Iterative Hard Thresholding: Theory and Practice

T. Blumensath
Institute for Digital Communications
Joint Research Institute for Signal and Image Processing
The University of Edinburgh

February, 2009

What to Expect

I) Iterative Hard Thresholding for Sparse Inverse Problems

What's the probelm?

II) Theory and Practice

Nice theory, bad attitude

III) Stability Beyond the Grave (RIP)

Operating beyond the limits

IV) Results

Up there with the best

PART 1

Iterative Hard Thresholding for Sparse Inverse Problems

The Problem and Solution

THE PROBLEM: Given

$$\mathbf{y} = \mathbf{\Phi x} + \mathbf{e},$$

where $\mathrm{y} \in \mathbb{R}^M$, $\mathrm{x} \in \mathbb{R}^N$, $\mathbf{\Phi} \in \mathbb{R}^{M \times N}$ and e is observation noise, estimate x given y and $\mathbf{\Phi}$ when $M << N$ but x is approximately K-sparse.

THE SOLUTION: The Iterative Hard Thresholding (IHT) algorithm uses the iteration

$$\mathbf{x}^{n+1} = P_K(\mathbf{x}^n + \mathbf{\Phi}^T(\mathbf{y} - \mathbf{\Phi x}^n)),$$

where P_K is a hard thresholding operator that keeps the largest (in magnitude) K elements of a vector (Or, more generally, a projector onto the closest element in the model).

PART 2

Theory and Practice

Convergence and Recovery Performance

CONVERGENCE: IHT is guaranteed to converge to a local minimum of $\|\mathbf{y} - \mathbf{\Phi x}\|_2^2 \text{ s. t. } \|\mathbf{x}\|_0 \leq K$ whenever $\|\mathbf{\Phi}\|_2 \leq 1$.

RECOVERY: If $\delta_{3K}(\mathbf{\Phi}) \leq 1/\sqrt{32}$, then after at most $\left\lceil \log_2 \left(\frac{\|\mathbf{x}_K\|_2}{\tilde{\epsilon}_K} \right) \right\rceil$ iterations, the IHT approximation $\mathbf{x}^\star$ satisfies

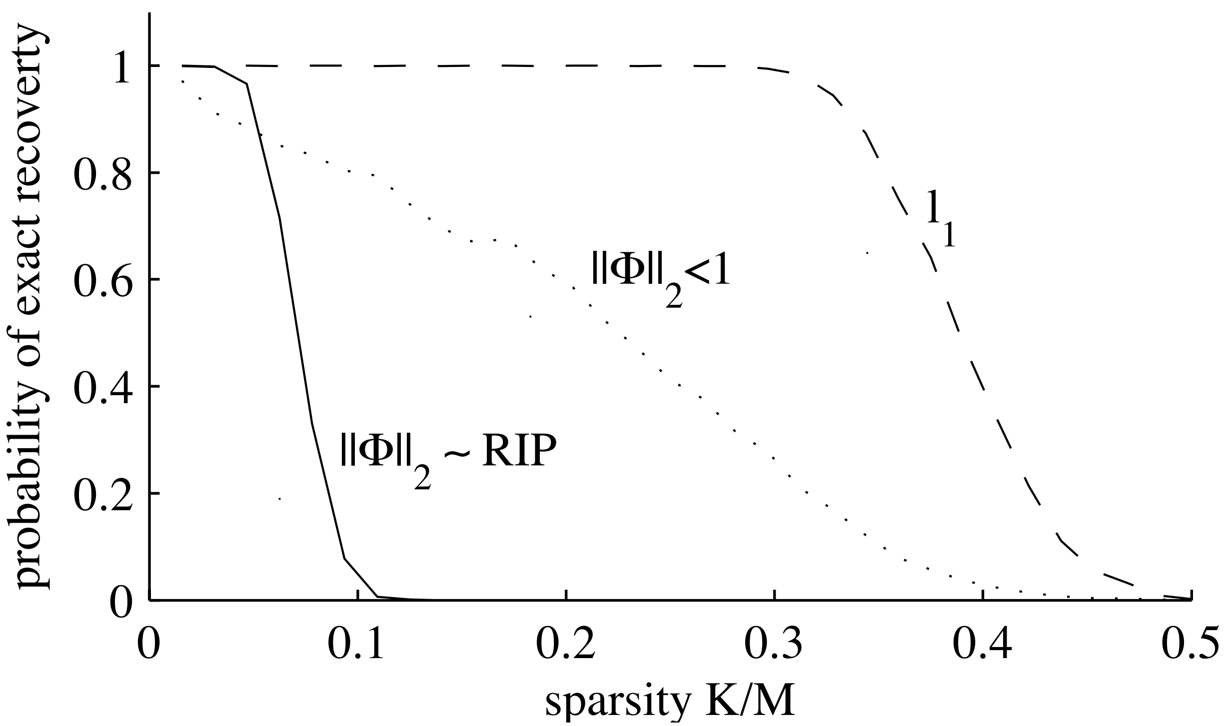
$$\|\mathbf{x}^\star - \mathbf{x}\|_2 \leq 7\tilde{\epsilon}_K.$$

where $\tilde{\epsilon}_K = \|\mathbf{x} - \mathbf{x}_K\|_2 + \frac{\|\mathbf{x} - \mathbf{x}_K\|_1}{\sqrt{K}} + \|\mathbf{e}\|_2$ and where $\delta_K(\mathbf{\Phi})$ is the smallest constant for which

$$(1 - \delta_K(\mathbf{\Phi}))\|\mathbf{x}\|_2^2 \leq \|\mathbf{\Phi x}\|_2^2 \leq (1 + \delta_K(\mathbf{\Phi}))\|\mathbf{x}\|_2^2$$

holds for all K sparse $\mathbf{x}$.

But


probability of exact recoverty
1
0.8
0.6
0.4
0.2
0
0
0.1
0.2
0.3
0.4
0.5
sparsity K/M
||Φ||2 ~ RIP
||Φ||2<1
l1

PART 3

Stability Beyond RIP

The Normalised Iterative Hard Thresholding Algorithm

The Normalised Iterative Hard Thresholding (NIHT) algorithm uses the iteration

$$\mathbf{x}^{n+1} = P_K(\mathbf{x}^n + \mu^n \mathbf{\Phi}^T(\mathbf{y} - \mathbf{\Phi x}^n)),$$

where P_K is a hard thresholding operator that keeps the largest (in magnitude) K elements of a vector (or, more generally, a projector onto the closest element in the model) and μ^n is a step-size.

Calculating the step size

Assume the support of $\mathbf{x}^n = \Gamma^n$ and that the support of $\mathbf{x}^{n+1} = \Gamma^{n+1} = \Gamma^n$, then the optimal step-size is (in terms of reduction in squared approximation error)

$$\mu^n = \frac{\mathbf{g}_{\Gamma^n}^T \mathbf{g}_{\Gamma^n}}{\mathbf{g}_{\Gamma^n}^T \mathbf{\Phi}_{\Gamma^n}^T \mathbf{\Phi}_{\Gamma^n} \mathbf{g}_{\Gamma^n}},$$

where $\mathbf{g} = \mathbf{\Phi}^T(\mathbf{y} - \mathbf{\Phi x}^n)$.

However, if $\Gamma^{n+1} \neq \Gamma^n$, this step-size might not be optimal and does not guarantee convergence. For guaranteed convergence we require that:

$$\mu \leq (1-c)\frac{\|\mathbf{x}^{n+1} - \mathbf{x}^n\|_2^2}{\|\mathbf{\Phi}(\mathbf{x}^{n+1} - \mathbf{x}^n)\|_2^2},$$

for some $c > 0$.

Hence, if $\Gamma^{n+1} \neq \Gamma^n$, calculate $\omega = (1-c)\frac{\|\mathbf{x}^{n+1} - \mathbf{x}^n\|_2^2}{\|\mathbf{\Phi}(\mathbf{x}^{n+1} - \mathbf{x}^n)\|_2^2}$ and, if $\mu^n > \omega$, set $\mu^n \leftarrow \mu^n / (\kappa(1-c))$ or, alternatively, set $\mu^n \leftarrow \omega$.

Why the Hassle?

CONVERGENCE: NIHT is guaranteed to converge to a local minimum of $\|\mathbf{y} - \mathbf{\Phi x}\|_2^2 \text{ s. t. } \|\mathbf{x}\|_0 \leq K$**.**

RECOVERY: If Φ satisfies $0 < \alpha_{2K} \leq \frac{\|\mathbf{\Phi x}\|_2}{\|\mathbf{x}\|_2} \leq \beta_{2K}$ for all $\mathbf{x} : \|\mathbf{x}\|_0 \leq 2K$. Given a noisy observation $\mathbf{y} = \mathbf{\Phi x} + \mathbf{e}$, where $\mathbf{x}$ is an arbitrary vector, let $\mathbf{x}^K$ be the best K-term approximation to $\mathbf{x}$.

If $\gamma_{2K} = \max\{1 - \frac{\alpha_{2K}^2}{\kappa\beta_{2K}^2}, \frac{\beta_{2K}^2}{\alpha_{2K}^2} - 1\} < 1/8$, then after at most $n^\star = \left\lceil \log_2 \left(\|\mathbf{x}^K\|_2 / \tilde{\epsilon}_K \right) \right\rceil$ iterations, IHT_K estimates $\mathbf{x}$ with accuracy given by

$$\|\mathbf{x} - \mathbf{x}^{n^\star}\|_2 \leq 9\tilde{\epsilon}_K, \tag{1}$$

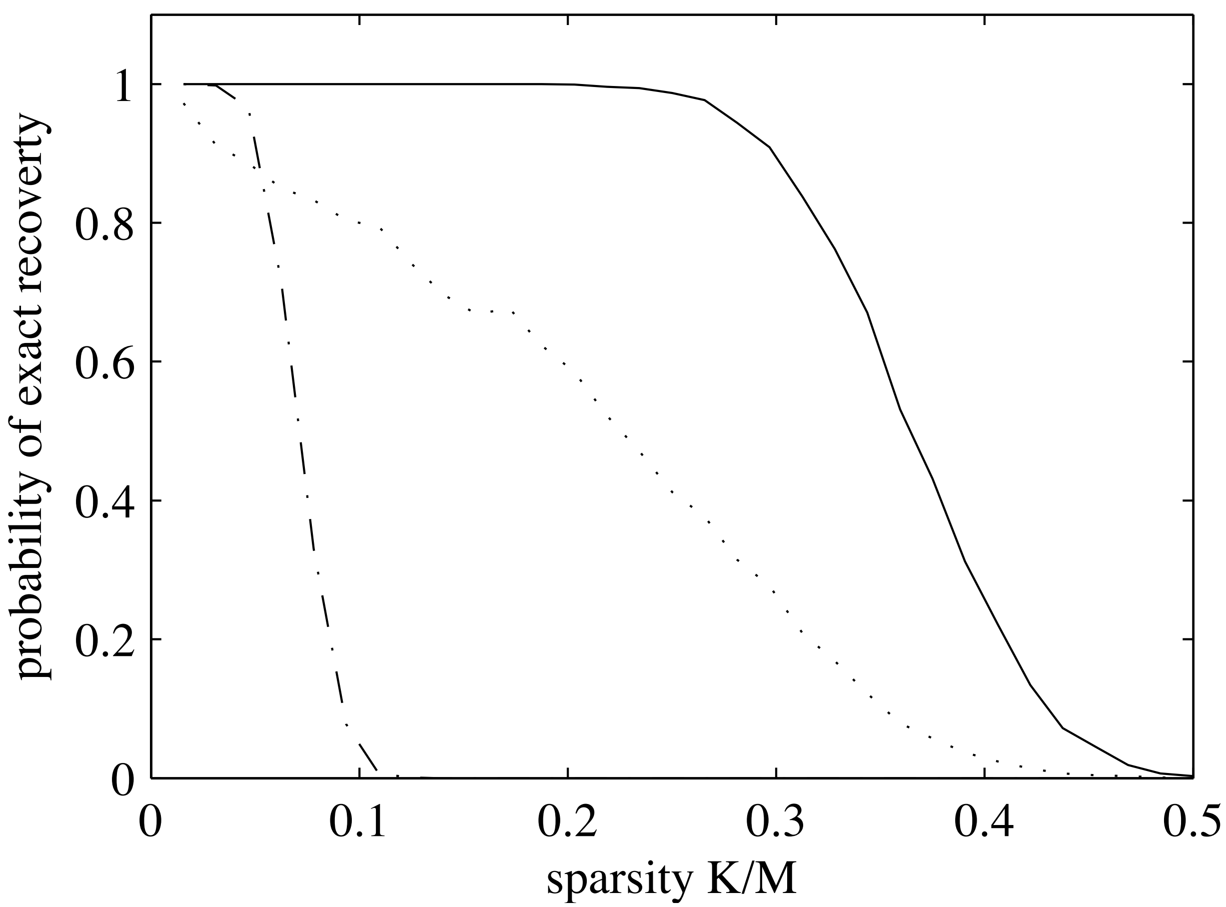
where

$$\tilde{\epsilon}_K = \|\mathbf{x} - \mathbf{x}^K\|_2 + \frac{\|\mathbf{x} - \mathbf{x}^K\|_1}{\sqrt{K}} + \frac{1}{\alpha_{2K}}\|\mathbf{e}\|_2. \tag{2}$$

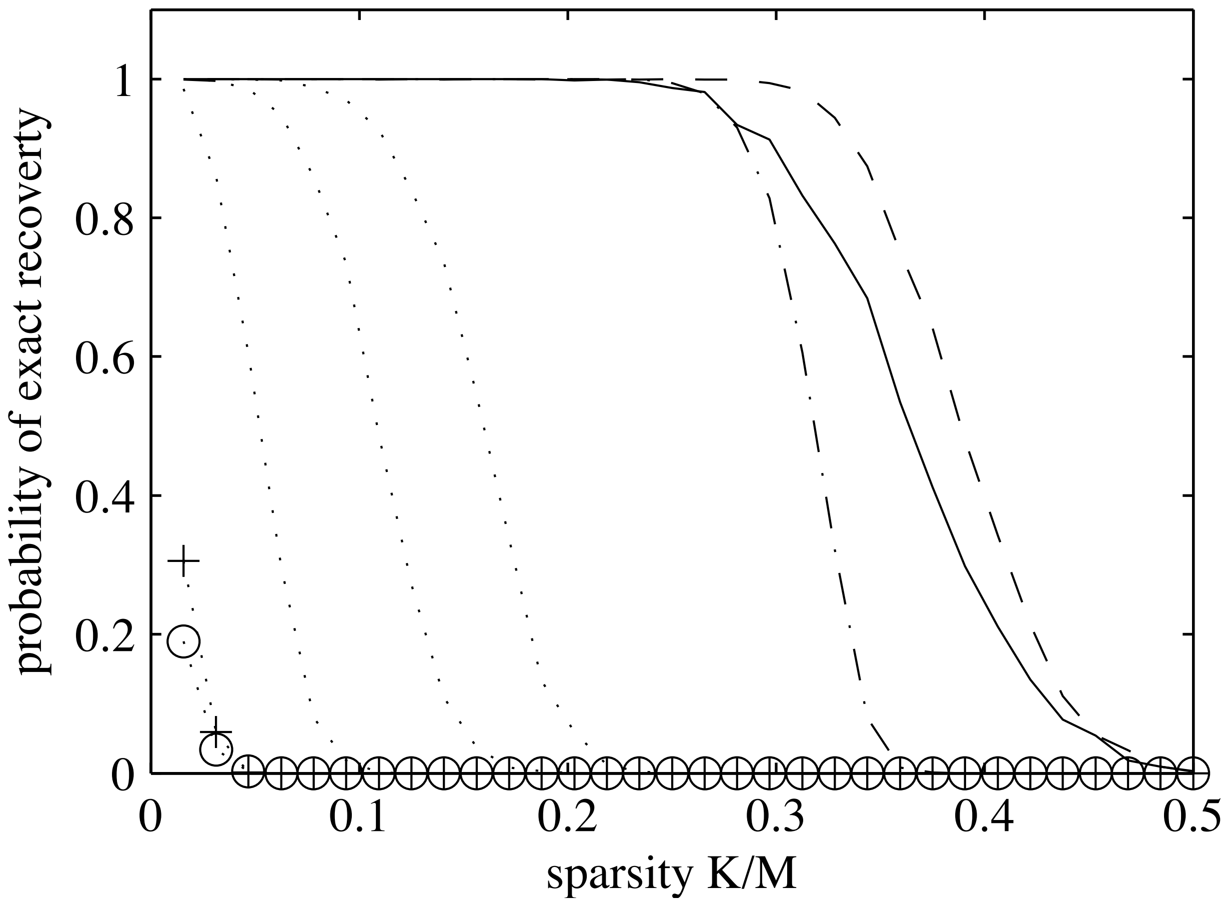
PART 4

Results

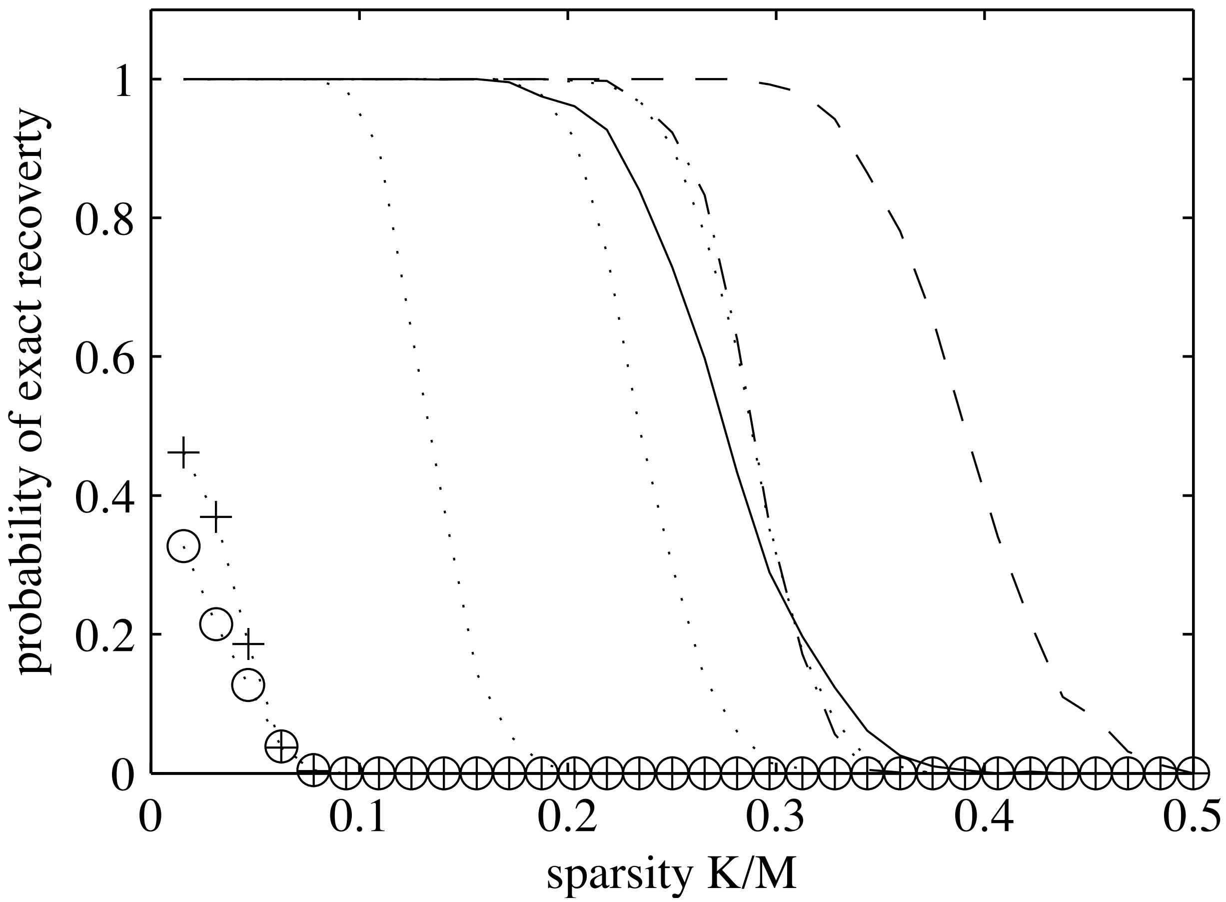
Before and After



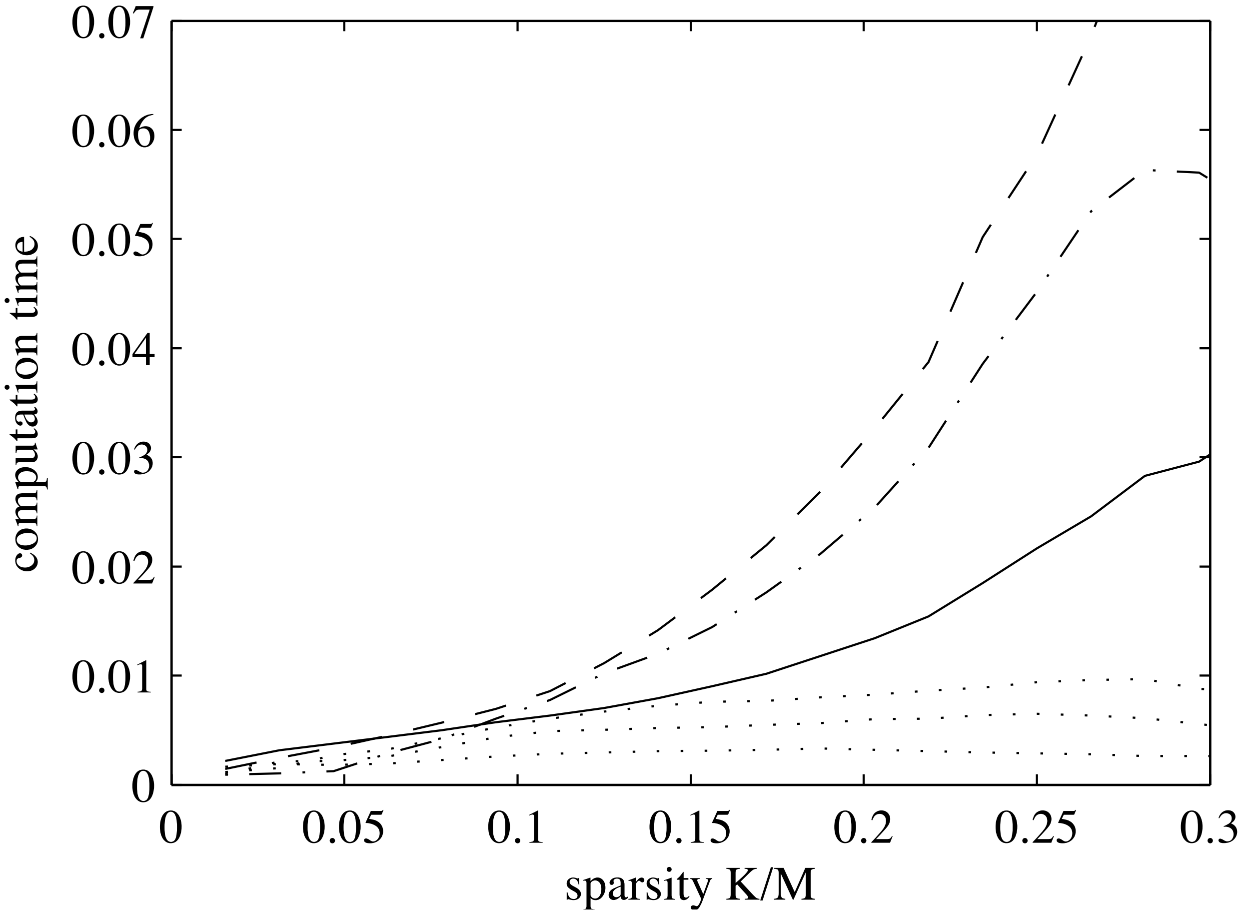
Comparison to other Algorithms



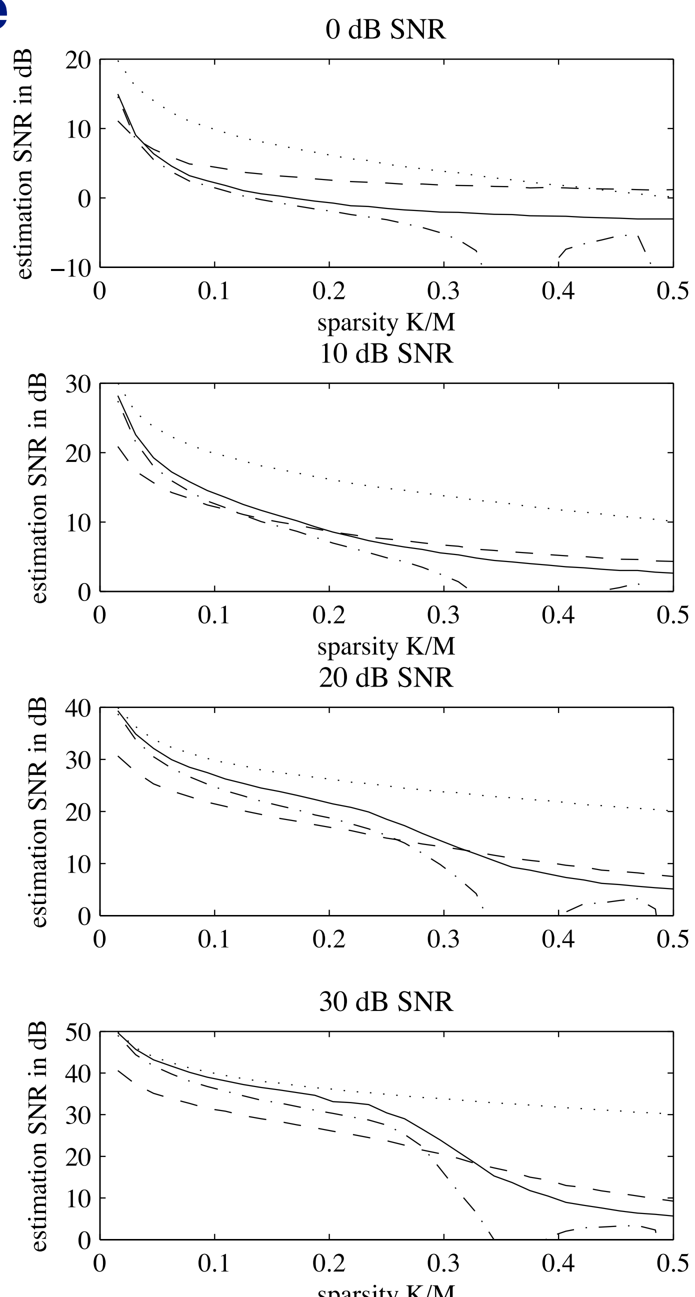
Comparison to other Algorithms



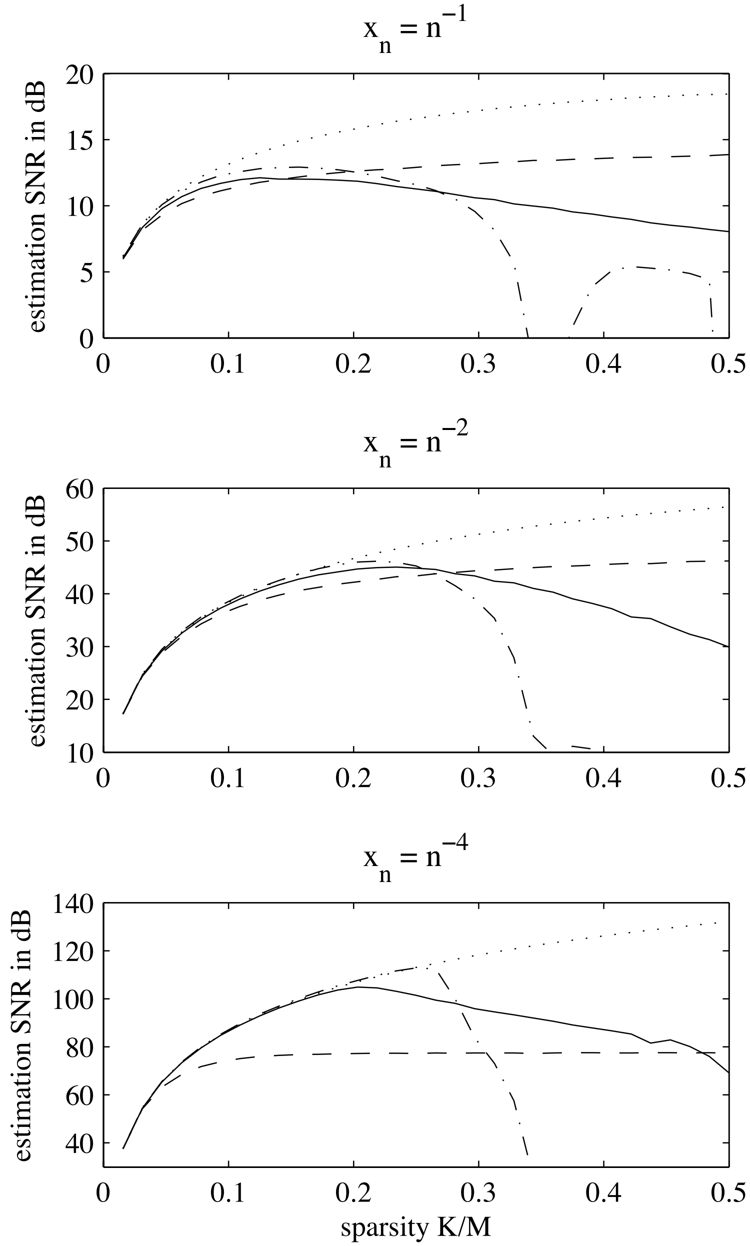
Speed Comparison



Robustness to Noise

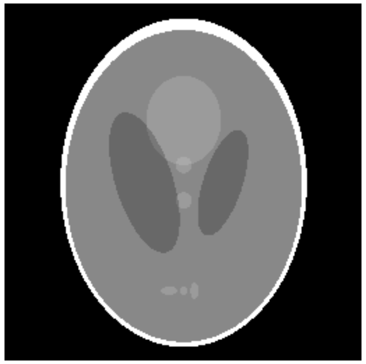


Robustness to Non-Exact-Sparsesness



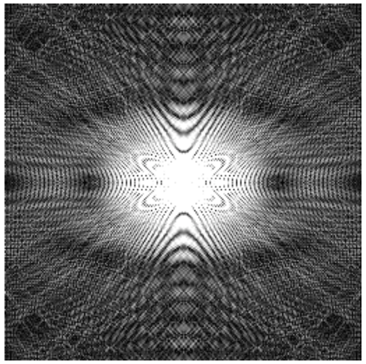
Larger Problems

Original / Reconstruction

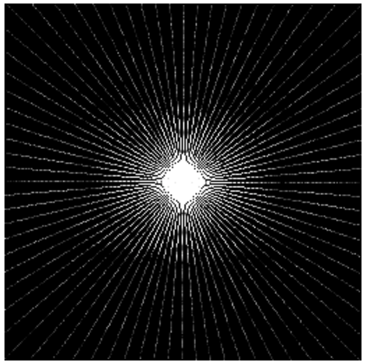


Haar Wavelet Transform

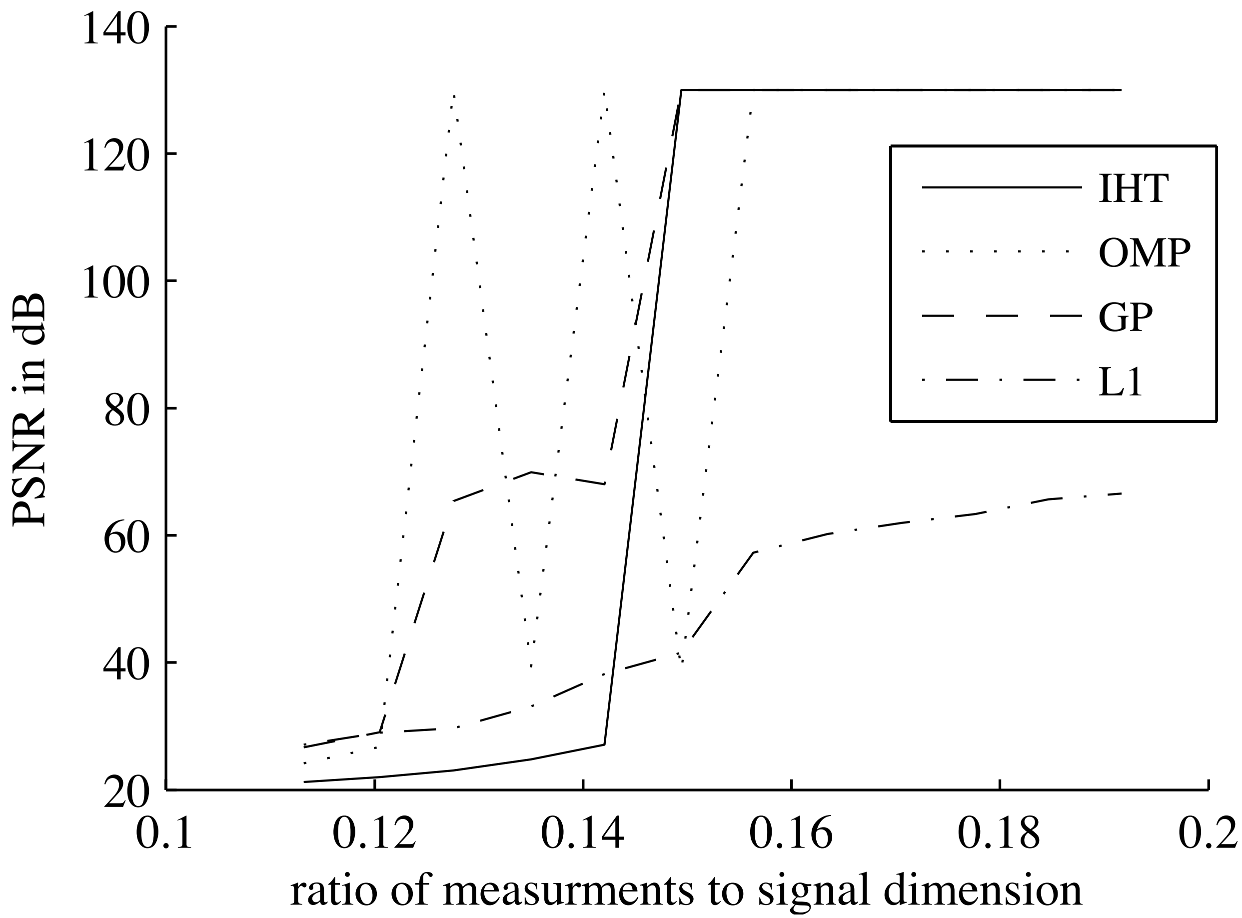
Frequency Domain



Observation



Larger Problems

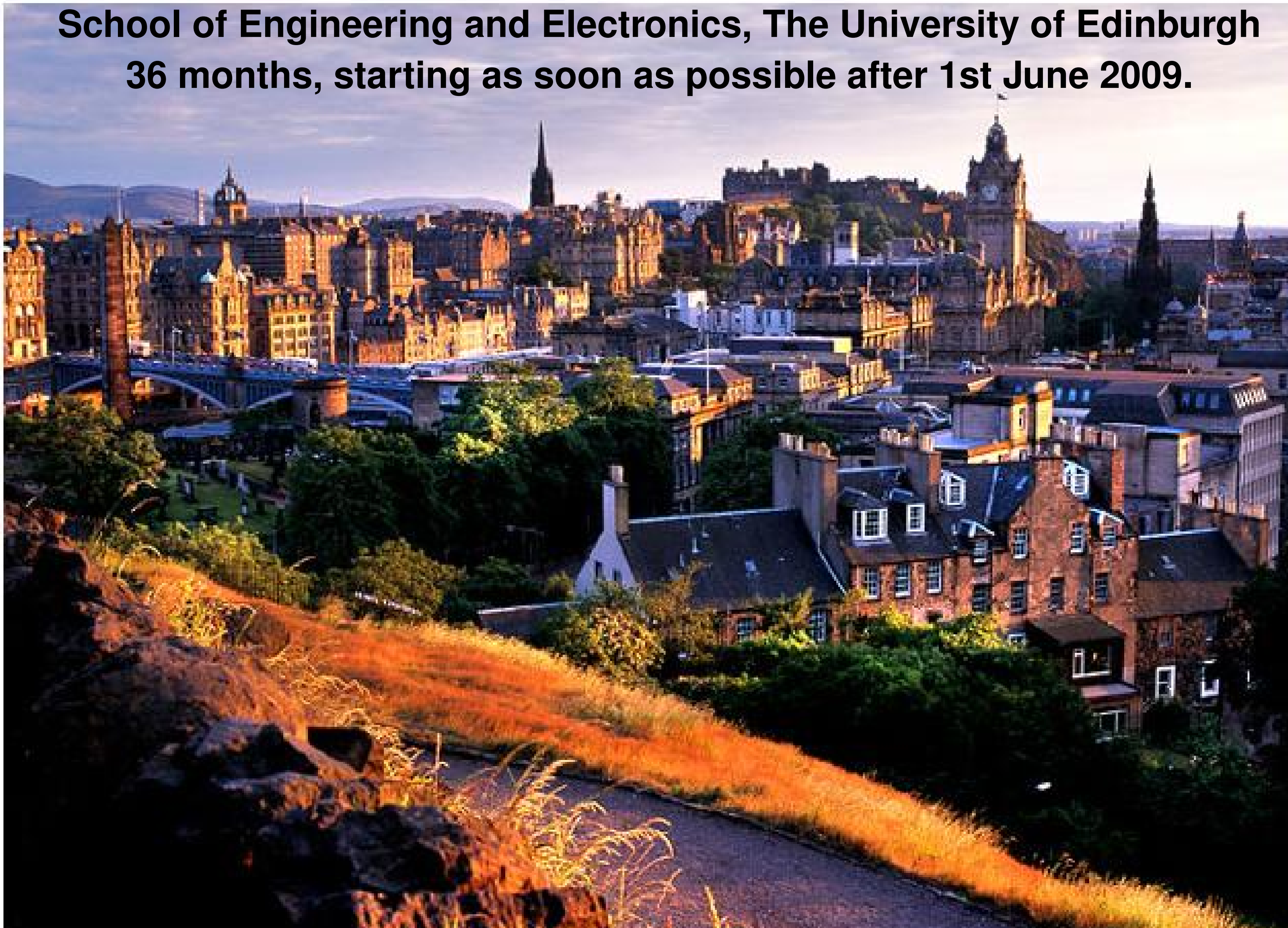


Sparse representations for the processing of large scale data

Compressed sensing theory and algorithms with particular application to dynamic MRI

School of Engineering and Electronics, The University of Edinburgh

36 months, starting as soon as possible after 1st June 2009.



Salary scale: 28,290 - 33,779 pa. (pounds sterling)

Informal enquiries to Prof. Mike Davies (mike.davies@ed.ac.uk)

For further details please go to: http://www.jobs.ed.ac.uk/