Why do Curvelets work?

Hassan Mansour

SINBAD Sponsor Meeting

Whistler - BC, December 2011

Outline

Part 1: Sparsity and randomized acquisition

- Joint work with Haneet Wason, Tim Lin, and Felix Herrmann

Outline

Part 1: Sparsity and randomized acquisition

- Joint work with Haneet Wason, Tim Lin, and Felix Herrmann

Part 2: Redundancy in curvelet support information

- Joint work with Rayan Saab, Özgür Yılmaz, Michael Friedlander, and Felix Herrmann

Outline

Part 1: Sparsity and randomized acquisition

- Joint work with Haneet Wason, Tim Lin, and Felix Herrmann

Part 2: Redundancy in curvelet support information

- Joint work with Rayan Saab, Özgür Yılmaz, Michael Friedlander, and Felix Herrmann

Part 3: Extensions in FWI, 3D acquisition and time-lapse imaging

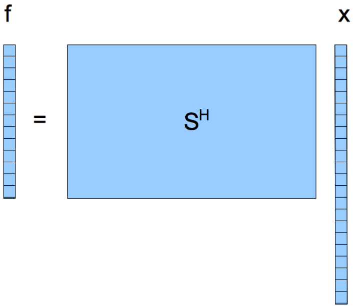
Part 1: Sparsity and randomized acquisition

Part 2: Redundancy in curvelet support information

Part 3: Applications in FWI, 3D acquisition and time-lapse imaging

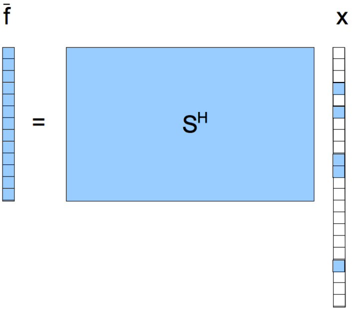
Curvelets and sparsity

- Consider a signal $f \in \mathbb{R}^N$, $f = S^H x$, where S is a curvelet transform matrix and x is the synthesis coefficient vector.
- We can approximate f by the signal $\bar{f}$ using the k-largest coefficients of x.
- For example: a 512×128 shot gather
  - Largest 4598 (0.95%) of curvelet coefficients produce SNR = 10.68dB .
  - Largest 6511 (9.93%) of Fourier coefficients produce SNR = 10.11dB.



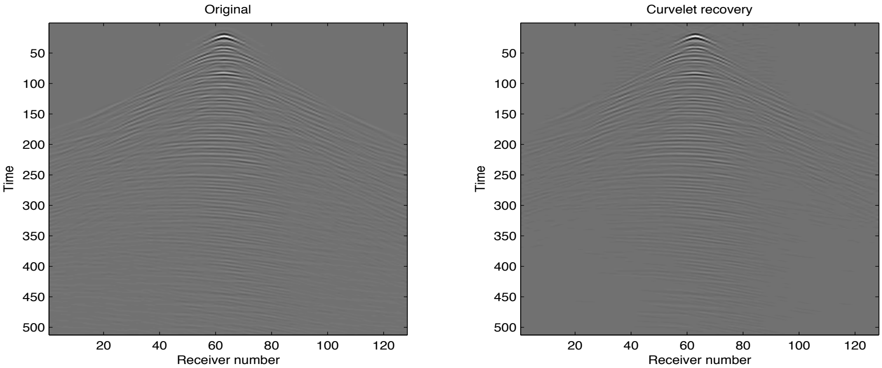
Curvelets and sparsity

- Consider a signal $f \in \mathbb{R}^N$, $f = S^H x$, where S is a curvelet transform matrix and x is the synthesis coefficient vector.
- We can approximate f by the signal $\bar{f}$ using the k-largest coefficients of x.
- For example: a 512×128 shot gather
  - Largest 4598 (0.95%) of curvelet coefficients produce SNR = 10.68dB .
  - Largest 6511 (9.93%) of Fourier coefficients produce SNR = 10.11dB.



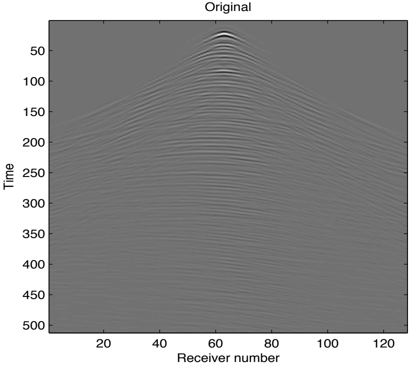
Curvelets and sparsity

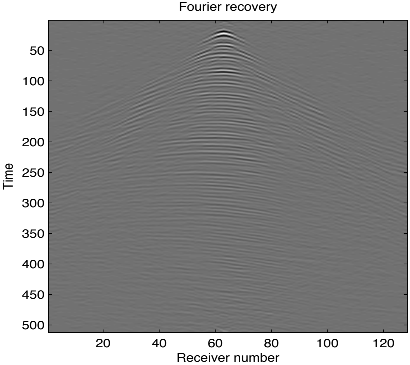
- Consider a signal $f \in \mathbb{R}^N$, $f = S^H x$, where S is a curvelet transform matrix and x is the synthesis coefficient vector.
- We can approximate f by the signal $\bar{f}$ using the k-largest coefficients of x.
- For example: a 512×128 shot gather
    - Largest 4598 (0.95%) of curvelet coefficients produce SNR $=$ 10.68dB .
    - Largest 6511 (9.93%) of Fourier coefficients produce SNR $=$ 10.11dB.



Curvelets and sparsity

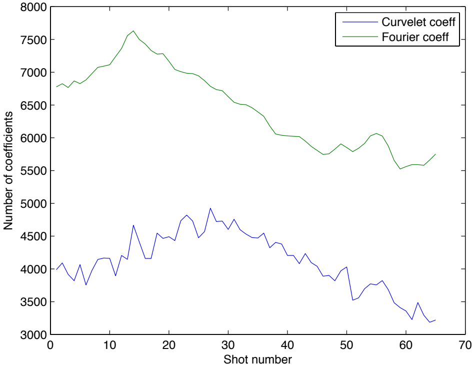
- Consider a signal $f \in \mathbb{R}^N$, $f = S^H x$, where S is a curvelet transform matrix and x is the synthesis coefficient vector.
- We can approximate f by the signal $\bar{f}$ using the k-largest coefficients of x.
- For example: a 512×128 shot gather
  - Largest 4598 (0.95%) of curvelet coefficients produce SNR = 10.68dB .
  - Largest 6511 (9.93%) of Fourier coefficients produce SNR = 10.11dB.





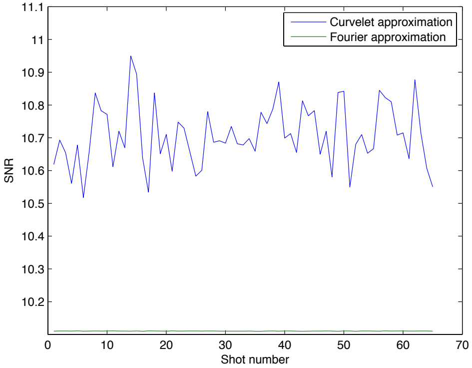
Curvelets and sparsity

- Consider a signal $f \in \mathbb{R}^N$, $f = S^H x$, where S is a curvelet transform matrix and x is the synthesis coefficient vector.
- We can approximate f by the signal $\bar{f}$ using the k-largest coefficients of x.
- For example: a 512×128 shot gather
    - Largest 4598 (0.95%) of curvelet coefficients produce SNR = 10.68dB .
    - Largest 6511 (9.93%) of Fourier coefficients produce SNR = 10.11dB.



Curvelets and sparsity

- Consider a signal $f \in \mathbb{R}^N$, $f = S^H x$, where S is a curvelet transform matrix and x is the synthesis coefficient vector.
- We can approximate f by the signal $\bar{f}$ using the k-largest coefficients of x.
- For example: a 512×128 shot gather
  - Largest 4598 (0.95%) of curvelet coefficients produce SNR = 10.68dB .
  - Largest 6511 (9.93%) of Fourier coefficients produce SNR = 10.11dB.

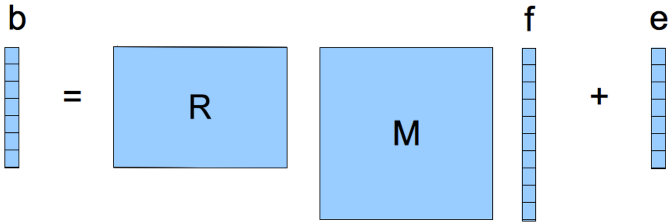


Compressed sensing

- Compressed Sensing is an acquisition paradigm for signals that admit *sparse* or nearly sparse representations in some transform domain.
- Given $n \ll N$ linear and noisy measurements $b = RMf + e$.
- Let $A = RMS^H$, it is possible to approximate x from the measurements b if
  - A obeys certain conditions: RIP, mutual coherence.
  - x is sufficiently sparse: depends on the recovery algorithm.

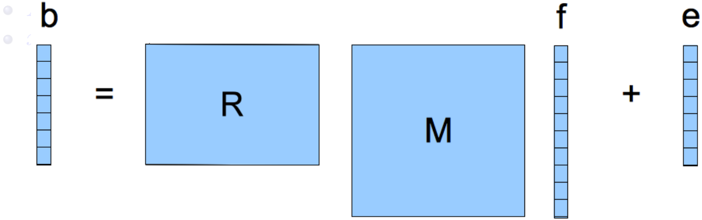
Compressed sensing

- Compressed Sensing is an acquisition paradigm for signals that admit *sparse* or nearly sparse representations in some transform domain.
- Given $n \ll N$ linear and noisy measurements $b = RMf + e$.
- Let $A = RMS^H$, it is possible to approximate x from the measurements b if



Compressed sensing

- Compressed Sensing is an acquisition paradigm for signals that admit *sparse* or nearly sparse representations in some transform domain.
- Given $n \ll N$ linear and noisy measurements $b = RMf + e$.
- Let $A = RMS^H$, it is possible to approximate x from the measurements b if



Compressed sensing

- Compressed Sensing is an acquisition paradigm for signals that admit *sparse* or nearly sparse representations in some transform domain.
- Given $n \ll N$ linear and noisy measurements $b = RMf + e$.
- Let $A = RMS^H$, it is possible to approximate x from the measurements b if
  - A obeys certain conditions: RIP, mutual coherence.
  - x is sufficiently sparse: depends on the recovery algorithm.

Compressed sensing

- Compressed Sensing is an acquisition paradigm for signals that admit *sparse* or nearly sparse representations in some transform domain.
- Given $n \ll N$ linear and noisy measurements $b = RMf + e$.
- Let $A = RMS^H$, it is possible to approximate x from the measurements b if
  - A obeys certain conditions: RIP, mutual coherence.
  - x is sufficiently sparse: depends on the recovery algorithm.

CS example: simultaneous/randomized acquisition

- Simultaneous acquisition is a perfect example of a compressed sensing problem.
- The objective is to recover the high-dimensional sequential shot record f from the lower-dimensional "supershot" record $b = RMf$.
- Formulate the acquisition process in terms of the sampling operator RM.
- We want to construct the sampling operator RM such that:
  1. $A = RMS^H$ satisfies the CS recovery conditions.
  2. RM is physically realizable.
- Recover the sequential shot record by finding $\tilde{f} = S^H\tilde{x}$, where

$$\tilde{x} = \arg\min_x \|x\|_1 \quad \text{s.t.} \quad Ax = b.$$

CS example: simultaneous/randomized acquisition

- Simultaneous acquisition is a perfect example of a compressed sensing problem.
- The objective is to recover the high-dimensional sequential shot record f from the lower-dimensional "supershot" record $b = RMf$.
- Formulate the acquisition process in terms of the sampling operator RM.
- We want to construct the sampling operator RM such that:
  1. $A = RMS^H$ satisfies the CS recovery conditions.
  2. RM is physically realizable.
- Recover the sequential shot record by finding $\tilde{f} = S^H\tilde{x}$, where

$$\tilde{x} = \arg\min_x \|x\|_1 \quad \text{s.t.} \quad Ax = b.$$

CS example: simultaneous/randomized acquisition

- Simultaneous acquisition is a perfect example of a compressed sensing problem.
- The objective is to recover the high-dimensional sequential shot record f from the lower-dimensional "supershot" record $b = RMf$.
- Formulate the acquisition process in terms of the sampling operator RM.
- We want to construct the sampling operator RM such that:
  1. $A = RMS^H$ satisfies the CS recovery conditions.
  2. RM is physically realizable.
- Recover the sequential shot record by finding $\tilde{f} = S^H\tilde{x}$, where

$$\tilde{x} = \arg\min_x \|x\|_1 \quad \text{s.t.} \quad Ax = b.$$

CS example: simultaneous/randomized acquisition

- Simultaneous acquisition is a perfect example of a compressed sensing problem.
- The objective is to recover the high-dimensional sequential shot record f from the lower-dimensional "supershot" record $b = RMf$.
- Formulate the acquisition process in terms of the sampling operator RM.
- We want to construct the sampling operator RM such that:
  1. $A = RMS^H$ satisfies the CS recovery conditions.
  2. RM is physically realizable.
- Recover the sequential shot record by finding $\tilde{f} = S^H\tilde{x}$, where

$$\tilde{x} = \arg\min_x \|x\|_1 \quad \text{s.t.} \quad Ax = b.$$

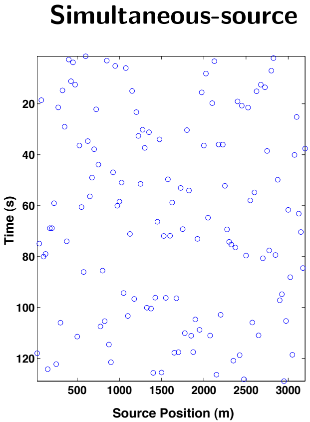
CS example: simultaneous/randomized acquisition

- Simultaneous acquisition is a perfect example of a compressed sensing problem.
- The objective is to recover the high-dimensional sequential shot record f from the lower-dimensional "supershot" record $b = RMf$.
- Formulate the acquisition process in terms of the sampling operator RM.
- We want to construct the sampling operator RM such that:
    1. $A = RMS^H$ satisfies the CS recovery conditions.
    2. RM is physically realizable.
- Recover the sequential shot record by finding $\tilde{f} = S^H \tilde{x}$, where

$$\tilde{x} = \arg\min_x \|x\|_1 \quad \text{s.t.} \quad Ax = b.$$

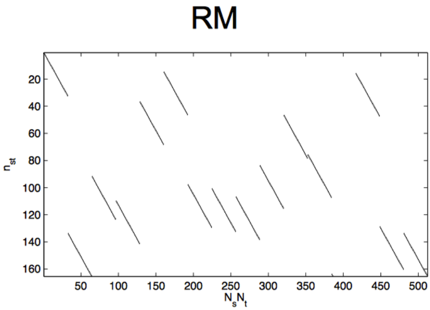
Simultaneous-source acquisition

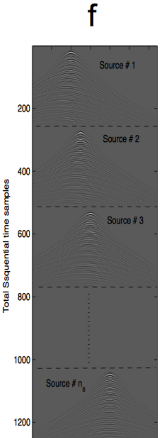
- Typically, we would like RM to be a matrix with i.i.d Gaussian random entries.
- In the marine impulsive airgun setting, only binary matrices with $0-1$ entries are possible.



Simultaneous-source acquisition

- Typically, we would like RM to be a matrix with i.i.d Gaussian random entries.
- Simultaneous acquisition is achieved by randomly dithering in time and across sources.

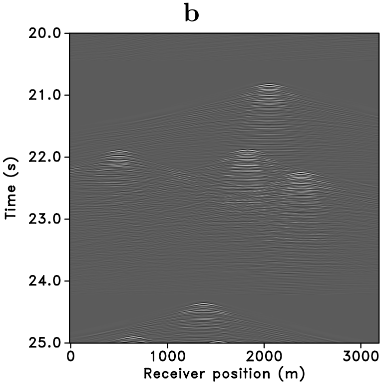




Simultaneous-source acquisition

- Typically, we would like RM to be a matrix with i.i.d Gaussian random entries.

- Simultaneous acquisition is achieved by randomly dithering in time and across sources.



Simultaneous-source acquisition

- Typically, we would like RM to be a matrix with i.i.d Gaussian random entries.
- With a 50% subsampling ratio, we can achieve a 10.5dB SNR using sparse recovery.
- Although this simultaneous-source scenario can be achieved physically, it requires an airgun located at each source location, which can be costly if not practically infeasible.

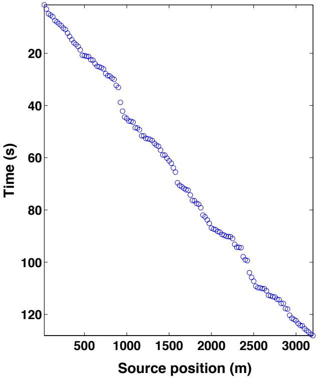
Simultaneous-source acquisition

- Typically, we would like RM to be a matrix with i.i.d Gaussian random entries.
- With a 50% subsampling ratio, we can achieve a 10.5dB SNR using sparse recovery.
- Although this simultaneous-source scenario can be achieved physically, it requires an airgun located at each source location, which can be costly if not practically infeasible.

Random time-dithering

- Sort the random source positions so that we only dither in time ("jitter blending").
- It turns out the resulting measurement matrix is almost as good as "ideal" simultaneous-source acquisition (SNR = 8.06 dB).

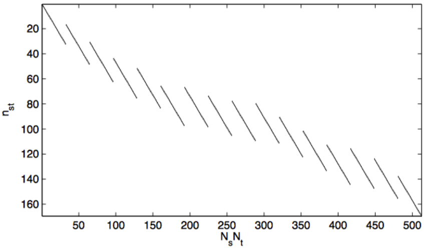
Random time-shifting



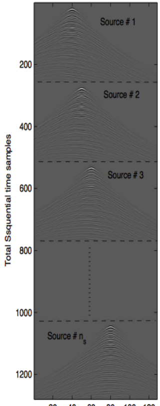
Random time-dithering

- Sort the random source positions so that we only dither in time ("jitter blending").
- It turns out the resulting measurement matrix is almost as good as "ideal" simultaneous-source acquisition (SNR = 8.06 dB).

RM

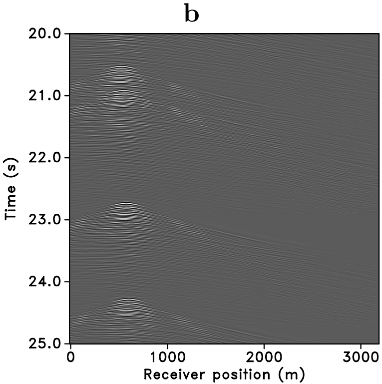


f



Random time-dithering

- Sort the random source positions so that we only dither in time ("jitter blending").
- It turns out the resulting measurement matrix is almost as good as "ideal" simultaneous-source acquisition (SNR = 8.06 dB).



Random time-dithering

- Sort the random source positions so that we only dither in time ("jitter blending").
- It turns out the resulting measurement matrix is almost as good as "ideal" simultaneous-source acquisition (SNR = 8.06 dB).

Measurement matrix conditions

- There are precise conditions on A that guarantee stable recovery.

Measurement matrix conditions

- There are precise conditions on A that guarantee stable recovery.
- The restricted isometry property (RIP) of order k indicates whether every group of k columns of A form a well conditioned submatrix.

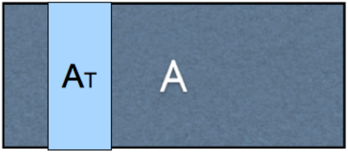
Definition: Restricted Isometry Property (RIP) (Candés and Tao)

The RIP constant $\delta_k \in (0, 1)$ is defined as the smallest constant such that $\forall x \in \Sigma_k^N$

$$(1 - \delta_k)\|x\|_2^2 \leq \|Ax\|_2^2 \leq (1 + \delta_k)\|x\|_2^2,$$

Measurement matrix conditions

- There are precise conditions on A that guarantee stable recovery.
- The restricted isometry property (RIP) of order k indicates whether every group of k columns of A form a well conditioned submatrix.



- RIP is equivalent to saying that for any set T of size k, the symmetric matrix $A_T^H A_T$ is positive definite with eigenvalues in $[1 - \delta_k, 1 + \delta_k]$.

Measurement matrix conditions

- There are precise conditions on A that guarantee stable recovery.

Stable recovery condition (Candés, Romberg, and Tao)

If A has RIP with constant $\delta_{(a+1)s} < \frac{a-1}{a+1}$ for some $a > 1$, then the signal x can be recovered using ℓ_1 minimization to produce an estimate $\tilde{x}$ with an error bounded by

$$\|x - \tilde{x}\|_2 \leq \frac{C}{\sqrt{s}} \|x - x_s\|_1.$$

Why do curvelets work?

- Monte Carlo estimation of the RIP constant δ_k of $A = RMS^H$, for some $k = |T|$

$$\delta_k = \sup_{T \in \{1, \ldots P\}} \max\{1 - \sigma_{\min}(A_T), \sigma_{\max}(A_T) - 1\}$$

- When S is the curvelet transform, we have stable recovery with respect to the best s-term approximation of the signal, where $s = k/8$.

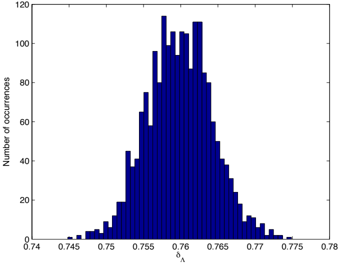
Why do curvelets work?

- Monte Carlo estimation of the RIP constant δ_k of $A = RMS^H$, for some $k = |T|$

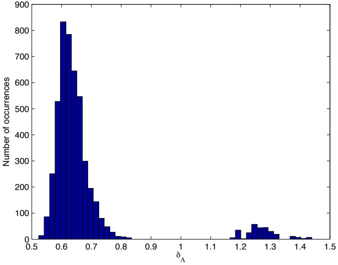
$$\delta_k = \sup_{T \in \{1,\ldots P\}} \max\{1 - \sigma_{\min}(A_T), \sigma_{\max}(A_T) - 1\}$$

- When S is the curvelet transform, we have stable recovery with respect to the best s-term approximation of the signal, where $s = k/8$.

Curvelet



Fourier

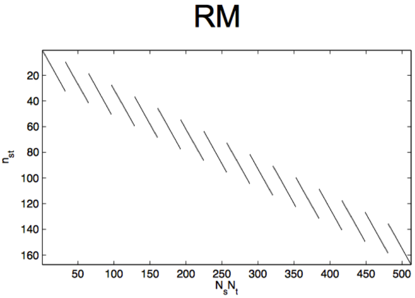


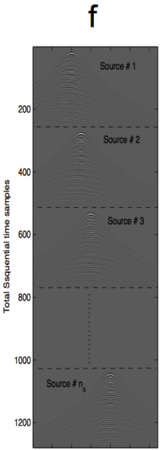
Importance of randomization

- Random time-dithering retains the randomness necessary for CS recovery, albeit at a lower order than simultaneous-source acquisition.
- To emphasize the importance of randomization, we include the case where we simply decrease the intershot time delays.

Importance of randomization

- Random time-dithering retains the randomness necessary for CS recovery, albeit at a lower order than simultaneous-source acquisition.
- To emphasize the importance of randomization, we include the case where we simply decrease the intershot time delays.





Importance of randomization

- Random time-dithering retains the randomness necessary for CS recovery, albeit at a lower order than simultaneous-source acquisition.
- To emphasize the importance of randomization, we include the case where we simply decrease the intershot time delays.

- Sparse recovery results in 4.8dB SNR.
- Using linear reconstruction followed by median filtering only produces 1.26dB SNR.

Importance of randomization

- Random time-dithering retains the randomness necessary for CS recovery, albeit at a lower order than simultaneous-source acquisition.
- To emphasize the importance of randomization, we include the case where we simply decrease the intershot time delays.

- Sparse recovery results in 4.8dB SNR.
- Using linear reconstruction followed by median filtering only produces 1.26dB SNR.

Measurement matrix conditions

- The mutual coherence measures the orthogonality of all columns of A.

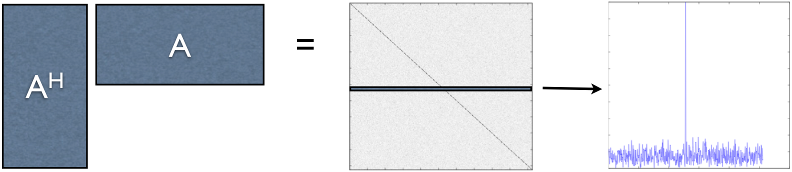
Definition: Mutual Coherence (Donoho and Elad; Bruckstein et al.)

The mutual coherence is equal to the largest inner product between between the normalized columns of A

$$\mu(A) = \max_{1 \leq i \neq j \leq P} \frac{|a_i^H a_j|}{(\|a_i\|_2 \cdot \|a_j\|_2)}.$$

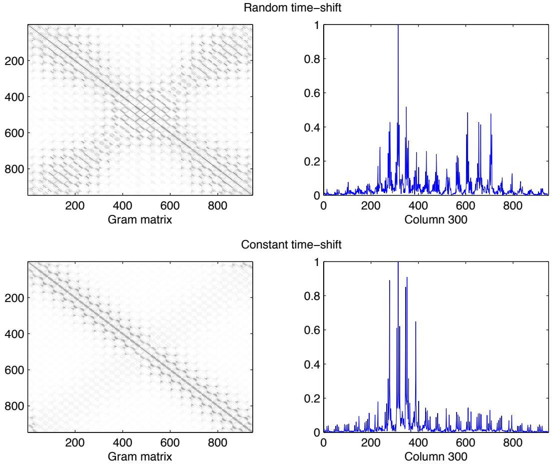
Measurement matrix conditions

- The mutual coherence measures the orthogonality of all columns of A.



Why does the random time-dithering operator work?

Mutual coherence of $A = RMS^H$ (curvelet), $k < \frac{1}{2}(1 + \frac{1}{\mu(A)})$



Steinhaus sequences

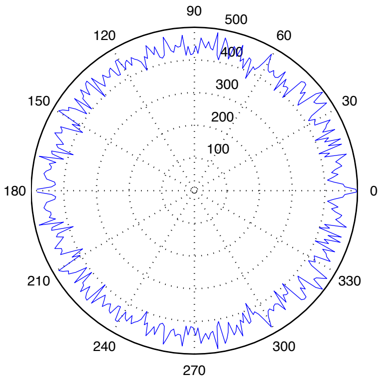
- A Steinhaus random variable v is a unit norm complex random variable uniformly distributed on the complex unit circle: $v = e^{i\theta}, \theta \in U(-\pi, \pi)$.
- A Stainhaus sequence is a collection of independent Steinhaus random variables.
- Do the curvelet coefficients of seismic shot gathers form instances of Steinhaus sequences?

Steinhaus sequences

- A Steinhaus random variable v is a unit norm complex random variable uniformly distributed on the complex unit circle: $v = e^{i\theta}, \theta \in U(-\pi, \pi)$.
- A Stainhaus sequence is a collection of independent Steinhaus random variables.
- Do the curvelet coefficients of seismic shot gathers form instances of Steinhaus sequences?

Steinhaus sequences

- A Steinhaus random variable v is a unit norm complex random variable uniformly distributed on the complex unit circle: $v = e^{i\theta}, \theta \in U(-\pi, \pi)$.
- A Stainhaus sequence is a collection of independent Steinhaus random variables.
- Do the curvelet coefficients of seismic shot gathers form instances of Steinhaus sequences? Distribution of $\mathrm{sgn}(x) = e^{i\theta}$, where $x = re^{i\theta}$.



Why may this be useful?

- An old result by Fuchs(2004) and Tropp(2005) states that if

$$|\langle A_T^{\dagger} a_j, \mathsf{sgn}(x_T) \rangle| < 1$$

then ℓ_1 minimization is guarantees to recovery a signal x supported on T.

Why may this be useful?

- Tropp (2007) then showed that random signals x that are supported on a set T and whose $\mathsf{sgn}(x)$ form Steinhaus sequences, satisfy the condition

$$|\langle A_T^\dagger a_j, \mathsf{sgn}(x_T)\rangle| < 1$$

with probability $1 - 2\rho$, when:

  - A has coherence μ
  - $\sigma_{\min}(A_T) > 1/\sqrt{2}$
  - $8\mu^2 k \leq \log(N/\rho)$

- This condition guarantees recovery and is less strict than the coherence condition for general signals.
- We have yet to find sampling operators RM that combined with a curvelet transform S produce an $A = RMS^H$ that satisfies this condition.

Why may this be useful?

- Tropp (2007) then showed that random signals x that are supported on a set T and whose $\mathsf{sgn}(x)$ form Steinhaus sequences, satisfy the condition

$$|\langle A_T^{\dagger} a_j, \mathsf{sgn}(x_T)\rangle| < 1$$

with probability $1 - 2\rho$, when:

  - A has coherence μ
  - $\sigma_{\min}(A_T) > 1/\sqrt{2}$
  - $8\mu^2 k \leq \log(N/\rho)$

- This condition guarantees recovery and is less strict than the coherence condition for general signals.
- We have yet to find sampling operators RM that combined with a curvelet transform S produce an $A = RMS^H$ that satisfies this condition.

Why may this be useful?

- Tropp (2007) then showed that random signals x that are supported on a set T and whose $\mathsf{sgn}(x)$ form Steinhaus sequences, satisfy the condition

$$|\langle A_T^{\dagger} a_j, \mathsf{sgn}(x_T)\rangle| < 1$$

with probability $1 - 2\rho$, when:

  - A has coherence μ
  - $\sigma_{\min}(A_T) > 1/\sqrt{2}$
  - $8\mu^2 k \leq \log(N/\rho)$

- This condition guarantees recovery and is less strict than the coherence condition for general signals.
- We have yet to find sampling operators RM that combined with a curvelet transform S produce an $A = RMS^H$ that satisfies this condition.

Part 1: Sparsity and randomized acquisition

Part 2: Redundancy in curvelet support information

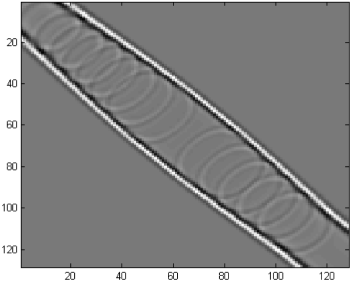
Part 3: Applications in FWI, 3D acquisition and time-lapse imaging

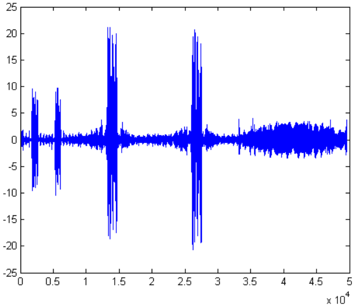
Redundancy in seismic data

- Seismic data are highly redundant.
- Examples from seismic lines:
  - the support of curvelet coefficient of time slices
  - the support of curvelet coefficients of offset slices
- How do we take advantage of this redundancy to reduce the acquisition cost?

Redundancy in seismic data

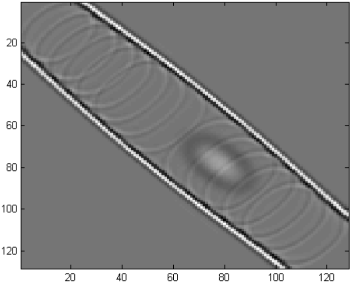
- Seismic data are highly redundant.
- Examples from seismic lines:
  - the support of curvelet coefficient of time slices
  - the support of curvelet coefficients of offset slices
- How do we take advantage of this redundancy to reduce the acquisition cost?

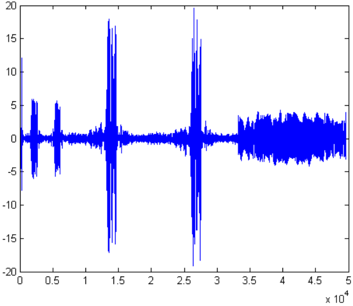




Redundancy in seismic data

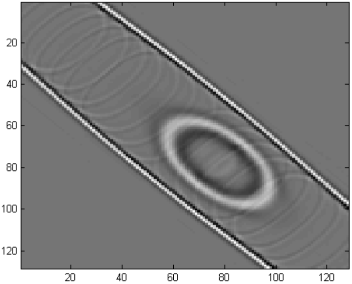
- Seismic data are highly redundant.
- Examples from seismic lines:
  - the support of curvelet coefficient of time slices
  - the support of curvelet coefficients of offset slices
- How do we take advantage of this redundancy to reduce the acquisition cost?

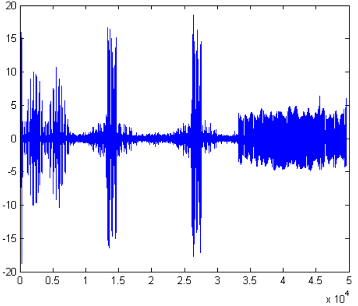




Redundancy in seismic data

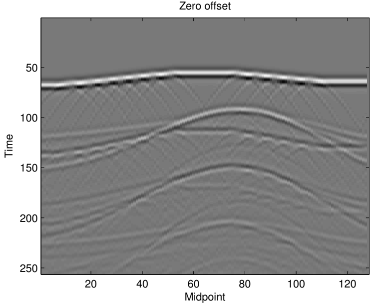
- Seismic data are highly redundant.
- Examples from seismic lines:
  - the support of curvelet coefficient of time slices
  - the support of curvelet coefficients of offset slices
- How do we take advantage of this redundancy to reduce the acquisition cost?

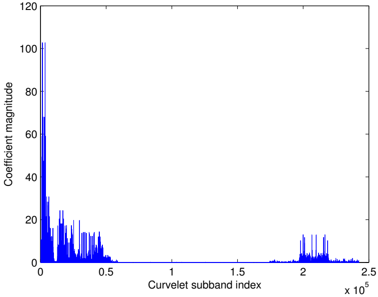




Redundancy in seismic data

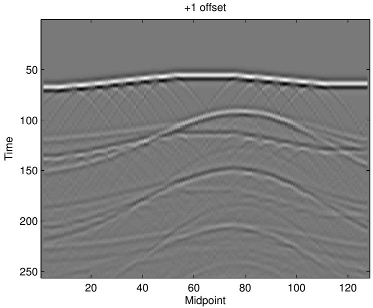
- Seismic data are highly redundant.
- Examples from seismic lines:
  - the support of curvelet coefficient of time slices
  - the support of curvelet coefficients of offset slices
- How do we take advantage of this redundancy to reduce the acquisition cost?

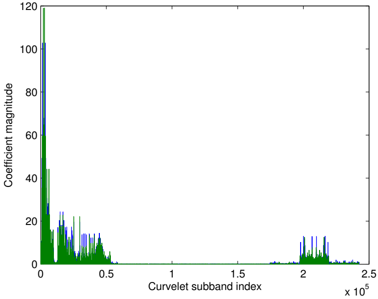




Redundancy in seismic data

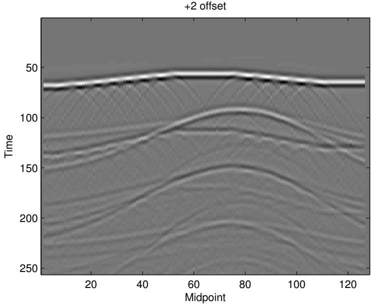
- Seismic data are highly redundant.
- Examples from seismic lines:
  - the support of curvelet coefficient of time slices
  - the support of curvelet coefficients of offset slices
- How do we take advantage of this redundancy to reduce the acquisition cost?

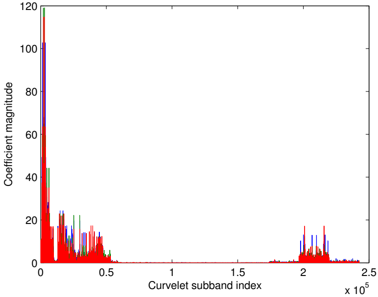




Redundancy in seismic data

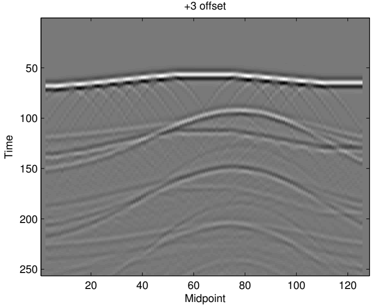
- Seismic data are highly redundant.
- Examples from seismic lines:
  - the support of curvelet coefficient of time slices
  - the support of curvelet coefficients of offset slices
- How do we take advantage of this redundancy to reduce the acquisition cost?

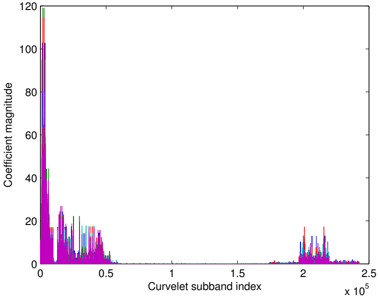




Redundancy in seismic data

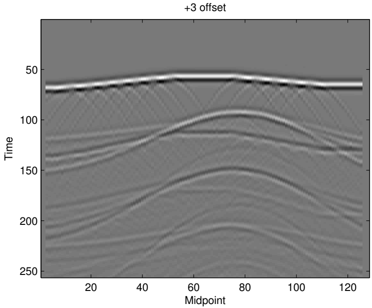
- Seismic data are highly redundant.
- Examples from seismic lines:
  - the support of curvelet coefficient of time slices
  - the support of curvelet coefficients of offset slices
- How do we take advantage of this redundancy to reduce the acquisition cost?

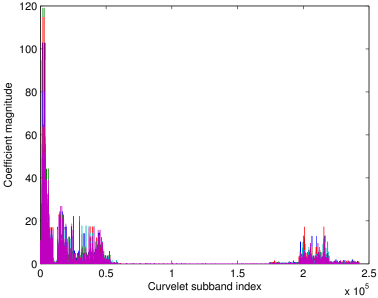




Redundancy in seismic data

- Seismic data are highly redundant.
- Examples from seismic lines:
  - the support of curvelet coefficient of time slices
  - the support of curvelet coefficients of offset slices
- How do we take advantage of this redundancy to reduce the acquisition cost?



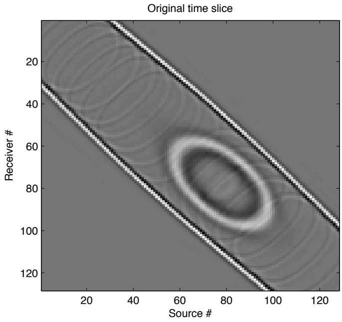


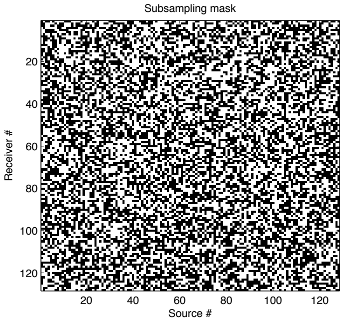
Random subsampling of seismic lines

- Randomize the locations of receivers and shots in the survey design $b = Ax$.
- Translates to random subsampling of a high resolution receiver and shot grid.
- Results in missing data along entire time axis.

Random subsampling of seismic lines

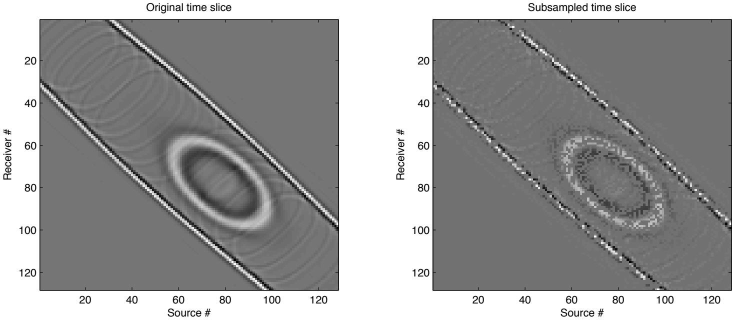
- Randomize the locations of receivers and shots in the survey design $b = Ax$.
- Translates to random subsampling of a high resolution receiver and shot grid.
- Results in missing data along entire time axis.





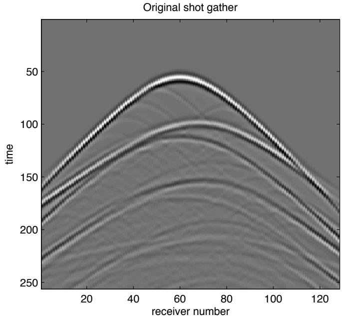
Random subsampling of seismic lines

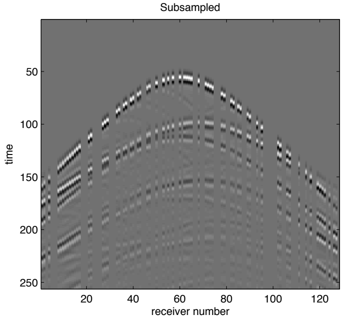
- Randomize the locations of receivers and shots in the survey design $b = Ax$.
- Translates to random subsampling of a high resolution receiver and shot grid.
- Results in missing data along entire time axis.



Random subsampling of seismic lines

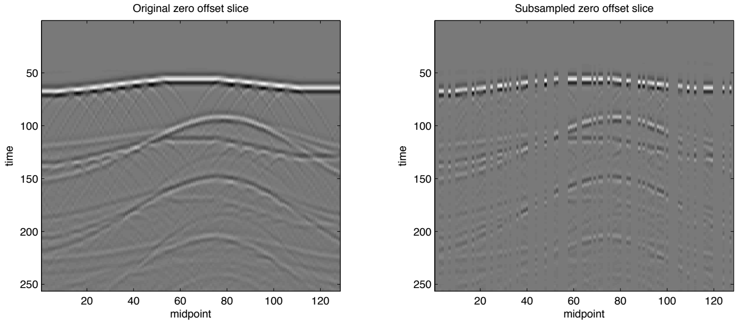
- Randomize the locations of receivers and shots in the survey design $b = Ax$.
- Translates to random subsampling of a high resolution receiver and shot grid.
- Results in missing data along entire time axis.





Random subsampling of seismic lines

- Randomize the locations of receivers and shots in the survey design $b = Ax$.
- Translates to random subsampling of a high resolution receiver and shot grid.
- Results in missing data along entire time axis.



Recovery from random subsampling

- Lessons from the morning talks:
- When sparsity and random subsampling are combined
  - Use ℓ_1 minimization to recover the signal.

$$\min_{x \in \mathbb{R}^N} \|x\|_1 \quad \text{subject to } b = Ax$$

Recovery from random subsampling

- Lessons from the morning talks:

- When sparsity and random subsampling are combined
  - Use ℓ_1 minimization to recover the signal.

$$\min_{x \in \mathbb{R}^N} \|x\|_1 \quad \text{subject to } b = Ax$$

Recovery from random subsampling

- Lessons from the morning talks:

- When additional support information is available (e.g. set $\widetilde{T}$)
  - Use weighted ℓ_1 minimization to recover the signal.

$$\min_x \|x\|_{1,\mathrm{w}} \text{subject to } b = Ax \quad \text{with} \quad \mathrm{w}_i = \begin{cases} 1, & i \in \widetilde{T}^c, \\ \omega, & i \in \widetilde{T}. \end{cases}$$

where $0 \le \omega \le 1$ and $\|x\|_{1,\mathrm{w}} := \sum_i \mathrm{w}_i |x_i|$.

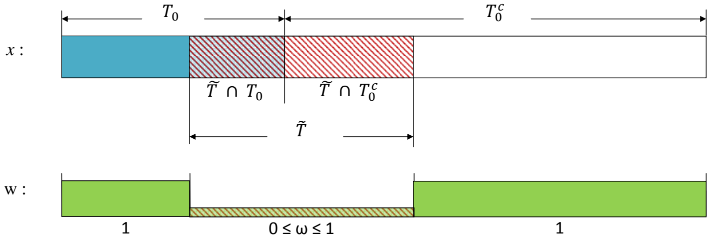
Recovery from random subsampling

- Lessons from the morning talks:

- When additional support information is available (e.g. set $\widetilde{T}$)
  - Use weighted ℓ_1 minimization to recover the signal.

$$\min_x \ \|x\|_{1,\mathrm{w}} \text{subject to } b = Ax \quad \text{with} \quad \mathrm{w}_i = \begin{cases} 1, & i \in \widetilde{T}^c, \\ \omega, & i \in \widetilde{T}. \end{cases}$$

where $0 \leq \omega \leq 1$ and $\|x\|_{1,\mathrm{w}} := \sum_i \mathrm{w}_i |x_i|$.



Practical considerations

- How do we find the set $\widetilde{T}$?
  - Solve the standard ℓ_1 problem for the zero-offset slice.
  - Find the support of the recovered coefficients that contribute 90% of the curvelet coefficient energy.
- Use $\widetilde{T}$ to assign weights $\omega \approx 0.3$ when recovering the adjacent offets.
- Use the support of the recovered adjacent offsets to find new sets $\widetilde{T}$.

Practical considerations

- How do we find the set $\widetilde{T}$?
  - Solve the standard ℓ_1 problem for the zero-offset slice.
  - Find the support of the recovered coefficients that contribute 90% of the curvelet coefficient energy.
- Use $\widetilde{T}$ to assign weights $\omega \approx 0.3$ when recovering the adjacent offets.
- Use the support of the recovered adjacent offsets to find new sets $\widetilde{T}$.

Practical considerations

- How do we find the set $\widetilde{T}$?
  - Solve the standard ℓ_1 problem for the zero-offset slice.
  - Find the support of the recovered coefficients that contribute 90% of the curvelet coefficient energy.
- Use $\widetilde{T}$ to assign weights $\omega \approx 0.3$ when recovering the adjacent offets.
- Use the support of the recovered adjacent offsets to find new sets $\widetilde{T}$.

Practical considerations

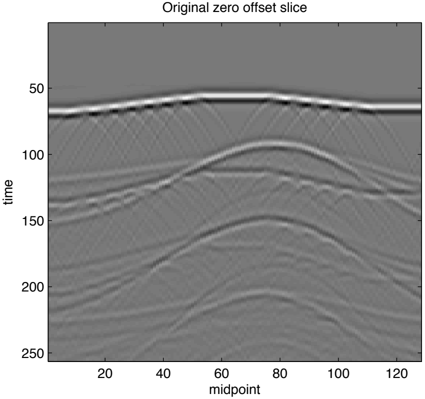
- How do we find the set $\widetilde{T}$?
    - Solve the standard ℓ_1 problem for the zero-offset slice.
    - Find the support of the recovered coefficients that contribute 90% of the curvelet coefficient energy.
- Use $\widetilde{T}$ to assign weights $\omega \approx 0.3$ when recovering the adjacent offets.
- Use the support of the recovered adjacent offsets to find new sets $\widetilde{T}$.

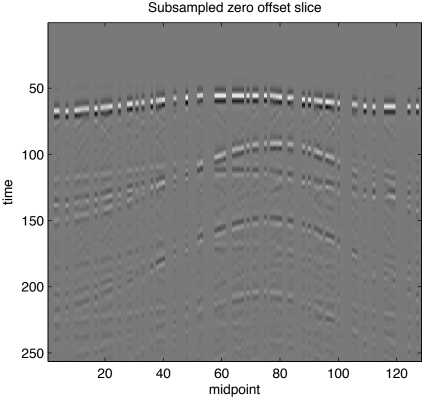
Practical considerations

- How do we find the set $\widetilde{T}$?
  - Solve the standard ℓ_1 problem for the zero-offset slice.
  - Find the support of the recovered coefficients that contribute 90% of the curvelet coefficient energy.
- Use $\widetilde{T}$ to assign weights $\omega \approx 0.3$ when recovering the adjacent offets.
- Use the support of the recovered adjacent offsets to find new sets $\widetilde{T}$.

Recovery using weighted ℓ_1

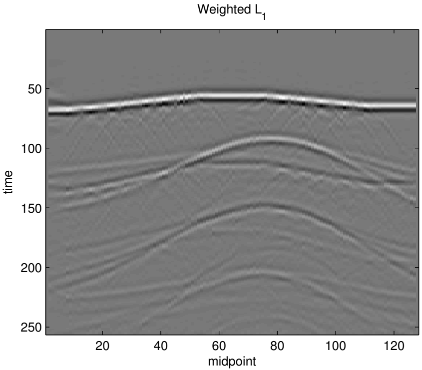
- Offset slices recovered using standard ℓ_1 vs weighted ℓ_1 minimization.
- Average of 2.13dB improvement in shot gather SNR.

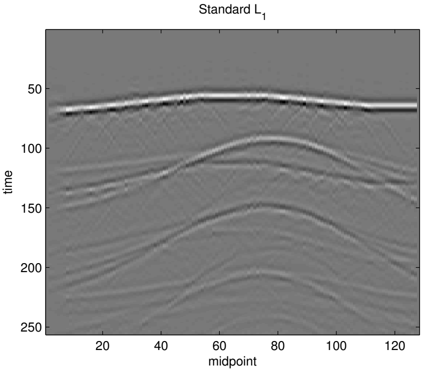




Recovery using weighted ℓ_1

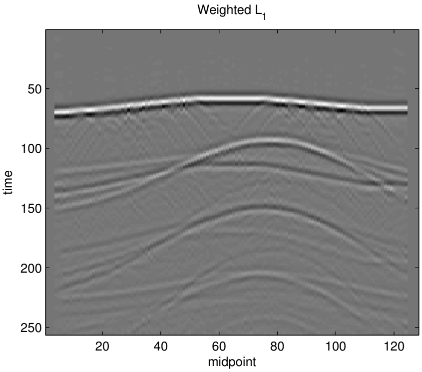
- Offset slices recovered using standard ℓ_1 vs weighted ℓ_1 minimization.
- Average of 2.13dB improvement in shot gather SNR.

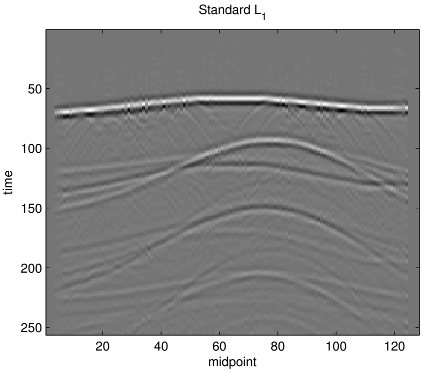




Recovery using weighted ℓ_1

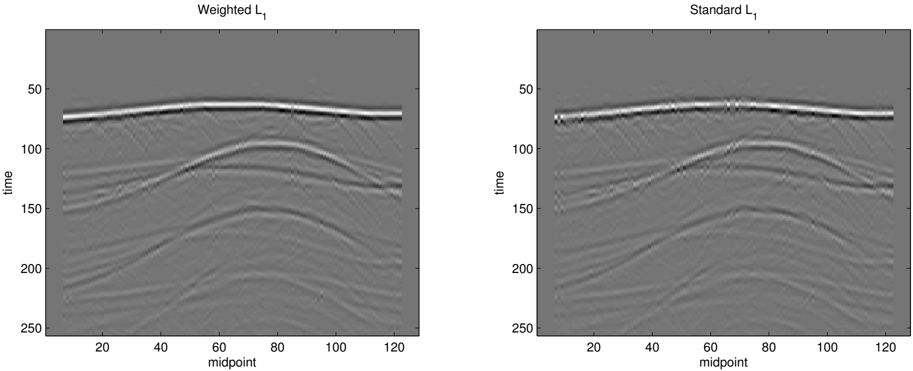
- Offset slices recovered using standard ℓ_1 vs weighted ℓ_1 minimization.
- Average of 2.13dB improvement in shot gather SNR.





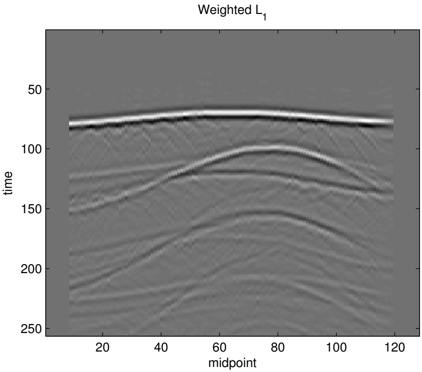
Recovery using weighted ℓ_1

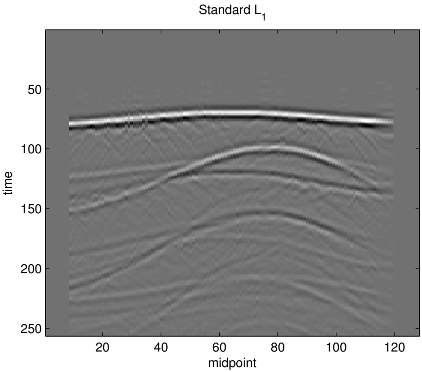
- Offset slices recovered using standard ℓ_1 vs weighted ℓ_1 minimization.
- Average of 2.13dB improvement in shot gather SNR.



Recovery using weighted ℓ_1

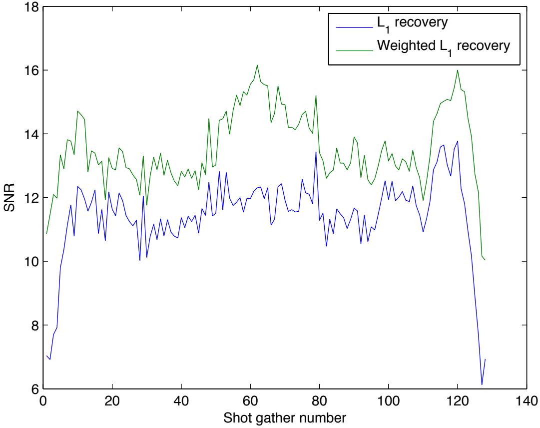
- Offset slices recovered using standard ℓ_1 vs weighted ℓ_1 minimization.
- Average of 2.13dB improvement in shot gather SNR.





Recovery using weighted ℓ_1

- Offset slices recovered using standard ℓ_1 vs weighted ℓ_1 minimization.
- Average of 2.13dB improvement in shot gather SNR.



Part 1: Sparsity and randomized acquisition

Part 2: Redundancy in curvelet support information

Part 3: Applications in FWI, 3D acquisition and time-lapse imaging

Where else can we use weighted ℓ_1?

- So far we have addressed the 2D inpainting problem.

- Other areas that can benefit from redundant support information:
  - Weighted recovery across adjacent azimuths.
  - Weighted recovery across frequency slices.

- How does iterative weighted ℓ_1 affect seismic recovery?

Where else can we use weighted ℓ_1?

- So far we have addressed the 2D inpainting problem.
- Other areas that can benefit from redundant support information:
  - Weighted recovery across adjacent azimuths.
  - Weighted recovery across frequency slices.
- How does iterative weighted ℓ_1 affect seismic recovery?

Where else can we use weighted ℓ_1?

- So far we have addressed the 2D inpainting problem.

- Other areas that can benefit from redundant support information:
  - Weighted recovery across adjacent azimuths.
  - Weighted recovery across frequency slices.

- How does iterative weighted ℓ_1 affect seismic recovery?

Where else can we use weighted ℓ_1?

- So far we have addressed the 2D inpainting problem.
- Other areas that can benefit from redundant support information:
  - Weighted recovery across adjacent azimuths.
  - Weighted recovery across frequency slices.
- How does iterative weighted ℓ_1 affect seismic recovery?

Least-squares migration with sparsity promotion

- Xiang Li will talk about the least-squares migration problem:

$$\delta\tilde{m} = \arg\min_{\delta m} \frac{1}{2}\|\delta d - \nabla\mathcal{F}[m_0, Q]\delta m\|_2^2$$

- Huge overdetermined system, solve over smaller batches iteratively with warm-starting.

δm: model update

δd: multi-source multi-frequency data residue

m_0: background velocity model

Q: sources

$\nabla\mathcal{F}[m_0, Q]$: linearized Born-scattering operator

Least-squares migration with sparsity promotion

- Xiang Li will talk about the least-squares migration problem:

$$\delta\tilde{m} = \arg\min_{\delta m} \frac{1}{2}\|\delta d - \nabla\mathcal{F}[m_0, Q]\delta m\|_2^2$$

- Huge overdetermined system, solve over smaller batches iteratively with warm-starting.

Least-squares migration with sparsity promotion

- Xiang Li will talk about the least-squares migration problem:

$$\delta\tilde{m} = \arg\min_{\delta m} \frac{1}{2}\|\delta d - \nabla\mathcal{F}[m_0, Q]\delta m\|_2^2$$

- Huge overdetermined system, solve over smaller batches iteratively with warm-starting.

- Least-squares migration with sparsity promotion:

$$\delta\tilde{x} = \arg\min_{\delta x} \frac{1}{2}\|\delta x\|_1 \quad \text{subject to} \quad \|\delta\tilde{d} - \nabla\mathcal{F}[m_0, \tilde{Q}]S^*\delta x\|_2^2 \leq \sigma$$

$$\delta\tilde{m} = S^*\delta\tilde{x}, \quad S \text{ is the curvelet transform}$$

Least-squares migration with sparsity promotion

- Xiang Li will talk about the least-squares migration problem:

$$\delta\tilde{m} = \arg\min_{\delta m} \frac{1}{2}\|\delta d - \nabla\mathcal{F}[m_0, Q]\delta m\|_2^2$$

- Huge overdetermined system, solve over smaller batches iteratively with warm-starting.

- Least-squares migration with weighted sparsity promotion:

$$\delta\tilde{x}_{k+1} = \arg\min_{\delta x} \frac{1}{2}\|W_k\delta x\|_1 \quad \text{subject to} \quad \|\delta\tilde{d} - \nabla\mathcal{F}[m_0, \tilde{Q}]S^*\delta x\|_2^2 \leq \sigma$$

W_k : diagonal weighting matrix using the support of $\delta\tilde{x}_k$

3D marine acquisition

- Recover an initial simultaneous/randomized marine seismic line.
- Use the support of the initial seismic line to recover adjacent lines.

Time-lapse imaging

- Use the support of the past survey to emphasize/de-emphasize the past artifacts from the new survey.

Conclusion

- We saw that curvelets sparsify seismic data more effectively.
- The curvelet transform combined with not-so-random sampling operators maintain good compressed sensing qualities.
- The redundancy that exists in seismic data is preserved in subsets of the support of curvelet coefficients.
- We will investigate how to include redundant support information in other recovery algorithms.

Thank you

Questions?

We would like to thank the authors of CurveLab (curvelet.org), a toolbox implementing the Fast Discrete Curvelet Transform, Madagascar (rsf.sf.net), a package for reproducible computational experiments, SPG'1 (cs.ubc.ca/labs/scl/spgl1), SPOT (http: //www.cs.ubc.ca/labs/scl/spot/), a suite of linear operators and problems for testing algorithms for sparse signal reconstruction, and pSPOT, SLIMs parallel extension of SPOT. The Gulf of Suez dataset was generously provided by Eric Verschuur. This work was in part financially supported by the Natural Sciences and Engineering Research Council of Canada Discovery Grant (22R81254) and the Collaborative Research and Development Grant DNOISE II (375142-08). This research was carried out as part of the SINBAD II project with support from the following organizations: BG Group, BGP, BP, Chevron, ConocoPhillips, Petrobras, PGS, Total SA, WesternGeco (Schlumberger)