Sparsity

- emerging role over last $\sim 4$ decades
- powerful tool
  - for data analysis
  - for computation
- better understanding will have enormous impact
What is an image?
Data analysis: Images

Sparsity in Data Analysis and Computation
Data analysis: Images

\[ f: [a, b] \times [c, d] \rightarrow \mathbb{R}_+ \]
Data analysis: Images

\[ f : [a, b] \times [c, d] \rightarrow \mathbb{R}_+ \]
Data analysis: Images

\[ f : [a, b] \times [c, d] \rightarrow \mathbb{R}_+ \]
Data analysis: Images

\[ f : [a, b] \times [c, d] \rightarrow \mathbb{R}_+ \]
Data analysis: Images

\[ f : [a, b] \times [c, d] \rightarrow \mathbb{R}_+ \]
Data analysis: Images

\[ f : [a, b] \times [c, d] \rightarrow [0, 1] \]

Sampled:

\[ f \left( a + \frac{b-a}{M} m, c + \frac{d-c}{N} n \right) \]

\[ m = 0, 1, \ldots, M \]
\[ n = 0, 1, \ldots, N \]

Or averaged:

\[ \int_{a - \frac{b-a}{M}}^{a + \frac{b-a}{M}} \int_{c - \frac{d-c}{N}}^{c + \frac{d-c}{N}} g(s, t) \, ds \, dt \]
In any case: object in high-dimensional space.

For practical purposes: need compression, storage, transmission, analysis.
In any case: object in high-dimensional space.

For practical purposes: need compression

How?

\[ \text{exploit mathematical properties of the class} \]

Translation invariance!

Each of these "snapshots" should be in the class
Invariance under Translation Group

\[ \Rightarrow \text{use irreducible representations to decompose the class} \]

\[ \Rightarrow \text{Fourier analysis!} \]

Indeed: JPEG standard for image compression uses DCT (discrete cosine transform)
Data analysis: Images

Sparsity in Data Analysis and Computation
JPEG standard:

uses DCT on $8 \times 8$ blocks

- technical reasons
  - in early 80s: $16 \times 16$
  - expected to go to even larger...
In 1980s: start of use of wavelet transform for images.

$\rightarrow$ decomposition of images into different types of building blocks.
Data analysis: Images
Wavelets

- High frequency wavelets much more "narrow" than low frequency wavelets

$\Rightarrow$ Need many more fine scale wavelets to cover the image domain than coarse scale wavelets

$\Rightarrow$ Traditional representation of wavelet decompositions of an image.
Compression Ratio: 3.3%
In JPEG-2000 image standard:

- wavelets instead of DCT.

Major reasons:
- graceful degradation as rate drops
- ease of implementing lossy/lossless compr.

Impact:
- none really on consumer products
- digital movies, sports reporting
Why are wavelets a good idea for images? What was "wrong" with the Fourier analysis argument?

Really the difference between Linear and Non-linear approximation.
Consider a simple class of functions on $\mathbb{T}$

$$f \in \mathcal{C}$$

$$f : \mathbb{T} \rightarrow \mathbb{C} \quad \text{“nice”}$$

on $\mathcal{C}$: probability measure

invariant under translations

Then one can prove that the “best” basis in which the $f \in \mathcal{C}$ can be decomposed is the Fourier basis
Namely:

If one wants to find the basis \( \varphi_1, \varphi_2, \ldots, \varphi_n \) of functions such that

\[
\mathbb{E} \left( \int \left| g(t) - \sum_{n=1}^{N} \langle g, \varphi_n \rangle \varphi_n(t) \right|^2 dt \right)
\]

are minimal, then these must be the Fourier exponentials \( e^{\pm \pi int} \)
However, consider the following example:
However, consider the following example:
Data analysis: Images

However, consider the following example:

Clearly translation invariant process...
Yet, one can prove that

\[
E \left( \frac{1}{T} \int_{T} |g(t) - \sum_{|n| \leq N} \langle g, e_n \rangle e_n(t) |^2 \, dt \right) \geq C \frac{1}{N}
\]

But with a wavelet expansion it is very simple to find a strategy that does better ...
Data analysis: Images

Sparsity in Data Analysis and Computation
One easily proves that with this strategy,

\[
\mathbb{E} \left( \frac{1}{T} \int | \hat{f}(t) - A_{2N+1} f(t) |^2 dt \right) \leq C N^{-2}
\]

where \( A_{2N+1} f \) is an approximation to \( f \) that uses only \( 2N+1 \) coefficients.
Data analysis: Images

But there is an enormous difference.

In 1 case
\[ E(\|f - \sum_{\ell=1}^{L} \langle f, \varphi_{\ell} \rangle \varphi_{\ell}\|^{2}) \]
is minimal

In the other case,
\[ E(\|f - \sum_{\ell \in \Lambda_{L}(f)} \langle f, \varphi_{\ell} \rangle \varphi_{\ell}\|^{2}) \]
is considered, with \# \Lambda_{L}(f) = L

In both cases, \( L \) coeffs allowed, but in 2nd case their choice can depend on \( f \).
Data analysis: Images

**Linear approximation:**

\[ \varphi_1, \varphi_2, \ldots, \varphi_n, \ldots \Rightarrow \text{Span}(\varphi_1, \ldots, \varphi_n) = V_n \]

and study \( \text{dist}(f, V_n) \)

**Nonlinear approximation:**

\[ \sum_n = \{ \sum_{\ell \in \mathbb{N}} c_\ell \varphi_\ell : \#\{ \ell ; c_\ell \neq 0 \} < n \} \]

now study \( \text{dist}(f, \sum_n) \).
Wavelets are a good basis for nonlinear approximation of images, because images have sparse wavelet expansions.

With hindsight: first example of benefit of sparse expansions.

Why do wavelets have this property?
Wavelets are connected with beautiful and strong theorems in harmonic analysis. Calderón-Zygmund theory

In fact, wavelets are not even the best basis for 2D-images.

Images really need curvelets (or shearlets).

For wavelets, we were lucky: we "guessed" a good basis.

Can we search for a good basis for sparse expansions?
Find good basis for sparse expansions?

- Search within "dictionaries" union of many bases.
- Nonlinear (adaptive) singular value decompositions
Each column is a vector in the dictionary.

Same kind of situation as in compressed sensing!
Compressed sensing.

Back to images, for a moment.

Images are sparse when expressed as a combination of wavelets.

For compression applications:
- use fast transform to decompose into wavelets
- retain only the significant coeffs. (identity depends on image)

Why bother first getting all these coeffs?
Why not "acquire" image sparsely?
In other words, if we know $x \in \mathbb{R}^n$ is a sparse vector, i.e., $\sum \neq 0 \leq k \ll N$, can we then determine $x$ by making fewer than $N$ measurements?

Answer: yes!
Compressive sensing is related to results in theoretical computer science.

Use Johnson-Lindenstrauss Lemma.

\[ \mathbf{x}_1, \ldots, \mathbf{x}_L \] vectors in \( V \).
\[ \text{dim } V = D \]

Consider projections of \( \mathbf{x}_i \) on randomly picked d-dim. subspace of \( V \).

Compare \( \langle \mathbf{P}_s \mathbf{x}_i , \mathbf{P}_s \mathbf{x}_j \rangle \equiv \frac{D}{d} \) with \( \langle \mathbf{x}_i, \mathbf{x}_j \rangle \).

How large should \( d \) be for these \( \mathbf{P} \) matrices to be close with high probability?

**Basically**: \( \log L \)
This result in CS has had a tremendous impact on:

- verify that proofs are correct with high probability by "random sampling"
- fast computation algorithms (with small probability of failure)
Fast computations: example.

\[ f \in \mathbb{C}^N \quad N \text{ huge} \]

There exists \( x \in \mathbb{C}^N \), with only \( K \ll N \) non-zero entries, that is close to \( f \).

\[ \Rightarrow \quad \text{To get a good approximation to } f, \]

one needs to take only \( O(K \log N) \) random samples of \( f \)

and algorithm runs in \( O(K \log N) \) time as well.
Finding good ways to represent data.

Knowing (or "believing") that there is a sparse expansion can be exploited to reconstruct from seemingly very insufficient data.

Search in a dictionary

$\leftrightarrow$ compressed sensing.

Find the dictionary if given a class of objects?
Johnson–Lindenstrauss  \[ \rightarrow \] dimension reduction.

Compressed sensing  \[ \rightarrow \] dimension reduction.

One last salvo about computation made feasible by “dimension reduction”

\[ \rightarrow \] comparing surfaces with applications to biology.
Sparsity in Data Analysis and Computation

\[ d_p \left( L, L' \right) \]

\[ = \inf_{m : L \to L', \text{ matching}} \left[ \min_{R \in \text{Euclidean gp.}} \sum_{p \in L} \| m(p) - R p \|^2 \right]^{1/2} \]

\[ D_p \left( S, S' \right) \]

\[ = \inf_{C : S \to S', \text{ area-preserving}} \left[ \min_{R \in \text{Euclidean gp.}} \int_{S} \| C(x) - R x \|^2 \ dA_x \right]^{1/2} \]
If \( \mathcal{D}^p(S, S') \) is small,
then \( \exists \) conformal map \( m : S \rightarrow S' \)
so that
\[
\min_R \int_R \| m(x) - R \|^2 \, dA_S \leq C \mathcal{D}^p(S, S')^{1/2}
\]

\[\text{use this to compute approx. to } \mathcal{D}^p(S, S')\]
by searching "deformations of conformal maps"
A. Observer Placed Landmarks

B. cP determined correspondence map between two structures

C. Propagated Landmarks

D. Observer Placed Landmarks

Sparsity in Data Analysis and Computation
With apologies to

A. Cohen  
J.P. d'Ales  
Y. Meyer  
R. DeVore  
R. Coifman  
E. Candès  
D. Donoho  
J. Romberg  
T. Tao  
W. Dahmen

L. Carin  
A. Gilbert  
M. Strauss  
R. Calderbank  
Y. Lipman  
O. Yilmaz  
D. Boyer  
J. Jernvall

and many, many more...