

Sparsity in Data Analysis and Computation

Duke workshop on Sparsity, July 2011

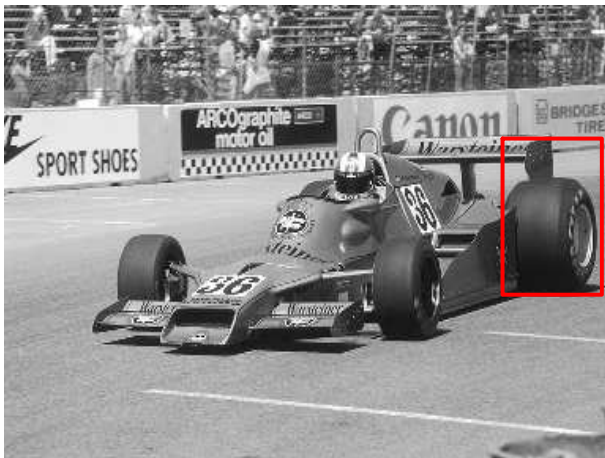
Sparsity

- emerging role over last ~ 4 decades
- powerful tool
 - for data analysis
 - for computation
- better understanding will have enormous impact

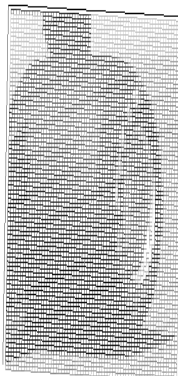
Data analysis: Images

What is an image?

Data analysis: Images

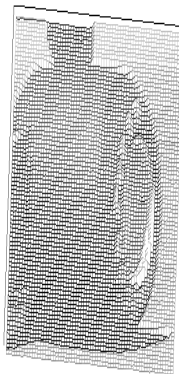


Data analysis: Images



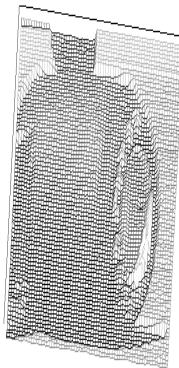
$$f : [a, b] \times [c, d] \longrightarrow \mathbb{R}_+$$

Data analysis: Images



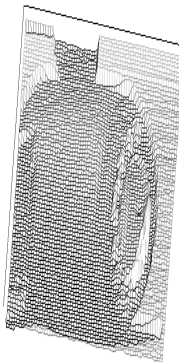
$$f : [a, b] \times [c, d] \longrightarrow \mathbb{R}_+$$

Data analysis: Images



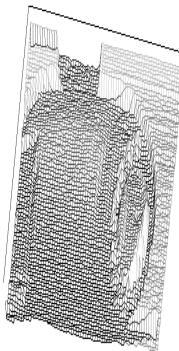
$$f : [a, b] \times [c, d] \longrightarrow \mathbb{R}_+$$

Data analysis: Images



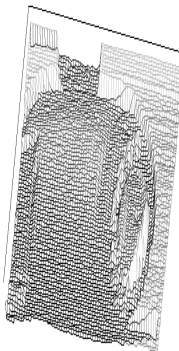
$$f : [a, b] \times [c, d] \longrightarrow \mathbb{R}_+$$

Data analysis: Images



$$f : [a, b] \times [c, d] \longrightarrow \mathbb{R}_+$$

Data analysis: Images



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

$$\text{sampled: } f\left(a + \frac{b-a}{M} m, c + \frac{d-c}{N} n\right)$$

$$m = 0, 1, \dots, M$$

$$n = 0, 1, \dots, N$$

or averaged :

$$\int_{\left|s - \frac{b-a}{M} m\right| < \frac{b-a}{2M}} \int_{\left|t - \frac{d-c}{N} n\right| < \frac{d-c}{2N}} f(s, t) \, ds \, dt$$

Data analysis: Images

In any case : object in high-dimensional space.

For practical purposes : need compression

storage , transmission, analysis

Data analysis: Images

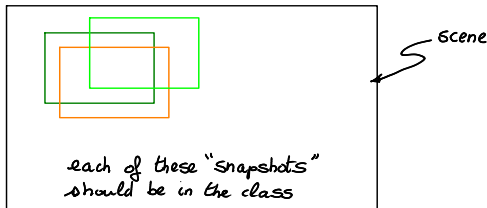
In any case : object in high-dimensional space.

For practical purposes : need compression

How ?

→ exploit mathematical properties
of the class

Translation invariance !



Invariance under Translation Group

⇒ use irreducible representations to decompose the class

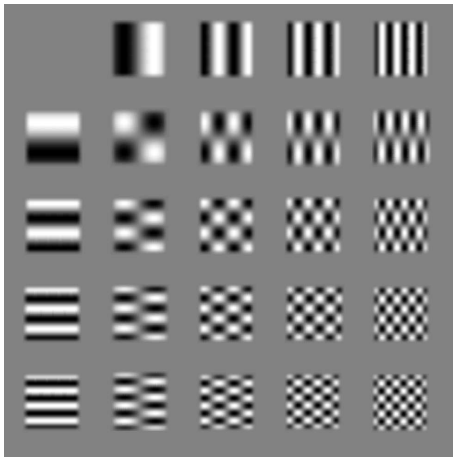
⇒ Fourier analysis!

Indeed:

JPEG standard for image compression
uses DCT

(discrete cosine transform)

Data analysis: Images



JPEG standard :

uses DCT on 8×8 blocks



technical reasons

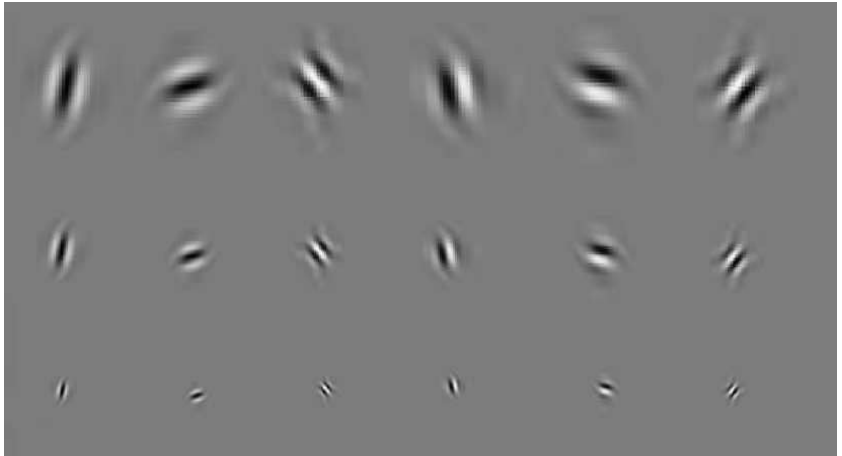
in early 80s: $\rightarrow 16 \times 16$

expected to go to even larger...

In 1980s : start of use of
Wavelet transform for images.

→ decomposition of images
into different type of building blocks.

Data analysis: Images



Wavelets

- high frequency wavelets much more "narrow" than low frequency wavelets
- ⇒ need many more fine scale wavelets to cover the image domain than coarse scale wavelets
- ⇒ traditional representation of wavelet decompositions of an image.



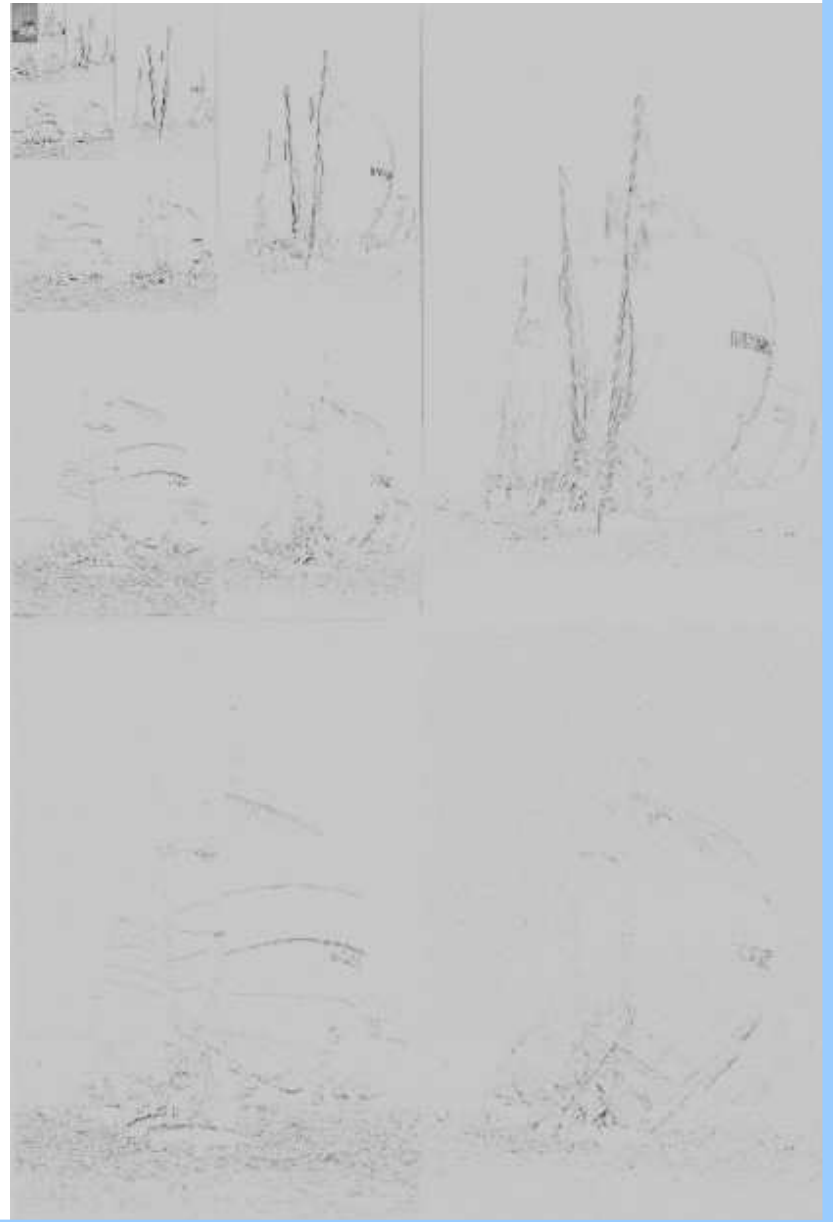










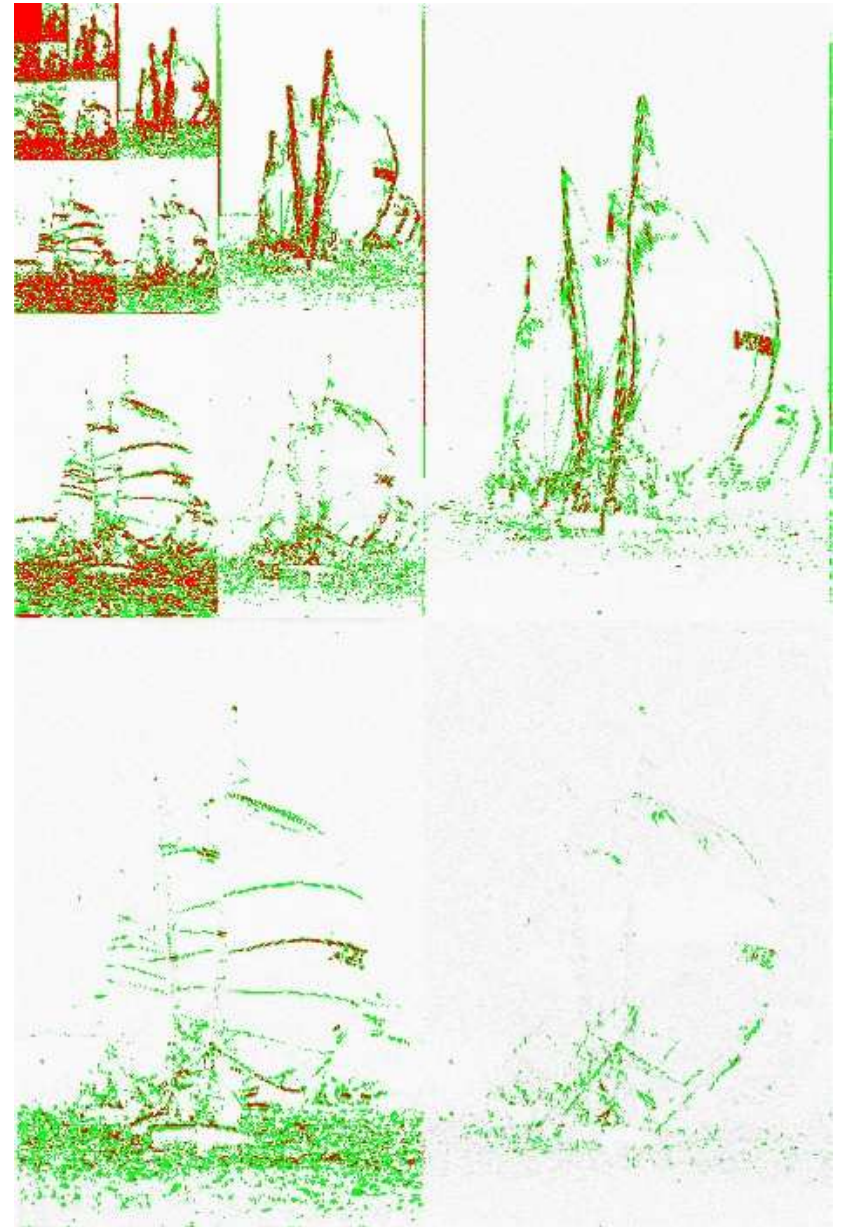


Compression





Compression Ratio: 3.3%



Compression Ratio: 10%

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

Data analysis: Images

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}

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

$f: \mathbb{T} \longrightarrow \mathbb{C}$ "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

Data analysis: Images

Namely :

If one wants to find the basis $\varphi_1, \varphi_2, \dots, \varphi_n, \dots$ of functions such that

$$\mathbb{E} \left(\int_{\mathbb{T}} |f(t) - \langle f, \varphi_1 \rangle \varphi_1(t)|^2 dt \right)$$

\vdots

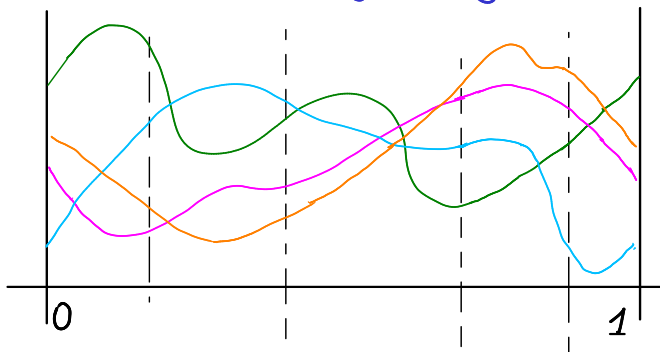
$$\mathbb{E} \left(\int_{\mathbb{T}} |f(t) - \sum_{n=1}^N \langle f, \varphi_n \rangle \varphi_n(t)|^2 dt \right)$$

\vdots

are minimal, then these must be the Fourier exponentials $e^{2\pi i n t}$

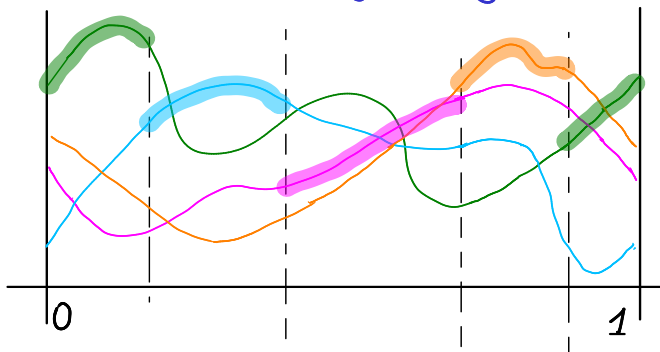
Data analysis: Images

However, consider the following example:



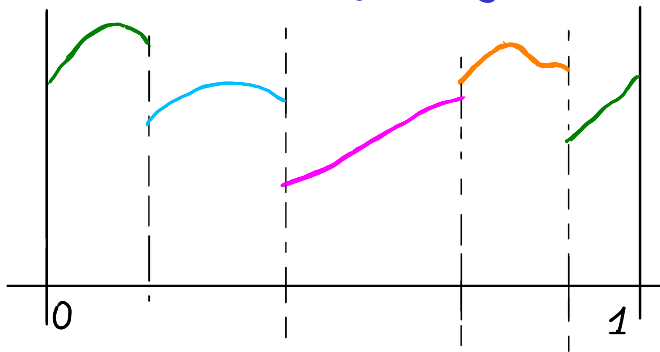
Data analysis: Images

However, consider the following example:



Data analysis: Images

However, consider the following example:



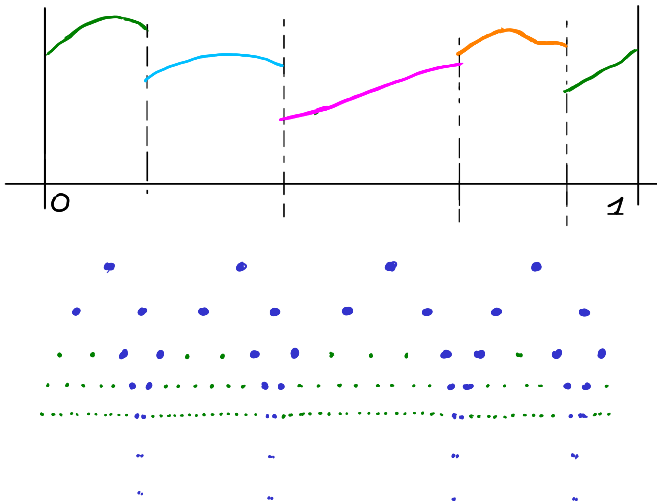
Clearly translation invariant process..

Yet, one can prove that

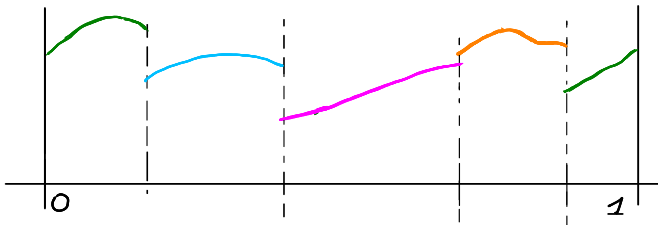
$$\mathbb{E} \left(\int_{\mathbb{T}} \left| f(t) - \sum_{|n| \leq N} \langle f, e_n \rangle e_n(t) \right|^2 dt \right) \geq C \frac{1}{N}$$

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

Data analysis: Images



Data analysis: Images



One easily proves that with this strategy,

$$\mathbb{E} \left(\int_{\mathbb{T}} |f(t) - A_{2N+1}(f)(t)|^2 dt \right) \leq C N^{-2}$$

where $A_{2N+1}(f)$ is an approximation to f that uses only $2N+1$ coeffs.

But there is an enormous difference.

In 1 case

$$\mathbb{E} \left(\left\| f - \sum_{\ell=1}^L \langle f, \varphi_{\ell} \rangle \varphi_{\ell} \right\|^2 \right)$$

is minimal

In the other case,

$$\mathbb{E} \left(\left\| f - \sum_{\ell \in \Lambda_L(f)} \langle f, \varphi_{\ell} \rangle \varphi_{\ell} \right\|^2 \right)$$

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:

$$\varphi_1, \varphi_2, \dots, \varphi_n, \dots \Rightarrow \text{Span}(\varphi_1, \dots, \varphi_n) = V_n$$

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

Nonlinear approximation:

$$\varphi_1, \varphi_2, \dots, \varphi_n, \dots$$

$$\Sigma_n = \left\{ \sum_{\ell \in \mathbb{N}} c_\ell \varphi_\ell : \#\{\ell; c_\ell \neq 0\} \leq n \right\}$$

now study $\text{dist}(f, \Sigma_n)$.

Data analysis: Images

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.

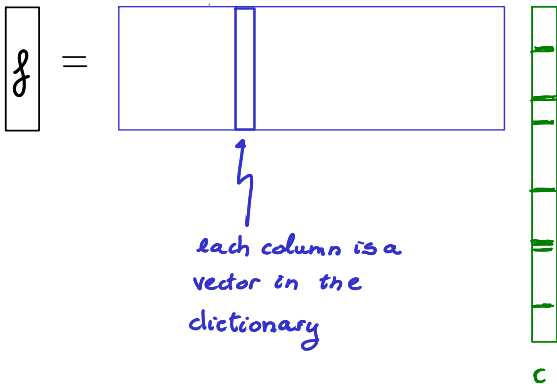
Can we search for a good basis for sparse
expansions?

Find good basis for sparse expansions?

- Search within "dictionaries"

↓
union of many bases.

- Nonlinear (adaptive) singular value decompositions



Same kind of situation as in
compressed sensing !

Compressed Sensing.

Back to images, for a moment.

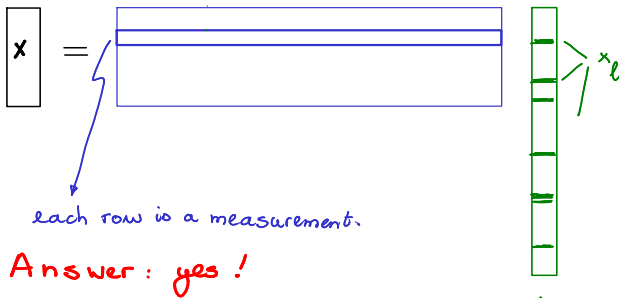
Images are sparse when expressed as a combination of wavelets.

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. $\#\{l; x_l \neq 0\} \leq K \ll N$, can we then determine x by making fewer than N measurements?



Compressive sensing is related to results in theoretical computer science

use Johnson-Lindenstrauss Lemma.

x_1, \dots, x_L vectors in V .
↑
high-dim. space
 $\dim V = D$

Consider projections of x_i on randomly picked d -dim. subspace of V .

Compare $\langle P x_i, P x_j \rangle \frac{D}{d}$ with $\langle x_i, x_j \rangle$
How large should d be for these 2 matrices to be close with high probability?

Basically: $\log L$

This result in CS has had a tremendous impact

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

Fast computations: example.

$f \in \mathbb{C}^N$ N huge

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

\Rightarrow To get a good approximation to f ,
one needs to take only (non-adaptively!)

$O(K \log N)$ 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
↔ compressed sensing.

Find the dictionary if given a class of objects?



Johnson - Lindenstrauss
Compressed sensing } dimension reduction.

One last salvo about computation made feasible
by "dimension reduction"

↳ comparing surfaces with applications
to biology.

$$d_P(\mathcal{L}, \mathcal{L}')$$

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

$$\mathbb{D}_P(S, S')$$

$$= \inf_{C: S \rightarrow S'} \left[\min_{\substack{R \in \text{Euclidean gp.} \\ \text{area-preserving}}} \int_S \|C(x) - R_x\|^2 dA_S \right]^{1/2}$$

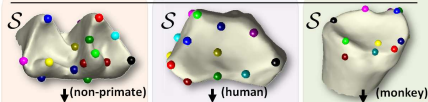
If $\mathcal{D}_{\mathbb{P}}(S, S')$ is small,

then \exists conformal map $m: S \rightarrow S'$
so that

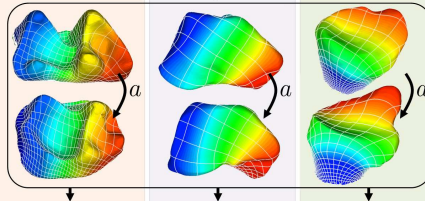
$$\min_R \int_S \|m(x) - Rx\|^2 dA_S \leq C \mathcal{D}_{\mathbb{P}}(S, S')^{1/2}.$$

→ use this to compute approx. to $\mathcal{D}_{\mathbb{P}}(S, S')$
by searching "deformations of conformal maps"

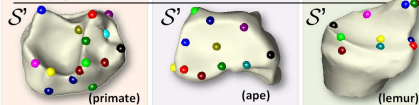
A. Observer Placed Landmarks



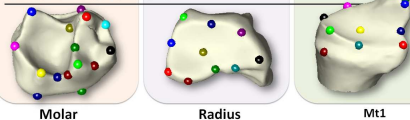
B. cP determined correspondence map between two structures



C. Propagated Landmarks



D. Observer Placed Landmarks



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 ...