for $d > 1$ ...

Reconstruction from samples

- least squares approximation of $(x_j, y_j) \in \mathbb{T}^d \times \mathbb{C}$

$$\min_f \sum_{j=0}^{M-1} |f(x_j) - y_j|^2$$

well conditioned if $\max_j |x_j - x_{j+1}| < 1/N$ (Gröchenig 1992)

- minimal norm interpolation of $(x_j, y_j) \in \mathbb{T}^d \times \mathbb{C}$

$$\min_f \|f\|_2 \quad \text{s.t.} \quad f(x_j) = y_j$$

well conditioned if $\min_j |x_j - x_{j+1}| > 1.6/N$ (Potts, K. 2006)

Reconstruction from samples

- for support $T \subset I_N = \{-\frac{N}{2}, \ldots, \frac{N}{2} - 1\}$, $S = |T| \ll N$, we consider sparse trigonometric polynomials

$$f : \mathbb{T} \to \mathbb{C}, \quad f(x) = \sum_{k \in T} \hat{f}_k \mathrm{e}^{-2\pi \mathrm{i} kx}$$

- (non)linear spaces of trigonometric polynomials

$$\Pi_T \subset \Pi_{I_N}, \quad \text{vs.} \quad \Pi_{I_N}(S) = \bigcup_{T \subset I_N : |T| \leq S} \Pi_T$$

- reconstruct $f \in \Pi_{I_N}(S)$ from samples y_j at nodes $x_j \in \mathbb{T}$, i.e.,

$$y_j = f(x_j) = \sum_{k \in I_N} \hat{f}_k \mathrm{e}^{-2\pi \mathrm{i} k x_j}, \quad j = 0, \ldots, M-1$$

Reconstruction from samples

- dimension N, Fourier coefficients $\hat{\mathbf{f}} \in \mathbb{C}^N$
- sparsity $S = |T|$, support $T = \operatorname{supp}(\hat{\mathbf{f}})$
- number of samples M, samples $(x_j, y_j) \in \mathbb{T} \times \mathbb{C}$
- interesting case

$$S \sim M \ll N$$

- nonequispaced Fourier matrix and its T_s-restriction

$$\mathbf{A} = (\mathrm{e}^{-2\pi \mathrm{i} k x_j})_{j=0,\ldots,M-1; k \in I_N} = (\ldots \phi_k | \phi_{k+1} \ldots) \in \mathbb{C}^{M \times N}$$
$$\mathbf{A}_{T_s} = (\phi_k)_{k \in T_s} \in \mathbb{C}^{M \times |T_s|}$$

- sampling a trigonometric polynomial

$$\mathbf{y} = \mathbf{A}\hat{\mathbf{f}}$$

Algorithms and analysis - Thresholding

Input: $\mathbf{y} \in \mathbb{C}^M$, maximum sparsity $S \in \mathbb{N}$

1: find $T \subset I_N$ to the S largest inner products $\{|\langle \mathbf{y}, \phi_l \rangle|\}_{l \in I_N}$
2: solve $\|\mathbf{A}_T \mathbf{c} - \mathbf{y}\|_2 \xrightarrow{\mathbf{c}} \min$
3: $(\hat{f}_k)_{k \in T} = \mathbf{c}$

Output: $\hat{\mathbf{f}} \in \mathbb{C}^N$, $T \subset I_N$

Remark:

- we might hope that $M^{-1}\langle \mathbf{y}, \phi_l \rangle = M^{-1} \sum_{j=0}^{M-1} f(x_j) \mathrm{e}^{2\pi \mathrm{i} l x_j} \approx \int_{\mathbb{T}} f(x) \mathrm{e}^{2\pi \mathrm{i} l x} \mathrm{d}x = \hat{f}_l$
- computation of the inner products by $(\langle \mathbf{y}, \phi_l \rangle)_{l \in I_N} = \mathbf{A}^{\mathsf{H}} \mathbf{y}$ in $\mathcal{O}(N \log N + M)$ floating point operations

Algorithms and analysis - Orthogonal Matching Pursuit

Input: $\mathbf{y} \in \mathbb{C}^M$, $\varepsilon > 0$

1: $s = 0$, $\mathbf{r}_0 = y$, $T_0 = \emptyset$
2: **repeat**
3: $\quad s = s + 1$
4: $\quad T_s = T_{s-1} \cup \{\arg\max_{k \in I_N} |\langle \mathbf{r}_{s-1}, \phi_k \rangle|\}$
5: $\quad$ solve $\|\mathbf{A}_{T_s}\mathbf{d}_s - \mathbf{y}\|_2 \xrightarrow{\mathbf{d}_s} \min$
6: $\quad \mathbf{r}_s = \mathbf{y} - \mathbf{A}_{T_s}\mathbf{d}_s$
7: **until** $s = M$ or $\|\mathbf{r}_s\| \leq \varepsilon$
8: $T = T_s$, $(\hat{f}_k)_{k \in T} = \mathbf{d}_s$

Output: $\hat{\mathbf{f}} \in \mathbb{C}^N$, $T \subset I_N$

Algorithms and analysis - Thresholding

Theorem 1. [Rauhut,K.]

- fix $f \in \Pi_{I_N}(S)$
- define its dynamic range by

$$R = \frac{\max_{k \in T} |\hat{f}_k|}{\min_{k \in T} |\hat{f}_k|}$$

- choose sampling nodes $x_0, \ldots, x_{M-1}$ independently and uniformly at random on $\mathbb{T}$ or on the grid $\frac{1}{N} I_N$
- if for some $\epsilon > 0$

$$M \geq C R^2 \cdot S \cdot \log(N/\epsilon)$$

- then with probability at least $1 - \epsilon$ thresholding recovers T and $\hat{\mathbf{f}}$

Algorithms and analysis - OMP

Theorem 2. [Rauhut,K.]

- fix $f \in \Pi_{I_N}(S)$, choose sampling nodes as before
- if for some $\epsilon > 0$

$$M \geq C \cdot S \cdot \log(N/\epsilon),$$

then with probability at least $1 - \epsilon$ orthogonal matching pursuit selects $k_1 \in \text{supp}(\hat{\mathbf{f}})$

Theorem 3. [Rauhut,K.]

- choose sampling nodes as before
- if for some $\epsilon > 0$

$$M \geq C \cdot S^2 \cdot \log(N/\epsilon)$$

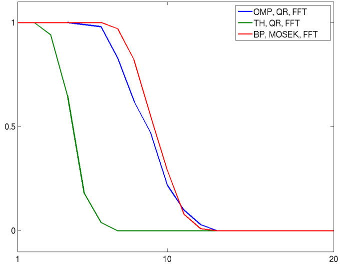
then with probability at least $1 - \epsilon$ OMP recovers every $f \in \Pi_{I_N}(S)$

Algorithms and analysis- recent results

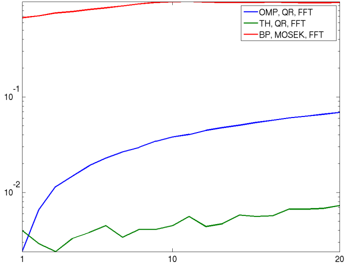
- Rauhut: If
$$M \leq C \cdot S^2/\sigma,$$
then with probability exceeding $1 - c_1/S - c_2/\sigma^2$ there exists an $f \in \Pi_{I_N}(S)$ on which tresholding fails. Similar result for OMP with S iterations.
- Needell, Vershynin: If
$$M \geq C \cdot S \cdot \log^*(N) \log(1/\epsilon),$$
then with probability at least $1 - \epsilon$ a slighly more expensive greedy method recovers every $f \in \Pi_{I_N}(S)$.
- Donoho, Tsaig: greedy methods for basis pursuit (2).

Numerical results

fixed dimension $N = 1000$, fixed number of samples $M = 40$, normalised Fourier coefficients $|\hat{f}_k| = 1$



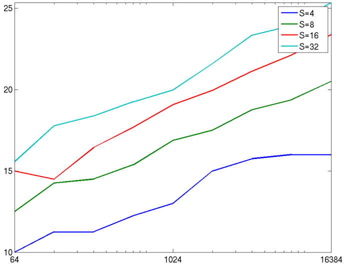
reconstruction rate vs. sparsity S

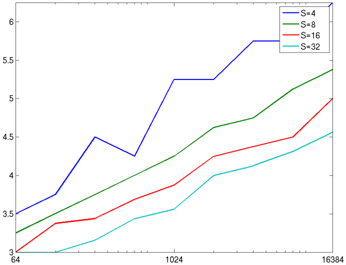


computation time vs. sparsity S

Numerical results

fixed reconstruction rate 90%, fixed sparsity S





Thresholding ($|\hat{f}_k| = 1$) OMP

generalised oversampling factor M/S vs. dimension N

Numerical results - Implementation

- fast matrix vector arithmetic with **A**
- least squares solver
  - QR factorisation with insert
  - LSQR (uniformly bounded condition number, Rauhut)

- available
  - Random sampling of sparse trigonometric polynomials II - orthogonal matching pursuit versus basis pursuit.

    (with Holger Rauhut, Found. Comput. Math., to appear)
  - MATLAB toolbox OMP4NFFT (with Holger Rauhut)
  - C subroutine library NFFT3 (with Jens Keiner, Daniel Potts)

www.tu-chemnitz.de/∼skunis

... - Matching Pursuit (Pure Greedy)

Input: $\mathbf{y} \in \mathbb{C}^M$, $\varepsilon > 0$, maximum number of iterations $L \in \mathbb{N}$

$s = 0$, $\mathbf{r}_0 = \mathbf{y}$, $T_0 = \emptyset$, $\hat{\mathbf{f}} = 0$
repeat
  $s = s + 1$
  $k_s = \arg\max_{k \in I_N} |\langle \mathbf{r}_{s-1}, \phi_k \rangle|$
  $\hat{f}_{k_s} = \hat{f}_{k_s} + \langle \mathbf{r}_{s-1}, \phi_{k_s} \rangle$
  $\mathbf{r}_s = \mathbf{r}_{s-1} - \langle \mathbf{r}_{s-1}, \phi_{k_s} \rangle \phi_{k_s}$
  $T_s = T_{s-1} \cup \{k_s\}$
until $s = L$ or $\|\mathbf{r}_s\| \leq \varepsilon$
$T = T_s$

Output: $\hat{\mathbf{f}} \in \mathbb{C}^N$, $T \subset I_N$

... - Sketch of proof

- fix $T \subset I_N$, $\mathbf{c} \in \mathbb{C}^S$, and choose M sampling nodes independently and uniformly at random on $\mathbb{T}$ or on $\frac{1}{N} I_N$
- for $k \notin T$ and $\delta > 0$ holds

$$\mathbb{P}\left(|\langle \mathbf{A}_T \mathbf{c}, \phi_k \rangle| \geq M\delta\right) \leq 4 \exp\left(-\frac{M\delta^2}{4\|\mathbf{c}\|_2^2 + \frac{4}{3\sqrt{2}}\|\mathbf{c}\|_1 \delta}\right)$$

- Remark: this quantifies the "quadrature rule"

$$\begin{aligned}\langle \mathbf{A}_T \mathbf{c}, \phi_k \rangle = \langle \mathbf{y}, \phi_k \rangle &= \sum_{j=0}^{M-1} f(x_j) \mathrm{e}^{2\pi \mathrm{i} k x_j} \\ &\approx M \cdot \int_{\mathbb{T}} f(x) \mathrm{e}^{2\pi \mathrm{i} k x} \mathrm{d}x = 0\end{aligned}$$

... - Sketch of proof

- thresholding recovers the correct support if

$$\min_{j \in T} |\langle \phi_j, \mathbf{A}_T \mathbf{c} \rangle| > \max_{k \notin T} |\langle \phi_k, \mathbf{A}_T \mathbf{c} \rangle|$$

- for $l \in T$, the triangle inequality yields

$$\begin{aligned} |M^{-1}\langle \phi_l, \mathbf{A}_T \mathbf{c} \rangle| &= |c_l + M^{-1}\langle \phi_l, \mathbf{A}_{T\setminus\{l\}} \mathbf{c}_{T\setminus\{l\}} \rangle| \\ &\geq \min_{j \in T} |c_j| - \max_{j \in T} |M^{-1}\langle \phi_j, \mathbf{A}_{T\setminus\{j\}} \mathbf{c}_{T\setminus\{j\}} \rangle| \end{aligned}$$

- hence, thresholding succeeds if

$$\begin{aligned} \max_{j \in T} |M^{-1}\langle \phi_j, \mathbf{A}_{T\setminus\{j\}} \mathbf{c}_{T\setminus\{j\}} \rangle| &< \min_{j \in T} |c_j|/2 \\ \max_{k \notin T} |M^{-1}\langle \phi_k, \mathbf{A}_T \mathbf{c} \rangle| &< \min_{j \in T} |c_j|/2 \end{aligned}$$