SINBAD Consortium 2010

Sparse approximations and compressed sensing: an overview

Özgür Yılmaz

SLIM
University of British Columbia

A revolution in sampling theory

- During the last 6 years, we have been witnessing a **revolution in sampling theory**.
- Initiated by the works of Donoho (plenary talk on Monday at the CMS meeting) and of Candès and Tao (Fields medalist), around 2004.
- Opened up a new field called **compressed sensing** or **compressive sampling**: Very active area. "Special Session on Compressed Sensing" at CMS Winter Meeting co-organized by Friedlander, Herrmann, OY. BIRS Workshop in March co-organized by OY.
- Relies heavily on the theory of **sparse approximations** that has been around for more than two decades (transforms such as wavelets, curvelets, Gabor,...).
- Various projects under DNOISE II (aim to) leverage and improve these new techniques.

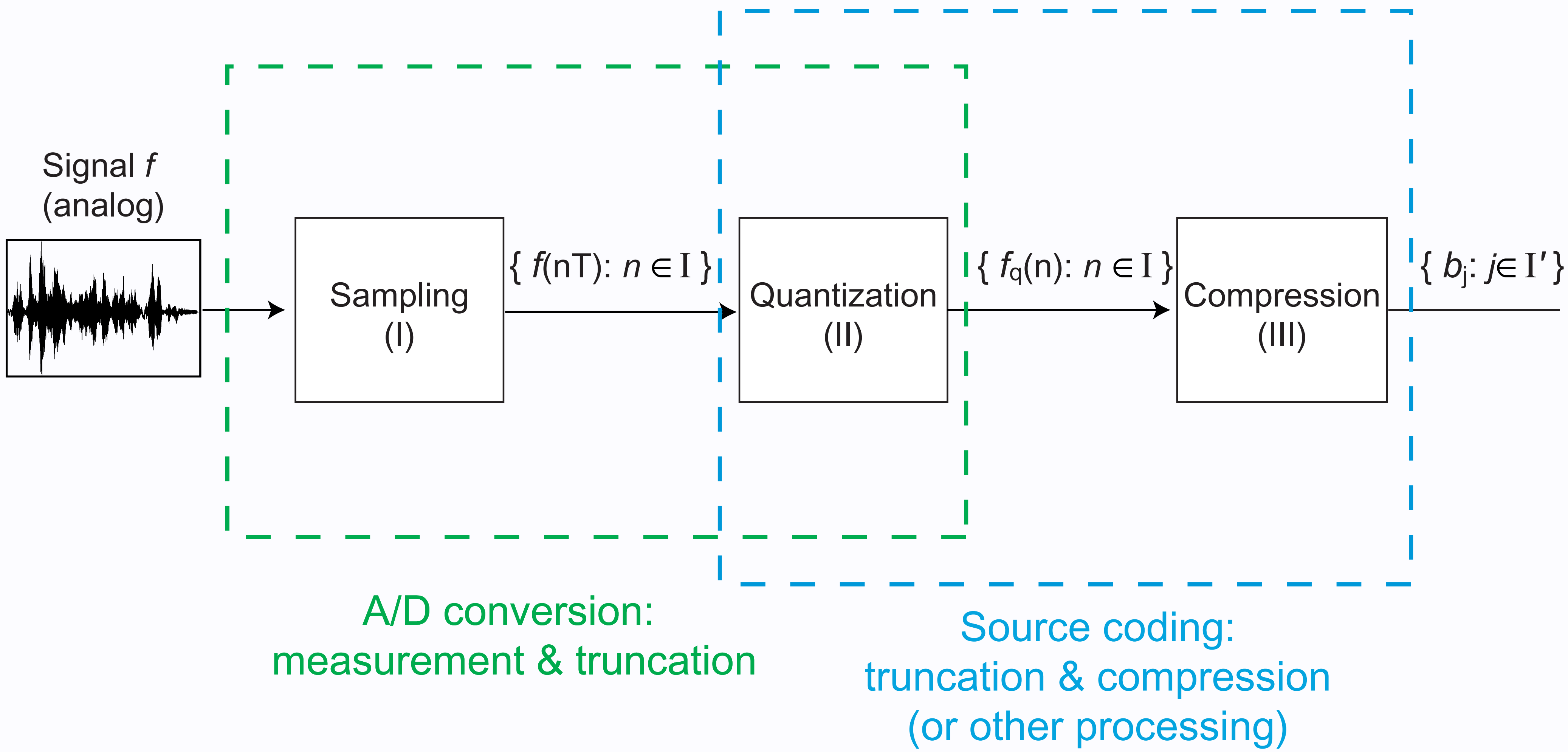
Motivation: Signal Acquisition and Processing

Inherently analog signals: Audio, images, seismic, etc.
Objective: Use digital technology to store and process analog signals – find efficient digital representation of analog signals.

Motivation: Signal Acquisition and Processing

Inherently analog signals: Audio, images, seismic, etc.
Objective: Use digital technology to store and process analog signals – find efficient digital representation of analog signals.

How is this done - classical approach



Classical Approach

Stage I (Sampling)

- samples obtained on a dense temporal/spatial grid,
- an appropriate sampling theorem ties resolution of "reconstruction" with the grid density.

Classical Approach

Stage I (Sampling)

- samples obtained on a dense temporal/spatial grid,
- an appropriate sampling theorem ties resolution of "reconstruction" with the grid density.

Example: Shannon-Nyquist Sampling Theorem.

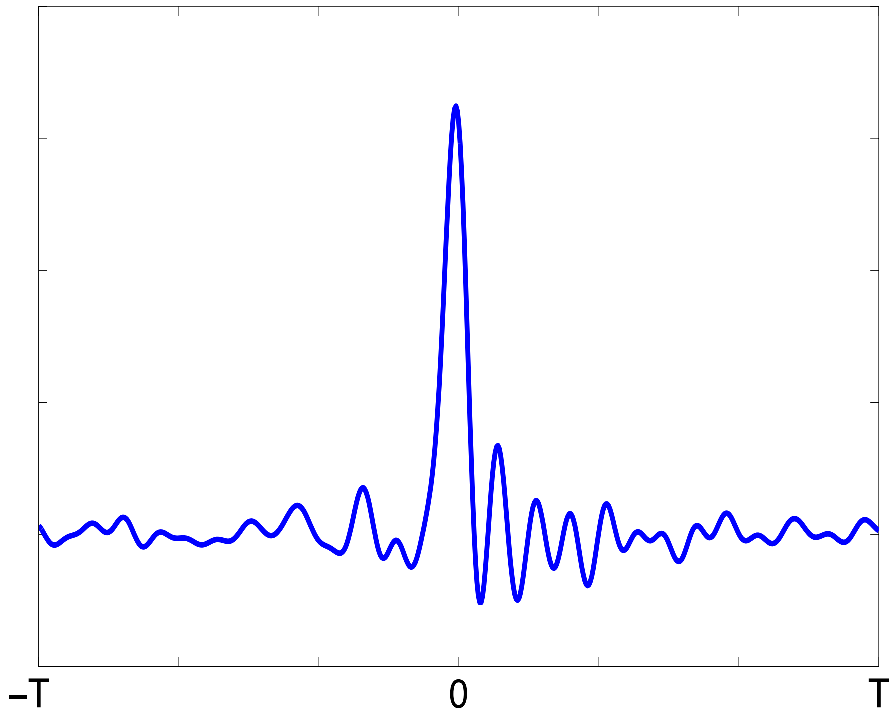
Suppose f is bandlimited with bandlimit Ω, i.e., $f \in B_\Omega$. Then for $\tau < \frac{1}{2\Omega}$, we have

$$f(t) = \sum_n f(n\tau)\phi(t - n\tau), \quad \forall t.$$

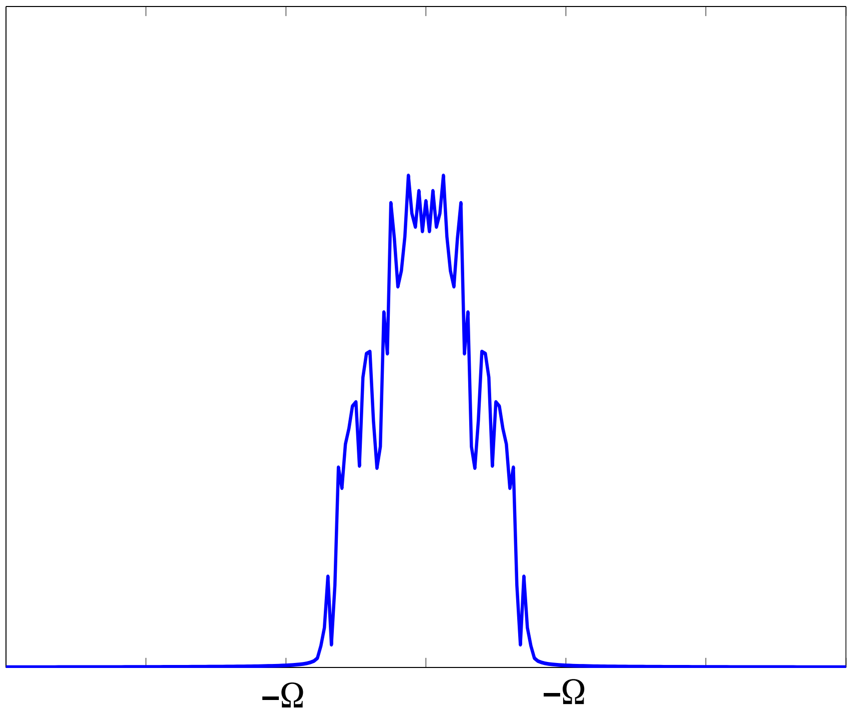
Above ϕ can be chosen with fast decay so the reconstruction is local.

Classical Sampling Theorem: The picture

A bandlimited f

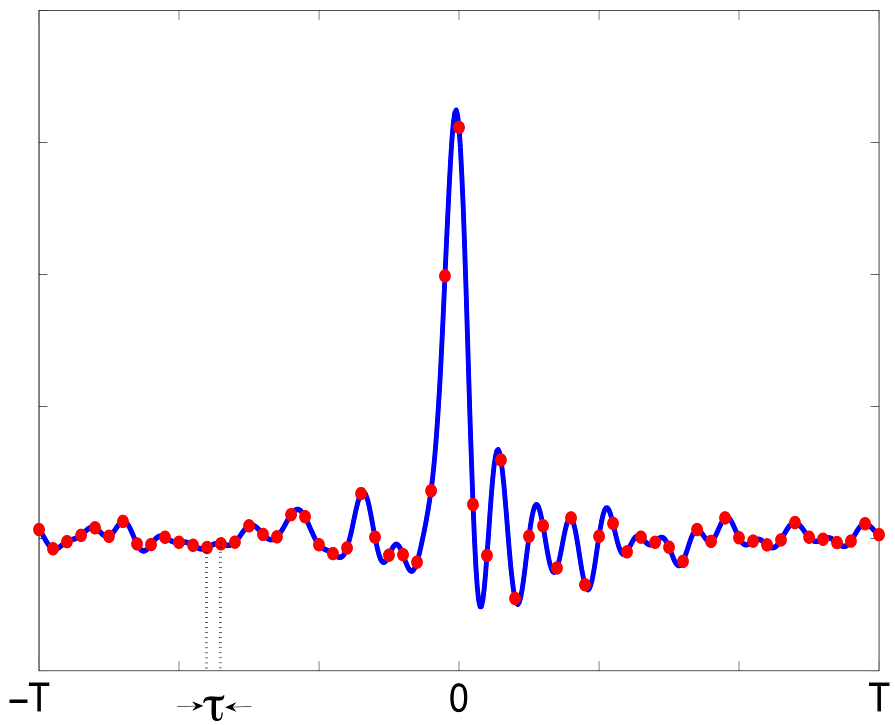


Fourier transform of f

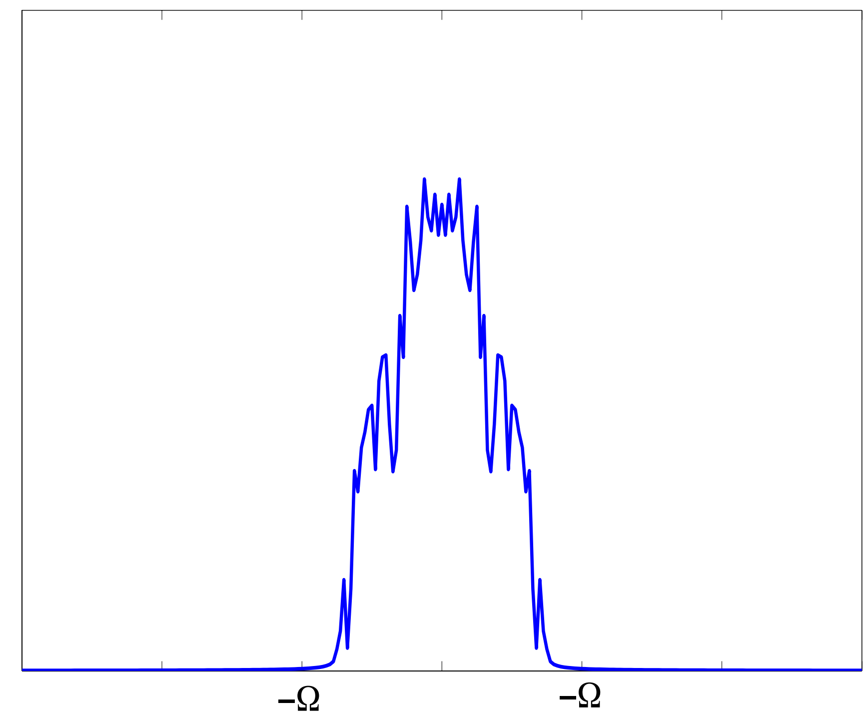


Classical Sampling Theorem: The picture

A bandlimited f



Fourier transform of f



Need $N \approx 2\Omega \times 2T$ samples to reconstruct f on $[-T, T]$.
Equivalently: Every bandlimited function $f \in B_\Omega$, on $[-T, T]$ can be represented by a vector $\mathbf{f} \in \mathbb{R}^N$ which we obtain by collecting N **measurements.**

What makes the classical sampling approach work?

1. f, the signal of interest, is structured $\longleftarrow$ "model the signal class".
2. We measure f by obtaining its samples on a regular grid $\longleftarrow$ "specify the measurement scheme".
3. Use Shannon-Nyquist sampling theorem to reconstruct $\longleftarrow$ "find a reconstruction method".

Note that:

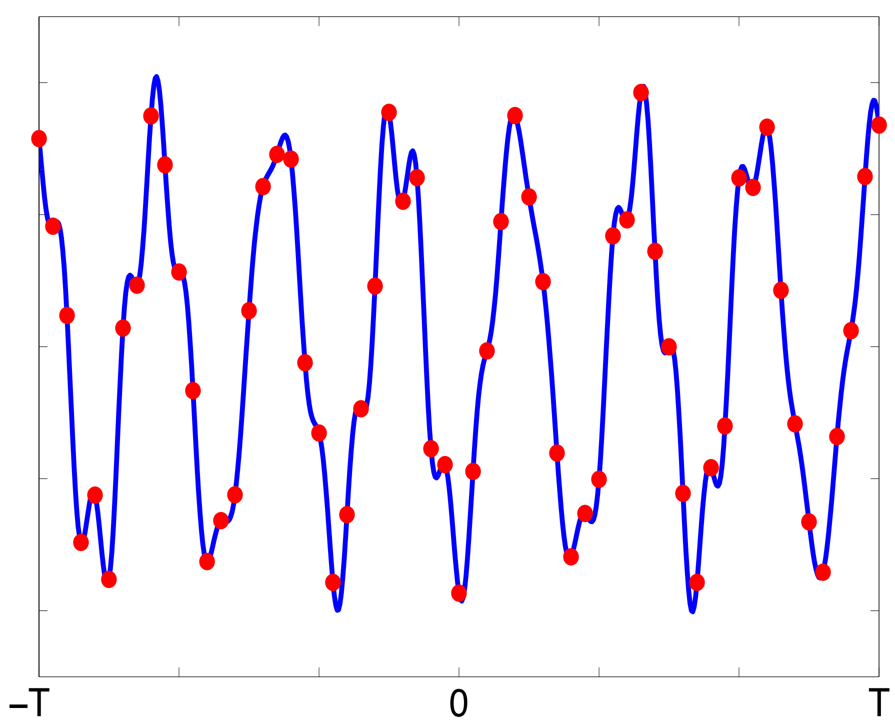
- ambient dimension of the corresponding representation is $N \sim \Omega T$.
- Different N-dimensional vectors correspond to samples of different bandlimited functions – so **no hope for dimension reduction**—i.e., we need N independent measurements— **under this signal model**.

Compressive Sampling Theory

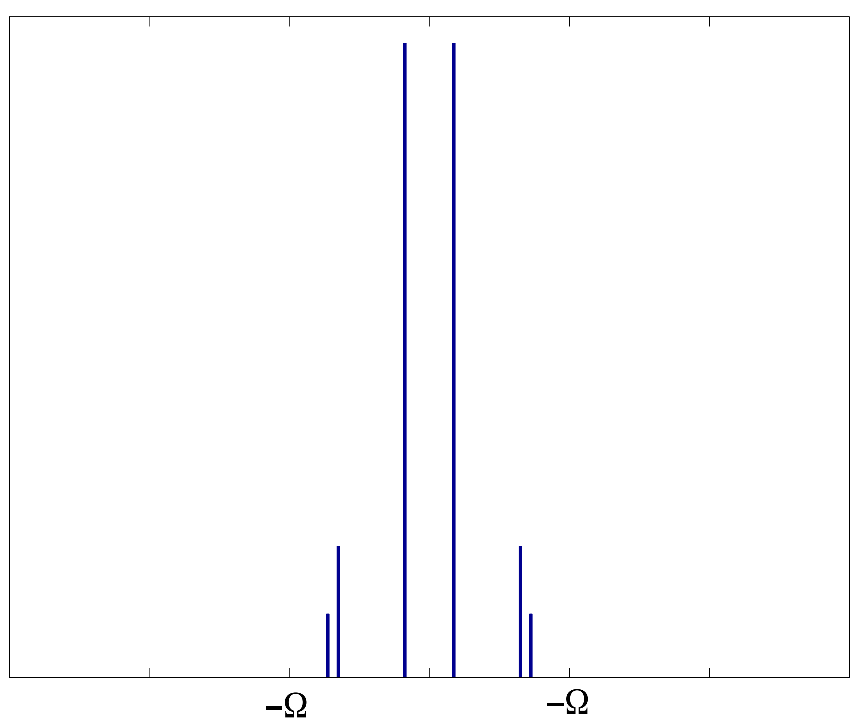
Above: Reduced a bandlimited function f to a vector $\mathbf{f}$ in $\mathbb{R}^N$.

Question: Can we reduce the dimensionality of the problem by **restricting the signal class further**?

Another bandlimited f



Fourier transform of f



An additional constraint of f: its **Fourier transform is sparse**.

Do we still need $N \approx 4\Omega T$ samples to reconstruct $\mathbf{f} \in \mathbb{R}^N$?

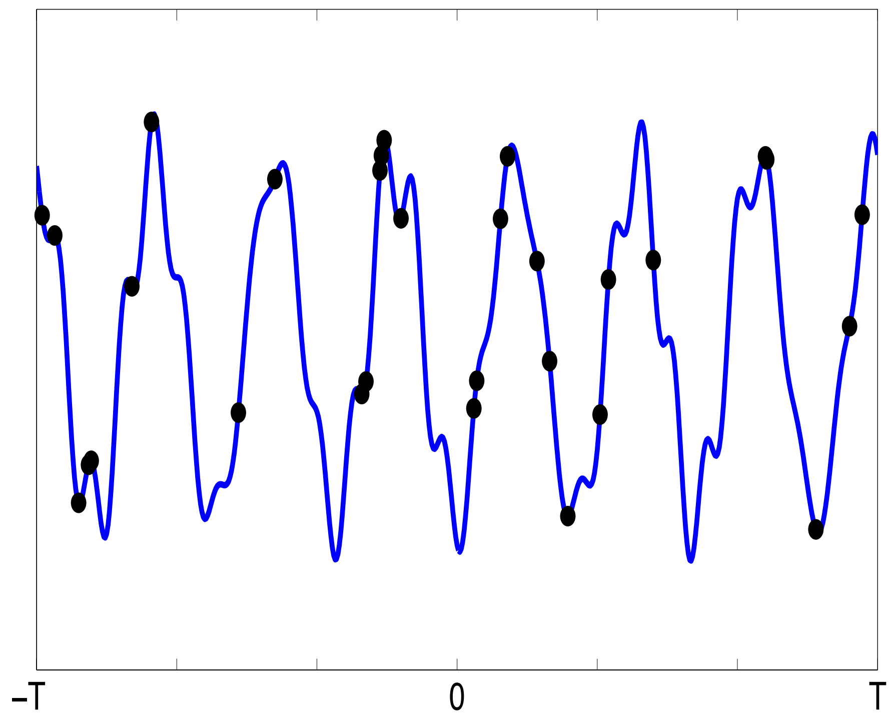
Compressive Sampling Theory

Rephrase the question: Suppose we know that

$$f \in B_\Omega \text{ and } f \text{ has a } \textbf{sparse Fourier transform}.$$

Do we still need to sample at the Nyquist rate for a good (perfect) reconstruction?

Consider the following set of samples (at irregular points):



Here: average sampling density is only 50% of Nyquist rate.
Claim: We can recover f from these samples!

Compressive Sampling Theory – using sparsity

Given:

1. Fourier transform of f is sparse.
2. We only know a few irregular samples, say n, that we showed in the previous slide. Using linear algebra, write:

$$\mathbf{f}_{\text{samples}} = Rf$$

   where R is an $n \times N$ "restriction matrix" (with $n \ll N$).

3. We also know that

$$f = F^* x, \quad \text{where } x \text{ is the Fourier transform of } f.$$

Combining these:

$$\mathbf{f}_{\text{samples}} = RF^* x, \text{ where the only unknown is } x.$$

Can we solve for x? If yes, we recover f via $f = F^* x$.

Compressive Sampling Theory – imposing sparsity

Recall: We have $\mathbf{f}_{\text{samples}} = \underbrace{RF^*}_{A} x$ which we want to solve for x.

Notes:

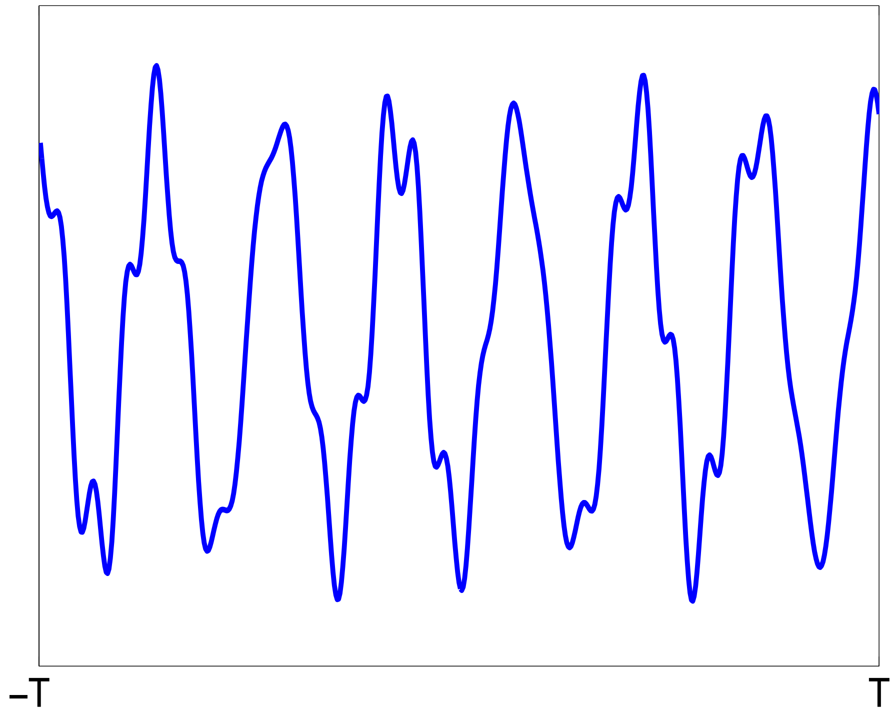
- $A = RF^*$ is an $n \times N$ (short) matrix – i.e., the system is underdetermined and has infinitely many solutions.
- However, we also know that the solution we seek is the Fourier transform of f, thus sparse.
- So, how about the "sparsest" solution of the system above?

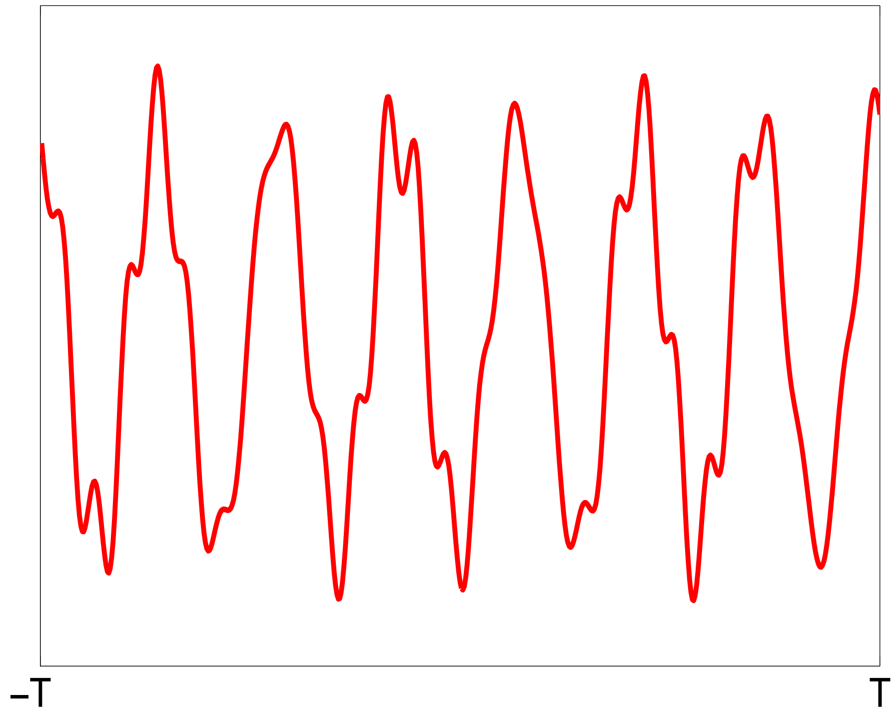
Solve

$$x_{\text{approx}} = \arg\min \|z\|_0 \text{ subject to } Az = \mathbf{f}_{\text{samples}}.$$

Compressive Sampling Theory – imposing sparsity

Here is the reconstruction obtained from the above samples (approx. 50% of Nyquist rate)





- We get essentially perfect reconstruction!
- How did we solve the combinatorial optimization problem:

$$\min \|z\|_0 \text{ subject to } \; Az = \mathbf{f}_{\text{samples}}?$$

We will come back to this later.

Compressive Sampling Theory – general framework

- Signal $f \in \mathbb{R}^N$, want to collect information on f.
- **Model the signal class:** f admits a sparse representation w.r.t. a known basis B: $f = B^*x$ where x is sparse.
- **Specify a measurement scheme:** Construct an $m \times N$ measurement matrix M with $m \ll N$ (above this was the restriction matrix R)
$$\mathbf{f}_{\text{meas}} = Mf = MB^*x$$
- **Reconstruction method:** Solve the underdetermined sparse recovery problem:

  $x_{\text{approx}} =$ "sparsest" z such that $\mathbf{f}_{\text{meas}} = MB^*z$.

Compressive Sampling Theory: main questions

Sparse recovery problem:

$$x_{\text{approx}} = \text{"sparsest" } z \text{ such that } \mathbf{f}_{\text{meas}} = MB^* z.$$

Main questions:

1. How do we find the sparsifying basis B?
2. How do we construct the measurement matrix M?
3. How many measurements do we need to have $x_{\text{approx}} = x$?
4. How do we solve the sparse recovery problem?

Compressive Sampling Theory: main questions

Sparse recovery problem:

$$x_{\text{approx}} = \text{“sparsest” } z \text{ such that } \mathbf{f}_{\text{meas}} = MB^* z.$$

Main questions:

1. How do we find the sparsifying basis B?
2. How do we construct the measurement matrix M?
3. How many measurements do we need to have $x_{\text{approx}} = x$?
4. How do we solve the sparse recovery problem?

Compressive Sampling Theory - sparsity transforms

First address question 1: How do we find sparsity transforms?

- Note that this is dependent heavily on the class of signals of interest.
- In the above example, the sparsity transform was Fourier transform.
- Applied and computational harmonic analysis community has been developing such transforms during the last three decades that are tailored to important signal classes such as: audio, natural images, seismic data and images.
- Rich area with interesting mathematics, directly applicable constructive results such as wavelet transform, curvelet transform etc.
- Next, we give examples of some important sparsity transforms.

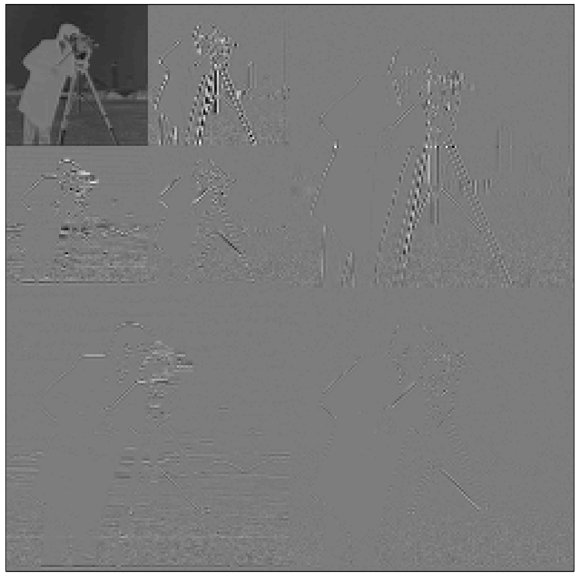
Sparsity transform - natural images

Wavelet transform sparsifies natural images.

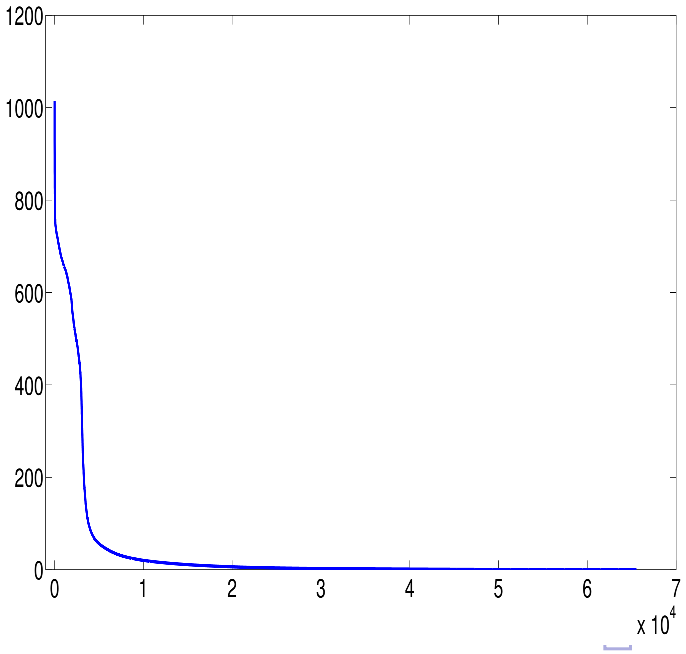
image

a wavelet atom

wavelet transform

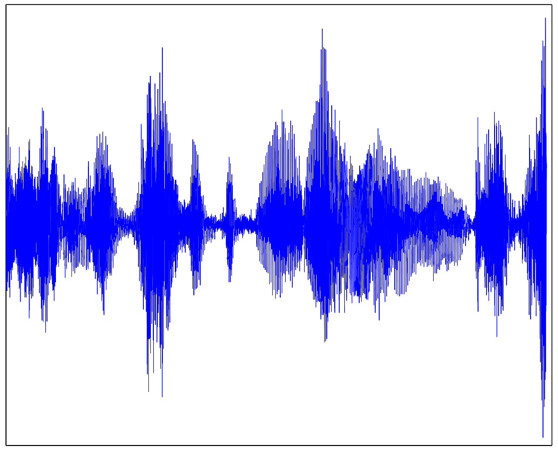


sorted coefficients

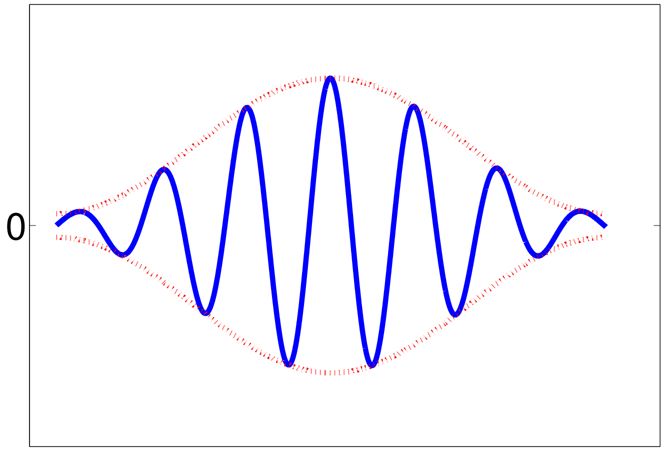


Short-time Fourier (Gabor) transform sparsifies audio signals.

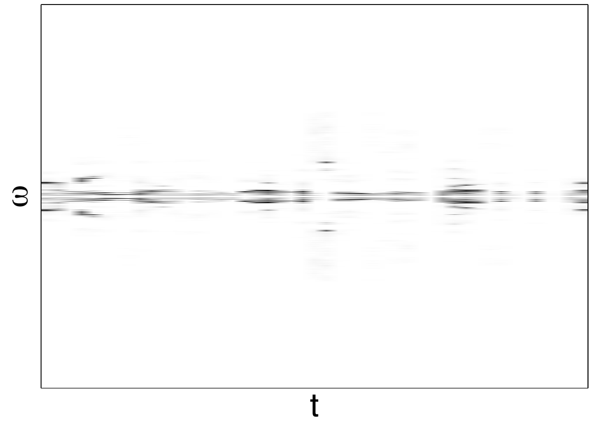
audio signal



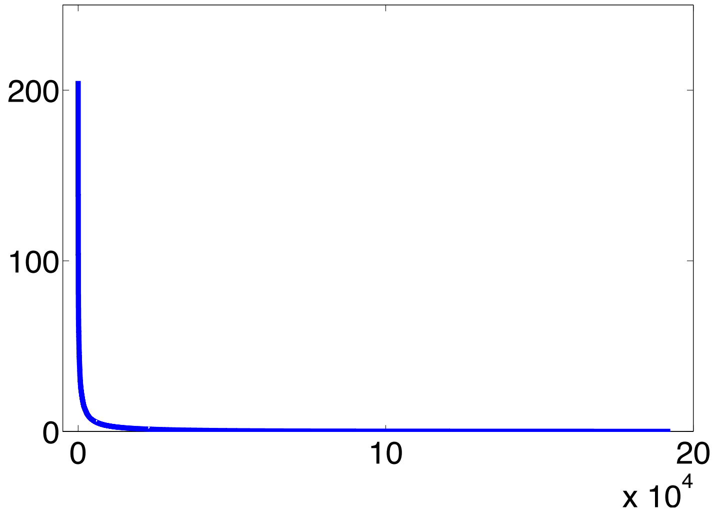
a Gabor atom



STFT transform



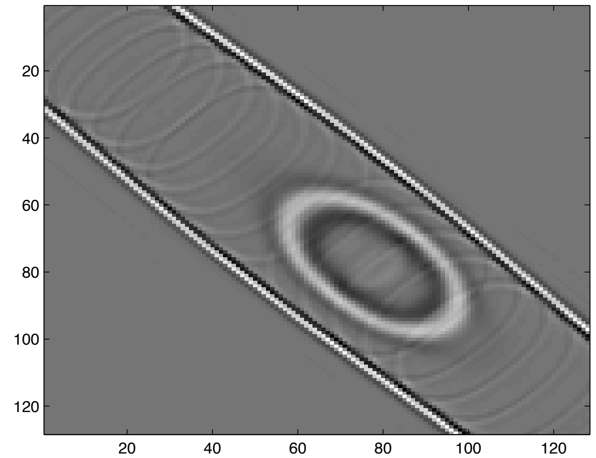
sorted coefficients



Sparsity transform - seismic

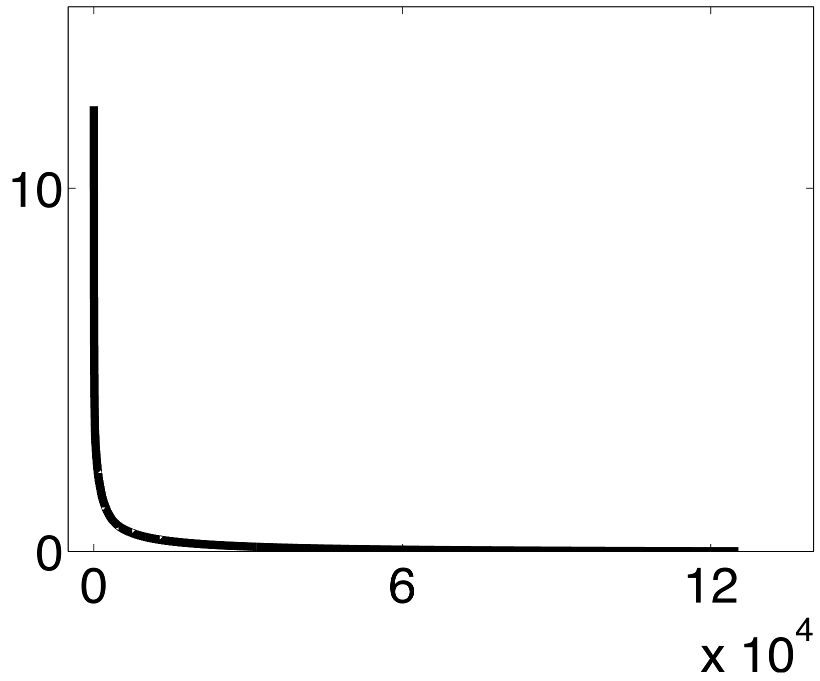
Curvelet transform sparsifies seismic data and images.

sampled Green's function



a curvelet atom

sorted coefficients



Compressive Sampling Theory: main questions

Sparse recovery problem:

$$x_{\text{approx}} = \text{"sparsest" } z \text{ such that } \mathbf{f}_{\text{meas}} = MB^* z.$$

Main questions:

1. How do we find the sparsifying basis B?
2. How do we construct the measurement matrix M?
3. How many measurements do we need to have $x_{\text{approx}} = x$?
4. How do we solve the sparse recovery problem?

Sparse recovery problem

Fix a sparsity basis B and a measurement matrix M (more later).
Set $A = MB^*$. We need to solve:

Recall: A is $n \times N$, $n \ll N$. We are after reconstruction algorithms, i.e., decoders, $\Delta : \mathbb{R}^n \mapsto \mathbb{R}^N$, with the following properties:

C1. $\Delta(Ax) = x$ whenever x is k-sparse (**exact reconstruction for sufficiently small** k).

C2. $\|x - \Delta(Ax + e)\| \lesssim \|e\| + \|x - x_k\|$. Here e: measurement error, e.g., thermal and computational noise. **Reconstruction works for noisy measurements and approx. sparse signals.**

C3. $\Delta(\cdot)$ can be **computed efficiently** (in some sense).

Whether we can achieve C1–C3 will depend on the choice of the measurement matrix A and dimensionality relations between n, N, and k.

Sparse recovery problem – the decoder Δ_0

Given $y = Ax$, the (noise-free) encoding of x, we want to find x. Clearly, this problem is non-trivial:

- **underdetermined** system $y = Az$ has **infinitely** many solutions (provided A is full-rank).
- x is one of these! Decoder must **choose the correct solution**.
- An intuitive decoder: choose the sparsest solution.

$$\Delta_0(b) := \arg\min_z \|z\|_0 \text{ subject to } y = Az.$$

Sparse recovery problem – the decoder Δ_0

Given $y = Ax$, the (noise-free) encoding of x, we want to find x. Clearly, this problem is non-trivial:

- **underdetermined** system $y = Az$ has **infinitely** many solutions (provided A is full-rank).
- x is one of these! Decoder must **choose the correct solution**.
- An intuitive decoder: choose the sparsest solution.

$$\Delta_0(b) := \arg\min_z \|z\|_0 \text{ subject to } y = Az.$$

Theorem (Donoho et al.)

If A is in general position (i.e., its Kruskal rank or its "spark" is n), then $\Delta_0(Ax) = x$ for $x \in \Sigma_s^N$ with $s < n/2$.

Note: Δ_0 is not stable or robust. More importantly, we need to solve a combinatorial optimization problem.

Sparse recovery problem – convex relaxation

The optimization problem for Δ_0 is combinatorial. Need alternatives.

How about ℓ_2 minimization? Choose the solution with smallest 2-norm:

$$\Delta_2(y) := \arg\min_z \|z\|_2 \text{ subject to } y = Az.$$

$\Delta_2 = A^*(AA^*)^{-1}$. The solution is not sparse.

Sparse recovery problem – convex relaxation

The optimization problem for Δ_0 is combinatorial. Need alternatives.

How about ℓ_2 minimization? Choose the solution with smallest 2-norm:

$$\Delta_2(y) := \arg\min_z \|z\|_2 \text{ subject to } y = Az.$$

$\Delta_2 = A^*(AA^*)^{-1}$. The solution is not sparse.

A much better alternative: ℓ_1 minimization. Choose the solution with smallest 1-norm:

$$\Delta_1(y) := \arg\min_z \|z\|_1 \text{ subject to } y = Az.$$

This can be formulated as a convex program. Moreover, unlike 2-norm, 1-norm promotes sparsity. (See talks by M. Friedlander and T. Lin.)

Sparse recovery by 1-norm minimization

Recent exciting developments show that Δ_1 satisfies the conditions (C1)-(C3), thus "equivalent to Δ_0", under certain conditions.

Theorem (Candès-Romberg-Tao, Donoho)

Suppose that A is "sufficiently similar" to an orthonormal matrix. Then, $\exists\ k_{\max}(A)$ such that

$$\|x - \Delta_1(Ax + e)\|_2 \lesssim \|e\|_2 + k^{-1/2}\|x - x_k\|_1$$

for all $k \leq k_{\max}$. In particular,

$$x \text{ is } k\text{-sparse } \Rightarrow \Delta_1(Ax) = x.$$

Remark. Δ_1 satisfies (C1)-(C3) if x is sufficiently sparse. Next, we investigate the dependence of $k_{\max}$ to A.

How to choose the measurement matrix

- There are precise conditions on A (in terms of its RIP constants) that guarantee that the theorem holds.
- For example, if A is a random matrix with iid Gaussian entries, then

$$n \gtrsim k \log(N/k)$$

  will suffice. **Number of measurements scale only logarithmically with the ambient dimension: grid size in our previous example.**
- This is theoretically optimal (deep results in geometric functional analysis).
- Other classes (Bernoulli, partial Fourier, ...) of random matrices will do, too!

How to choose the measurement matrix — more remarks

- Gaussian and sub-Gaussian matrices are unitarily invariant, so the dimension relation is independent of the sparsity basis. These are **universal measurement matrices**:

$$M \text{ is Gaussian and } B \text{ is unitary} \implies A = MB^* \text{ is Gaussian.}$$

- Ideal for dimension reduction in simulations. Also, acquisition with simultaneous sources.
- Difficult to implement depending on the physics—e.g.,, in the sampling example. In such cases:
  - sample in a domain that is incoherent with the sparsity domain: e.g.,

$$\text{sparse in Fourier} \implies \text{sample in time}$$

  - Randomly sub-sample (possibly on a jittered grid), i.e., “apply” a restriction matrix R.

  The corresponding $A = RM$ will be a “good” compressive sampling matrix.

Impact on applications

One of the main conclusions of compressive sampling theory: Accuracy scales logarithmically with the grid size.

Far-reaching implications:

- Digital camera technology (e.g., Baraniuk et al.)
- Medical imaging (e.g., Lustig and Donoho)
- Analog-to-digital conversion (e.g., Baraniuk et al.,)

Impact on applications

One of the main conclusions of compressive sampling theory: Accuracy scales logarithmically with the grid size.

Far-reaching implications:

- Digital camera technology (e.g., Baraniuk et al.)
- Medical imaging (e.g., Lustig and Donoho)
- Analog-to-digital conversion (e.g., Baraniuk et al.,)
- **Seismic imaging and interpolation (SLIM)**
  - Denoising
  - Interpolation – recovery from data with missing traces
  - Primary/multiple separation
  - Acquisition with simultaneous sources
  - "Compressed computation" and full wave inversion
  - ...

Our theoretical contributions

Past work:

- CS recovery algorithms via non-convex optimization (Saab and Y)
- A provably convergent algorithm for coherent source separation (Wang, Saab, Herrmann, Y)
- Efficient quantization (for A/D conversion) of compressive samples (Saab, Y, and collaborators)
- Improved compressive recovery when partial, relatively accurate support information is available via weighted ℓ_1 minimization (Mansour, Saab, Friedlander, Y) – see Mansour's talk.

Ongoing work:

- Leveraging our recent results on weighted ℓ_1 to obtain a "reweighted" version with provable recovery guarantees (Mansour).
- Leverage higher-dimensional structure in the sparsity pattern – measurement matrices that can be factored as Kroenecker products (Saab – see Saab's talk)
- ...

Concluding remarks

- Compressive sampling theory: number of samples scales only logarithmically with the grid size!
- Theory helps us design effective (optimal) acquisition geometries.
- Dimension reduction process is linear; reconstruction is non-linear. (In contrast to classical methods where reconstruction is linear, however dimension reduction is non-linear.)
- Transforming consequences for seismic (as well as other) signal acquisition and processing.
- Very active area of research — 100s of papers in the last few years.
- Active group at UBC, covering this area from all angles: theory, algorithms, and applications to various problems in exploration seismology – see the upcoming talks.

Acknowledgements

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, BP, Chevron, ConocoPhillips, Petrobras, Total SA, and WesternGeco.