Inverse problems and sparse models (1/6)

Rémi Gribonval
INRIA Rennes - Bretagne Atlantique, France
remi.gribonval@inria.fr

http://people.irisa.fr/Remi.Gribonvall

PDF of the slides

- http://www.irisa.fr/metiss/gribonval/Teaching/

Overview of the course

- **Introduction**
  - ✓ sparsity & data compression
  - ✓ inverse problems in signal and image processing
    - ✦ image deblurring, image inpainting,
    - ✦ channel equalization, signal separation,
    - ✦ tomography, MRI
  - ✓ sparsity & under-determined inverse problems
    - ✦ well-posedness

- **Complexity & Feasibility**
  - ✓ NP-completeness of ideal sparse approximation
  - ✓ Relaxations
  - ✓ L1 is *sparsity-inducing* and *convex*

Overview of the course

- **Pursuit Algorithms**
  - ✓ L1 has performance guarantees
  - ✓ L1 is computationally feasible: Basis Pursuit
  - ✓ Greedy algorithms: Matching Pursuit & al
  - ✓ Complexity of Pursuit Algorithms

- **Recovery guarantees**
  - ✓ Coherence *vs* Restricted Isometry Constant
  - ✓ Worked examples
  - ✓ Summary

Further material on sparsity

- **Books**
  - ✓ Signal Processing perspective
    - ✦ S. Mallat, «Wavelet Tour of Signal Processing», 3rd edition, 2008
    - ✦ M. Elad, «Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing», 2009.
  - ✓ Mathematical perspective
    - ✦ S. Foucart, H. Rauhut, «A Mathematical Introduction to Compressed Sensing», Springer, in preparation.

- **Review paper:**
 - ✦ Bruckstein, Donoho, Elad, SIAM Reviews, 2009

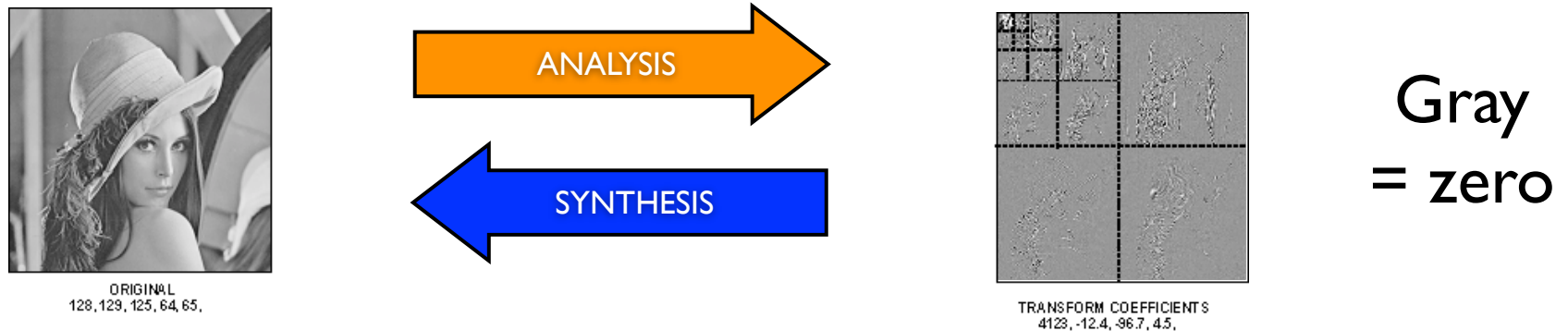# Sparse models & data compression

Large-scale data

- **Fact** : digital data = large volumes

  - ✓ 1 second stereo audio, CD quality   = 1,4 Mbit
  - ✓ 1 uncompressed 10 Mpixels picture = 240 Mbit

- **Need** : «concise» data representations

  - ✓ storage & transmission (volume / bandwidth) ...
  - ✓ manipulation & processing (algorithmic complexity)

Notion of sparse representation
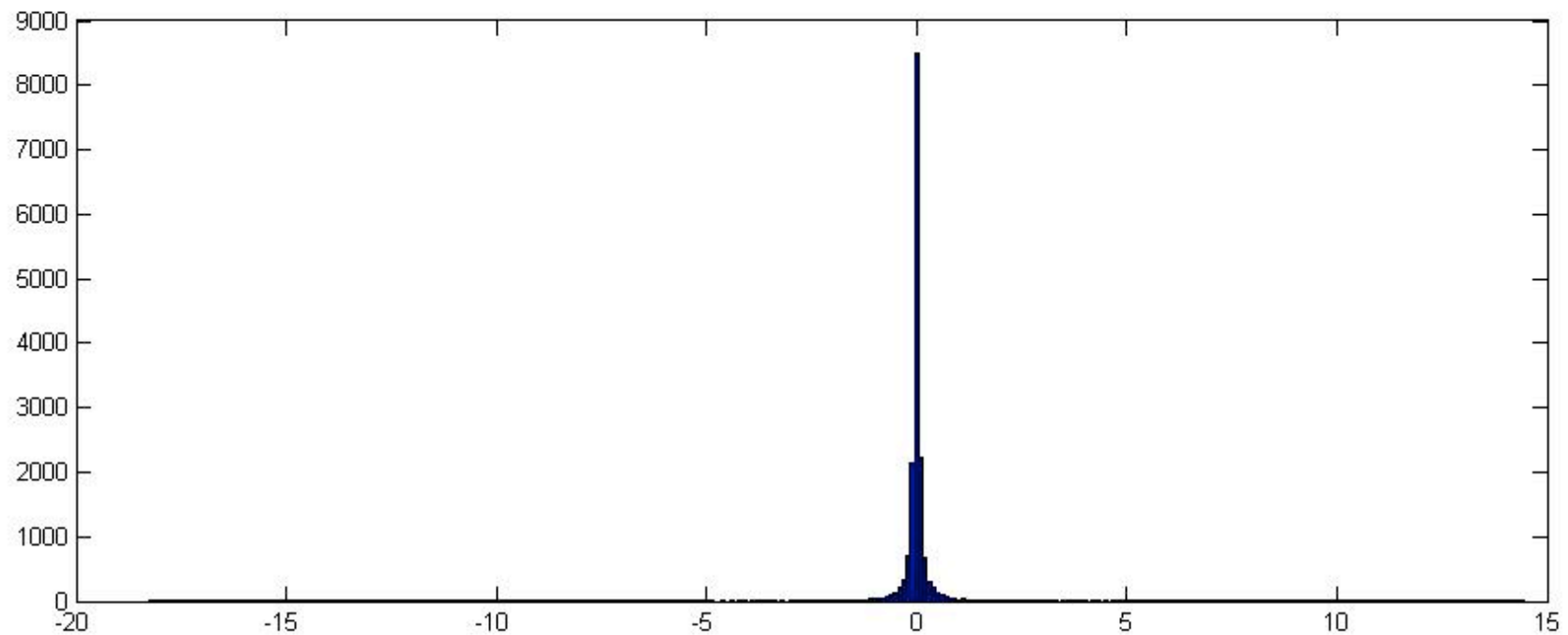
- Audio : time-frequency representations (MP3)

Black = zero
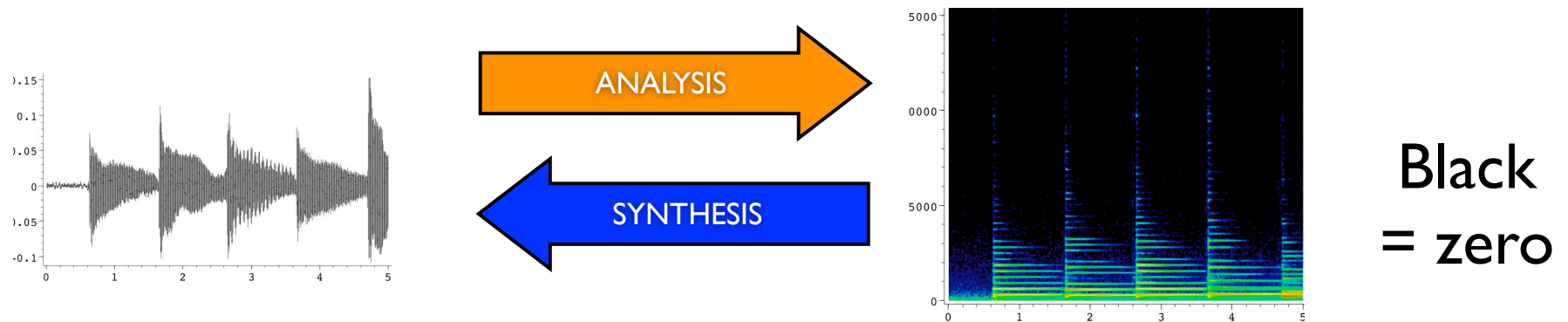
- Images : wavelet transform (JPEG2000)

ORIGINAL
128,129,125,64,65,

TRANSFORM COEFFICIENTS
4123, -12.4, -96.7, 4.5,

Gray = zero

Evidence of sparsity

- Histogram of MDCT coefficients of a musical sound

Mathematical expression of the sparsity assumption

- Signal / image = high dimensional vector

$$y \in \mathbb{R}^N$$

- Definition:
  - ✓ **Atoms**: basis vectors $\varphi_k \in \mathbb{R}^N$
    - ✦ ex: time-frequency atoms, wavelets
  - ✓ **Dictionary:**
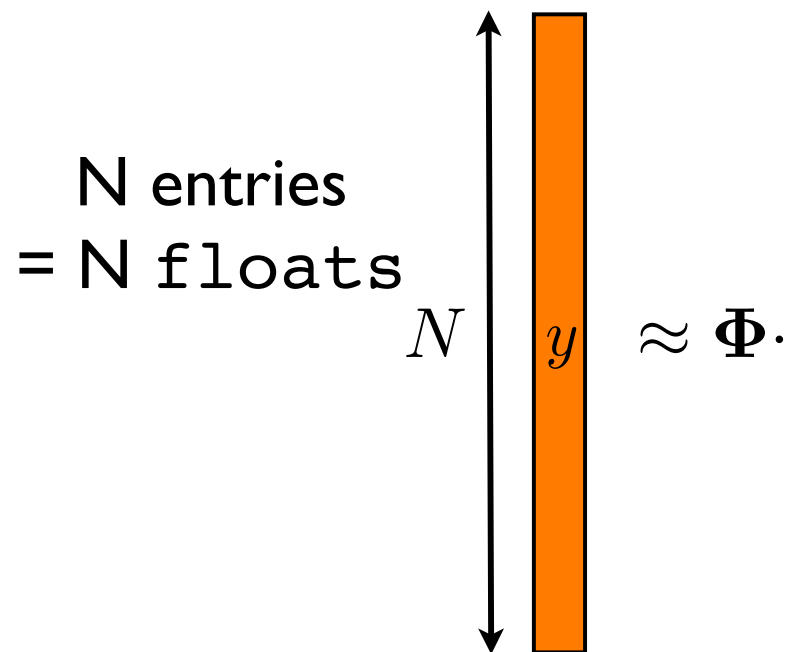   - ✦ collection of atoms $\{\varphi_k\}_{1 \leq k \leq K}$
    - ✦ matrix $\mathbf{\Phi} = [\varphi_k]_{1 \leq k \leq K}$ which columns are the atoms
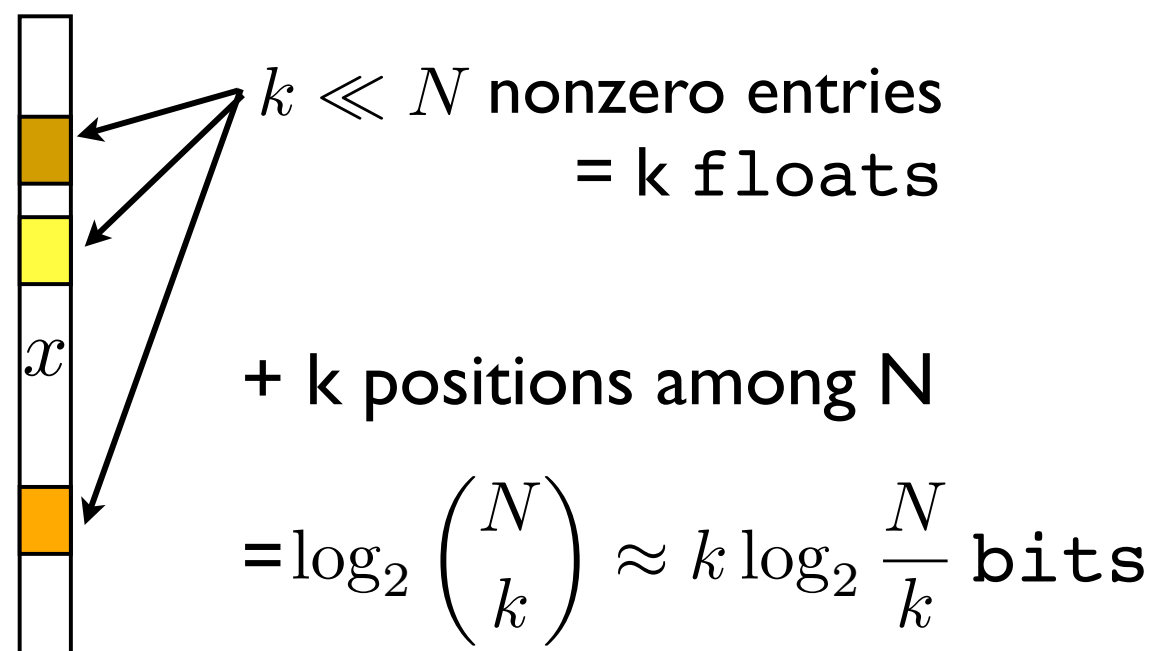
- Sparse **signal model** = combination of *few* atoms

$$y \approx \sum_k x_k \varphi_k = \mathbf{\Phi} x$$

Sparsity & compression

- Full vector

- Sparse vector

$k \ll N$ **nonzero entries** $= \texttt{k floats}$

N entries $= \texttt{N floats}$

$$N \begin{vmatrix} y \end{vmatrix} \approx \mathbf{\Phi} \cdot \begin{vmatrix} x \end{vmatrix}$$

+ k positions among N

$$= \log_2 \binom{N}{k} \approx k \log_2 \frac{N}{k} \ \texttt{bits}$$

Key practical issues: choose dictionary
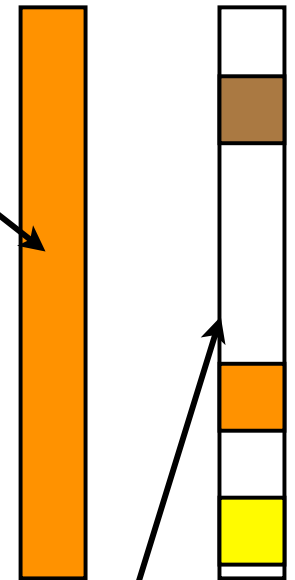
Sparsity: definition

Not sparse

- A vector is
  - ✓ **sparse** if it has (many) zero coefficients
  - ✓ *k*-**sparse** if it has *at most k* nonzero coefficients

- Symbolic representation as column vector

- **Support** = indices of nonzero components

- Sparsity measured with **L0 pseudo-norm**

$$\|x\|_0 := \sharp\{n, \ x_n \neq 0\} = \sum_n |x_n|^0$$

3-sparse

Convention here
$$a^0 = 1(a > 0); 0^0 = 0$$

- *In french:*
  - ✦ sparse   -> *«creux», «parcimonieux»*
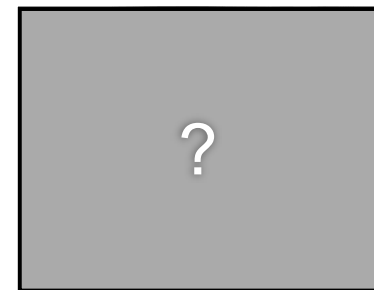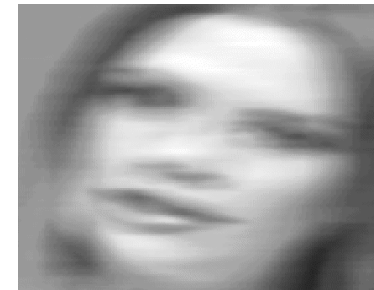 - ✦ sparsity, sparseness   -> *«parcimonie», ~~«sparsité»~~*

Inverse problems in
signal and image processing

Deconvolution problem
2D Example : deblurring problem

- **Given data:**
  - ✓ blurred image $\quad y[i,j]$

  - ✓ information on blurring process
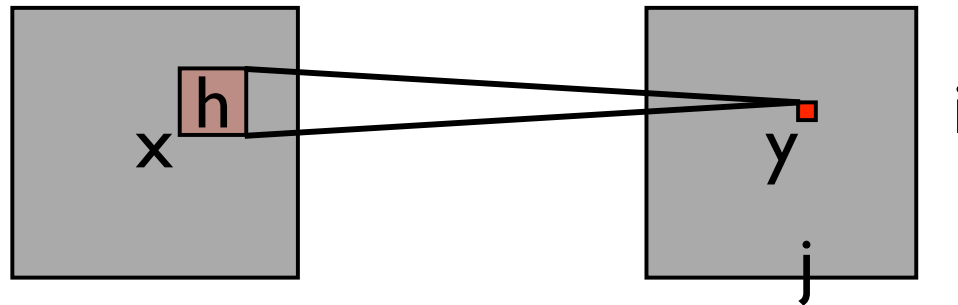
- **Desired estimate:**
  - ✓ deblurred image $\quad x[i,j]$

Blurring process = 2D Convolution

- Definition

$$y[i,j] = (h \star x)[i,j] :== \sum_{k,\ell} h[k,\ell]x[i-k, j-\ell]$$
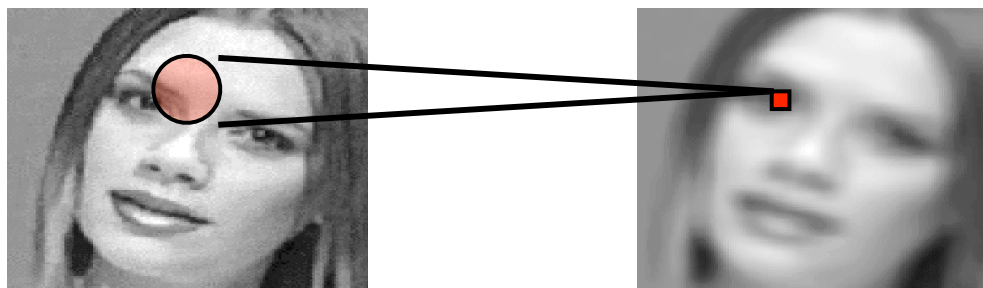
- Interpretation : local average

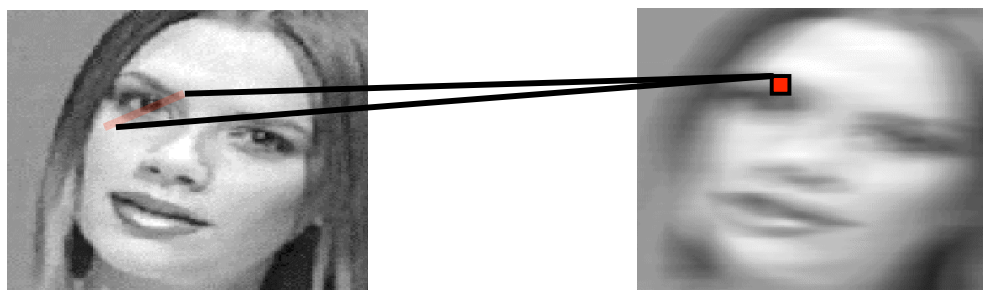Reproduced from http://www.robots.ox.ac.uk/~improofs/super-resolution/super-res1.html

Examples of 2D convolution

h =point spread function (PSF)

- ## Optical blur

- ## Motion blur

Reproduced from http://www.robots.ox.ac.uk/~improofs/super-resolution/super-res1.html

Example of point spread function

$$h[i, j]$$
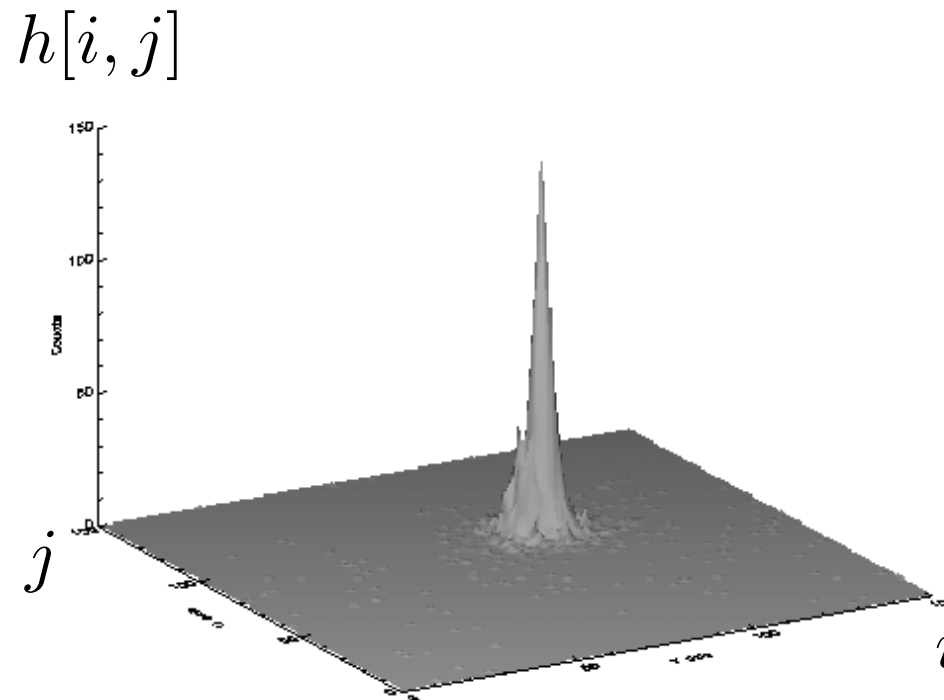
Reproduced from http://hea-www.harvard.edu/HRC/calib/hrci_qe.html

1D deconvolution problems

- **General form** $\quad y(t) = (h \star x)(t) := \int h(\tau) x(t - \tau) d\tau$

- **Telecom: channel equalization**
  - ✓ h = channel impulse response

- **Audio: de-reverberation (reflections on walls)**
  - ✓ h = room impulse response
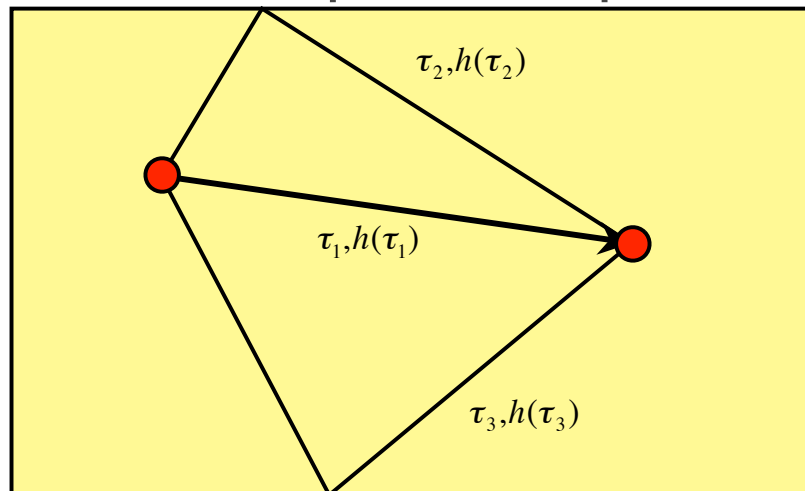
Example of room impulse response

Reproduced from http://www.am3d.com/technology/acoustical
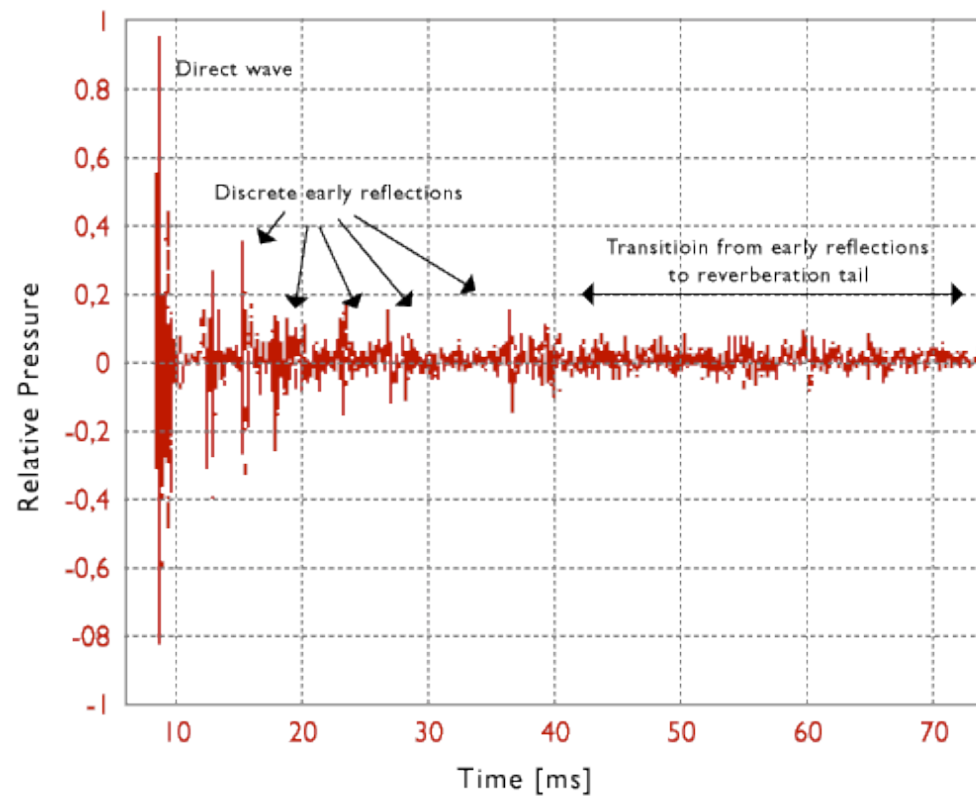
Deconvolution Problem

- **Given**
  - ✓ measured data *y*
  - ✓ known filter *h*

- **Find unknown *x* such that**

$$y = h \star x$$

Naive deconvolution in the Fourier domain

- Convolution and Fourier / inverse Fourier

$$y(t) = (h \star x)(t) \quad \xrightarrow{F\{\cdot\}} \quad Y(f) = H(f)X(f)$$

$$\hat{x}(t) = x(t) \quad \xleftarrow{F^{-1}\{\cdot\}} \quad \hat{X}(f) = \frac{Y(f)}{H(f)} = X(f)$$

- H(f) = transfer function of filter h

Issues with naive deconvolution

- **Presence of noise**

$$y(t) = (h \star x)(t) + n(t) \qquad \longleftrightarrow \qquad Y(f) = H(f)X(f) + N(f)$$

- **Smooth filter:**
  - ✓ fast decay of H(f)
  - ✓ small values in H(f)

$$\hat{X}(f) := \frac{Y(f)}{H(f)} = X(f) + \frac{N(f)}{H(f)}$$

  - ✓ division by small values = strong amplification of noise

- **Consequence = missing frequency information**
  - ✓ N frequency components to estimate $\qquad X \in \mathbb{R}^N$

  - ✓ *m* < N reliable frequency components $\qquad Y \in \mathbb{R}^m$

X

Y
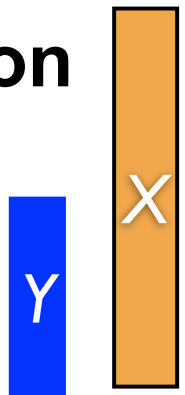
Inverse problems

Linear inverse problems: definition

- **Definition**: a problem where a high-dimensional vector must be estimated from its low dimensional projection

- **Generic form:**

$$\mathbf{b} = \mathbf{A}\mathbf{y} + \mathbf{e}$$

observation/measure   unknown   noise

projection matrix

- ✓ m observations / measures $\mathbf{b} \in \mathbb{R}^m$ $\mathbf{A} \in \mathbb{R}^{m \times N}$
- ✓ N unknowns $\mathbf{y} \in \mathbb{R}^N$

Example: Inpainting Problem

- **Unknown image with N pixels**
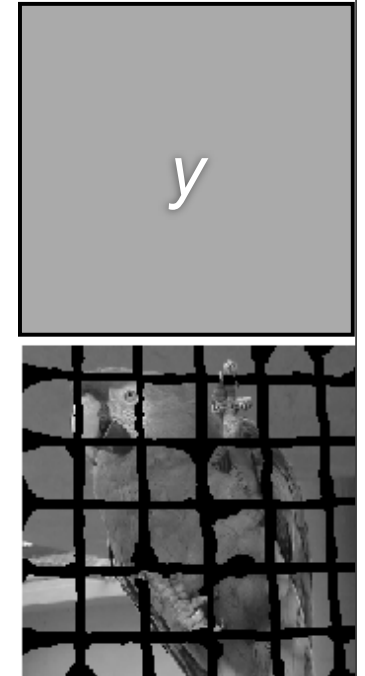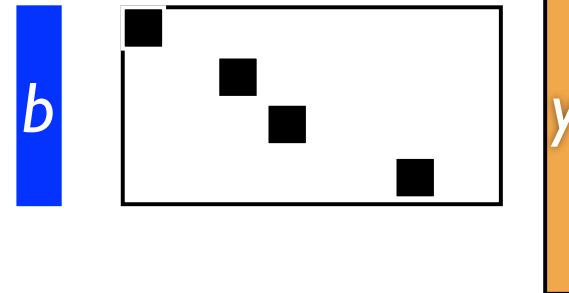
$$\mathbf{y} \in \mathbb{R}^N$$

- **Partially observed image:**
  - ✓ *m* < N observed pixels

$$b[\vec{p}] = y[\vec{p}], \ \vec{p} \in \mathrm{Observed}$$

- **Measurement matrix**

$$\mathbf{b} = \mathbf{My}$$

Classes of linear inverse problems

● **Determined:** the matrix **A** is square and invertible
- ✓ Unique solution to $\mathbf{b} = \mathbf{A}\mathbf{y}$
- ✓ Linear function of observations

$$\mathbf{y} = \mathbf{A}^{-1}\mathbf{b}$$

$$\boxed{\mathbf{A}}$$
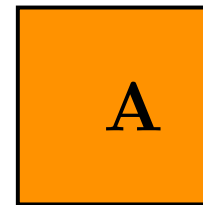
● **Over-determined:** more equations than unknowns
- ✓ Unique solution to $\mathbf{b} = \mathbf{A}\mathbf{y}$:
- ✓ Linear function of observations
- ✓ with pseudo-inverse $\quad \mathbf{y} = \mathbf{A}^{\dagger}\mathbf{b}$

$$\boxed{\mathbf{A}}$$

● **Under-determined:** fewer equations than unknowns
- ✓ **Infinitely many solutions to** $\mathbf{b} = \mathbf{A}\mathbf{y}$
- ✓ **Need to choose one?**

$$\boxed{\mathbf{A}}$$

Inverse problems

Signal space ~ R^N

Set of signals of interest

x

Nonlinear
Approximation =

Sparse recovery

Linear
projection $\mathbf{A}$

$\mathbf{b}$

Observation space ~ R^m
m<<N
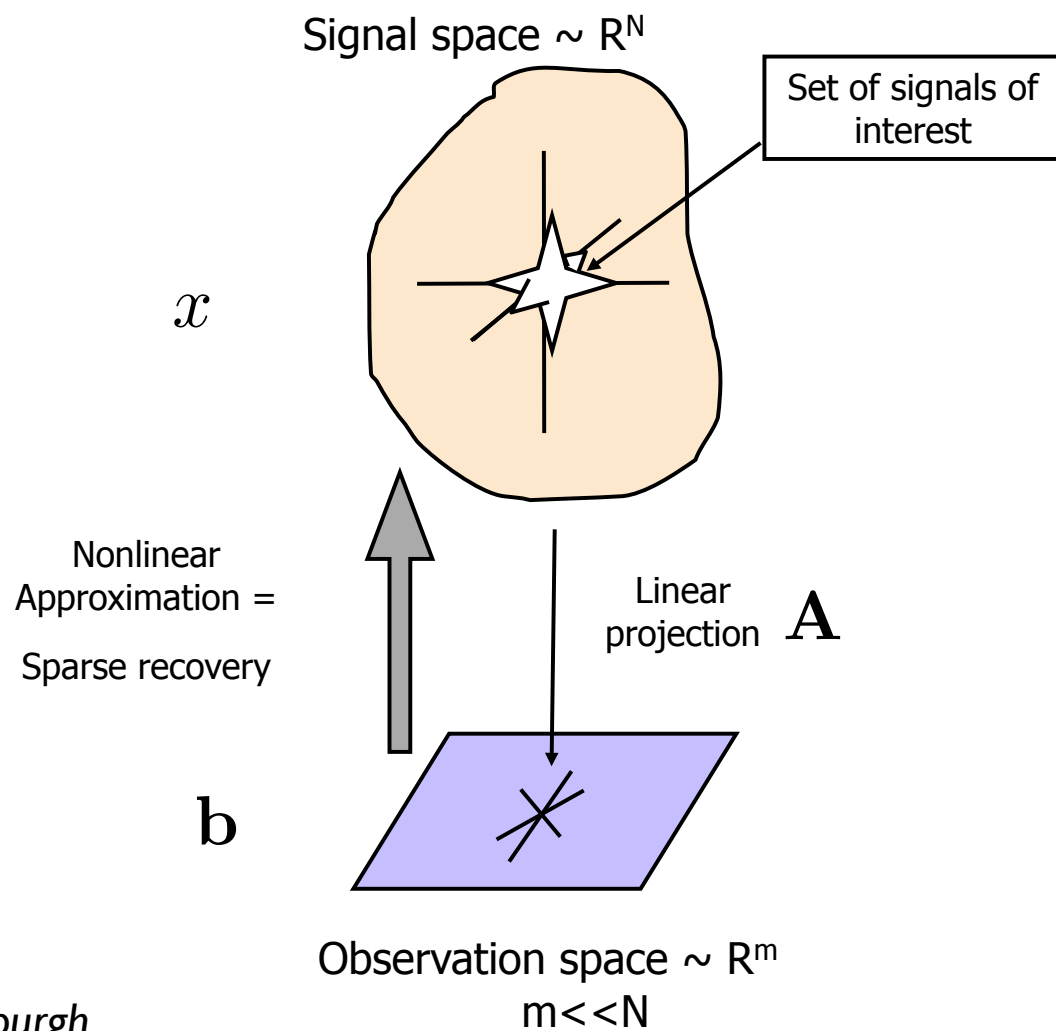
Courtesy: M. Davies, U. Edinburgh
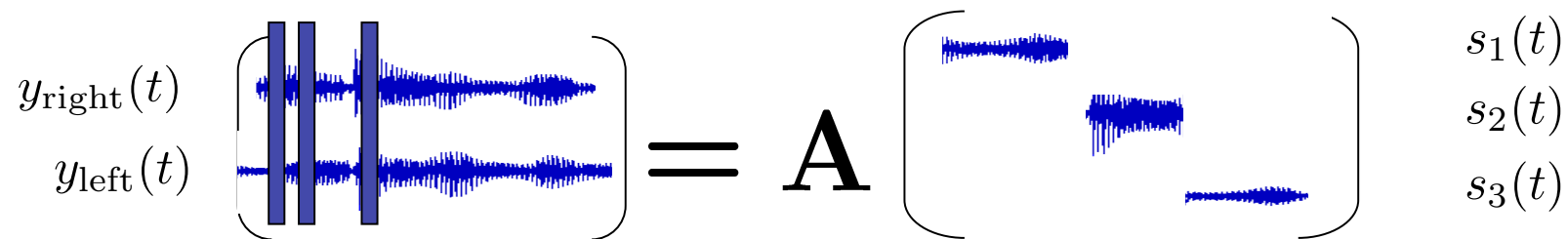
Example : audio source separation

- « Softly as in a morning sunrise »
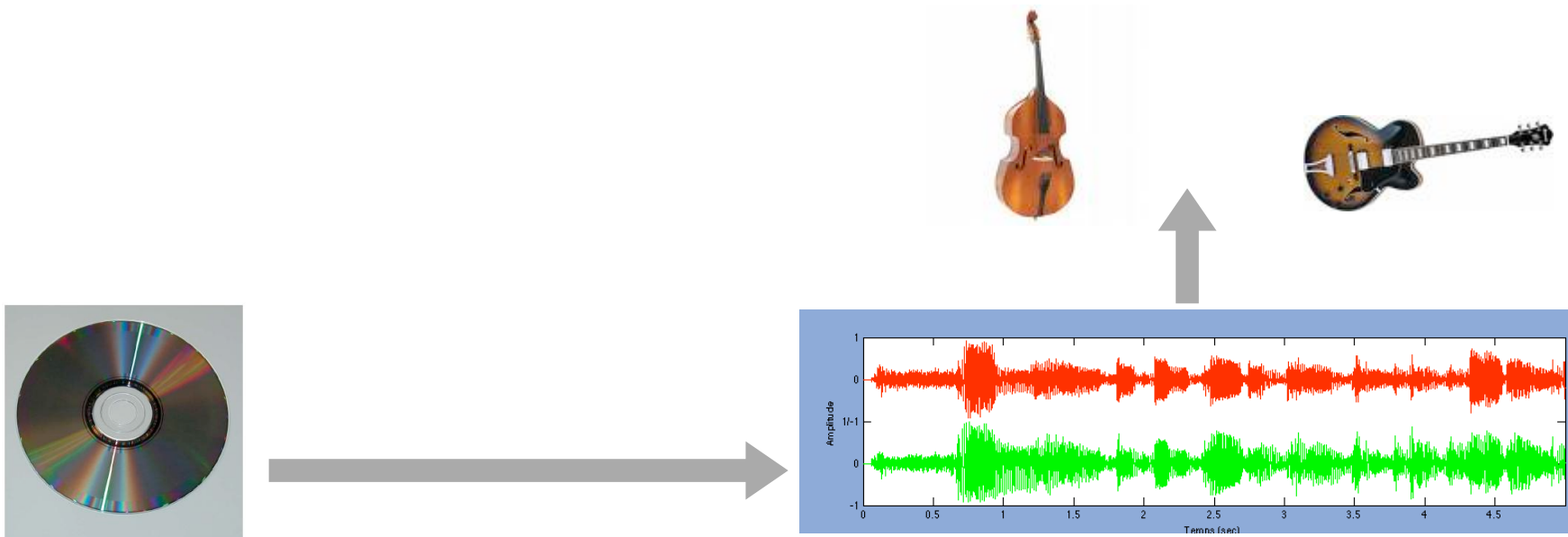
Blind Source Separation

- Mixing model : linear instantaneous mixture

$$\begin{matrix} y_{\text{right}}(t) \\ y_{\text{left}}(t) \end{matrix} \begin{pmatrix} & & \\ & & \end{pmatrix} = \mathbf{A} \begin{pmatrix} & & \\ & & \end{pmatrix} \begin{matrix} s_1(t) \\ s_2(t) \\ s_3(t) \end{matrix}$$

- Source model : if disjoint time-supports ...

$y_{\text{left}}(t)$

$y_{\text{right}}(t)$

... then clustering to :
1- identify (columns of) the mixing matrix
2- recover sources

Blind Source Separation

- Mixing model : linear instantaneous mixture
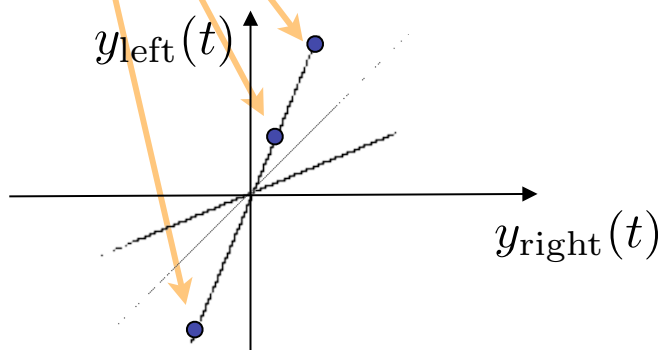
$$\begin{matrix} y_{\text{right}}(t) \\ y_{\text{left}}(t) \end{matrix} \left[\text{~~~~~~} \right] = \mathbf{A} \left[\text{~~~~~~} \right] \begin{matrix} s_1(t) \\ s_2(t) \\ s_3(t) \end{matrix}$$

- In practice ...
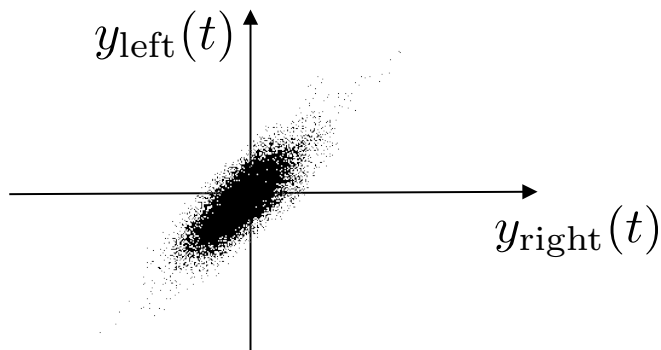
Time-Frequency Masking

- Mixing model in the time-frequency domain

$$\begin{matrix} Y_{\text{right}}(\tau, f) \\ Y_{\text{left}}(\tau, f) \end{matrix} \left[\begin{matrix} \\ \end{matrix} \right] = \mathbf{A} \, \mathbf{S}(\tau, f)$$

- And "miraculously" ...

$Y_{\text{left}}(\tau, f)$

$Y_{\text{right}}(\tau, f)$
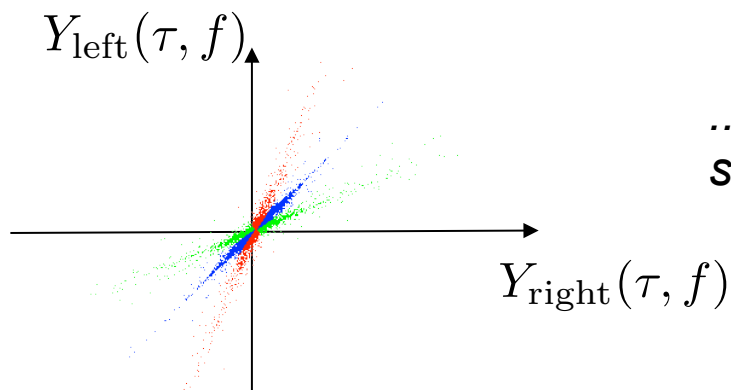
*... time-frequency representations of audio signals are (often) **almost disjoint**.*

Inverse problems

- **Inverse problem** : exploit indirect or incomplete obervation to recontruct some data

$$z = \mathbf{M}\mathbf{y}$$

↑ fewer equations than unknowns

- **Sparsity** : represent / approximate high-dimensional & complex data using few parameters

$$\mathbf{y} \approx \mathbf{\Phi}x$$

↑ few nonzero components

Observations → **Data**
Reconstruct

Data → **Representation**
Reduce the dimension
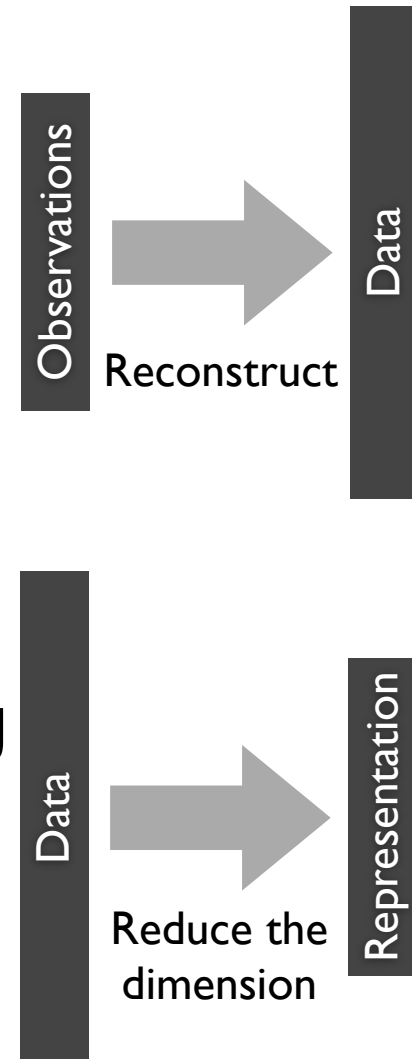
Signal Processing Vocabulary

Forward
linear model
$$\mathbf{b} \approx \mathbf{A}x$$

Unknown x
representation, sources, ...

Known linear system: $\mathbf{A}$
dictionary, mixing matrix, sensing system...

Observed data: $\mathbf{b}$
signal, image, mixture of sources,...

Decomposition
Reconstruction
Separation
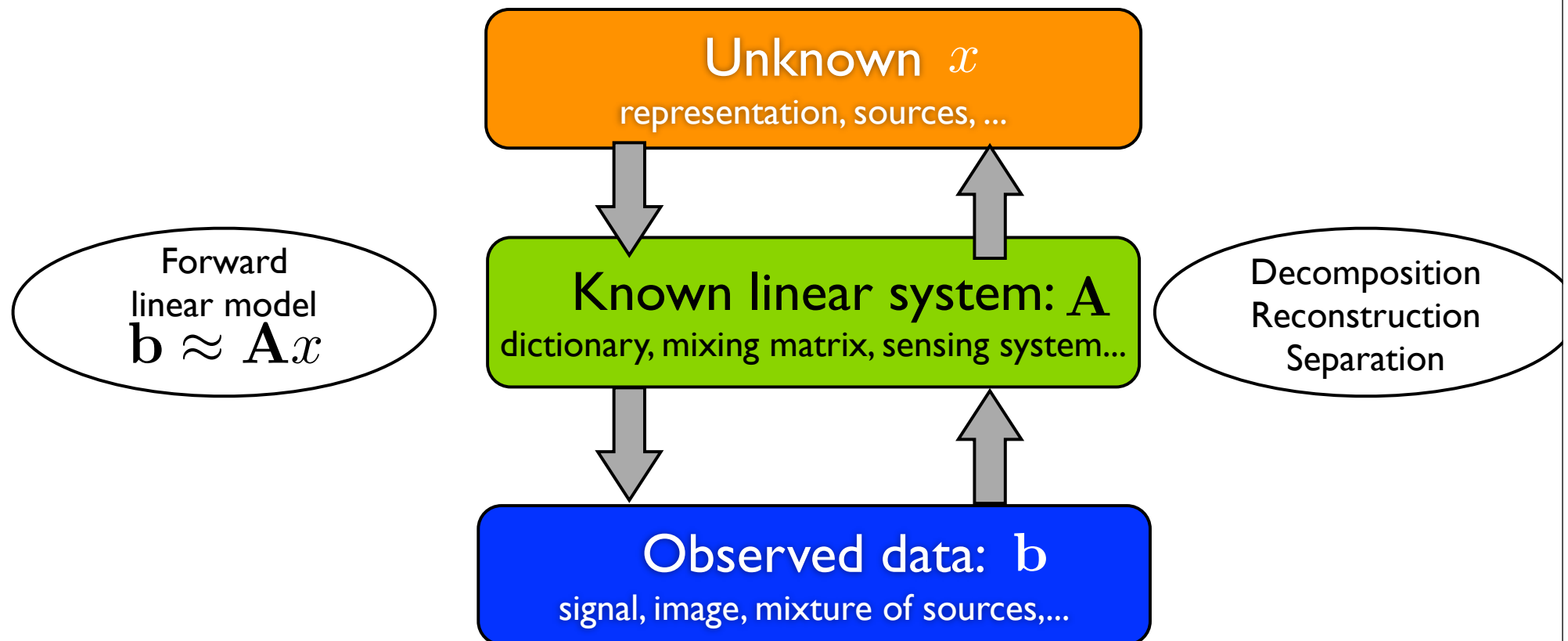
Machine Learning Vocabulary

Forward
linear model
$$\mathbf{b} \approx \mathbf{A}x$$

$$y = Xw$$

$\mathcal{W}$ **Regression coeffs**
Unknown x
representation, sources, ...

X **Design matrix** $\mathbf{A}$
Known linear system:
dictionary, mixing matrix, sensing system...

y **Observation**
Observed data: $\mathbf{b}$
signal, image, mixture of sources,...

Decomposition
Reconstruction
Separation

Statistics Vocabulary

β Regression coeffs
Unknown x
representation, sources, ...

Forward
linear model
$\mathbf{b} \approx \mathbf{A}x$

X **Design matrix** A
Known linear system: A
dictionary, mixing matrix, sensing system...

Decomposition
Reconstruction
Separation

$y = X\beta$

y **Observation**
Observed data: b
signal, image, mixture of sources,...

Inverse problems & Sparsity

Inverse Problems & Sparsity: Mathematical foundations

- ● **Bottleneck 1990-2000 :**
  - ✓ *Ill-posedness* when fewer equations than unknowns

$$\mathbf{A}x_0 = \mathbf{A}x_1 \not\Rightarrow x_0 = x_1$$

- ● **Novelty 2001-2006 :**
  - ✓ *Well-posedness* = uniqueness of sparse solution:

  - ✦ if x_0, x_1 are "sufficiently sparse",

  - ✦ then $\mathbf{A}x_0 = \mathbf{A}x_1 \Rightarrow x_0 = x_1$

  - ✓ *Recovery* of x_0 with practical *pursuit algorithms*
    - ✦ Thresholding, Matching Pursuits, Minimisation of Lp norms p<=1,...

Sparsity and subset selection

- **Under-determined system**
  - ✓ Infinitely many solutions

$$b \sim A x$$

- **If vector is sparse:**
  - ✓ If support is known (and columns independent)
    - ✦ nonzero values characterized by (over)determined linear problem
  - ✓ **If support is unknown**
    - ✦ Main issue = finding the support!
    - ✦ This is the **subset selection problem**

$$b \sim$$

- **Objectives of the course**
  - ✦ **Well-posedness** of subset selection
  - ✦ Efficient subset selection algorithms = **pursuit algorithms**
  - ✦ **Stability guarantees** of pursuits