Greedy Methods for Generalized Sparse Approximation

Deanna Needell

September 26, 2014

This work was supported by NSF grant DMS-1045536, NSF Career grant #1348721, and the Alfred P. Sloan Fellowship.

Collaborators

Joint work with Alison Kingman, Chris Garnatz, James LaManna, Shenyinying Tu, Xiaoyi Gu
(UCLA-Claremont Summer REU 2014)

What is Compressed Sensing?

- It is a method of compression that can compress data efficiently.

What is Compressed Sensing?

- It is a method of compression that can compress data efficiently.
- We can obtain this compression without having to acquire the entire object.

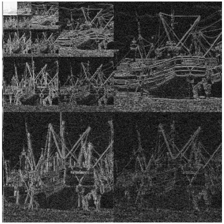
What is Compressed Sensing?

- It is a method of compression that can compress data efficiently.
- We can obtain this compression without having to acquire the entire object.
- CS works because most signals contain less information than their dimension would suggest.

What is Compressed Sensing?

- It is a method of compression that can compress data efficiently.
- We can obtain this compression without having to acquire the entire object.
- CS works because most signals contain less information than their dimension would suggest.
- In our model we will only be working with *sparse* signals

Why is compression possible?

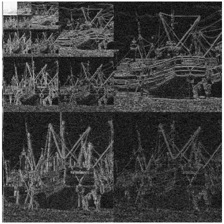


Assume f is s-sparse:

- In the coordinate basis: $\|f\|_0 \stackrel{\text{def}}{=} |\text{supp}(()f)| \leq s \ll d$

Why is compression possible?



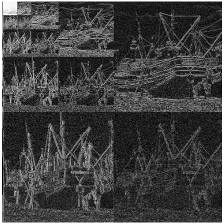


Assume f is s-sparse:

- In the coordinate basis: $\|f\|_0 \overset{\text{def}}{=} |\text{supp}(()f)| \leq s \ll d$
- In some orthonormal basis: $f = Dx$ where $\|x\|_0 \leq s \ll d$

Why is compression possible?





Assume f is s-sparse:

- In the coordinate basis: $\|f\|_0 \stackrel{\text{def}}{=} |\text{supp}(()f)| \leq s \ll d$
- In some orthonormal basis: $f = Dx$ where $\|x\|_0 \leq s \ll d$

In practice, we encounter compressible signals.

Mathematical Formulation

To compress a signal, we take a small number of measurements:

1. Signal of interest $x \in \mathbb{R}^n$

Mathematical Formulation

To compress a signal, we take a small number of measurements:

1. Signal of interest $x \in \mathbb{R}^n$
2. Measurement operator $A : \mathbb{R}^n \to \mathbb{R}^m$ $(m \ll n)$

Mathematical Formulation

To compress a signal, we take a small number of measurements:

1. Signal of interest $x \in \mathbb{R}^n$
2. Measurement operator $A : \mathbb{R}^n \to \mathbb{R}^m$ $(m \ll n)$
3. Measurements $y = Ax$.

$$\begin{bmatrix} y \end{bmatrix} = \begin{bmatrix} \quad A \quad \end{bmatrix} \begin{bmatrix} x \end{bmatrix}$$

Mathematical Formulation

To compress a signal, we take a small number of measurements:

1. Signal of interest $x \in \mathbb{R}^n$
2. Measurement operator $A : \mathbb{R}^n \to \mathbb{R}^m$ $(m \ll n)$
3. Measurements $y = Ax$.

$$\begin{bmatrix} y \end{bmatrix} = \begin{bmatrix} & & A & & \end{bmatrix} \begin{bmatrix} \\ x \\ \\ \end{bmatrix}$$

4. y is the compression of x!

Mathematical Formulation

To compress a signal, we take a small number of measurements:

1. Signal of interest $x \in \mathbb{R}^n$
2. Measurement operator $A : \mathbb{R}^n \to \mathbb{R}^m$ $(m \ll n)$
3. Measurements $y = Ax + \xi$.

$$\begin{bmatrix} y \end{bmatrix} = \begin{bmatrix} \quad & A & \quad \end{bmatrix} \begin{bmatrix} x \\ \\ \end{bmatrix} + \begin{bmatrix} \xi \end{bmatrix}$$

4. y is the compression of x!
5. And then the measurements get corrupted with noise.

MRI applications

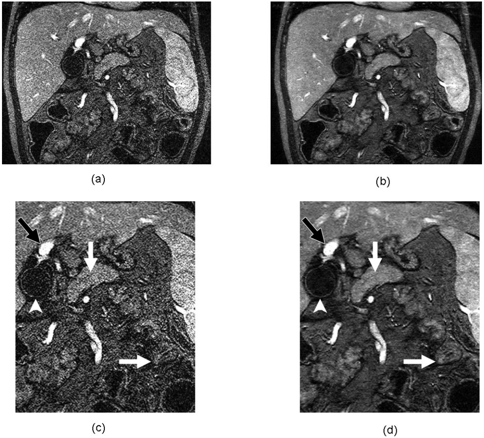


Figure: Two different MRIs done on a young child. The left figures took 45 minutes. The right only took 8 minutes using compressed sensing

Compressed Sensing Algorithms

- ℓ_1-minimization (Candès et. al., Donoho et. al.)

Compressed Sensing Algorithms

- ℓ_1-minimization (Candès et. al., Donoho et. al.)
- Orthogonal Matching Pursuit (OMP) (Gilbert-Tropp)

Compressed Sensing Algorithms

- ℓ_1-minimization (Candès et. al., Donoho et. al.)
- Orthogonal Matching Pursuit (OMP) (Gilbert-Tropp)
- Compressive Sampling Matching Pursuit (CoSaMP) (N-Tropp)

Compressed Sensing Algorithms

- ℓ_1-minimization (Candès et. al., Donoho et. al.)
- Orthogonal Matching Pursuit (OMP) (Gilbert-Tropp)
- Compressive Sampling Matching Pursuit (CoSaMP) (N-Tropp)
- Many others (IHT, SP, ...)

Compressive Sampling Matching Pursuit (CoSaMP) Algorithm

CoSaMP (N-Tropp)

---

input: Sampling operator A, measurements y, sparsity level s
initialize: Set $x^0 = 0$, $i = 0$.
repeat
signal proxy: Set $p = A^*(y - Ax^i)$, $\Omega = \mathrm{supp}(p_{2s})$, $T = \Omega \cup \mathrm{supp}(x^i)$.
signal estimation: Using least-squares, set $b|_T = A_T^\dagger y$ and $b|_{T^c} = 0$.
prune and update: Increment i and to obtain the next approximation, set $x^i = b_s$.
output: s-sparse reconstructed vector $\hat{x} = x^i$

Motivation

- Many CS algorithms can recover a signal with sparse representation in orthonormal dictionary

Motivation

- Many CS algorithms can recover a signal with sparse representation in orthonormal dictionary
- Unfortunately, most real world signals are sparse in non-orthonormal dictionaries

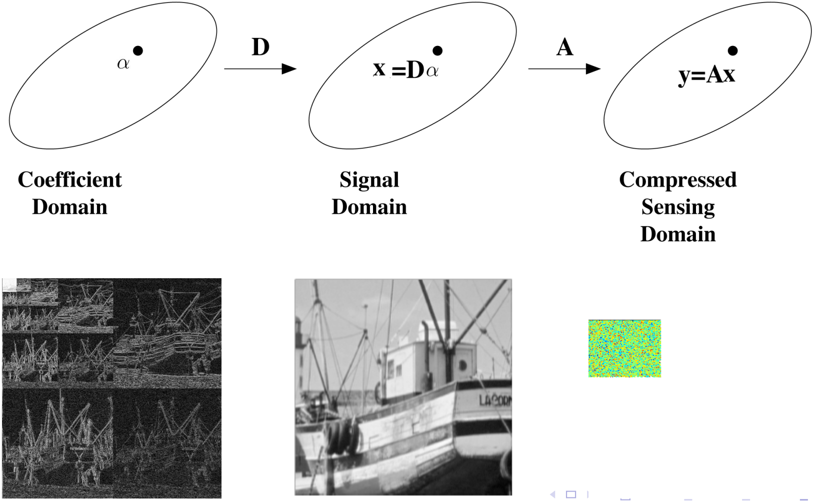
Motivation

- Many CS algorithms can recover a signal with sparse representation in orthonormal dictionary
- Unfortunately, most real world signals are sparse in non-orthonormal dictionaries
- Little to no theory exists guaranteeing recovery in the non-orthonormal case!

Motivation

- Many CS algorithms can recover a signal with sparse representation in orthonormal dictionary
- Unfortunately, most real world signals are sparse in non-orthonormal dictionaries
- Little to no theory exists guaranteeing recovery in the non-orthonormal case!
- Signal Space CoSaMP (Davenport-N-Wakin) attempts to provide a practical solution for poorly-behaved dictionaries

Compression



SSCoSaMP Algorithm

CoSaMP (N-Tropp)

---

input: Sampling operator A, measurements y, sparsity level s
initialize: Set $x^0 = 0$, $i = 0$.
repeat
signal proxy: Set $p = A^*(y - Ax^i)$, $\Omega = \mathrm{supp}(p_{2s})$, $T = \Omega \cup \mathrm{supp}(x^i)$.
signal estimation: Using least-squares, set $b|_T = A_T^\dagger y$ and $b|_{T^c} = 0$.
prune and update: Increment i and to obtain the next approximation, set $x^i = b_s$.
output: s-sparse reconstructed vector $\widehat{x} = x^i$

SSCoSaMP Details

- Unfortunately, finding best k- or $2k$-sparse representation is NP-hard

SSCoSaMP Details

- Unfortunately, finding best k- or $2k$-sparse representation is NP-hard
- Given only x, we need to find a sparse representation α such that $x = D\alpha$

SSCoSaMP Details

- Unfortunately, finding best k- or $2k$-sparse representation is NP-hard
- Given only x, we need to find a sparse representation α such that $x = D\alpha$
- This is actually its own compressed sensing problem!

SSCoSaMP Details

- Unfortunately, finding best k- or $2k$-sparse representation is NP-hard
- Given only x, we need to find a sparse representation α such that $x = D\alpha$
- This is actually its own compressed sensing problem!
- (A very ill-posed one)

SSCoSaMP Details

- Unfortunately, finding best k- or $2k$-sparse representation is NP-hard
- Given only x, we need to find a sparse representation α such that $x = D\alpha$
- This is actually its own compressed sensing problem!
- (A very ill-posed one)
- Instead, SSCoSaMP computes near-optimal sparse support using simpler CS algorithm such as L1, OMP, CoSaMP, etc...

SSCoSaMP Details

- Unfortunately, finding best k- or $2k$-sparse representation is NP-hard
- Given only x, we need to find a sparse representation α such that $x = D\alpha$
- This is actually its own compressed sensing problem!
- (A very ill-posed one)
- Instead, SSCoSaMP computes near-optimal sparse support using simpler CS algorithm such as L1, OMP, CoSaMP, etc...
- SSCoSaMP (Alg) denotes which algorithm is used for the identify/prune steps

SSCoSaMP vs. CoSaMP

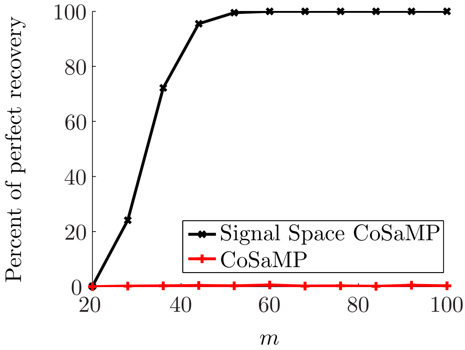


Figure: D is an orthogonal but non-normalized basis.

Difficulties with SSCoSaMP

- Near-optimal support has to satisfy certain conditions to guarantee accurate recovery of SSCoSaMP.

Difficulties with SSCoSaMP

- Near-optimal support has to satisfy certain conditions to guarantee accurate recovery of SSCoSaMP.
- Theroetical guarantees rely on strong conditions for the near-optimal approximation, and these conditions do *not* depend on the signal structure.

Difficulties with SSCoSaMP

- Near-optimal support has to satisfy certain conditions to guarantee accurate recovery of SSCoSaMP.
- Theroetical guarantees rely on strong conditions for the near-optimal approximation, and these conditions do *not* depend on the signal structure.
- However...

Behavior

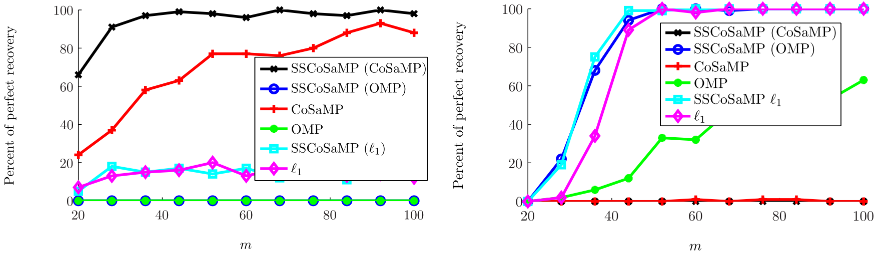


Figure: The figure on the left in the case where the non-zeros of α are clustered together. The figure on the left is the case where the non-zeros are well separated.

Signal Tests

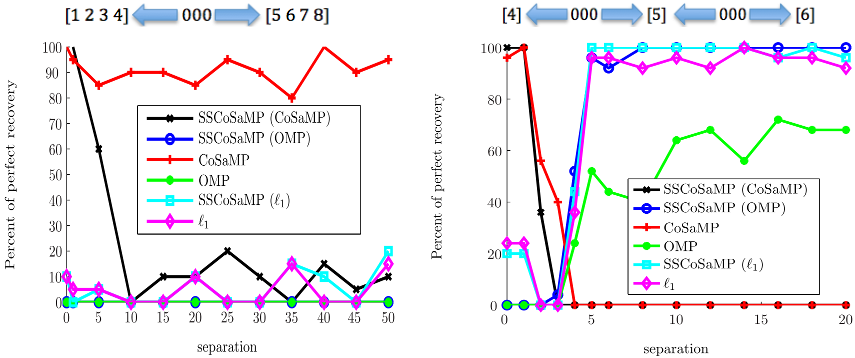


Figure: Left: separations represent the number of zeros between two clusters size $k/2$. Right: separations represent the number of zeros between each nonzero entry. Measurements and sparsity are $m = 100$ and $k = 8$, respectively with a $4\times$ overcomplete DFT dictionary.

Hybrid signal

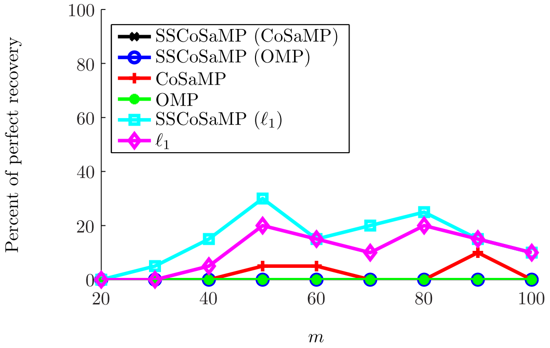


Figure: SSCoSaMP recovering a sparse vector with a hybrid sparse support: a block of $k/2$ nonzeros with the remaining $k/2$ nonzeros spaced at least 8 slots apart from all other nonzeros.

Neighborly Orthogonal Matching Pursuit (NOMP)

- In each iteration, OMP adds the largest coordinate of the proxy signal to the support set.
- Experimental results show that OMP only performs well when recovering well separated signals.
- NOMP is an alteration to OMP in that it takes coordinates adjacent to the largest one.
- In our experiments, we used NOMP with a window of six coordinates.

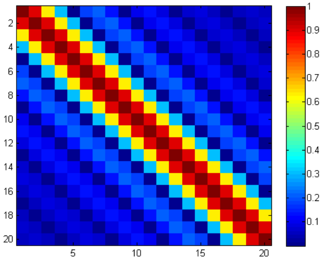


Figure: Section of a 1024 × 1024, 4× overcomplete DFT dictionary. NOMP takes advantage of the correlated columns.

NOMP

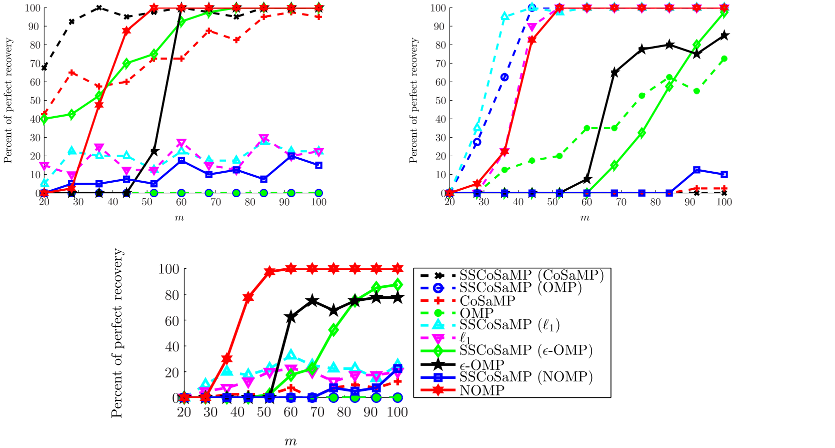


Figure: Percent perfect recovery of clustered signals (left) and well separated signals (right) and hybrid signals (bottom). NOMP is the only algorithm that performs well in all three cases.

NOMP

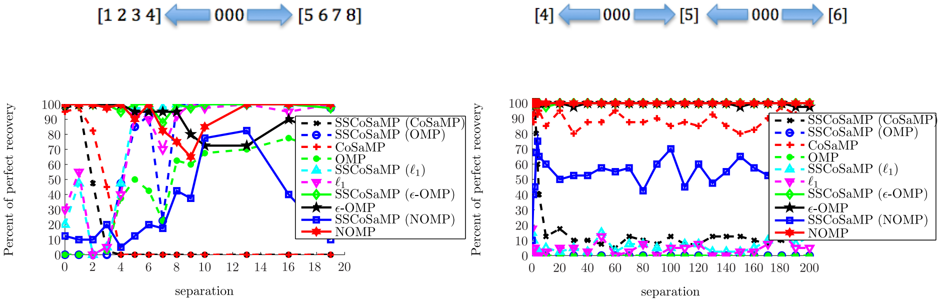


Figure: Left: Increasing the separation between single coefficients. Right: Increasing the separation between two clusters of coefficients.

NOMP

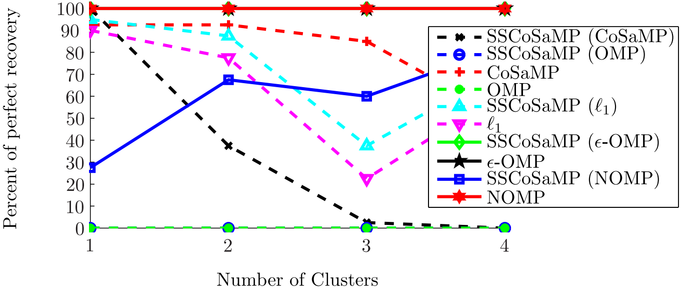


Figure: Percent perfect recovery as the number of clusters increases.

USSCoSaMP Algorithm

Input: $\boldsymbol{A}$, $\boldsymbol{D}$, $\boldsymbol{y}$, k, stopping criterion
Initialize: $\boldsymbol{r} = \boldsymbol{y}$, $\boldsymbol{x}_0 = 0$, $\ell = 0$, $\Gamma_{old} = \varnothing$
while not converged **do**

Proxy:	$\widetilde{\boldsymbol{v}} = \boldsymbol{A}^* \boldsymbol{r}$
Identify:	$\Omega = \mathcal{S}_{\boldsymbol{D}}(\widetilde{\boldsymbol{v}}, 2k)$
Merge:	$T = \Omega \cup \Gamma_{old}$
Least Squares:	$\widetilde{\boldsymbol{w}} = \underset{z}{\operatorname{argmin}} \lVert \boldsymbol{z} - \boldsymbol{A}\boldsymbol{x} \rVert_2 \quad \text{s.t.} \quad \boldsymbol{z} \in \mathcal{R}(\boldsymbol{D}_T)$
Prune:	$\Gamma_{omp} = \mathcal{S}_{\boldsymbol{D}}(\widetilde{\boldsymbol{w}}, k)$
	$\Gamma_{cosamp} = \mathcal{S}_{\boldsymbol{D}}(\widetilde{\boldsymbol{w}}, k)$
Union:	$\Gamma = \Gamma_{omp} \cup \Gamma_{cosamp}$
Least Squares:	$\widetilde{\boldsymbol{x}} = \underset{z}{\operatorname{argmin}} \lVert \boldsymbol{z} - \boldsymbol{A}\boldsymbol{x} \rVert_2 \quad \text{s.t.} \quad \boldsymbol{z} \in \mathcal{R}(\boldsymbol{D}_\Gamma)$
Update:	$\boldsymbol{x}_{\ell+1} = \mathcal{P}_\Gamma \widetilde{\boldsymbol{x}}$
	$\boldsymbol{r} = \boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}_{\ell+1}$
	$\ell = \ell + 1$
	$\Gamma_{old} = \Gamma$

end while
Output: $\hat{\boldsymbol{x}} = \boldsymbol{x}_\ell$

USSCoSaMP Performance

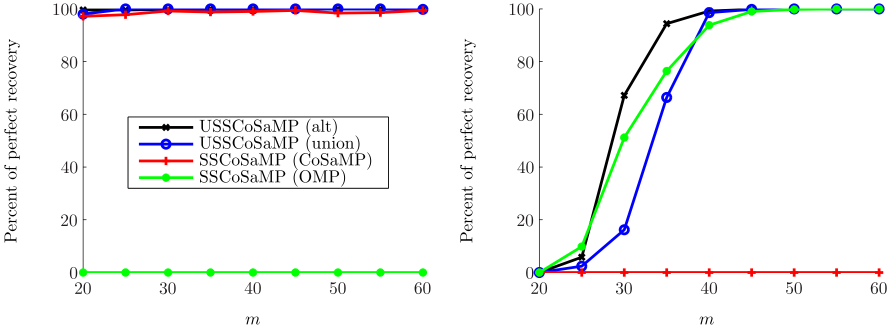


Figure: Left: Clustered. Right: Well-Separated. 500 trials with $k = 8$, $n = 256$, $d = 1024$.

USSCoSaMP succeeds simultaneously for both signal models!

USSCoSaMP Performance

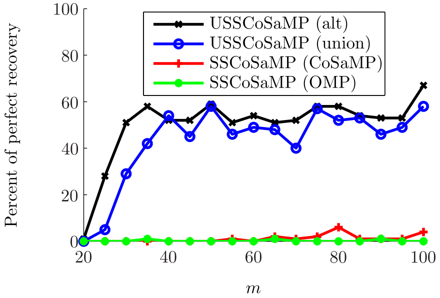


Figure: Hybrid signal. 500 trials with $k = 8$, $n = 256$, $d = 1024$.

Catalog of empirical results

Method	Clustered	Spread	Hybrid	2 Clust	4 Clust	Alternating	Pair Spread
SSCoSaMP (CoSaMP)	100	0	0	20	0	100	0
SSCoSaMP (ℓ_1)	20	100	30	10	60	25	35
SSCoSaMP (OMP)	0	100	0	0	0	0	5
CoSaMP	100	0	10	100	60	100	0
OMP	0	60	0	0	0	0	10
ℓ_1	20	100	20	10	50	20	25
USSCoSaMP	100	100	65	80	30	100	10
NOMP	100	100	100	100	100	100	100

Table: SSCoSaMP variants and new algorithms' performance on various types of sparse coefficient vectors. All of which are sparse with respect to a $4\times$ overcomplete DFT dictionary. A minimum of 40 trials were performed on each test.

SSCoSaMP(ℓ_1): Well-separated Signal (no noise)

- **Theorem** Let D be an $n \times d (n \leq d)$ overcomplete DFT dictionary. Let signal x have a k-sparse expansion in D, i.e. $x = D\alpha$. If T is the support of α and obeys

$$\min_{t,t' \in T : t \neq t'} |t - t'| \geq 4k/n \tag{1}$$

then the solution to

$$\min ||\alpha_{est}||_1 \text{ subject to } D\alpha_{est} = x \tag{2}$$

is exact.

SSCoSaMP(ℓ_1): Well-separated Signal (no noise)

- **Theorem** Let D be an $n \times d (n \leq d)$ overcomplete DFT dictionary. Let signal x have a k-sparse expansion in D, i.e. $x = D\alpha$. If T is the support of α and obeys

$$\min_{t,t' \in T: t \neq t'} |t - t'| \geq 4k/n \tag{1}$$

then the solution to

$$\min ||\alpha_{est}||_1 \text{ subject to } D\alpha_{est} = x \tag{2}$$

is exact.
- Therefore the near-optimal support is the same as the optimal support, which guarantees accurate recovery of SSCoSaMP(ℓ_1).

SSCoSaMP(ℓ_1): Well-separated Signal (with noise)

- **Corollary** Let D be an $n \times d (n \leq d)$ overcomplete DFT dictionary. Let signal x have a k-sparse expansion in D, i.e. $x = D\alpha$. Assume noise model $x = D\alpha + e$ where α is k-sparse and $||e||_2 \leq \varepsilon$. Let T be the support of α and $\{\alpha_t\}$ be the set of nonzeros in α. If $\{\alpha_t\}$ obeys

$$\min_{t,t' \in T : t \neq t'} |t - t'| \geq 4k/n \tag{3}$$

then the solution to

$$\min ||\alpha_{est}||_1 \text{ subject to } ||D\alpha_{est} - x||_2 \leq \varepsilon \tag{4}$$

obeys $||\alpha_{est} - \alpha||_2 \leq C\varepsilon$.

SSCoSaMP(ℓ_1): Well-separated Signal (with noise)

- **Corollary** Let D be an $n \times d (n \leq d)$ overcomplete DFT dictionary. Let signal x have a k-sparse expansion in D, i.e. $x = D\alpha$. Assume noise model $x = D\alpha + e$ where α is k-sparse and $||e||_2 \leq \varepsilon$. Let T be the support of α and $\{\alpha_t\}$ be the set of nonzeros in α. If $\{\alpha_t\}$ obeys

$$\min_{t,t' \in T : t \neq t'} |t - t'| \geq 4k/n \tag{3}$$

  then the solution to

$$\min ||\alpha_{est}||_1 \text{ subject to } ||D\alpha_{est} - x||_2 \leq \varepsilon \tag{4}$$

  obeys $||\alpha_{est} - \alpha||_2 \leq C\varepsilon$.
- Additionally, $||D(\alpha_{est,k} - \alpha_{opt})||_2$ is bounded by $C\varepsilon$.

SSCoSaMP(OMP): Well-seperated signal (noiseless)

- **Theorem** Let D be an $n \times d$ overcomplete DFT dictionary, α a k-sparse vector with support T. If α is well-seperated such that D_T is incoherent, then OMP recovers α exactly from $x = D\alpha$. In particular, OMP gives exact recovery if

$$B(d_{min}) := \frac{1}{n(1-\delta_k)} \sum_{\ell=0}^{k-\lfloor \frac{k}{2} \rfloor - 1} \left| \csc \frac{(\ell \cdot d_{min} + \mu)\pi}{d} \sin \frac{(\ell \cdot d_{min} + \mu)n\pi}{d} \right|$$

$$+ \sum_{\ell=1}^{\lfloor \frac{k}{2} \rfloor} \left| \csc \frac{(\ell \cdot d_{min} - \mu)\pi}{d} \sin \frac{(\ell \cdot d_{min} - \mu)n\pi}{d} \right| \leq 1$$

where d_{min} denotes the minimum distance between columns in T and μ denotes the minimum distance between columns in T and columns in T^c.

SSCoSaMP(OMP): Well-seperated signal (noiseless)

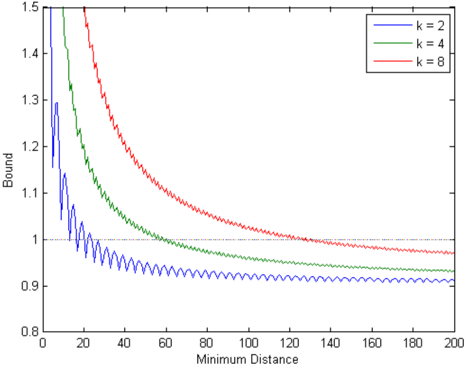


Figure: When $\mu = 1, n = 256, d = 1024$, the value of $B(d_{min})$ with increasing sparsity k.

SSCoSaMP(OMP): Well-seperated signal (with noise)

- **Theorem** Let D be an $n \times d$ overcomplete DFT dictionary, α a k-sparse vector with support T. Assume signal is corrupted by noise, $x = D\alpha + e$, where $||e||_2 \leq \epsilon$. If D satisfies

$$B(d_{min}) < 1$$

and the minimum magnitude of nonzero elements of α obeys

$$\min_{i \in T} |\alpha_i| \geq \frac{\epsilon(\sqrt{\frac{d}{n}} + \sqrt{1 + \delta_k})}{1 - B(d_{min})}$$

then OMP will exactly recover the support T (and thus $\|\hat{\alpha} - \alpha\|_2 \leq \epsilon$).

SSCoSaMP(OMP): Well-seperated signal (with noise)

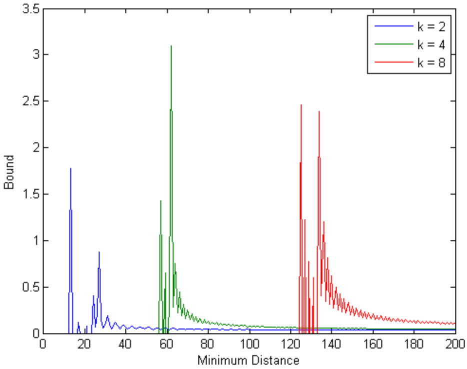


Figure: When $||e||_2 = 10^{-3}$, the minimum magnitude of nonzero elements of well-separated signal α.

Summary

- Rigorous empirical investigation into the recovery performance of several methods for several classes of sparse signals.

Summary

- Rigorous empirical investigation into the recovery performance of several methods for several classes of sparse signals.
- Algorithm design that successfully recovers arbitrary sparse signals in a DFT dictionary that improves upon existing methods.

Summary

- Rigorous empirical investigation into the recovery performance of several methods for several classes of sparse signals.
- Algorithm design that successfully recovers arbitrary sparse signals in a DFT dictionary that improves upon existing methods.
- Provid theoretical backing for the success of SSCoSaMP(OMP) and SSCoSaMP(ℓ_1) in the well-separated case.

Thank you!

References:

- M. A. Davenport and M. B. Wakin. Compressive sensing of analog signals using discrete prolate spheroidal sequences. Appl. Comput. Harmon. A., 2011, to appear.
- M. A. Davenport, D. Needell and M. B. Wakin. Signal Space CoSaMP for Sparse Recovery with Redundant Dictionaries, IEEE Trans. Info. Theory, vol. 59, iss. 10, 6820 - 6829, 2013.
- R. Giryes and D. Needell. Greedy Signal Space Methods for Incoherence and Beyond, Appl. Harmon. Analysis, to appear.
- D. Needell and J. A. Tropp. CoSaMP: Iterative signal recovery from incomplete and inaccurate samples. Applied and Computational Harmonic Analysis, 26(3):301-321, 2008.