## Decoding in Compressed Sensing

Albert Cohen, Wolfgang Dahmen, Ronald DeVore, Guergana Petrova, Przemek Wojtaszczek

## Discrete Compressed Sensing

- $x \in \mathbb{R}^{N}$ with $N$ large


# Discrete Compressed Sensing 

- $x \in \mathbb{R}^{N}$ with $N$ large
- We are able to ask $n$ questions about $x$


## Discrete Compressed Sensing

- $x \in \mathbb{R}^{N}$ with $N$ large
- We are able to ask $n$ questions about $x$
- Question means inner product $v \cdot x$ with $v \in \mathbb{R}^{N}$ - called sample


## Discrete Compressed Sensing

- $x \in \mathbb{R}^{N}$ with $N$ large
- We are able to ask $n$ questions about $x$
- Question means inner product $v \cdot x$ with $v \in \mathbb{R}^{N}$ - called sample
- What are the best questions to ask??


## Discrete Compressed Sensing

- $x \in \mathbb{R}^{N}$ with $N$ large
- We are able to ask $n$ questions about $x$
- Question means inner product $v \cdot x$ with $v \in \mathbb{R}^{N}$ - called sample
- What are the best questions to ask??
- Any such sampling is given by $\Phi x$ where $\Phi$ is an $n \times N$ matrix


## Discrete Compressed Sensing

- $x \in \mathbb{R}^{N}$ with $N$ large
- We are able to ask $n$ questions about $x$
- Question means inner product $v \cdot x$ with $v \in \mathbb{R}^{N}$ - called sample
- What are the best questions to ask??
- Any such sampling is given by $\Phi x$ where $\Phi$ is an $n \times N$ matrix
- We are interested in the good / best matrices $\Phi$


## Discrete Compressed Sensing

- $x \in \mathbb{R}^{N}$ with $N$ large
- We are able to ask $n$ questions about $x$
- Question means inner product $v \cdot x$ with $v \in \mathbb{R}^{N}$ - called sample
- What are the best questions to ask??
- Any such sampling is given by $\Phi x$ where $\Phi$ is an $n \times N$ matrix
- We are interested in the good / best matrices $\Phi$
- Here good means the samples $y=\Phi x$ contain enough information to approximate $x$ well


## Encoder/Decoder

- We view $\Phi$ as an encoder


## Encoder/Decoder

- We view $\Phi$ as an encoder
- Since $\Phi: \mathbb{R}^{N} \rightarrow \mathbb{R}^{n}$ many $x$ are encoded with same $y$


## Encoder/Decoder

- We view $\Phi$ as an encoder
- Since $\Phi: \mathbb{R}^{N} \rightarrow \mathbb{R}^{n}$ many $x$ are encoded with same $y$
- $\mathcal{N}:=\{\eta: \Phi \eta=0\}$ the null space of $\Phi$


## Encoder/Decoder

- We view $\Phi$ as an encoder
- Since $\Phi: \mathbb{R}^{N} \rightarrow \mathbb{R}^{n}$ many $x$ are encoded with same $y$
- $\mathcal{N}:=\{\eta: \Phi \eta=0\}$ the null space of $\Phi$
- $\mathcal{F}(y):=\{x: \Phi x=y\}=x_{0}+\mathcal{N}$ for any $x_{0} \in \mathcal{F}(y)$


## Encoder/Decoder

- We view $\Phi$ as an encoder
- Since $\Phi: \mathbb{R}^{N} \rightarrow \mathbb{R}^{n}$ many $x$ are encoded with same $y$
- $\mathcal{N}:=\{\eta: \Phi \eta=0\}$ the null space of $\Phi$
- $\mathcal{F}(y):=\{x: \Phi x=y\}=x_{0}+\mathcal{N}$ for any $x_{0} \in \mathcal{F}(y)$
- The hyperplanes $\mathcal{F}(y)$ with $y \in \mathbb{R}^{n}$ stratify $\mathbb{R}^{N}$


## The sets $\mathcal{F}(y)$



TAMU $2008-$ D. 4/2

## Encoder/Decoder

- We view $\Phi$ as an encoder
- Since $\Phi: \mathbb{R}^{N} \rightarrow \mathbb{R}^{n}$ many $x$ are encoded with same $y$
- $\mathcal{N}:=\{\eta: \Phi \eta=0\}$ the null space of $\Phi$
- $\mathcal{F}(y):=\{x: \Phi x=y\}=x_{0}+\mathcal{N}$ for any $x_{0} \in \mathcal{F}(y)$
- The hyperplanes $\mathcal{F}(y)$ with $y \in \mathbb{R}^{n}$ stratify $\mathbb{R}^{N}$
- Decoder is any (possibly nonlinear) mapping $\Delta$ from $\mathbb{R}^{n} \rightarrow \mathbb{R}^{N}$


## Encoder/Decoder

- We view $\Phi$ as an encoder
- Since $\Phi: \mathbb{R}^{N} \rightarrow \mathbb{R}^{n}$ many $x$ are encoded with same $y$
- $\mathcal{N}:=\{\eta: \Phi \eta=0\}$ the null space of $\Phi$
- $\mathcal{F}(y):=\{x: \Phi x=y\}=x_{0}+\mathcal{N}$ for any $x_{0} \in \mathcal{F}(y)$
- The hyperplanes $\mathcal{F}(y)$ with $y \in \mathbb{R}^{n}$ stratify $\mathbb{R}^{N}$
- Decoder is any (possibly nonlinear) mapping $\Delta$ from $\mathbb{R}^{n} \rightarrow \mathbb{R}^{N}$
- $\bar{x}:=\Delta(\Phi(x))$ is our approximation to $x$ from the information extracted


## Encoder/Decoder

- We view $\Phi$ as an encoder
- Since $\Phi: \mathbb{R}^{N} \rightarrow \mathbb{R}^{n}$ many $x$ are encoded with same $y$
- $\mathcal{N}:=\{\eta: \Phi \eta=0\}$ the null space of $\Phi$
- $\mathcal{F}(y):=\{x: \Phi x=y\}=x_{0}+\mathcal{N}$ for any $x_{0} \in \mathcal{F}(y)$
- The hyperplanes $\mathcal{F}(y)$ with $y \in \mathbb{R}^{n}$ stratify $\mathbb{R}^{N}$
- Decoder is any (possibly nonlinear) mapping $\Delta$ from $\mathbb{R}^{n} \rightarrow \mathbb{R}^{N}$
- $\bar{x}:=\Delta(\Phi(x))$ is our approximation to $x$ from the information extracted
- This is a typical inverse problem since $x$ is underdetermined by $y$


## Performance

- The only reason such a system might work is because we model signals as sparse or compressible


## Performance

- The only reason such a system might work is because we model signals as sparse or compressible
- How should we measure performance of the encoding-decoding?


## Performance

- The only reason such a system might work is because we model signals as sparse or compressible
- How should we measure performance of the encoding-decoding?
- Define $\Sigma_{k}:=\left\{x \in \mathbb{R}^{N}: \# \operatorname{supp}(\mathrm{x}) \leq \mathrm{k}\right\}$

$$
\sigma_{k}(x)_{X}:=\inf _{z \in \Sigma_{k}}\|x-z\|_{X}
$$

## Performance

- The only reason such a system might work is because we model signals as sparse or compressible
- How should we measure performance of the encoding-decoding?
- Define $\Sigma_{k}:=\left\{x \in \mathbb{R}^{N}: \# \operatorname{supp}(\mathrm{x}) \leq \mathrm{k}\right\}$

$$
\sigma_{k}(x)_{X}:=\inf _{z \in \Sigma_{k}}\|x-z\|_{X}
$$

- Given an encoding - decoding pair $(\Phi, \Delta)$, we say that this pair is Instance-Optimal of order $k$ for $X$ if for an absolute constant $C>0$

$$
\|x-\Delta(\Phi(x))\|_{X} \leq C \sigma_{k}(x)_{X}
$$

## Performance

- The only reason such a system might work is because we model signals as sparse or compressible
- How should we measure performance of the encoding-decoding?
- Define $\Sigma_{k}:=\left\{x \in \mathbb{R}^{N}: \# \operatorname{supp}(\mathrm{x}) \leq \mathrm{k}\right\}$

$$
\sigma_{k}(x)_{X}:=\inf _{z \in \Sigma_{k}}\|x-z\|_{X}
$$

- Given an encoding - decoding pair $(\Phi, \Delta)$, we say that this pair is Instance-Optimal of order $k$ for $X$ if for an absolute constant $C>0$

$$
\|x-\Delta(\Phi(x))\|_{X} \leq C \sigma_{k}(x)_{X}
$$

- Given $n, N$, the best encoding - decoding pairs are those which have the largest $k$.


## Restricted Isometry Property

- What properties determine whether $(\Phi, \Delta)$ is good?


## Restricted Isometry Property

- What properties determine whether $(\Phi, \Delta)$ is good?
- One property which is known to be sufficient in other results is the following Restricted Isometry Property (RIP) introduced by Candes-Romberg-Tao


## Restricted Isometry Property

- What properties determine whether $(\Phi, \Delta)$ is good?
- One property which is known to be sufficient in other results is the following Restricted Isometry Property (RIP) introduced by Candes-Romberg-Tao
- $\Phi$ has the RIP of order $k$ with constant $\delta \in(0,1)$ if

$$
(1-\delta)\|x\|_{\ell_{2}} \leq\|\Phi x\|_{\ell_{2}} \leq(1+\delta)\|x\|_{\ell_{2}} \quad x \in \Sigma_{k}
$$

## Restricted Isometry Property

- What properties determine whether $(\Phi, \Delta)$ is good?
- One property which is known to be sufficient in other results is the following Restricted Isometry Property (RIP) introduced by Candes-Romberg-Tao
- $\Phi$ has the RIP of order $k$ with constant $\delta \in(0,1)$ if

$$
(1-\delta)\|x\|_{\ell_{2}} \leq\|\Phi x\|_{\ell_{2}} \leq(1+\delta)\|x\|_{\ell_{2}} \quad x \in \Sigma_{k}
$$

- General Philosophy: the larger the value of $k$ the better the performance of $\Phi$


## Restricted Isometry Property

- What properties determine whether $(\Phi, \Delta)$ is good?
- One property which is known to be sufficient in other results is the following Restricted Isometry Property (RIP) introduced by Candes-Romberg-Tao
- $\Phi$ has the RIP of order $k$ with constant $\delta \in(0,1)$ if $(1-\delta)\|x\|_{\ell_{2}} \leq\|\Phi x\|_{\ell_{2}} \leq(1+\delta)\|x\|_{\ell_{2}} \quad x \in \Sigma_{k}$
- General Philosophy: the larger the value of $k$ the better the performance of $\Phi$
- Given $n, N$, we know there are matrices $\Phi$ which have RIP for $k \leq c_{0} n / \log (N / n)$


## Restricted Isometry Property

- What properties determine whether $(\Phi, \Delta)$ is good?
- One property which is known to be sufficient in other results is the following Restricted Isometry Property (RIP) introduced by Candes-Romberg-Tao
- $\Phi$ has the RIP of order $k$ with constant $\delta \in(0,1)$ if

$$
(1-\delta)\|x\|_{\ell_{2}} \leq\|\Phi x\|_{\ell_{2}} \leq(1+\delta)\|x\|_{\ell_{2}} \quad x \in \Sigma_{k}
$$

- General Philosophy: the larger the value of $k$ the better the performance of $\Phi$
- Given $n, N$, we know there are matrices $\Phi$ which have RIP for $k \leq c_{0} n / \log (N / n)$
- This range of $k$ cannot be improved: from now on we refer to this as the largest range of $k$


## Sample Results: $X=\ell_{1}$

- Cohen-Dahmen-DeVore If $\Phi$ has RIP for $2 k$ then $\Phi$ is instance-optimal of order $k$ in $\ell_{1}$ :

$$
\|x-\Delta(\Phi(x))\|_{\ell_{1}} \leq C \sigma_{k}(x)_{\ell_{1}}
$$

## Sample Results: $X=\ell_{1}$

- Cohen-Dahmen-DeVore If $\Phi$ has RIP for $2 k$ then $\Phi$ is instance-optimal of order $k$ in $\ell_{1}$ :

$$
\|x-\Delta(\Phi(x))\|_{\ell_{1}} \leq C \sigma_{k}(x)_{\ell_{1}}
$$

- This means that there is some decoder - not necessarily practical (we shall discuss practical decoders in a moment)


## Sample Results: $X=\ell_{1}$

- Cohen-Dahmen-DeVore If $\Phi$ has RIP for $2 k$ then $\Phi$ is instance-optimal of order $k$ in $\ell_{1}$ :

$$
\|x-\Delta(\Phi(x))\|_{\ell_{1}} \leq C \sigma_{k}(x)_{\ell_{1}}
$$

- This means that there is some decoder - not necessarily practical (we shall discuss practical decoders in a moment)
- Hence we know there are matrices $\Phi$ which have Instance Optimality in $\ell_{1}$ for $k \leq c_{0} n / \log (N / n)$


## Sample Results: $X=\ell_{1}$

- Cohen-Dahmen-DeVore If $\Phi$ has RIP for $2 k$ then $\Phi$ is instance-optimal of order $k$ in $\ell_{1}$ :

$$
\|x-\Delta(\Phi(x))\|_{\ell_{1}} \leq C \sigma_{k}(x)_{\ell_{1}}
$$

- This means that there is some decoder - not necessarily practical (we shall discuss practical decoders in a moment)
- Hence we know there are matrices $\Phi$ which have Instance Optimality in $\ell_{1}$ for $k \leq c_{0} n / \log (N / n)$
- This range of $k$ cannot be improved


## Sample Results: $\ell_{p}$

- Given $n, N$ and a constant $C_{0}$ then we have instance optimality in $\ell_{2}$ for $k$ and this $C_{0}$ only if $k \leq \frac{C_{0} n}{N}$


## Sample Results: $\ell_{p}$

- Given $n, N$ and a constant $C_{0}$ then we have instance optimality in $\ell_{2}$ for $k$ and this $C_{0}$ only if $k \leq \frac{C_{0} n}{N}$
- This shows that instance-optimal is not a viable concept for $\ell_{2}$


## Sample Results: $\ell_{p}$

- Given $n, N$ and a constant $C_{0}$ then we have instance optimality in $\ell_{2}$ for $k$ and this $C_{0}$ only if $k \leq \frac{C_{0} n}{N}$
- This shows that instance-optimal is not a viable concept for $\ell_{2}$
- We have to make $c N$ measurement to even get instance optimality for $k=1$


## Sample Results: $\ell_{p}$

- Given $n, N$ and a constant $C_{0}$ then we have instance optimality in $\ell_{2}$ for $k$ and this $C_{0}$ only if $k \leq \frac{C_{0} n}{N}$
- This shows that instance-optimal is not a viable concept for $\ell_{2}$
- We have to make $c N$ measurement to even get instance optimality for $k=1$
- This bound cannot be improved


## Sample Results: $\ell_{p}$

- Given $n, N$ and a constant $C_{0}$ then we have instance optimality in $\ell_{2}$ for $k$ and this $C_{0}$ only if $k \leq \frac{C_{0} n}{N}$
- This shows that instance-optimal is not a viable concept for $\ell_{2}$
- We have to make $c N$ measurement to even get instance optimality for $k=1$
- This bound cannot be improved
- For $1<p<2$ the range of $k$ is
$k \leq c_{0} N^{\frac{2-2 / p}{1-2 / p}}[n / \log (N / n)]^{\frac{p}{2-p}}$


## What are good matrices

- In other words: Which matrices give biggest range of $k$ ?


## What are good matrices

- In other words: Which matrices give biggest range of $k$ ?
- A sufficient condition is that the matrix satisfy RIP or order $2 k$ for $k$ in the large range


## What are good matrices

- In other words: Which matrices give biggest range of $k$ ?
- A sufficient condition is that the matrix satisfy RIP or order $2 k$ for $k$ in the large range
- With high probability on the draw of the matrix any of the following is known to give a best matrix


## What are good matrices

- In other words: Which matrices give biggest range of $k$ ?
- A sufficient condition is that the matrix satisfy RIP or order $2 k$ for $k$ in the large range
- With high probability on the draw of the matrix any of the following is known to give a best matrix
- Choose at random $N$ vectors from the unit sphere in $\mathbb{R}^{n}$ and use these as the columns of $\Phi$


## What are good matrices

- In other words: Which matrices give biggest range of $k$ ?
- A sufficient condition is that the matrix satisfy RIP or order $2 k$ for $k$ in the large range
- With high probability on the draw of the matrix any of the following is known to give a best matrix
- Choose at random $N$ vectors from the unit sphere in $\mathbb{R}^{n}$ and use these as the columns of $\Phi$
- Choose each entry of $\Phi$ independently and at random from the Gaussian distribution $\mathcal{N}(0,1 / \sqrt{n})$


## What are good matrices

- In other words: Which matrices give biggest range of $k$ ?
- A sufficient condition is that the matrix satisfy RIP or order $2 k$ for $k$ in the large range
- With high probability on the draw of the matrix any of the following is known to give a best matrix
- Choose at random $N$ vectors from the unit sphere in $\mathbb{R}^{n}$ and use these as the columns of $\Phi$
- Choose each entry of $\Phi$ independently and at random from the Gaussian distribution $\mathcal{N}(0,1 / \sqrt{n})$
- We choose each entry of $\Phi$ independently and at random from the Bernouli distribution and then normalize columns to have length one.


## What are good matrices

- In other words: Which matrices give biggest range of $k$ ?
- A sufficient condition is that the matrix satisfy RIP or order $2 k$ for $k$ in the large range
- With high probability on the draw of the matrix any of the following is known to give a best matrix
- Choose at random $N$ vectors from the unit sphere in $\mathbb{R}^{n}$ and use these as the columns of $\Phi$
- Choose each entry of $\Phi$ independently and at random from the Gaussian distribution $\mathcal{N}(0,1 / \sqrt{n})$
- We choose each entry of $\Phi$ independently and at random from the Bernouli distribution and then normalize columns to have length one.
- Problem: There are no known constructions. Moreover verifying RIP is not feasible computationally


## Instance-Optimality in Probability

- We saw that Instance-Optimality for $\ell_{2}^{N}$ is not viable
- We shall now see it is possible if you accept some probability of failure
- Suppose $\Phi(\omega)$ is a collection of random matrices
- We say this family satisfies RIP of order $k$ with probability $1-\epsilon$ if a random draw $\{\Phi(\omega)\}$ will satisfy RIP of order $k$ with probability $1-\epsilon$
- We say $\{\Phi(\omega)\}$ is bounded with probability $1-\epsilon$ if there is an absolute constant $C_{0}$ such that for each $x \in \mathbb{R}^{N}$

$$
\|\Phi(\omega)(x)\|_{\ell_{2}^{N}} \leq C_{0}\|x\|_{\ell_{2}^{N}}
$$

with probability $1-\epsilon$ a random draw $\{\Phi(\omega)\}$

## Theorem: Cohen-Dahmen-DeVore

- If $\{\Phi(\omega)\}$ satisfies RIP of order $3 k$ and boundedness each with probability $1-\epsilon$ then there are decoders $\Delta(\omega)$ such that given any $x \in \ell_{2}^{N}$ we have with probability $1-2 \epsilon$

$$
\|x-\Delta(\omega) \Phi(\omega)(x)\|_{\ell_{2}^{N}} \leq C_{0} \sigma_{k}(x)_{\ell_{2}^{N}}
$$

## Theorem: Cohen-Dahmen-DeVore

- If $\{\Phi(\omega)\}$ satisfies RIP of order $3 k$ and boundedness each with probability $1-\epsilon$ then there are decoders $\Delta(\omega)$ such that given any $x \in \ell_{2}^{N}$ we have with probability $1-2 \epsilon$

$$
\|x-\Delta(\omega) \Phi(\omega)(x)\|_{\ell_{2}^{N}} \leq C_{0} \sigma_{k}(x)_{\ell_{2}^{N}}
$$

- Instance-optimality in probability


## Theorem: Cohen-Dahmen-DeVore

- If $\{\Phi(\omega)\}$ satisfies RIP of order $3 k$ and boundedness each with probability $1-\epsilon$ then there are decoders $\Delta(\omega)$ such that given any $x \in \ell_{2}^{N}$ we have with probability $1-2 \epsilon$

$$
\|x-\Delta(\omega) \Phi(\omega)(x)\|_{\ell_{2}^{N}} \leq C_{0} \sigma_{k}(x)_{\ell_{2}^{N}}
$$

- Instance-optimality in probability
- Theorem holds for Gaussian, Bernouli and other random families of matrices with $\epsilon=e^{-c n}$


## Theorem: Cohen-Dahmen-DeVore

- If $\{\Phi(\omega)\}$ satisfies RIP of order $3 k$ and boundedness each with probability $1-\epsilon$ then there are decoders $\Delta(\omega)$ such that given any $x \in \ell_{2}^{N}$ we have with probability $1-2 \epsilon$

$$
\|x-\Delta(\omega) \Phi(\omega)(x)\|_{\ell_{2}^{N}} \leq C_{0} \sigma_{k}(x)_{\ell_{2}^{N}}
$$

- Instance-optimality in probability
- Theorem holds for Gaussian, Bernouli and other random families of matrices with $\epsilon=e^{-c n}$
- Range of $k$ is $k \leq c_{0} n / \log (N / n)$


## Theorem: Cohen-Dahmen-DeVore

- If $\{\Phi(\omega)\}$ satisfies RIP of order $3 k$ and boundedness each with probability $1-\epsilon$ then there are decoders $\Delta(\omega)$ such that given any $x \in \ell_{2}^{N}$ we have with probability $1-2 \epsilon$

$$
\|x-\Delta(\omega) \Phi(\omega)(x)\|_{\ell_{2}^{N}} \leq C_{0} \sigma_{k}(x)_{\ell_{2}^{N}}
$$

- Instance-optimality in probability
- Theorem holds for Gaussian, Bernouli and other random families of matrices with $\epsilon=e^{-c n}$
- Range of $k$ is $k \leq c_{0} n / \log (N / n)$
- Decoder is impractical


## Theorem: Cohen-Dahmen-DeVore

- If $\{\Phi(\omega)\}$ satisfies RIP of order $3 k$ and boundedness each with probability $1-\epsilon$ then there are decoders $\Delta(\omega)$ such that given any $x \in \ell_{2}^{N}$ we have with probability $1-2 \epsilon$

$$
\|x-\Delta(\omega) \Phi(\omega)(x)\|_{\ell_{2}^{N}} \leq C_{0} \sigma_{k}(x)_{\ell_{2}^{N}}
$$

- Instance-optimality in probability
- Theorem holds for Gaussian, Bernouli and other random families of matrices with $\epsilon=e^{-c n}$
- Range of $k$ is $k \leq c_{0} n / \log (N / n)$
- Decoder is impractical
- Notice probability is on the draw of $\Phi$ not on $x$


## Decoding

- By far the most intriguing part of Compressed Sensing is the decoding - Hence this Workshop


## Decoding

- By far the most intriguing part of Compressed Sensing is the decoding - Hence this Workshop
- There are continuing debates as to which decoding is numerically fastest


## Decoding

- By far the most intriguing part of Compressed Sensing is the decoding - Hence this Workshop
- There are continuing debates as to which decoding is numerically fastest
- Some common decoders


## Decoding

- By far the most intriguing part of Compressed Sensing is the decoding - Hence this Workshop
- There are continuing debates as to which decoding is numerically fastest
- Some common decoders
- $\ell_{1}$ minimization: Long history - most everyone in this workshop has contributed to its development


## Decoding

- By far the most intriguing part of Compressed Sensing is the decoding - Hence this Workshop
- There are continuing debates as to which decoding is numerically fastest
- Some common decoders
- $\ell_{1}$ minimization: Long history - most everyone in this workshop has contributed to its development
- Greedy algorithms - find support of a good approximation vector and then decode using $\ell_{2}$ minimization (Gilbert-Tropp; Needell-Vershynin(ROMP), Donoho (STOMP)), Needell - Tropp (CoSAMP)


## Decoding

- By far the most intriguing part of Compressed Sensing is the decoding - Hence this Workshop
- There are continuing debates as to which decoding is numerically fastest
- Some common decoders
- $\ell_{1}$ minimization: Long history - most everyone in this workshop has contributed to its development
- Greedy algorithms - find support of a good approximation vector and then decode using $\ell_{2}$ minimization (Gilbert-Tropp; Needell-Vershynin(ROMP), Donoho (STOMP)), Needell - Tropp (CoSAMP)
- Iterative Reweighted Least Squares (Osborne, Daubechies-DeVore-Fornasier-Gunturk)


## Issues in Decoding

- Range of Instance Optimality: When combined with encoder does it give full range of instance optimality?


## Issues in Decoding

- Range of Instance Optimality: When combined with encoder does it give full range of instance optimality?
- Number of computations to decode?


## Issues in Decoding

- Range of Instance Optimality: When combined with encoder does it give full range of instance optimality?
- Number of computations to decode?
- Robustness to noise?


## Issues in Decoding

- Range of Instance Optimality: When combined with encoder does it give full range of instance optimality?
- Number of computations to decode?
- Robustness to noise?
- Theorems versus numerical examples


## Issues in Decoding

- Range of Instance Optimality: When combined with encoder does it give full range of instance optimality?
- Number of computations to decode?
- Robustness to noise?
- Theorems versus numerical examples
- Instance Optimality in Probability


## Issues in Decoding

- Range of Instance Optimality: When combined with encoder does it give full range of instance optimality?
- Number of computations to decode?
- Robustness to noise?
- Theorems versus numerical examples
- Instance Optimality in Probability
- Given that we cant construct best encoding matrices it seems that the best results would correspond to random draws of matrices


## Issues in Decoding

- Range of Instance Optimality: When combined with encoder does it give full range of instance optimality?
- Number of computations to decode?
- Robustness to noise?
- Theorems versus numerical examples
- Instance Optimality in Probability
- Given that we cant construct best encoding matrices it seems that the best results would correspond to random draws of matrices
- We shall concentrate the rest of the talk on which decoders give instance optimality in $\ell_{2}$ with high probability for a large range of $k$


## Results for Decoding for Random Draws

- Gilbert and Tropp proved that for Bernouli random matrices, OMP captures a $k$ sparse vector with high probability for the large range of $k$


## Results for Decoding for Random Draws

- Gilbert and Tropp proved that for Bernouli random matrices, OMP captures a $k$ sparse vector with high probability for the large range of $k$
- Are there practical decoders that give instance optimality in $\ell_{2}$ ?


## Results for Decoding for Random Draws

- Gilbert and Tropp proved that for Bernouli random matrices, OMP captures a $k$ sparse vector with high probability for the large range of $k$
- Are there practical decoders that give instance optimality in $\ell_{2}$ ?
- Wojtaszczek has shown that for Gaussian matrices $\ell_{1}$ minimization does the job for the large range of $k$


## Results for Decoding for Random Draws

- Gilbert and Tropp proved that for Bernouli random matrices, OMP captures a $k$ sparse vector with high probability for the large range of $k$
- Are there practical decoders that give instance optimality in $\ell_{2}$ ?
- Wojtaszczek has shown that for Gaussian matrices $\ell_{1}$ minimization does the job for the large range of $k$
- The proof is nontrivial and rests on a geometric property of Gaussian matrices: with high probability on the draw $\Phi(\omega)$ the image of the unit $\ell_{1}$ ball under such matrices will with high probability contain an $\ell_{2}$ ball of radius $\frac{c_{0} \sqrt{L}}{\sqrt{n}}, \quad L:=n / \log (N / n)$


## Results for Decoding for Random Draws

- Gilbert and Tropp proved that for Bernouli random matrices, OMP captures a $k$ sparse vector with high probability for the large range of $k$
- Are there practical decoders that give instance optimality in $\ell_{2}$ ?
- Wojtaszczek has shown that for Gaussian matrices $\ell_{1}$ minimization does the job for the large range of $k$
- The proof is nontrivial and rests on a geometric property of Gaussian matrices: with high probability on the draw $\Phi(\omega)$ the image of the unit $\ell_{1}$ ball under such matrices will with high probability contain an $\ell_{2}$ ball of radius $\frac{c_{0} \sqrt{L}}{\sqrt{n}}, L:=n / \log (N / n)$
- This Geometric Property does not hold for more general random families, e.g. Bernoulli.


## Extension to more general families

- DeVore-Petrova-Wojtaszczek


## Extension to more general families

- DeVore-Petrova-Wojtaszczek
- We introduce new geometry to handle general random families (including Bernoulli)


## Extension to more general families

- DeVore-Petrova-Wojtaszczek
- We introduce new geometry to handle general random families (including Bernoulli)
- $\|y\|_{J}:=\max \left(\left\{\|y\|_{\ell_{\infty}\left(\mathbb{R}^{n}\right)}, \sqrt{L}\|y\|_{\ell_{2}\left(\mathbb{R}^{n}\right)}\right\}, \quad L:=\log (N / n)\right.$.


## Extension to more general families

- DeVore-Petrova-Wojtaszczek
- We introduce new geometry to handle general random families (including Bernoulli)
- $\|y\|_{J}:=\max \left(\left\{\|y\|_{\ell_{\infty}\left(\mathbb{R}^{n}\right)}, \sqrt{L}\|y\|_{\ell_{2}\left(\mathbb{R}^{n}\right)}\right\}, \quad L:=\log (N / n)\right.$.
- The unit ball $U_{J}$ under this norm consists of all $y \in \mathbb{R}^{n}$ such that $\|y\|_{\ell_{\infty}} \leq 1 / \sqrt{n}$ and $\|y\|_{\ell_{2}} \leq \sqrt{L} / \sqrt{n}$


## Extension to more general families

- DeVore-Petrova-Wojtaszczek
- We introduce new geometry to handle general random families (including Bernoulli)
- $\|y\|_{J}:=\max \left(\left\{\|y\|_{\ell_{\infty}\left(\mathbb{R}^{n}\right)}, \sqrt{L}\|y\|_{\ell_{2}\left(\mathbb{R}^{n}\right)}\right\}, \quad L:=\log (N / n)\right.$.
- The unit ball $U_{J}$ under this norm consists of all $y \in \mathbb{R}^{n}$ such that $\|y\|_{\ell_{\infty}} \leq 1 / \sqrt{n}$ and $\|y\|_{\ell_{2}} \leq \sqrt{L} / \sqrt{n}$
- $U_{J}$ is a trimmed $\ell_{2}$ ball


## Extension to more general families

- DeVore-Petrova-Wojtaszczek
- We introduce new geometry to handle general random families (including Bernoulli)
- $\|y\|_{J}:=\max \left(\left\{\|y\|_{\ell_{\infty}\left(\mathbb{R}^{n}\right)}, \sqrt{L}\|y\|_{\ell_{2}\left(\mathbb{R}^{n}\right)}\right\}, \quad L:=\log (N / n)\right.$.
- The unit ball $U_{J}$ under this norm consists of all $y \in \mathbb{R}^{n}$ such that $\|y\|_{\ell_{\infty}} \leq 1 / \sqrt{n}$ and $\|y\|_{\ell_{2}} \leq \sqrt{L} / \sqrt{n}$
- $U_{J}$ is a trimmed $\ell_{2}$ ball
- With high probability on the draw $\Phi(\omega)$ we have $\Phi\left(U\left(\ell_{1}^{N}\right)\right) \supset c_{0} U_{J}$


## Extension to more general families

- DeVore-Petrova-Wojtaszczek
- We introduce new geometry to handle general random families (including Bernoulli)
- $\|y\|_{J}:=\max \left(\left\{\|y\|_{\ell_{\infty}\left(\mathbb{R}^{n}\right)}, \sqrt{L}\|y\|_{\ell_{2}\left(\mathbb{R}^{n}\right)}\right\}, \quad L:=\log (N / n)\right.$.
- The unit ball $U_{J}$ under this norm consists of all $y \in \mathbb{R}^{n}$ such that $\|y\|_{\ell_{\infty}} \leq 1 / \sqrt{n}$ and $\|y\|_{\ell_{2}} \leq \sqrt{L} / \sqrt{n}$
- $U_{J}$ is a trimmed $\ell_{2}$ ball
- With high probability on the draw $\Phi(\omega)$ we have $\Phi\left(U\left(\ell_{1}^{N}\right)\right) \supset c_{0} U_{J}$
- Using this result we can prove that encoding with Bernouli (and more general random families) decoding with $\ell_{1}$ minimization is instance optimal for the large range of $k$


## A Greedy Decoder

- Let $\Phi(\omega)$ be a random family of $n \times N$ compressed sensing matrices


## A Greedy Decoder

- Let $\Phi(\omega)$ be a random family of $n \times N$ compressed sensing matrices
- We suppose $n=a m$ with $a$ and $m$ integers


## A Greedy Decoder

- Let $\Phi(\omega)$ be a random family of $n \times N$ compressed sensing matrices
- We suppose $n=a m$ with $a$ and $m$ integers
- Let $\Phi_{j}:=\left[\phi_{1}^{j}, \ldots, \phi_{N}^{1}\right], j=1, \ldots, a$, be obtained from the rows $i=j m+1, \ldots, j(m+1)$ of $\Phi$


## A Greedy Decoder

- Let $\Phi(\omega)$ be a random family of $n \times N$ compressed sensing matrices
- We suppose $n=a m$ with $a$ and $m$ integers
- Let $\Phi_{j}:=\left[\phi_{1}^{j}, \ldots, \phi_{N}^{1}\right], j=1, \ldots, a$, be obtained from the rows $i=j m+1, \ldots, j(m+1)$ of $\Phi$
- We shall give an iterative thresholding algorithm for decoding


## A Greedy Decoder

- Let $\Phi(\omega)$ be a random family of $n \times N$ compressed sensing matrices
- We suppose $n=a m$ with $a$ and $m$ integers
- Let $\Phi_{j}:=\left[\phi_{1}^{j}, \ldots, \phi_{N}^{1}\right], j=1, \ldots, a$, be obtained from the rows $i=j m+1, \ldots, j(m+1)$ of $\Phi$
- We shall give an iterative thresholding algorithm for decoding
- At each step it will choose a set $\bar{\Lambda}_{j}$ of coordinates (of $x$ ) where inner products with residuals are large


## A Greedy Decoder

- Let $\Phi(\omega)$ be a random family of $n \times N$ compressed sensing matrices
- We suppose $n=a m$ with $a$ and $m$ integers
- Let $\Phi_{j}:=\left[\phi_{1}^{j}, \ldots, \phi_{N}^{1}\right], j=1, \ldots, a$, be obtained from the rows $i=j m+1, \ldots, j(m+1)$ of $\Phi$
- We shall give an iterative thresholding algorithm for decoding
- At each step it will choose a set $\bar{\Lambda}_{j}$ of coordinates (of $x$ ) where inner products with residuals are large
- These coordinates are added to all previously selected coordinates to give the new set $\Lambda_{j}:=\bar{\Lambda}_{1} \cup \cdots \cup \bar{\Lambda}_{j}$ which is our current set of coordinates where we think $x$ is big


## A Greedy Decoder

- Let $\Phi(\omega)$ be a random family of $n \times N$ compressed sensing matrices
- We suppose $n=a m$ with $a$ and $m$ integers
- Let $\Phi_{j}:=\left[\phi_{1}^{j}, \ldots, \phi_{N}^{1}\right], j=1, \ldots, a$, be obtained from the rows $i=j m+1, \ldots, j(m+1)$ of $\Phi$
- We shall give an iterative thresholding algorithm for decoding
- At each step it will choose a set $\bar{\Lambda}_{j}$ of coordinates (of $x$ ) where inner products with residuals are large
- These coordinates are added to all previously selected coordinates to give the new set $\Lambda_{j}:=\bar{\Lambda}_{1} \cup \cdots \cup \bar{\Lambda}_{j}$ which is our current set of coordinates where we think $x$ is big
- There is a threshold parameter $\delta>0$ which can be chosen for examole as $1 / 8$


## The Greedy Thresholding Algorithm

- Fix a value of $k$


## The Greedy Thresholding Algorithm

- Fix a value of $k$
- Initial Step: Let $r^{1}:=y^{1}=\Phi_{1}(x)$ be the initial residual


## The Greedy Thresholding Algorithm

- Fix a value of $k$
- Initial Step: Let $r^{1}:=y^{1}=\Phi_{1}(x)$ be the initial residual
- $\bar{\Lambda}_{1}:=\Lambda_{1}$ is defined as the set of all coordinates $\nu$ where

$$
\left|\left\langle y^{1}, \phi_{\nu}^{1}\right\rangle\right|>\delta k^{-1 / 2}\left\|y^{1}\right\|_{\ell_{2}}
$$

## The Greedy Thresholding Algorithm

- Fix a value of $k$
- Initial Step: Let $r^{1}:=y^{1}=\Phi_{1}(x)$ be the initial residual
- $\bar{\Lambda}_{1}:=\Lambda_{1}$ is defined as the set of all coordinates $\nu$ where

$$
\left|\left\langle y^{1}, \phi_{\nu}^{1}\right\rangle\right|>\delta k^{-1 / 2}\left\|y^{1}\right\|_{\ell_{2}}
$$

- Define $x^{1}$ as the least squares minimizer

$$
x^{1}:=\underset{\operatorname{supp}(\mathrm{z}) \subset \Lambda_{1}}{\operatorname{Argmin}}\left\|y^{1}-\Phi^{1} z\right\|_{\ell^{2}}
$$

## The Greedy Thresholding Algorithm

- Fix a value of $k$
- Initial Step: Let $r^{1}:=y^{1}=\Phi_{1}(x)$ be the initial residual
- $\bar{\Lambda}_{1}:=\Lambda_{1}$ is defined as the set of all coordinates $\nu$ where

$$
\left|\left\langle y^{1}, \phi_{\nu}^{1}\right\rangle\right|>\delta k^{-1 / 2}\left\|y^{1}\right\|_{\ell_{2}}
$$

- Define $x^{1}$ as the least squares minimizer

$$
x^{1}:=\underset{\operatorname{supp}(z) \subset \Lambda_{1}}{\operatorname{Argmin}}\left\|y^{1}-\Phi^{1} z\right\|_{\ell^{2}}
$$

- New residual $r^{1}:=\Phi_{2}\left(x-x^{1}\right)=y^{2}-\Phi_{2}\left(x^{1}\right)$


## The Greedy Thresholding Algorithm

- Fix a value of $k$
- Initial Step: Let $r^{1}:=y^{1}=\Phi_{1}(x)$ be the initial residual
- $\bar{\Lambda}_{1}:=\Lambda_{1}$ is defined as the set of all coordinates $\nu$ where

$$
\left|\left\langle y^{1}, \phi_{\nu}^{1}\right\rangle\right|>\delta k^{-1 / 2}\left\|y^{1}\right\|_{\ell_{2}}
$$

- Define $x^{1}$ as the least squares minimizer

$$
x^{1}:=\underset{\operatorname{supp}(z) \subset \Lambda_{1}}{\operatorname{Argmin}}\left\|y^{1}-\Phi^{1} z\right\|_{\ell^{2}}
$$

- New residual $r^{1}:=\Phi_{2}\left(x-x^{1}\right)=y^{2}-\Phi_{2}\left(x^{1}\right)$
- General Step: Repeat the above with $r^{1}$ replaced by $r^{j}$ and $\Phi_{1}$ replaced by $\Phi_{j}$


## The Stopping Criteria

- We do not want sets with cardinality larger than $2 k$


## The Stopping Criteria

- We do not want sets with cardinality larger than $2 k$
- If the size of $\Lambda_{j}:=\bar{\Lambda}_{1} \cup \ldots \cup \bar{\Lambda}_{j}$ exceeds $2 k$ trim back to $2 k$ coordinates by removing elements from $\bar{\Lambda}_{j}$


## The Stopping Criteria

- We do not want sets with cardinality larger than $2 k$
- If the size of $\Lambda_{j}:=\bar{\Lambda}_{1} \cup \ldots \cup \bar{\Lambda}_{j}$ exceeds $2 k$ trim back to $2 k$ coordinates by removing elements from $\bar{\Lambda}_{j}$
- This gives the output set $\Lambda$


## The Stopping Criteria

- We do not want sets with cardinality larger than $2 k$
- If the size of $\Lambda_{j}:=\bar{\Lambda}_{1} \cup \ldots \cup \bar{\Lambda}_{j}$ exceeds $2 k$ trim back to $2 k$ coordinates by removing elements from $\bar{\Lambda}_{j}$
- This gives the output set $\Lambda$
- $\bar{x}$ is the least squares solution for this $\Lambda$ using $y^{j}$ and $\Phi_{j}$


## The Stopping Criteria

- We do not want sets with cardinality larger than $2 k$
- If the size of $\Lambda_{j}:=\bar{\Lambda}_{1} \cup \ldots \cup \bar{\Lambda}_{j}$ exceeds $2 k$ trim back to $2 k$ coordinates by removing elements from $\bar{\Lambda}_{j}$
- This gives the output set $\Lambda$
- $\bar{x}$ is the least squares solution for this $\Lambda$ using $y^{j}$ and $\Phi_{j}$
- otherwise stop at step $j=a$ and output $\bar{x}:=x^{a}$


## Assumptions on the Random Matrices

- The next slide will give an instance optimality for this algorithm


## Assumptions on the Random Matrices

- The next slide will give an instance optimality for this algorithm
- Here are the assumptions on the random family $\Phi(\omega)$, $\omega \in \Omega$ for this theorem to hold


## Assumptions on the Random Matrices

- The next slide will give an instance optimality for this algorithm
- Here are the assumptions on the random family $\Phi(\omega)$, $\omega \in \Omega$ for this theorem to hold
- The entries of $\Phi$ are independent


## Assumptions on the Random Matrices

- The next slide will give an instance optimality for this algorithm
- Here are the assumptions on the random family $\Phi(\omega)$, $\omega \in \Omega$ for this theorem to hold
- The entries of $\Phi$ are independent
- Given $\delta>0$ and $x \in \mathbb{R}^{N}$ with probability

$$
\geq 1-C_{1} e^{-c_{1} n \delta^{2}} \text { we have }
$$

$$
\left|\|\Phi(x)\|_{\ell_{2}}-\|x\|_{\ell_{2}}\right| \leq \delta\|x\|_{\ell_{2}}
$$

## Assumptions on the Random Matrices

- The next slide will give an instance optimality for this algorithm
- Here are the assumptions on the random family $\Phi(\omega)$, $\omega \in \Omega$ for this theorem to hold
- The entries of $\Phi$ are independent
- Given $\delta>0$ and $x \in \mathbb{R}^{N}$ with probability

$$
\geq 1-C_{1} e^{-c_{1} n \delta^{2}} \text { we have }
$$

$$
\left|\|\Phi(x)\|_{\ell_{2}}-\|x\|_{\ell_{2}}\right| \leq \delta\|x\|_{\ell_{2}}
$$

- Given $\delta>0, l \in\{1, \ldots, N\}$ and $z \in \mathbb{R}^{n}$ with probability $\geq 1-C_{1} e^{-c_{1} n \delta^{2}}$ we have

$$
\left|\left\langle z, \phi_{l}\right\rangle\right| \leq \delta\|z\|_{\ell_{2}}
$$

## Assumptions on the Random Matrices

- The next slide will give an instance optimality for this algorithm
- Here are the assumptions on the random family $\Phi(\omega)$, $\omega \in \Omega$ for this theorem to hold
- The entries of $\Phi$ are independent
- Given $\delta>0$ and $x \in \mathbb{R}^{N}$ with probability

$$
\geq 1-C_{1} e^{-c_{1} n \delta^{2}} \text { we have }
$$

$$
\left|\|\Phi(x)\|_{\ell_{2}}-\|x\|_{\ell_{2}}\right| \leq \delta\|x\|_{\ell_{2}}
$$

- Given $\delta>0, l \in\{1, \ldots, N\}$ and $z \in \mathbb{R}^{n}$ with probability $\geq 1-C_{1} e^{-c_{1} n \delta^{2}}$ we have

$$
\left|\left\langle z, \phi_{l}\right\rangle\right| \leq \delta\|z\|_{\ell_{2}}
$$

- These properties hold for Gaussian and Bernouli


## Thresholding Theorem

THEOREM: Given any $0<\delta \leq 1 / 8 \sqrt{3}$ and given any $r, s>0$. The thresholding decoder applied with this choice of $\delta$ where the random matrices are of size $n \times N$ with $n=a m$ and $a:=\lceil r \log N\rceil$ gives the following. For any $x \in \mathbb{R}^{N}$, with probability $\geq 1-n^{-s}$ the decoded vector $\bar{x}$ of the above greedy decoder satisfies

$$
\|x-\bar{x}\|_{\ell_{2}} \leq n^{-r}\|x\|+198 \sigma_{k}(x)
$$

for any $k \leq c(\delta, s) n /(\log N)^{2}$.

- Loss of a logarithm in the range of $k$


## Thresholding Theorem

THEOREM: Given any $0<\delta \leq 1 / 8 \sqrt{3}$ and given any $r, s>0$. The thresholding decoder applied with this choice of $\delta$ where the random matrices are of size $n \times N$ with $n=a m$ and $a:=\lceil r \log N\rceil$ gives the following. For any $x \in \mathbb{R}^{N}$, with probability $\geq 1-n^{-s}$ the decoded vector $\bar{x}$ of the above greedy decoder satisfies

$$
\|x-\bar{x}\|_{\ell_{2}} \leq n^{-r}\|x\|+198 \sigma_{k}(x)
$$

for any $k \leq c(\delta, s) n /(\log N)^{2}$.

- Loss of a logarithm in the range of $k$
- Do not quite have instance optimality


## Thresholding Theorem

THEOREM: Given any $0<\delta \leq 1 / 8 \sqrt{3}$ and given any $r, s>0$. The thresholding decoder applied with this choice of $\delta$ where the random matrices are of size $n \times N$ with $n=a m$ and $a:=\lceil r \log N\rceil$ gives the following. For any $x \in \mathbb{R}^{N}$, with probability $\geq 1-n^{-s}$ the decoded vector $\bar{x}$ of the above greedy decoder satisfies

$$
\|x-\bar{x}\|_{\ell_{2}} \leq n^{-r}\|x\|+198 \sigma_{k}(x)
$$

for any $k \leq c(\delta, s) n /(\log N)^{2}$.

- Loss of a logarithm in the range of $k$
- Do not quite have instance optimality
- Fine for any numerical purpose

