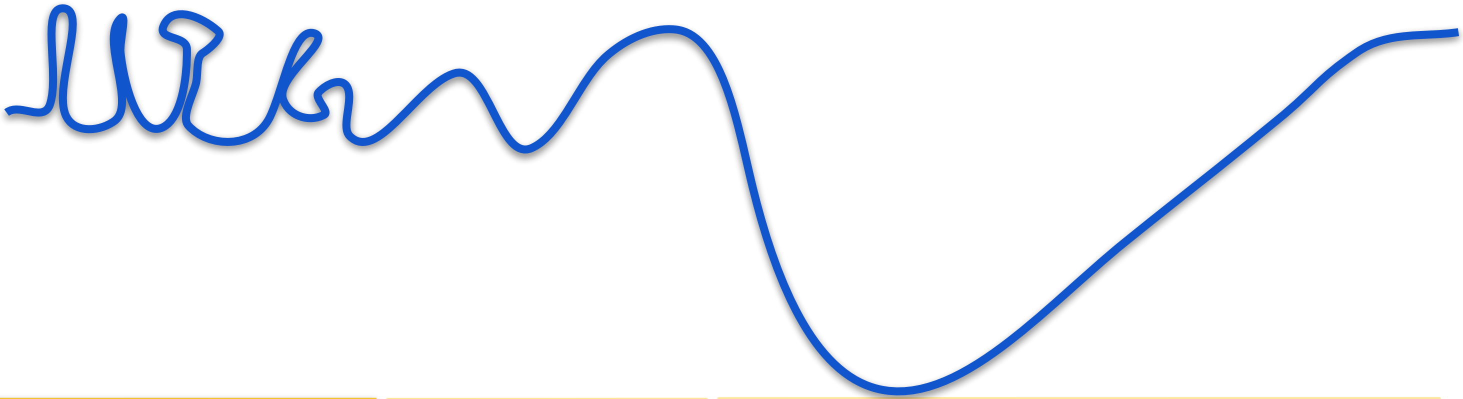
Computing with Signals



DA 623

Jan - May 2023

IIT Guwahati

Instructors: Neeraj Sharma

Lecture-33_34

Basis Representation

Sparsity

Sparse Signal Recovery Algorithms

Basis Representation

Sparsity

Sparse Signal Recovery Algorithms

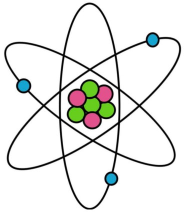
SIAM REVIEW
Vol. 43, No. 1, pp. 129–159

Atomic Decomposition by Basis Pursuit*

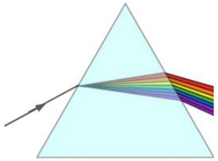
Scott Shaobing Chen†
David L. Donoho‡
Michael A. Saunders§

Lecture Notes: Sparsity and Compressive Sensing, Justin Romberg, Georgia Tech Uni.

Basis Representation



Taking the signal apart.
Writing it as a discrete linear combinations of "atoms".



$$x(t) = \sum_{\gamma \in \Gamma} \alpha(\gamma) \psi_\gamma(t)$$

for some fixed set of *basis* signals $\{\psi_\gamma(t)\}_{\gamma \in \Gamma}$. Here Γ is a discrete index set (for example $\mathbb{Z}$, $\mathbb{N}$, $\mathbb{Z} \times \mathbb{Z}$, $\mathbb{N} \times \mathbb{Z}$ etc.) which will be different depending on the application.

Translate (linearly) the signal into into a discrete list of numbers in such a way that it can be reconstructed (i.e. the translation is lossless). Linear transform = series of inner products, so this mapping looks like:

$$x(t) \longrightarrow \left\{ \begin{array}{c} \langle x(t), \psi_1(t) \rangle \\ \langle x(t), \psi_2(t) \rangle \\ \vdots \\ \langle x(t), \psi_\gamma(t) \rangle \\ \vdots \end{array} \right\}$$

for some fixed set of signals $\{\psi_\gamma(t)\}_{\gamma \in \Gamma}$.

Basis Representation

Fourier series

Let $x(t) \in L_2([0,1])$. Then we can build up $x(t)$ using harmonic complex sinusoids:

$$x(t) = \sum_{k \in \mathbb{Z}} \alpha(k)\, \mathrm{e}^{\mathrm{j}2\pi kt}$$

where

$$\begin{aligned}\alpha(k) &= \int_0^1 x(t)\, \mathrm{e}^{-\mathrm{j}2\pi kt}\, \mathrm{d}t \\ &= \langle x(t), \mathrm{e}^{\mathrm{j}2\pi kt} \rangle.\end{aligned}$$

Basis Representation

Fourier series: properties

1. The $\{\alpha(k)\}$ carry semantic information about which frequencies are in the signal.
2. If $x(t)$ is smooth, the magnitudes $|\alpha(k)|$ fall off quickly as k increases. This energy compaction provides a kind of implicit *compression*.

Basis Representation

Sinc interpolation

$$x[n] = x(nT),$$
$$x(t) = \sum_{n=-\infty}^{\infty} x[n] \frac{\sin(\pi(t - nT))}{\pi(t - nT)/T}.$$

Basis Representation

Sinc interpolation

$$x[n] = x(nT),$$
$$x(t) = \sum_{n=-\infty}^{\infty} x[n] \frac{\sin(\pi(t - nT))}{\pi(t - nT)/T}.$$

$$x(t) = \sum_{n=\infty}^{\infty} \alpha(n)\, \psi_n(t)$$

$$\psi_n(t) = \sqrt{T}\, \frac{\sin(\pi(t - nT))}{\pi(t - nT)}$$
$$\alpha(n) = \sqrt{T}\, x(nT).$$

Basis Representation

Ortho-basis expansion

If $\{\psi_\gamma\}_{\gamma\in\Gamma}$ is an orthobasis for H, then every $x(t) \in H$ can be written as

$$x(t) = \sum_{\gamma\in\Gamma} \langle x(t), \psi_\gamma(t)\rangle \, \psi_\gamma(t).$$

$$\langle \psi_\gamma, \psi_{\gamma'}\rangle = \begin{cases} 1 & \gamma = \gamma' \\ 0 & \gamma \neq \gamma' \end{cases}.$$

Basis Representation

Ortho-basis expansion

Analysis step $$\Psi^*[x(t)] \;=\; \{\langle x(t), \psi_\gamma(t)\rangle\}_{\gamma\in\Gamma} = \{\alpha(\gamma)\}_{\gamma\in\Gamma}.$$

Synthesis step $$\Psi[\{\alpha(\gamma)\}_{\gamma\in\Gamma}] \;=\; \sum_{\gamma\in\Gamma} \alpha(\gamma)\, \psi_\gamma(t).$$

Basis Representation

Parseval's Theorem

Theorem. Let $\{\psi_\gamma\}_{\gamma\in\Gamma}$ be an orthobasis for a space H. Then for any two signals $x, \in H$

$$\langle x, y\rangle_H = \sum_{\gamma\in\Gamma} \alpha(\gamma)\beta(\gamma)^*$$

where

$$\alpha(\gamma) \;=\; \langle x, \psi_\gamma\rangle_H \quad \text{and} \quad \beta(\gamma) \;=\; \langle y, \psi_\gamma\rangle_H.$$

Basis Representation

Parseval's Theorem

- every space of signals for which we can find any ortho-basis can be discretized
- mapping from (continuous) signal space into (discrete) coefficient space preserves inner products
  - it preserves all of the geometrical relationships between the signals (i.e. distances and angles).

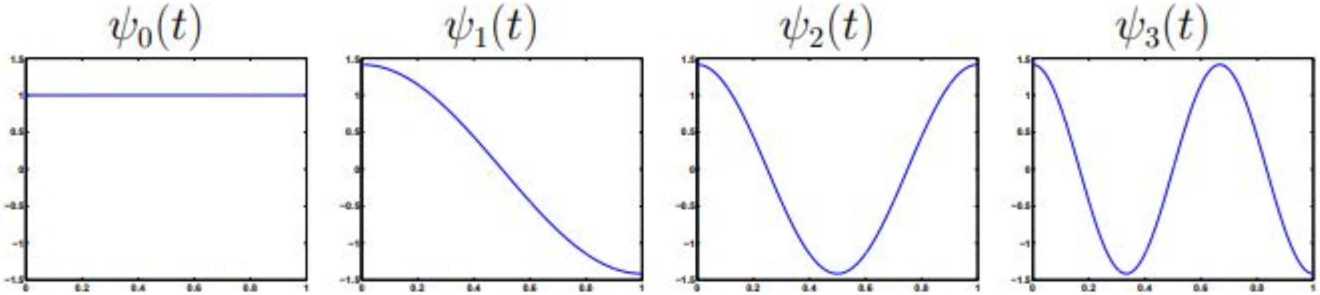
- in some sense, this means that all signal processing can be done by manipulating discrete sequences of numbers.

Basis Representation

Cosine Transform (CT)

The cosine-I basis functions for $t \in [0, 1]$ are

$$\psi_k(t) = \begin{cases} 1 & k = 0 \\ \sqrt{2}\cos(\pi k t) & k > 0 \end{cases}.$$



Basis Representation

Discrete Cosine Transform (CT)

Definition: The DCT basis functions for $\mathbb{R}^N$ are

$$\psi_k[n] = \begin{cases} \sqrt{\frac{1}{N}} & k = 0 \\ \sqrt{\frac{2}{N}} \cos\left(\frac{\pi k}{N}\left(n + \frac{1}{2}\right)\right) & k = 1, \ldots, N-1 \end{cases}, \quad n = 0, 1, \ldots, N-1.$$

Basis Representation

PCA

Formally, let $\psi_1(t), \ldots, \psi_N(t)$ be a finite set of orthogonal vectors in H, and set

$$\mathcal{V} = \text{span}\{\psi_1, \ldots, \psi_N\}.$$

Given a fixed signal $x_0(t) \in H$, the solution $\tilde{x}_0(t)$ to

$$\min_{x \in \mathcal{V}} \|x_0(t) - x(t)\|_2^2 \tag{1}$$

is given by

$$\tilde{x}_0(t) = \sum_{k=1}^{N} \langle x_0(t), \psi_k(t) \rangle \psi_k(t).$$

Basis Representation

Non-orthogonal basis

$$\begin{bmatrix} \alpha_0 \\ \alpha_1 \\ \vdots \\ \alpha_{N-1} \end{bmatrix} = \begin{bmatrix} \langle x, \psi_0 \rangle \\ \langle x, \psi_1 \rangle \\ \vdots \\ \langle x, \psi_{N-1} \rangle \end{bmatrix}.$$

Stacking up the (transposed) ψ_k as rows in an $N \times N$ matrix Ψ^*,

$$\Psi^* = \begin{bmatrix} \text{---} & \psi_0^* & \text{---} \\ \text{---} & \psi_1^* & \text{---} \\ \vdots & \vdots & \vdots \\ \text{---} & \psi_{N-1}^* & \text{---} \end{bmatrix},$$

we have the straightforward relationships

$$\alpha = \Psi^* x, \quad \text{and} \quad x = \Psi^{*-1} \alpha.$$

Basis Representation

Non-orthogonal basis

$$\begin{bmatrix} \alpha_0 \\ \alpha_1 \\ \vdots \\ \alpha_{N-1} \end{bmatrix} = \begin{bmatrix} \langle x, \psi_0 \rangle \\ \langle x, \psi_1 \rangle \\ \vdots \\ \langle x, \psi_{N-1} \rangle \end{bmatrix}.$$

Stacking up the (transposed) ψ_k as rows in an $N \times N$ matrix Ψ^*,

$$\Psi^* = \begin{bmatrix} \text{---} & \psi_0^* & \text{---} \\ \text{---} & \psi_1^* & \text{---} \\ \vdots & \vdots & \vdots \\ \text{---} & \psi_{N-1}^* & \text{---} \end{bmatrix},$$

we have the straightforward relationships

$$\alpha = \Psi^* x, \quad \text{and} \quad x = \Psi^{*-1}\alpha.$$

$$\boxed{x[n] = \sum_{k=0}^{N-1} \langle x, \psi_k \rangle \tilde{\psi}_k[n].}$$

$$\boxed{\Psi^{*-1} = \begin{bmatrix} | & | & \cdots & | \\ \tilde{\psi}_0 & \tilde{\psi}_1 & \cdots & \tilde{\psi}_{N-1} \\ | & | & \cdots & | \end{bmatrix}}$$

$$\boxed{\sigma_1^2 \|x\|_2^2 \ \le \ \|\alpha\|_2^2 \ \le \ \sigma_N^2 \|x\|_2^2,}$$

Basis Representation

Over-complete frames: Fat matrix

$$\begin{bmatrix} \alpha_0 \\ \alpha_1 \\ \vdots \\ \alpha_{M-1} \end{bmatrix} = \begin{bmatrix} \langle x, \psi_0 \rangle \\ \langle x, \psi_1 \rangle \\ \vdots \\ \langle x, \psi_{M-1} \rangle \end{bmatrix}$$

$$\Psi^* = \begin{bmatrix} \text{---} & \psi_0^* & \text{---} \\ \text{---} & \psi_1^* & \text{---} \\ \vdots & \vdots & \vdots \\ \text{---} & \psi_{M-1}^* & \text{---} \end{bmatrix}$$

$$x = (\Psi\Psi^*)^{-1}\Psi\Psi^* x.$$

$$\tilde{\psi}_k = (\Psi\Psi^*)^{-1}\psi_k.$$

$$x[n] = \sum_{k=0}^{M-1} \langle x, \psi_k \rangle \tilde{\psi}_k[n].$$

Basis Representation

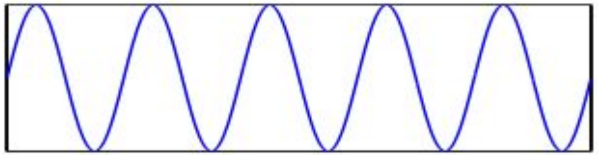
- Signal/image $f(t)$ in the time/spatial domain
- Decompose f as a *superposition of atoms*

$$f(t) = \sum_i \alpha_i \psi_i(t)$$

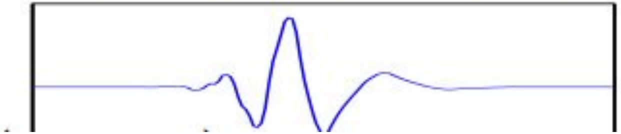
$$\psi_i = \text{basis functions}$$

$$\alpha_i = \text{expansion coefficients in } \psi\text{-domain}$$

- Classical example: Fourier series
  ψ_i = complex sinusoids
  α_i = Fourier coefficients



- Modern example: wavelets
  ψ_i = "little waves"
  α_i = wavelet coefficients



Basis Representation

Two sequences of functions: $\{\psi_i(t)\}, \{\tilde{\psi}(t)\}$

Analysis (inner products):

$$\alpha = \tilde{\Psi}^*[f], \qquad \alpha_i = \langle \tilde{\psi}_i, f \rangle$$

Synthesis (superposition):

$$f = \Psi[\alpha], \qquad f = \sum_i \alpha_i \psi_i(t)$$

- If $\{\psi_i(t)\}$ is an orthobasis, then

$$\|\alpha\|_{\ell_2}^2 = \|f\|_{L_2}^2 \qquad \text{(Parseval)}$$

$$\sum_i \alpha_i \beta_i = \int f(t) g(t)\, dt \qquad \text{(where } \beta = \tilde{\Psi}[g]\text{)}$$

Basis
Representation

Sparsity

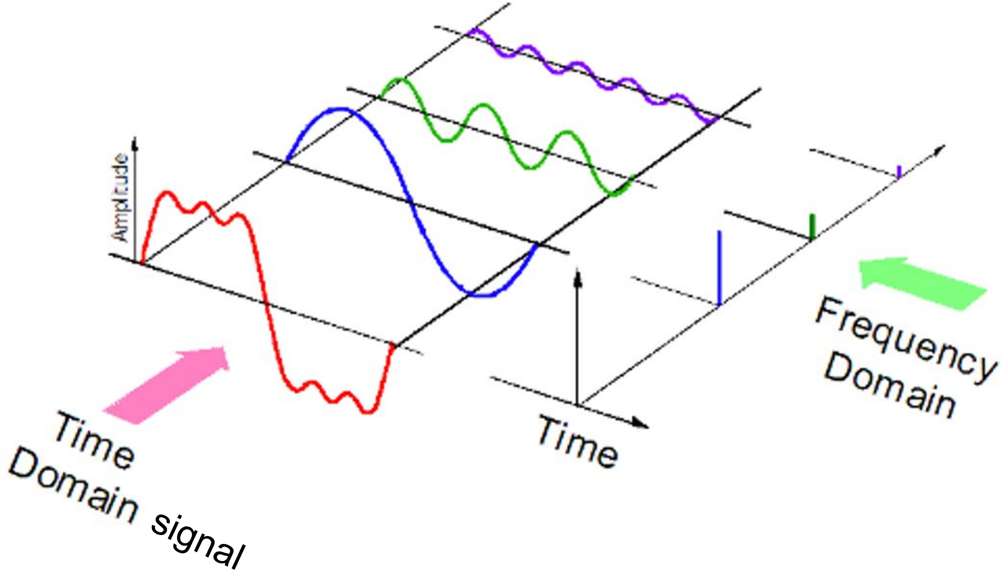
Sparse Signal
Recovery
Algorithms

SIAM REVIEW
Vol. 43, No. 1, pp. 129–159

Atomic Decomposition by Basis Pursuit*

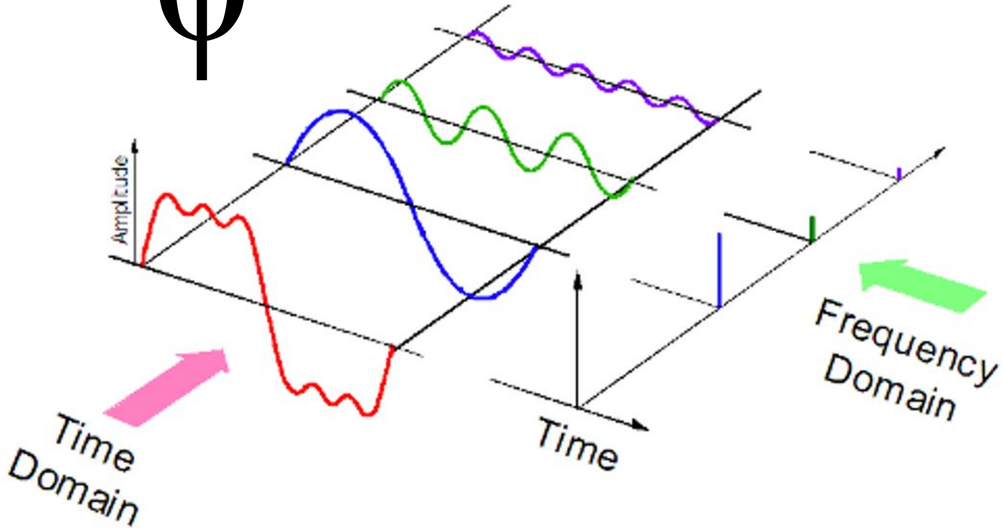
Scott Shaobing Chen†
David L. Donoho‡
Michael A. Saunders§

Lecture Notes: Sparsity and Compressive Sensing,
Justin Romberg, Georgia Tech Uni.



Amplitude
Time
Frequency Domain
Time Domain signal

ψ



α

x

Two sequences of functions: $\{\psi_i(t)\}, \{\tilde{\psi}(t)\}$

Analysis (inner products):

$$\alpha = \tilde{\Psi}^*[f], \qquad \alpha_i = \langle \tilde{\psi}_i, f \rangle$$

Synthesis (superposition):

$$f = \Psi[\alpha], \qquad f = \sum_i \alpha_i \psi_i(t)$$

Two sequences of functions: $\{\psi_i(t)\}, \{\tilde{\psi}(t)\}$

Analysis (inner products):

$$\alpha = \tilde{\Psi}^*[f], \qquad \alpha_i = \langle \tilde{\psi}_i, f \rangle$$

Synthesis (superposition):

$$f = \Psi[\alpha], \qquad f = \sum_i \alpha_i \psi_i(t)$$

signal/data   Synthesis dictionary   Coefficients/ Representation

$$x = \Psi \alpha$$

- Classical: signal/image is "bandlimited" or "low-pass"
- Modern: smooth between isolated singularities (e.g. 1D piecewise poly)
- Postmodern: 2D image is smooth between smooth edge contours

signal/data — Synthesis dictionary — Coefficients/ Representation

$$x = \Psi \alpha$$

- Classical: signal/image is "bandlimited" or "low-pass"
- Modern: smooth between isolated singularities (e.g. 1D piecewise poly)
- Postmodern: 2D image is smooth between smooth edge contours

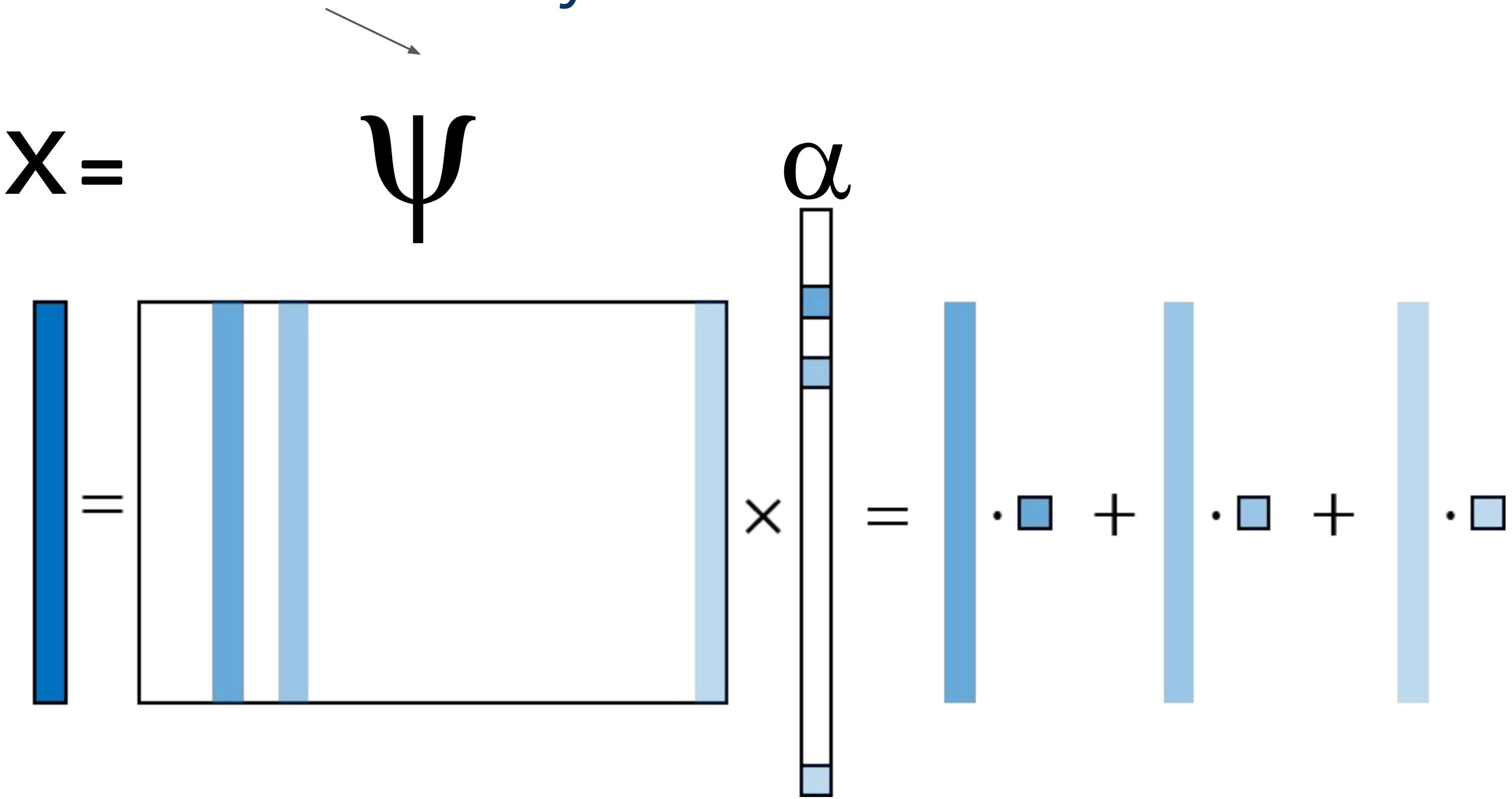
---

Ψ

- Ortho-basis (NxN)
- Basis (NxN)
- Overcomplete (NxM, M>>N)

Given x, choice of \Psi determines the behavior of \alpha.

Notion of Dictionary



An overcomplete dictionary (more columns than rows) can help in obtaining a representation \alpha which is sparse.

signal/data

Synthesis dictionary

Coefficients/ Representation

$$\mathbf{x} = \Psi \alpha$$

An pursued goal is Construct "good representation"

- sparsifies signals/images of interest
- can be computed using fast algorithms ($O(N)$ or $O(N \log N)$ — think of the FFT)

Linear approximation

- Linear S-term approximation: keep S coefficients in fixed locations

$$f_S(t) = \sum_{m=1}^{S} \alpha_m \psi_m(t)$$

  - projection onto fixed subspace
  - lowpass filtering, principle components, etc.

- Fast coefficient decay $\Rightarrow$ good approximation

$$|\alpha_m| \lesssim m^{-r} \quad \Rightarrow \quad \|f - f_S\|_2^2 \lesssim S^{-2r+1}$$

- Take $f(t)$ periodic, d-times continuously differentiable, $\Psi =$ Fourier series:

$$\|f - f_S\|_2^2 \ \lesssim \ S^{-2d}$$

The smoother the function, the better the approximation

Something similar is true for wavelets ...

Image approximation using DCT

Take 1% of "low pass" coefficients, set the rest to zero

original

approximated

rel. error = 0.075

Adapted from: Lecture Notes on Sparsity and Compressive Sensing,
Justin Romberg, Georgia Tech Uni.

Non-linear Approximation

$$\min_{\beta \in \mathbb{R}^n} \|f - \Psi\beta\|_2^2 \quad \text{subject to} \quad \#\{\gamma : \beta[\gamma] \neq 0\} \leq S.$$

1. Compute $\alpha = \Psi^* f$.

2. Find the locations of the S-largest terms in α; call this set Γ.

3. Set

$$\tilde{\beta}_S[\gamma] = \begin{cases} \alpha[\gamma] & \gamma \in \Gamma \\ 0 & \gamma \notin \Gamma \end{cases}$$

4. Compute $\tilde{f}_S = \Psi \tilde{\beta}_S$.

Image approximation using DCT

Take 1% of "low pass" coefficients, set the rest to zero

original

approximated

rel. error = 0.075

Image approximation using DCT

Take 1% of *largest* coefficients, set the rest to zero (adaptive)

original

approximated

rel. error = 0.057

Image approximation using DCT

Take 1% of *largest* coefficients, set the rest to zero (adaptive)

original

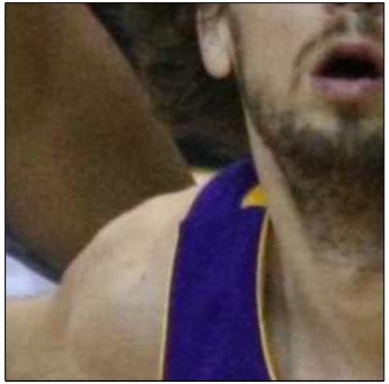
approximated

rel. error $= 0.057$

Image approximation using DCT

Take 1% of "low pass" coefficients, set the rest to zero

original



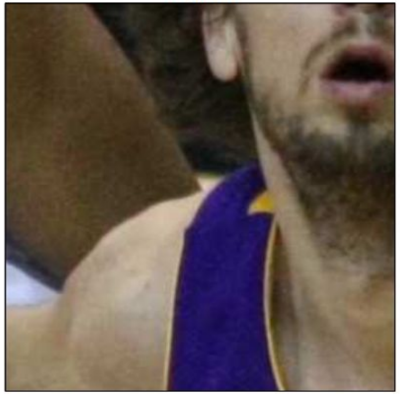
approximated

rel. error = 0.075

Image approximation using DCT

Take 1% of *largest* coefficients, set the rest to zero (adaptive)

original



approximated

rel. error = 0.057

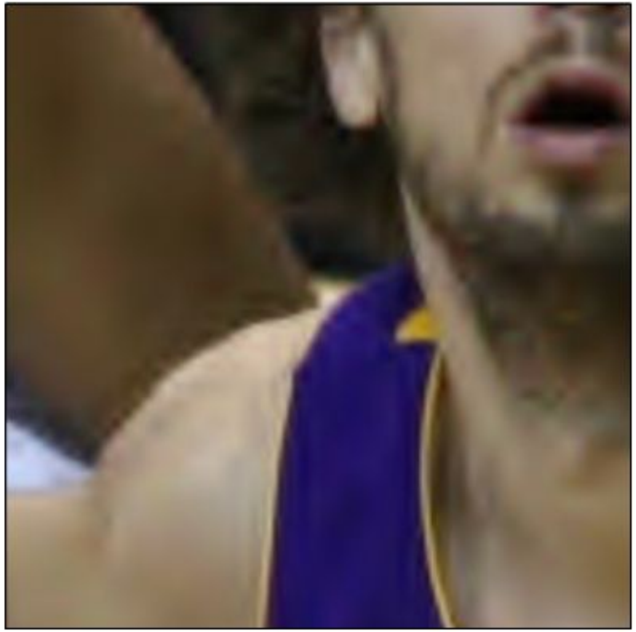
DCT/wavelets comparison

Take 1% of *largest* coefficients, set the rest to zero (adaptive)

DCT

rel. error = 0.057

wavelets



rel. error = 0.031

Nonlinear approximation

- Nonlinear S-term approximation: keep S *largest* coefficients

$$f_S(t) = \sum_{\gamma \in \Gamma_S} \alpha_\gamma \psi_\gamma(t), \qquad \Gamma_S = \text{locations of } S \text{ largest } |\alpha_m|$$

- Fast decay of sorted coefficients $\Rightarrow$ good approximation

$$|\alpha|_{(m)} \lesssim m^{-r} \quad \Rightarrow \quad \|f - f_S\|_2^2 \lesssim S^{-2r+1}$$

$|\alpha|_{(m)} = m$th largest coefficient

Linear v. nonlinear approximation

- For $f(t)$ *uniformly smooth* with d "derivatives"

	S-term approx. error
Fourier, linear	S^{-2d+1}
Fourier, nonlinear	S^{-2d+1}
wavelets, linear	S^{-2d+1}
wavelets, nonlinear	S^{-2d+1}

- For $f(t)$ *piecewise smooth*

	S-term approx. error
Fourier, linear	S^{-1}
Fourier, nonlinear	S^{-1}
wavelets, linear	S^{-1}
wavelets, nonlinear	S^{-2d+1}

Nonlinear wavelet approximations *adapt* to singularities

Sparse representation - a “good representation”

Sparse representations yield algorithms for (among other things)

1. compression,
2. estimation in the presence of noise (“denoising”),
3. inverse problems (e.g. tomography),
4. acquisition (compressed sensing)

that are

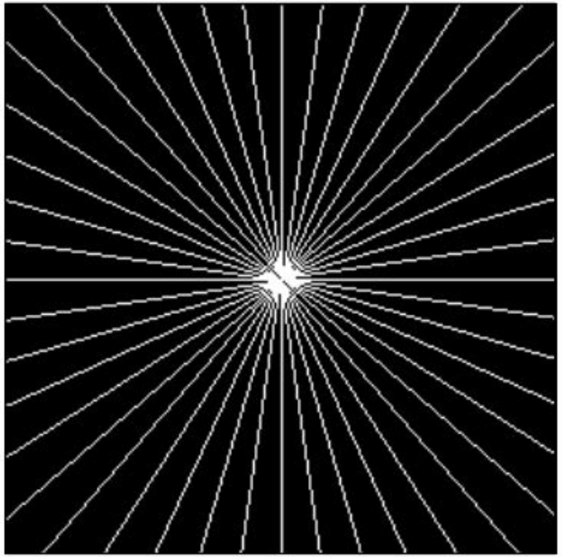
- fast,
- relatively simple,
- and produce (nearly) optimal results

$$\overset{\text{signal/data}}{\downarrow \atop x} = \overset{\text{Synthesis dictionary}}{\downarrow \atop \psi} \; \overset{\text{Coefficients/ Representation}}{\downarrow \atop \alpha}$$

A simple underdetermined inverse problem

Observe a subset Ω of the 2D discrete Fourier plane

phantom (hidden)



white star = sample locations

$N := 512^2 = 262,144$ pixel image

observations on 22 radial lines, $10,486$ samples, $\approx 4\%$ coverage

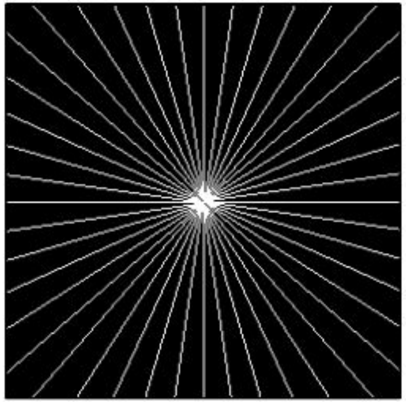
Minimum energy reconstruction

Reconstruct g^* with

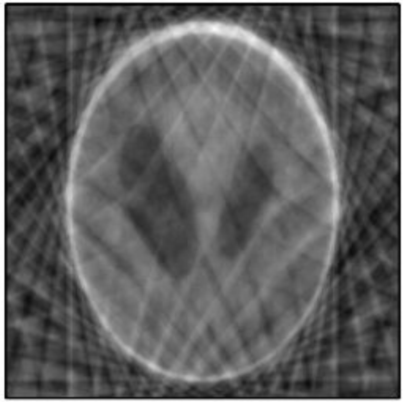
$$\hat{g}^*(\omega_1, \omega_2) = \begin{cases} \hat{f}(\omega_1, \omega_2) & (\omega_1, \omega_2) \in \Omega \\ 0 & (\omega_1, \omega_2) \notin \Omega \end{cases}$$

Set unknown Fourier coeffs to zero, and inverse transform

original



Fourier samples



g^*

Total-variation reconstruction

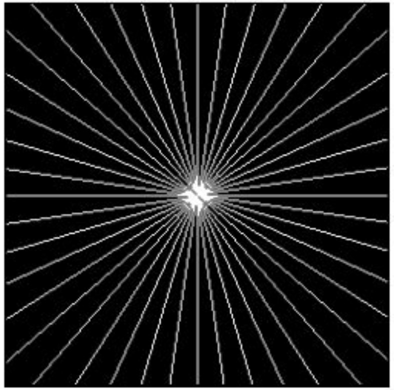
Find an image that

- Fourier domain: *matches observations*
- Spatial domain: has a *minimal amount of oscillation*

Reconstruct g^* by solving:

$$\min_g \sum_{i,j} |(\nabla g)_{i,j}| \quad \text{s.t.} \quad \hat{g}(\omega_1, \omega_2) = \hat{f}(\omega_1, \omega_2), \quad (\omega_1, \omega_2) \in \Omega$$

original



Fourier samples

$g^* =$ original
perfect reconstruction

Total-variation reconstruction

$\|.\|_1$ l_1-norm induces sparsity

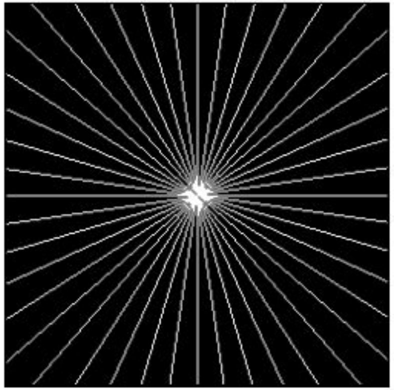
Find an image that

- Fourier domain: *matches observations*
- Spatial domain: has a *minimal amount of oscillation*

Reconstruct g^* by solving:

$$\min_g \boxed{\sum_{i,j} |(\nabla g)_{i,j}|} \quad \text{s.t.} \quad \hat{g}(\omega_1, \omega_2) = \hat{f}(\omega_1, \omega_2), \quad (\omega_1, \omega_2) \in \Omega$$

original



Fourier samples

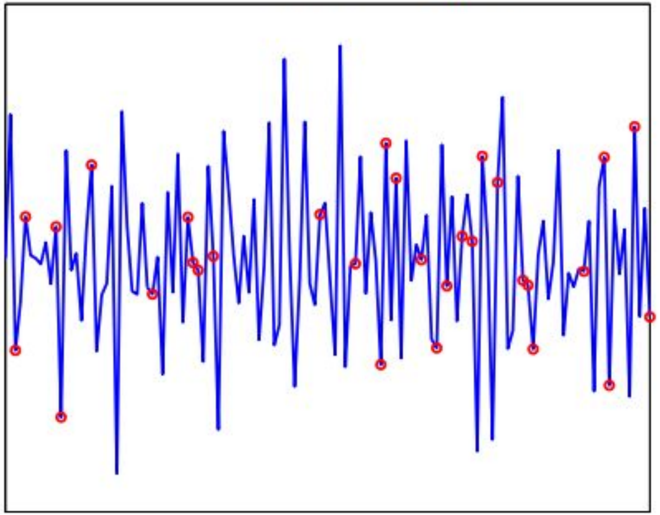
g^* = original

perfect reconstruction

Sampling a superposition of sinusoids

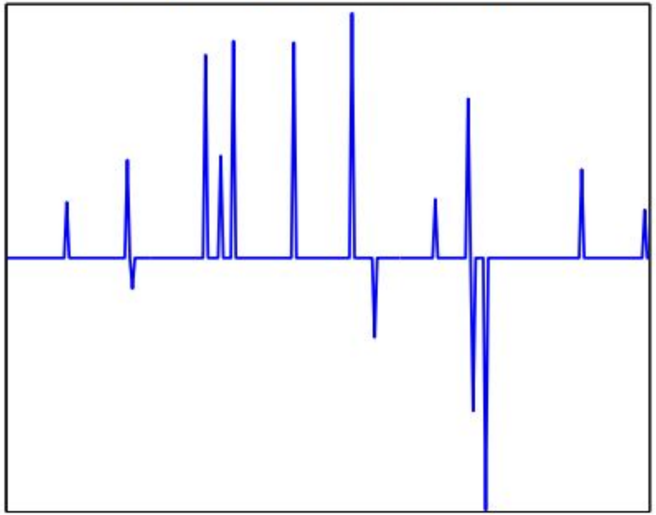
We take M samples of a superposition of S sinusoids:

Time domain $x_0(t)$



Measure M samples
(red circles = samples)

Frequency domain $\hat{x}_0(\omega)$

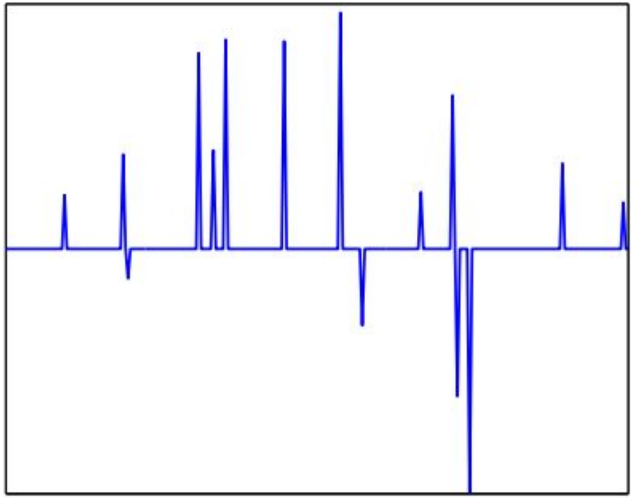


S nonzero components

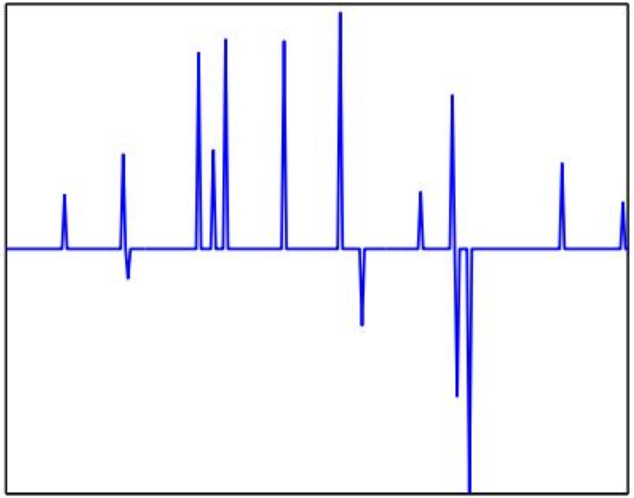
Sampling a superposition of sinusoids

Reconstruct by solving

$$\min_x \boxed{\|\hat{x}\|_{\ell_1}} \quad \text{subject to} \quad x(t_m) = x_0(t_m), \quad m = 1, \ldots, M$$



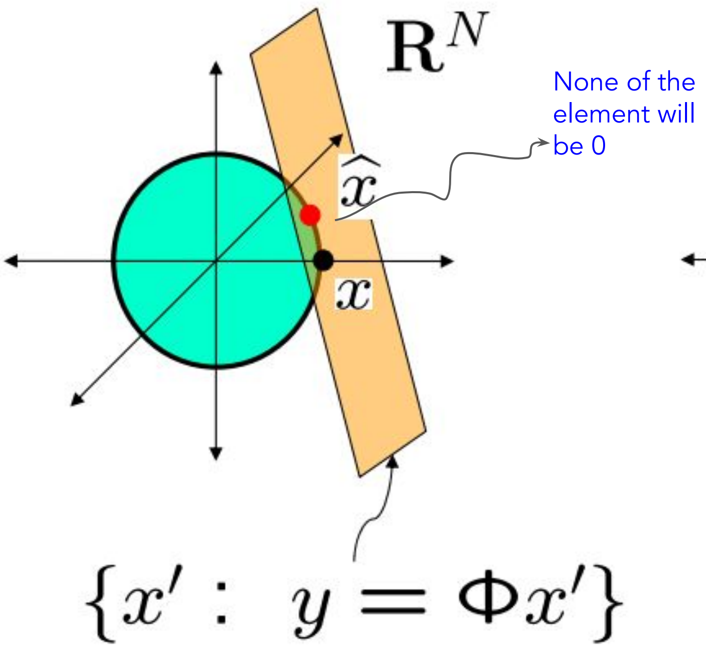
original $\hat{x}_0$, $S = 15$

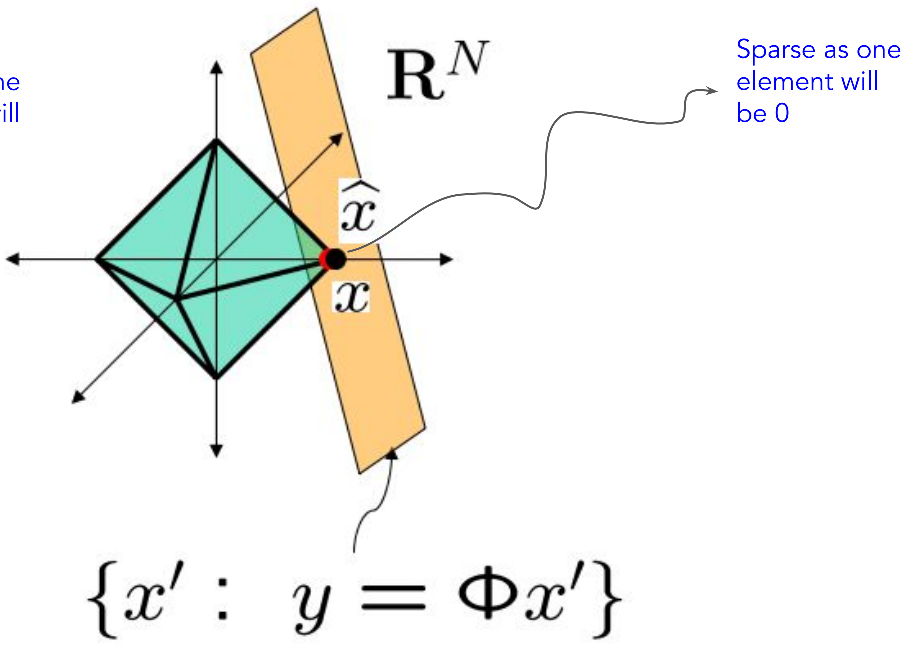


perfect recovery from 30 samples

Graphical intuition for ℓ_1

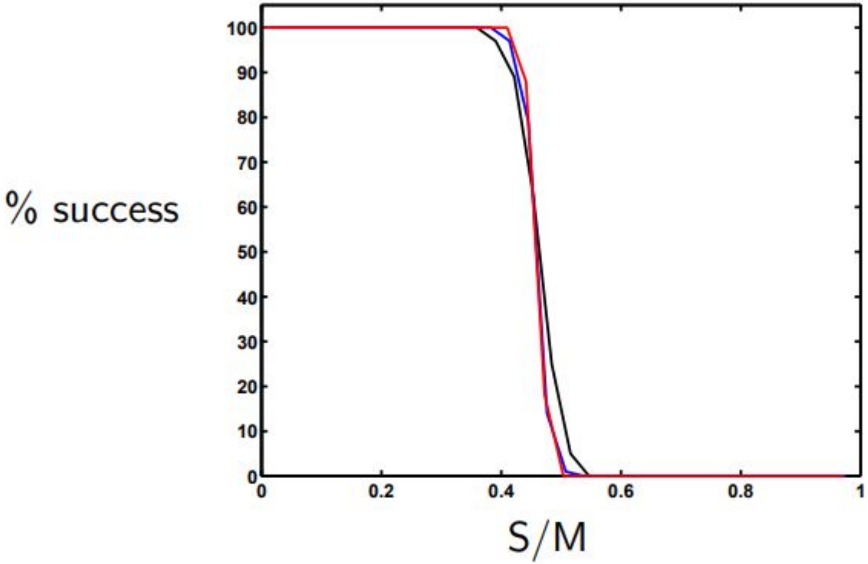
$\min_x \|x\|_2$ s.t. $\Phi x = y$ $\min_x \|x\|_1$ s.t. $\Phi x = y$





Numerical recovery curves

- Resolutions $N = 256, 512, 1024$ (black, blue, red)
- Signal composed of S randomly selected sinusoids
- Sample at M randomly selected locations



- In practice, perfect recovery occurs when $M \approx 2S$ for $N \approx 1000$

A nonlinear sampling theorem

Exact Recovery Theorem (Candès, R, Tao, 2004):

- Unknown $\hat{x}_0$ is supported on set of size S
- Select M sample locations $\{t_m\}$ "at random" with

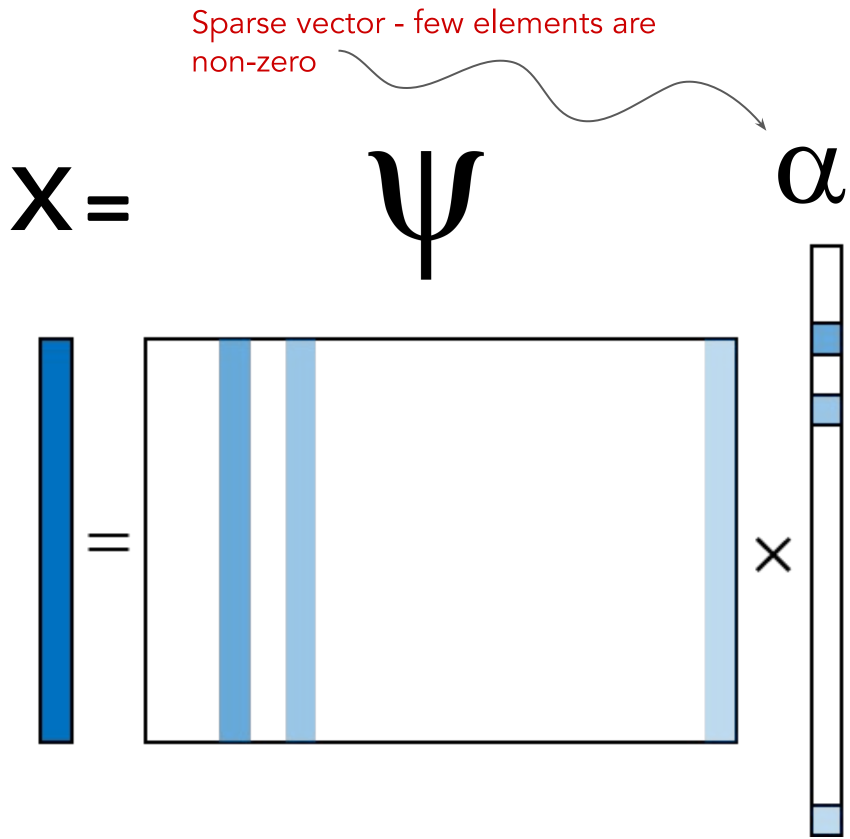
$$M \geq \mathrm{Const} \cdot S \log N$$

- Take time-domain samples (measurements) $y_m = x_0(t_m)$
- Solve

$$\min_x \|\hat{x}\|_{\ell_1} \quad \text{subject to} \quad x(t_m) = y_m, \quad m = 1, \ldots, M$$

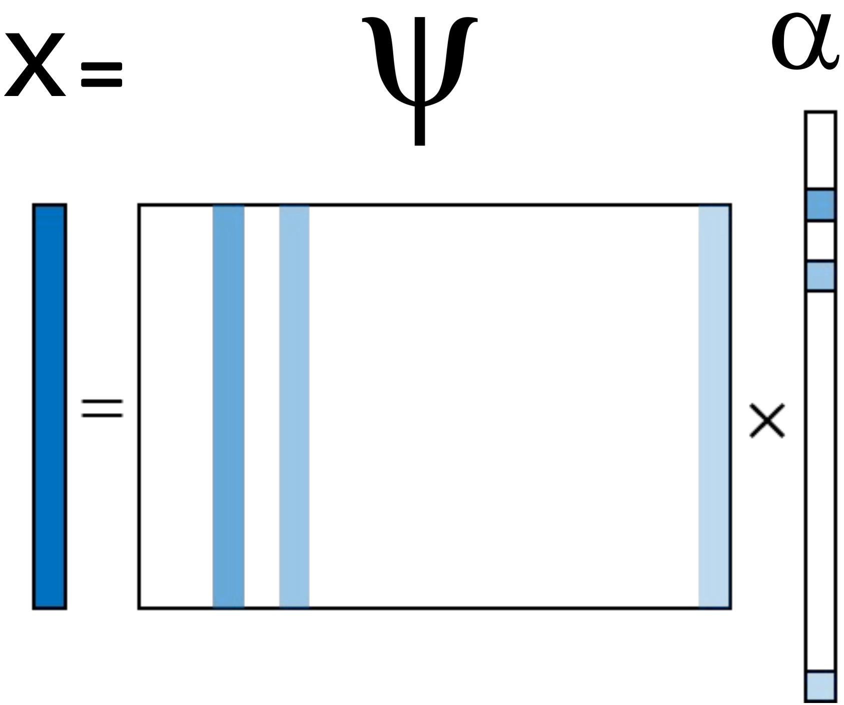
- Solution is *exactly* f with extremely high probability
- In total-variation/phantom example, S=number of jumps

Sparse representations are representations that account for most or all information of a signal with a linear combination of a small number of atoms.



Sparse representations are representations that account for most or all information of a signal with a linear combination of a small number of atoms.

Sparse vector - few elements are non-zero



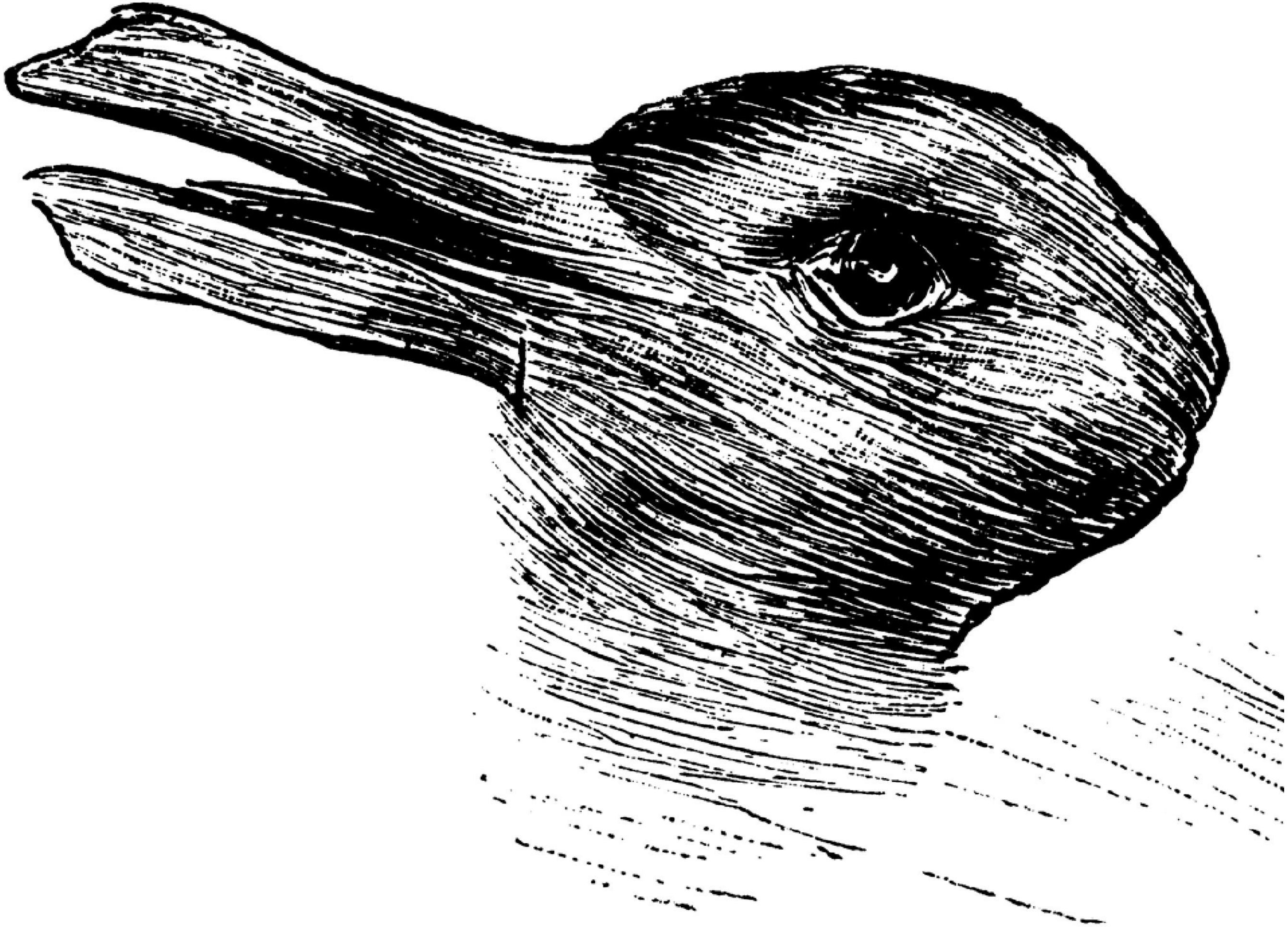
Given x and \Psi with more columns than rows, solving for a sparse \alpha is non-trivial and a challenging problem.

Greedy algorithms Matching pursuit (MP) and the closely related Orthogonal Matching Pursuit (OMP) operate by iterative choosing columns of the matrix. At each iteration, the column that reduces the approximation error the most is chosen.

Convex programming Relaxes the combinatorial problem into a closely related convex program, and minimizes a global cost function. The particular program, based on ℓ_1 minimization, we will look at has been given the name Basis Pursuit in the literature.

Course summary

- Signals
  - Types of signal - time, space, applications
- Signal Models
  - Polynomials
  - Sines and cosines
- Representations
  - Fourier series
  - Fourier transform
  - Convolution
  - Filtering
  - Linear Systems: Impulse response and head related transfer function
- DFT
  - Computation
  - Neural network
- Time-frequency representation
  - spectrum varies with time
  - Instantaneous frequency
  - STFT and spectrogram
- Clustering
  - k-means
  - Distance measure: DP and DWT
- Dimensionality Reduction
  - Linear spaces
  - PCA
  - LDA
- Sparse representations
  - Introduction
  - Basis and representations
  - L2, L1 and L0 norm
- and other things we discussed in class



Thank you!