SPARSITY IN SIGNAL AND IMAGE PROCESSING

Compressive Sensing: Theory and applications

INSA - GM - 5th year

Aline Roumy

informatics mathematics Inria

November-December 2020

Outline

1. Part 1 - Why compressive sensing?
2. Part 2 - Compressive sensing: how it works?
   - Notations (Reminder)
   - Problem formulation
   - Compressive sensing vs other schemes
3. Part 3 - Compressive sensing: good sensing matrix?
   - First insights
   - Reconstruction guarantee: Restricted Isometry Property
   - RIP and Iterative Hard Thresholding
   - Which matrices satisfy the RIP?
4. Part 4 - Compressive sensing: what it is good for?
5. Part 5 - Compressive sensing: summary

About me

Aline Roumy
Researcher at Inria, Rennes
Expertise: compression for video streaming
image/signal processing, information theory, machine learning

Web: http://people.rennes.inria.fr/Aline.Roumy/
email: aline.roumy@inria.fr

Course schedule (tentative)

Compressive sensing (CS): a self-sufficient course with a lot of connections to sparse approximations

- Nov. 30, 3 hours, lecture (why CS?)+(how it works?)
- Dec. 7, 3 hours, lecture (**why it works?**)
- Dec. 8, **4 hours**, lecture (what it is good for?) + (lab)

Tools

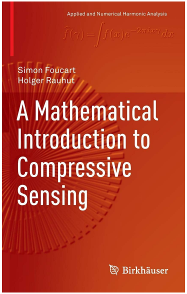
- Quiz: socrative RoomName: ALINER
- Computer Lab: collaborative jupyter notebook.

Course grading

- Project:
  - group of 3 persons
  - choose a paper within the list
  - write a report ($\sim$ 4 to 8 pages)
    implementation, further reading more than welcomed
  - oral presentation: 15 min + 5 min (questions) /group.
  - You will get a course grade from the evaluation of your report+oral.
  - Date: Paper, group repartition: Dec 11th (email)
    Report deadline: Jan. xxth (email)
    Slide deadline: Jan. xxth (email)
    Presentation: Jan. 25th, 8am.
- Final Exam:
  - (individual) oral exam: questions de synthèse de cours
  - 15 min preparation (with documents) / 15 min oral
  - Draw with mouse and share your screen, may be also take a picture of your preparation
  - Date: Jan. 18th, 8am.

Course material

S. Foucart, H. Rauhut, A mathematical introduction to compressive sensing, Birkhaüser, 2013.

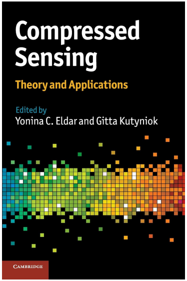


Early and short version:

S. Foucart, Notes on compressed sensing, 2009. (pdf)

Course material

Compressed Sensing: Theory and Applications, Edited by Y.C. Eldar and G. Kutyniok, Cambridge University Press, 2012.



- Chapter 1:
  M.A. Davenport, M.F. Duarte, Y.C. Eldar, G. Kutyniok Introduction to compressed sensing. **(pdf)**
- Short version:
  G. Kutyniok, Theory and Applications of Compressed Sensing, GAMM Mitteilungen 36 (2013), 79-101.

Outline

1. **Part 1 - Why compressive sensing?**
2. **Part 2 - Compressive sensing: how it works?**
   - Notations (Reminder)
   - Problem formulation
   - Compressive sensing vs other schemes
3. **Part 3 - Compressive sensing: good sensing matrix?**
   - First insights
   - Reconstruction guarantee: Restricted Isometry Property
   - RIP and Iterative Hard Thresholding
   - Which matrices satisfy the RIP?
4. **Part 4 - Compressive sensing: what it is good for?**
5. **Part 5 - Compressive sensing: summary**

Part 1 - Why compressive sensing?

What is compressive sensing?

Compressive sensing:
is a novel way to acquire (or sense or sample) and compress data.

Classical = sampling then compression

Compressive sensing = sampling **AND** compression

Several names exist:

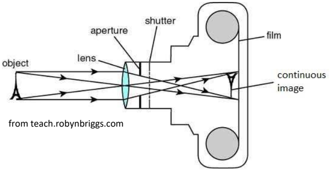
- compressed sensing
- compressed sampling
- compressive sampling
- **compressive sensing**. More accurate. Chosen in this course.
  The one of the reference book.

Part 1 - Why compressive sensing?

Review of **classical** digital acquisition:
classical=sampling + compression

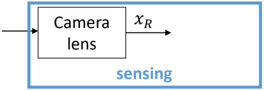
Film camera

Film camera: records images passing through the camera's lens.



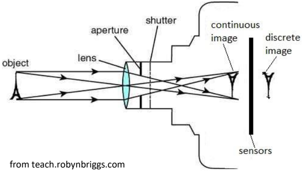
$x_R \colon [0,1]^2 \to \mathbb{R}^3$

continuous image



Digital camera

Digital camera: converts an image into **digital** data and **compress** it.

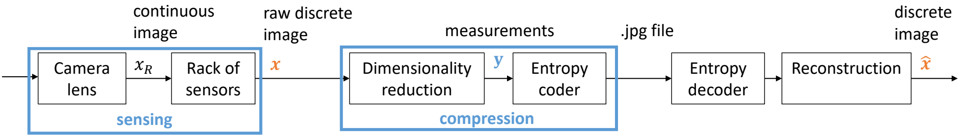


from teach.robynbriggs.com

$x_R : [0,1]^2 \to \mathbb{R}^3$

$x : \{1, N_a\} \times \{1, N_b\} \to \{0{,}255\}^3$

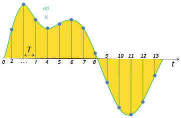
$y : \{1, M_a\} \times \{1, M_b\} \to \{0{,}255\}^3$



In this course: focus on the signal processing processes i.e. sensing + dimensionality reduction
entropy coder is an invertible process (from samples to bit)

Questions related to Digital camera

Question related to sampling:
is it possible to recover a continuous signal from its sampled (discrete) version?



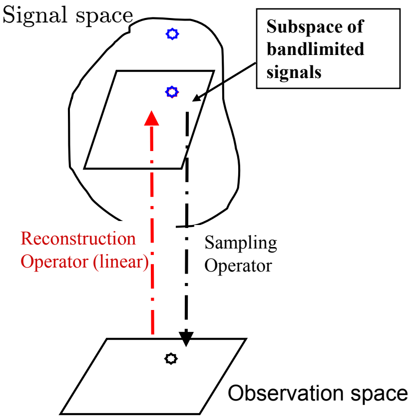
Wikipedia.

cf. course of Clément Elvira

Question related to compression:
is it possible to reduce the size of a discrete image?

Sampling: (1) optimal sampling rate

Nyquist–Shannon sampling theorem: "Exact reconstruction of a continuous-time signal from discrete samples is possible if the signal is bandlimited and the sampling frequency is greater than twice the highest frequency."

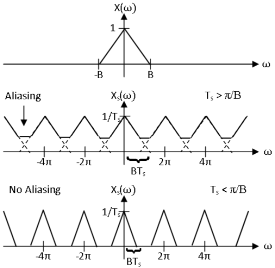


Mike Davies.

Sampling: (2) degradation if “slow” sampling

Sampling below the optimal rate introduces:

(1) aliasing

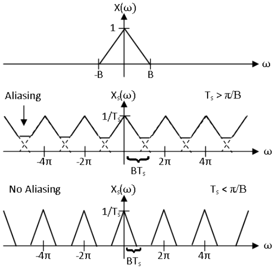


SVI.nl

Sampling: (2) degradation if "slow" sampling

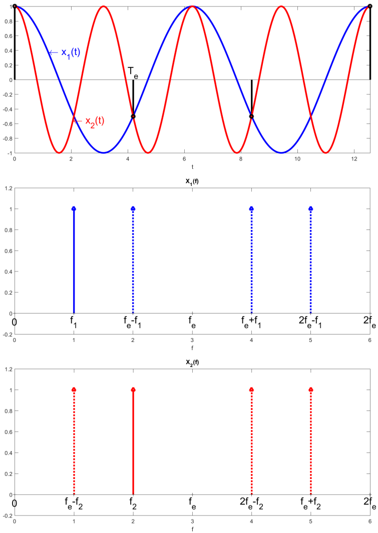
Sampling below the optimal rate introduces:

(1) aliasing



SVI.nl

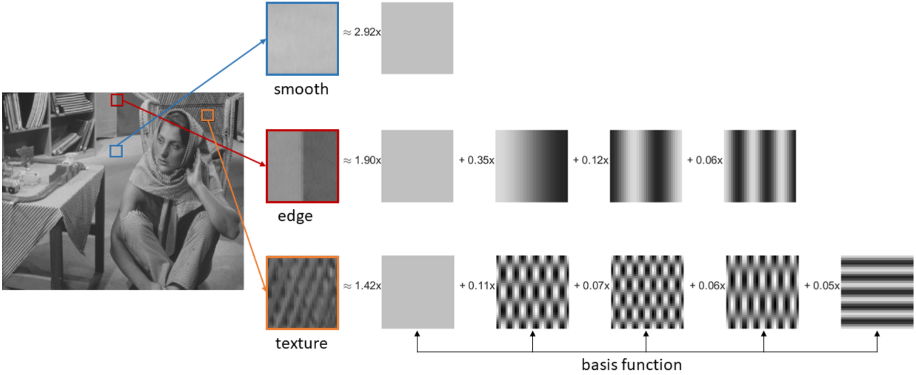
(2) signal ambiguity



Compression: (1) image decomposition principle

1- Split image into blocks of size $N_1 \times N_2$ each

2- Decompose each $N_1 \times N_2$ image block as:



How to choose the basis functions? How to compute the coefficients?

Compression: (2) image decomposition example

with 2D-discrete cosine transform (DCT) **(orthogonal basis)**

1- Split image into blocks of size $N_1 \times N_2$ each

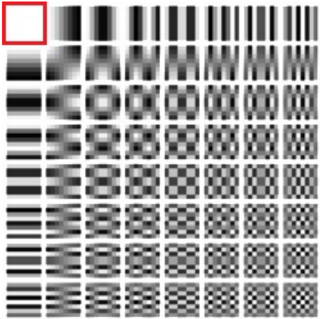
2- For each $N_1 \times N_2$ image block (x_{n_1,n_2})
compute the $N_1 \times N_2$ block of transformed image (c_{k_1,k_2}) with:

$$c_{k_1,k_2} = \sum_{n_1=0}^{N_1-1} \sum_{n_2=0}^{N_2-1} x_{n_1,n_2} \underbrace{\cos\left[\frac{\pi}{N_1}\left(n_1+\frac{1}{2}\right)k_1\right]\cos\left[\frac{\pi}{N_2}\left(n_2+\frac{1}{2}\right)k_2\right]}_{\Phi_{n_1,n_2}(k_1,k_2)}$$

Example: 8x8 DCT transform

Top-left matrix is $(\Phi_{n_1,n_2}(k_1=0,k_2=0))_{n_1,n_2}$

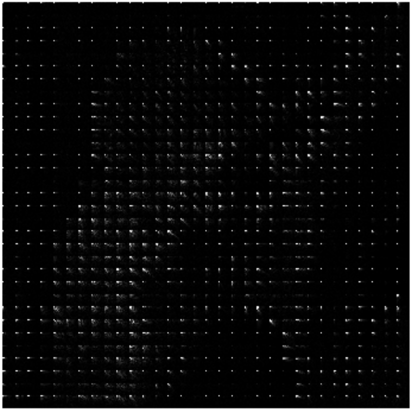
Quiz 1, 2, 3



Compression: (3) image decomposition result

Left: image

Right: discrete cosine transform of image



Key concept: few degrees of freedom in the transform domain

Compression: (4) dimensionality reduction with s-term approximation

1. Dimensionality reduction:

   keep the s coefficients c_s with largest absolute value

2. Reconstruction: $\hat{x} = \Phi^{-1} c_s$

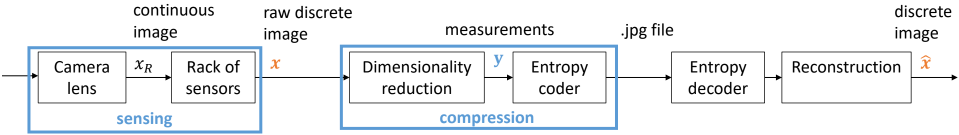
Left: 1% kept

Right: 5% kept

Quiz 4

Summary on classical sensing

Classical



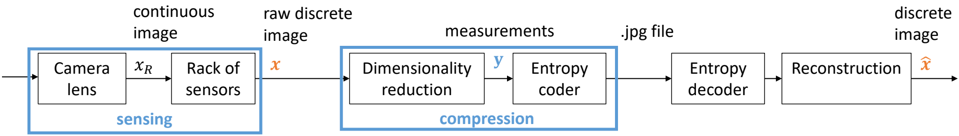
Sampling raw discrete HD video
1920x1080=2.07 M pixels/image
25Hz: images/s,
12(=8+2+2) bits/pixel
$\rightarrow$ 0.6 Gbit/s

Compression
For instance, HEVC (2013)
0.6 Gbit/s $\rightarrow$ 2Mbit/s

compression ratio 300:1!!!

Classical vs compressive sensing

Classical

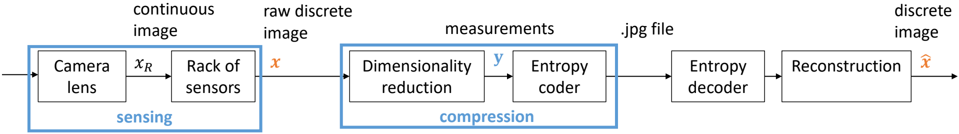


lots of samples,
throw most of the coefficients away

x: $[1, N_a] \times [1, N_b] \to \{0, 255\}^3$
y: $[1, M_a] \times [1, M_b] \to \{0, 255\}^3$
$(M_a M_b \ll N_a N_b)$

Classical vs compressive sensing

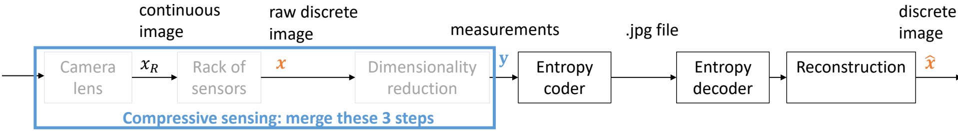
Classical



lots of samples,
throw most of the coefficients away

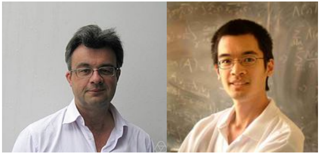
x: $[1, N_a] \times [1, N_b] \to \{0, 255\}^3$
y: $[1, M_a] \times [1, M_b] \to \{0, 255\}^3$
$(M_a M_b \ll N_a N_b)$

Compressive sensing: can we acquire less data in the first place?
and still recover $\hat{x}$?



Can we sample signals at the "Information Rate"?

Yes, we can!



Wikipedia.

Wikipedia.

E. J. Candes and T. Tao, 2005
"Decoding by linear programming"

D. L. Donoho, 2006
"Compressed sensing"

Outline

1. Part 1 - Why compressive sensing?
2. Part 2 - Compressive sensing: how it works?
   - Notations (Reminder)
   - Problem formulation
   - Compressive sensing vs other schemes
3. Part 3 - Compressive sensing: good sensing matrix?
   - First insights
   - Reconstruction guarantee: Restricted Isometry Property
   - RIP and Iterative Hard Thresholding
   - Which matrices satisfy the RIP?
4. Part 4 - Compressive sensing: what it is good for?
5. Part 5 - Compressive sensing: summary

Part 2 - Maths of compressive sensing - how it works?

Notations (Reminder)

Norms

Definition (l_p-norm)

The l_p**-norm** of $x \in \mathbb{R}^n$, $p > 1$ is defined as

$$||x||_p = \begin{cases} \left(\sum_{i=1}^{n} |x_i|^p \right)^{1/p} & p \in [1, \infty) \\ \max_i |x_i| & p = \infty \end{cases}$$

If $p < 1$, definition still valid, but triangle inequality not satisfied
$\Rightarrow$ **quasi-norm**.

Definition (inner product)

$$\langle x, z \rangle = z^T x = \sum_{i=1}^{n} x_i z_i$$

See textbook F.R. for extension to $\mathbb{C}^n$.

Definition (support and l_0-norm)

The **support** of a vector x is the index set of its non-zero entries, i.e.

$$\text{supp }(x) = \{j \in [n] : x_j \neq 0\}, \text{ where } [n] = \{1, 2, ..., n\}$$

The l_0-**norm** of x is defined as

$$||x||_0 = \text{ card } (\text{ supp } (x) \text{ })$$

$||x||_0$ counts the number of non-zero entries of x.
$||.||_0$ is not even a quasi-norm.

Sparsity definition

Definition (s-sparse)

A signal $x \in \mathbb{R}^n$ is said to be s-sparse if it has at most s non-zero entries, i.e. $||x||_0 \leq s$.

Definition (Σ_s)

We define Σ_s as the set containing all s-sparse signals, i.e. $\Sigma_s = \{x \in \mathbb{R}^n : ||x||_0 \leq s\}$.

Quiz 5

Note 1: Sparsity is a highly nonlinear model (Σ_s is not a linear space)

Note 2: in many practical cases, x is not sparse itself, but it has a sparse representation in some basis Φ. We still say that x is s-sparse, with the understanding that we can write $x = \Phi u$, and $||u||_0 \leq s$.

Approximate sparsity

- A sparse signal can be represented exactly giving the positions and values of its s nonzero components
- Real-world signals are rarely exactly sparse.
  We need to
  - generalize the def: from "sparse" to "compressible" signals,
  - describe the representation error i.e. the error incurred representing just s components of the signal.

Best s-term approximation

The best s-term approximation picks the s components that minimize the representation error

Definition (best s-term approximation)

For $p > 0$, the l_p-error incurred by the best s-term approximation to a vector $x \in \mathbb{R}^n$ is given by

$$\sigma_s(x)_p = \min_{\hat{x} \in \Sigma_s} ||x - \hat{x}||_p$$

- If $x \in \Sigma_s$, then $\sigma_s(x)_p = 0$ for any p.

Compressible signal

Optimal strategy to compute the best s-term approximation:
thresholding

- Reorder the elements of x by decreasing magnitude
- Pick the first s elements, set all others to zero.

Definition (compressible signal)

a signal $x \in \mathbb{R}^n$ is said to be compressible if the error of its best s-term approximation decays quickly in s
i.e. if $\exists C_1, q > 0$ such that $|x_i| \leq C_1 i^{-q}$., when the coefficients have been ordered such that $|x_1| \geq |x_2| ... \geq |x_n|$.

Sparsity support

Suppose $x \in R^n$. Let $S \subset [n]$ and $S^c \subset [n] \backslash S$

- S: sparsity support of x, i.e. the locations of the nonzero coefficients of x
- S^c: set of locations of the 0 coefficients
- S for compressible signal: set of locations of the coefficients belonging to the best s-term approximation of x.

Notation

x_S vector obtained by setting the entries of x indexed by S^c to 0.

M_S matrix obtained by setting the columns of M indexed by S^c to 0.

- Same notation to denote vectors/matrices where the elements/columns have been removed, instead of being set to 0

Outline

1. Part 1 - Why compressive sensing?
2. Part 2 - Compressive sensing: how it works?
   - Notations (Reminder)
   - Problem formulation
   - Compressive sensing vs other schemes
3. Part 3 - Compressive sensing: good sensing matrix?
   - First insights
   - Reconstruction guarantee: Restricted Isometry Property
   - RIP and Iterative Hard Thresholding
   - Which matrices satisfy the RIP?
4. Part 4 - Compressive sensing: what it is good for?
5. Part 5 - Compressive sensing: summary

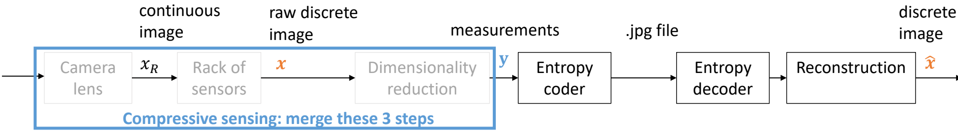
Part 2 - Maths of compressive sensing - how it works?

Problem formulation

Compressive sensing

Goal of Compressive sensing (CS):

- achieve the same reconstruction quality on $\hat{x}$ as the best s-term approximation
- from the measurement y acquired with a nonadaptive encoder.



To achieve this, we need to

1. model the dependency between signal x and measurement y
2. formulate the reconstruction problem

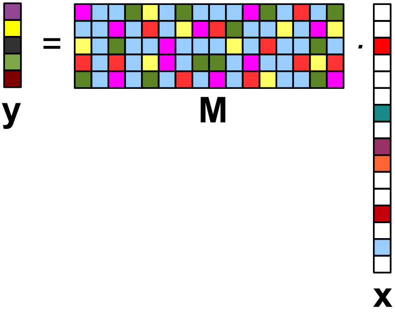
Sensing process model

(Modeling the dependency between signal and measurement)

Let $x \in R^{nx1}$ *be a s-sparse signal to be recovered.*
Let $y \in R^{mx1}$, $m < n$, *be linear measurements of the signal as*

$$y = Mx$$

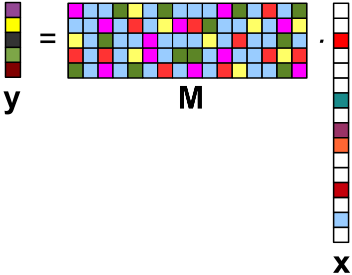
with $M \in R^{mxn}$, *being the sensing matrix.*



Reconstruction: problem formulation

(problem formulation)

Given measurement y*, sensing matrix* M *and the model* $y = Mx$*, Recover* x*,* s*-sparse.*

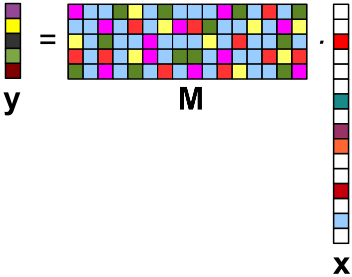


Difficulties?

Reconstruction: problem formulation

(problem formulation)

Given measurement y, sensing matrix M and the model $y = Mx$, Recover x, s-sparse.



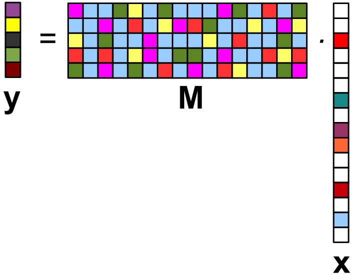
Difficulties?

- Underdetermined system $\Rightarrow$ infinitely many solutions.

Reconstruction: problem formulation

(problem formulation)

Given measurement y, sensing matrix M and the model $y = Mx$, Recover x, s-sparse.



Difficulties?

- Underdetermined system $\Rightarrow$ infinitely many solutions.
- **Idea** exploit the sparsity assumption of x.

Minimum l_0-norm solution

$$\hat{x} = \arg\min_{z \in \mathbb{R}^n} ||z||_0 \textit{ subject to } Mz = y$$

Complexity?

- Problem is non-convex
- Problem is NP-hard:
  for a given s, try all possible $\binom{n}{s}$ supports, estimate the s nonzero values of x, check if constraint is satisfied
  $\Rightarrow$ **infeasible** for practical problem sizes

Practical philosophies

$$\hat{x} = \arg\min_{z \in \mathbb{R}^n} ||z||_0 \textit{ subject to } Mz = y$$

Greedy algorithms

Focus on $||x||_0$

Thresholding algorithms

Focus on $y \sim Mx$

Convex relaxation algorithms

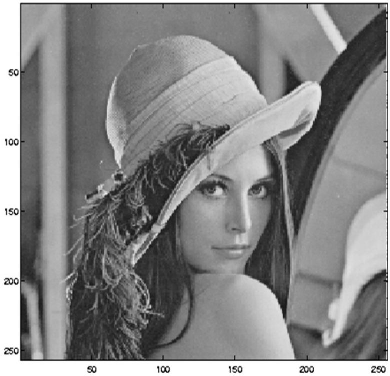
Solve a nicer problem

see course C. Elvira

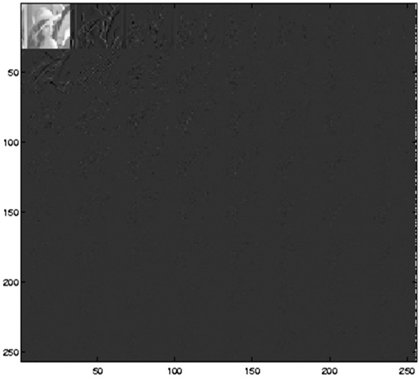
Signal sparse in transform domain

Real signals are rarely directly sparse...

but rather sparse in a transform domain



original image



DCT coefficients of the image
in the transform domain

Signal sparse vs signal sparse in transform domain

x sparse

SENSING
$y = Mx$

RECONSTRUCTION
$\hat{x} = \arg\min_{z \in \mathbb{R}^n} ||z||_1$
subject to $Mz = y$
.

$x = \Phi u$, u sparse

SENSING
$y = Mx$

RECONSTRUCTION
$\hat{u} = \arg\min_{z \in \mathbb{R}^n} ||z||_1$
subject to $M\Phi z = y$
$\hat{x} = \Phi \hat{u}$

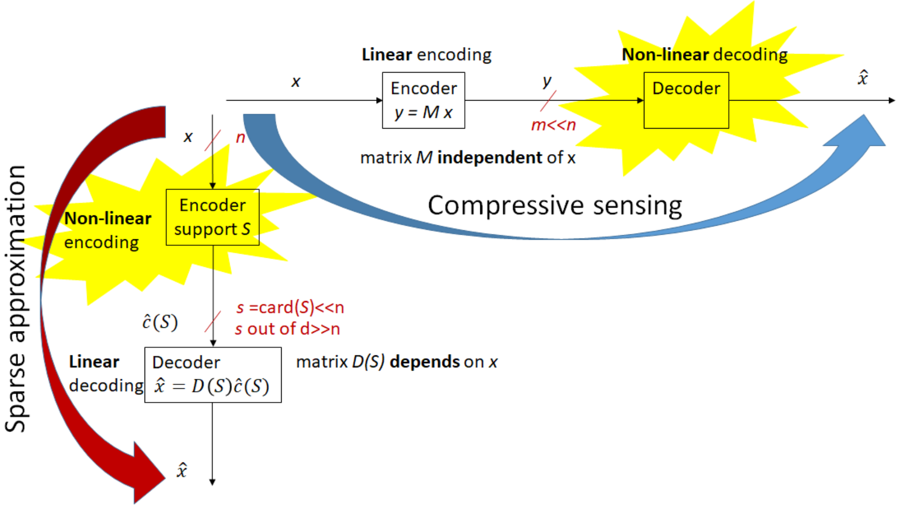
In conclusion: sparse vs sparse in the transform domain

- same sensing
- similar reconstruction problem
- Make sure that $M\Phi$ (and not M) is a “good” sensing matrix

Part 2 - Maths of compressive sensing - how it works?

Compressive sensing vs other schemes

Compressive sensing (CS) vs Sparse approximation (SA)



CS vs SA (con't)

Non-linear solvers:

CS Given y and M, find $\hat{x}$ sparse such that $M\hat{x} \approx y$.

Return $\hat{x}$ with guarantee that

$$||\hat{x} - x|| \quad \text{small}$$

SA Given x and D, find $\hat{c}$ sparse such that $\hat{x} = D\hat{c} \approx x$.

Return $\hat{x}$ with guarantee that

$$||\hat{x} - x|| = ||D(\hat{c} - c)|| \quad \text{small}$$

CS vs SA (con't)

Non-linear solvers:

CS Given y and M, find $\hat{x}$ sparse such that $M\hat{x} \approx y$.

Return $\hat{x}$ with guarantee that

$$||\hat{x} - x|| \quad \text{small}$$

SA Given x and D, find $\hat{c}$ sparse such that $\hat{x} = D\hat{c} \approx x$.

Return $\hat{x}$ with guarantee that

$$||\hat{x} - x|| = ||D(\hat{c} - c)|| \quad \text{small}$$

Same decomposition algorithms

Different criteria

CS vs SA (con't)

Non-linear solvers:

CS Given y and M, find $\hat{x}$ sparse such that $M\hat{x} \approx y$.

Return $\hat{x}$ with guarantee that

$$||\hat{x} - x|| \quad \text{small}$$

SA Given x and D, find $\hat{c}$ sparse such that $\hat{x} = D\hat{c} \approx x$.

Return $\hat{x}$ with guarantee that

$$||\hat{x} - x|| = ||D(\hat{c} - c)|| \quad \text{small}$$

Root-finding algorithm:

CS Given $y = 0$ and f, find $\hat{x}$ such that $y = 0 \approx f(\hat{x})$.

Return $\hat{x}$ with guarantee that

$$||\hat{x} - x|| \quad \text{small}$$

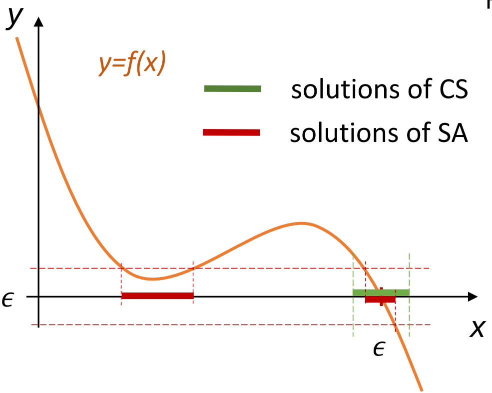
SA Given $y = 0$ and f, find $\hat{x}$ such that $y = 0 \approx \hat{y} = f(\hat{x})$.

Return $\hat{y}$ with guarantee that

$$||f(\hat{x}) - 0|| \quad \text{small}$$

CS: proximity to the true root
SA: proximity to zero in the range of the function

CS vs SA (con't)



Root-finding algorithm:

CS Given $y = 0$ and f, find $\hat{x}$ such that $y = 0 \approx f(\hat{x})$.
Return $\hat{x}$ with guarantee that

$||\hat{x} - x||$ small

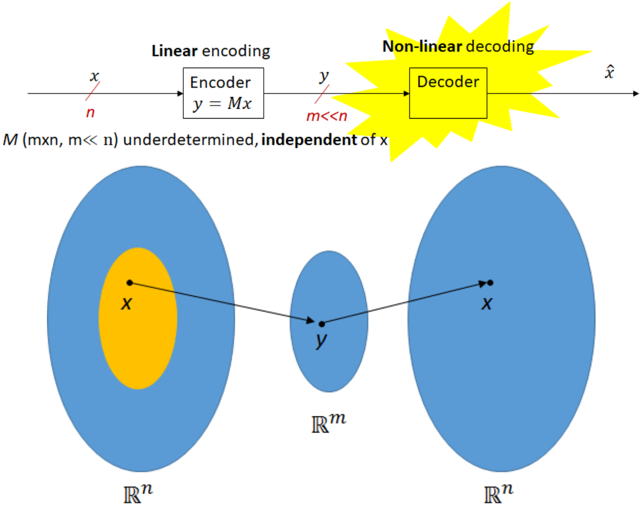
SA Given $y = 0$ and f, find $\hat{x}$ such that $y = 0 \approx \hat{y} = f(\hat{x})$.
Return $\hat{y}$ with guarantee that

$||f(\hat{x}) - 0||$ small

CS: proximity to the true root
SA: proximity to zero in the range of the function

Part 3 - Compressive sensing - good sensing matrix?

Compressive sensing: summary of what seen so far

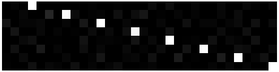


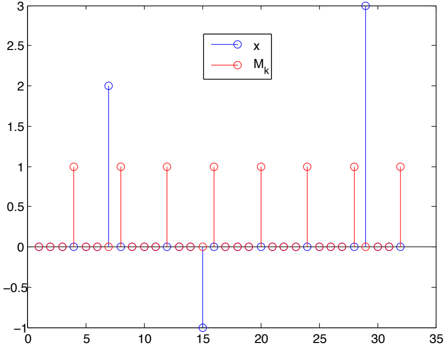
How should we choose a “good” matrix M with $m \ll n$?

Part 3 - Compressive sensing - good sensing matrices?

First insights

Sensing matrices that are not good





Vector y is all zero!
→ If x sparse, M must be **non-sparse**

Part 3 - Compressive sensing - good sensing matrices?

Reconstruction guarantee: Restricted Isometry Property

The problem: invert $y = Mx$

M square 

$\exists$ a reconstruction map:

$$\mathbb{R}^m \to \mathbb{R}^n$$

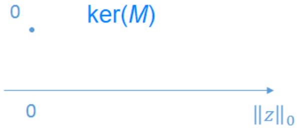
$$y \mapsto x = M^{-1}y$$

$$\Updownarrow$$

condition on the matrix

$$\mathrm{rank}(M) = m = n$$

$$\ker(M) = \{z : Mz = 0\} = \{0\}$$



The problem: invert $y = Mx$

M square 

$\exists$ a reconstruction map:

$$\mathbb{R}^m \to \mathbb{R}^n$$

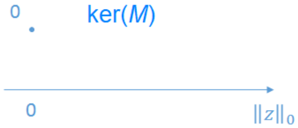
$$y \mapsto x = M^{-1}y$$

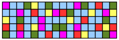
$$\Updownarrow$$

condition on the matrix

$\text{rank}(M) = m = n$

$\ker(M) = \{z : Mz = 0\} = \{0\}$



M fat 

$\exists$ a reconstruction map:

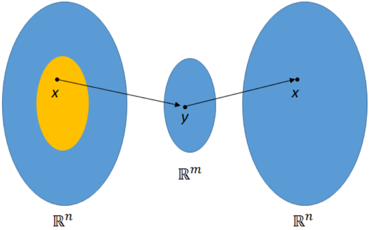
$$\mathbb{R}^m \to \mathbb{R}^n$$

$$y \mapsto x =??$$

$$\Updownarrow$$

condition on the matrix ???

NEW Reduce the domain of definition of M: s-sparse



The restricted isometry property (RIP) : definition

Definition (RIP)

Let $\epsilon > 0$, $s \in \mathbb{N}$. A matrix $M \in \mathbb{R}^{m,n}$ with $m \ll n$ is (ϵ, s)-RIP if

$$\forall x \in \Sigma_s, \ (1-\epsilon)||x||_2^2 \leq ||Mx||_2^2 \leq (1+\epsilon)||x||_2^2 \tag{1}$$

Interpretation of (ϵ, s)-RIP:

- M preserves the Euclidean norm of s-sparse vectors

The restricted isometry property (RIP) : definition

Definition (RIP)

Let $\epsilon > 0$, $s \in \mathbb{N}$. A matrix $M \in \mathbb{R}^{m,n}$ with $m \ll n$ is (ϵ, s)-RIP if

$$\forall x \in \Sigma_s, \; (1-\epsilon)||x||_2^2 \leq ||Mx||_2^2 \leq (1+\epsilon)||x||_2^2 \tag{1}$$

Interpretation of (ϵ, s)-RIP:

- M preserves the Euclidean norm of s-sparse vectors
- $M \in \mathbb{R}^{m,n}$ with $m \ll n$ is (ϵ, s)-RIP if

$$\forall x \in \Sigma_s \setminus \{0\}, \; \left| \frac{||Mx||_2^2 - ||x||_2^2}{||x||_2^2} \right| \leq \epsilon \tag{2}$$

The restricted isometry property (RIP) : definition

Definition (RIP)

Let $\epsilon > 0$, $s \in \mathbb{N}$. A matrix $M \in \mathbb{R}^{m,n}$ with $m \ll n$ is (ϵ, s)-RIP if

$$\forall x \in \Sigma_s, \ (1-\epsilon)||x||_2^2 \leq ||Mx||_2^2 \leq (1+\epsilon)||x||_2^2 \tag{1}$$

Interpretation of (ϵ, s)-RIP:

- M preserves the Euclidean norm of s-sparse vectors
- $M \in \mathbb{R}^{m,n}$ with $m \ll n$ is (ϵ, s)-RIP if

$$\forall x \in \Sigma_s \setminus \{0\}, \ \left| \frac{||Mx||_2^2 - ||x||_2^2}{||x||_2^2} \right| \leq \epsilon \tag{2}$$

Quiz 1, 2

RIP: l_0 reconstruction

Proposition (RIP and l_0 reconstruction)

Let $M \in \mathbb{R}^{m,n}$ with $m \ll n$. Let $0 < \epsilon < 1$. If M is $(\epsilon, 2s)$-RIP, then

$$\forall x \in \Sigma_s, \quad \hat{x} = x, \quad \text{with } \hat{x} \in \arg \min_{z:Mz=y} ||z||_0.$$

Proof. *Quiz 2+ Th 2.13 of Foucart-Rauhut (course of N. Bertin).*

Interpretation:
"a $(\epsilon, 2s)$-RIP matrix is a good sensing matrix for l_0 reconstruction."

THE Question: *is a $(\epsilon, 2s)$-RIP matrix a good sensing matrix for practical reconstruction algorithms?*

RIP: operator norm

Lemma (RIP and operator norm)

Let $M \in \mathbb{R}^{m,n}$ with $m \ll n$. If M is (ϵ, s)-RIP, then

$$\forall S \subset [\![1, n]\!], |S| \leq s, \quad ||M_S^T M_S - I_S||_{op} \leq \epsilon. \tag{3}$$

Recall (note in the definition below $||.||_2$ not $||.||_2^2$)

$$||M_S^T M_S - I||_{op} = \max_{x_S \neq 0} \frac{||(M_S^T M_S - I)x_S||_2}{||x_S||_2}.$$

Proof. **Quiz 3**

Interpretation: **Quiz 4**

$\forall S, M_S^T M_S \approx I_S$ when applied to any x_S (vector of size S)

Part 3 - Compressive sensing - good sensing matrices?

RIP and Iterative Hard Thresholding

A practical algorithm

Definition (Iterative Hard Thresholding (IHT))

$$x^0 = 0$$

$$x^{l+1} = H_s\left(x^l + M^T(y - Mx^l)\right)$$

$$\text{output: } \hat{x}_{IHT} = \lim_{l \to +\infty} x^l$$

H_s: Hard Thresholding
keeps the s coefficients with largest absolute value.

Justification: $x^{l+1} = H_s\left(x^l + \text{ error}(x - x^l)\right)$

RIP is good for IHT

Theorem (Optimality of IHT for RIP matrices)

Let $M \in \mathbb{R}^{m,n}$ *with* $m \ll n$*. Let* $\epsilon > 0$*.*
If M is $(\epsilon, \mathbf{3}s)$*-RIP, then*

$$||x^{l+1} - x|| \leq 2\epsilon ||x^{l} - x|| \tag{4}$$

Interpretation: **Quiz 5**

RIP is good for IHT

Theorem (Optimality of IHT for RIP matrices)

Let $M \in \mathbb{R}^{m,n}$ *with* $m \ll n$. *Let* $\epsilon > 0$.
If M is $(\epsilon, \mathbf{3}s)$*-RIP, then*

$$||x^{l+1} - x|| \leq 2\epsilon ||x^l - x|| \quad (4)$$

In particular, if $\epsilon < \frac{1}{2}$, $\quad x^l \xrightarrow[l \to +\infty]{} x.$

Interpretation: **Quiz 5**

Proof. **Quiz 6**

Summary: if M is $(\epsilon, 3s)$-RIP, with $\epsilon < 1/2$, then $\hat{x}_{IHT} = x$

Similarly: if M is $(\epsilon, 2s)$-RIP, with $\epsilon < 1/3$, then $\hat{x}_{BP} = x$ [FR, Th 6.9]

if M is $(\epsilon, 13s)$-RIP, with $\epsilon < 1/6$, then $\hat{x}_{OMP} = x$ [FR, Th 6.25]

THE question: how to construct a matrix M that is $(1/2, 3s)$-RIP?

Part 3 - Compressive sensing - good sensing matrices?

Which matrices satisfy the RIP?

Concentration inequality

Theorem (Concentration of Gaussian Matrices)

Let $x \in \mathbb{R}^n$. Let $M \in \mathbb{R}^{m,n}$ s.t. $M_{i,j} \sim \mathcal{N}(0, 1/m)$ i.i.d.

$$\forall 0 \leq t \leq 3, \quad \mathbb{P}_M \left(\underbrace{\left| \frac{||Mx||_2^2}{||x||_2^2} - 1 \right| > t}_{(*)} \right) \leq 2e^{-\frac{mt^2}{6}} \tag{5}$$

Interpretation:

- $\text{Neg}(*) \Leftrightarrow (1-t)||x||_2^2 \leq ||Mx||_2^2 \leq (1+t)||x||_2^2 \Leftrightarrow M$ is good for this x
- **Quiz 7**

$$(5) \;\; \Leftrightarrow \underbrace{\exists \alpha, \delta \text{ s.t. } \mathbb{P}_M \left(\left| ||Mx||_2^2 - \mathbb{E}[||Mx||_2^2] \right| > \alpha \right) \leq \delta}_{\text{concentration (around the mean) inequality}}$$

Other concentration inequalities

Markov's inequality (due to Chebyshev (Markov's teacher)):
Given a non-negative random variable X with finite mean

$$\mathbb{P}(X \geq t) \leq \frac{\mathbb{E}[X]}{t}, \quad \forall t > 0. \qquad \text{Decay in } \mathcal{O}(\frac{1}{t}) \tag{6}$$

Other concentration inequalities

Markov's inequality (due to Chebyshev (Markov's teacher)):
Given a non-negative random variable X with finite mean

$$\mathbb{P}(X \geq t) \leq \frac{\mathbb{E}[X]}{t}, \quad \forall t > 0. \qquad \text{Decay in } \mathcal{O}(\frac{1}{t}) \tag{6}$$

Chebyshev's inequality: Given a random variable X with mean μ and finite variance (denoted $\text{var}(X) < \infty$)

$$\mathbb{P}(|X - \mathbb{E}(X)| \geq t) \leq \frac{\text{var}(X)}{t^2}, \quad \forall t > 0. \qquad \text{Decay in } \mathcal{O}(\frac{1}{t^2}) \tag{7}$$

Other concentration inequalities

Chernoff bound: (due to Herman Rubin)
Given a random variable X with mean μ and finite variance

$$\mathbb{P}(X - \mu \geq t) \leq \frac{\mathbb{E}[e^{\lambda|X-\mu|}]}{e^{\lambda t}}, \quad \forall t, \lambda > 0. \text{ Decay in } \mathcal{O}(e^{-\lambda t}) \tag{8}$$

Other concentration inequalities

Chernoff bound: (due to Herman Rubin)
Given a random variable X with mean μ and finite variance

$$\mathbb{P}(X-\mu \geq t) \leq \frac{\mathbb{E}[e^{\lambda|X-\mu|}]}{e^{\lambda t}}, \quad \forall t, \lambda > 0. \quad \text{Decay in } \mathcal{O}(e^{-\lambda t}) \tag{8}$$

Cramer-Chernoff method:

step 1 Apply Chernoff bound

step 2 Bound optimization

$$\inf_{\lambda > 0} \frac{\mathbb{E}[e^{\lambda|X-\mu|}]}{e^{\lambda t}}$$

step 3 Repeat with $X' := -X$.

Difference between RIP and concentration

Concentration inequality for Gaussian matrices (5) means
Given x

$$\mathbb{P}_M\left(\left|||Mx||_2^2 - ||x||_2^2\right| > t||x||_2^2\right) \leq 2e^{-\frac{mt^2}{6}}$$

RIP means
For all x s-sparse

$$(1-t)||x||_2^2 \leq ||Mx||_2^2 \leq (1+t)||x||_2^2$$

Quiz 8

Condition for "RIP" over FINITE set

Lemma (Johnson-Lindenstrauss)

Let $M \in \mathbb{R}^{m,n}$ s.t. $M_{i,j} \sim \mathcal{N}(0, 1/m)$. Let $t > 0, \delta > 0$.
Let $\mathcal{Q}$ a **finite set** *of vectors $\subset \mathbb{R}^n$.*

If $m \geq \dfrac{6}{t^2} \log \dfrac{2|\mathcal{Q}|}{\delta}$, then

$$\mathbb{P}_M \left(\sup_{x \in \mathcal{Q}} \left| \frac{||Mx||_2^2}{||x||_2^2} - 1 \right| \leq t \right) \geq 1 - \delta \tag{9}$$

Interpretation: with probability at least $1 - \delta$, the norm of the vectors is preserved (precision t).

Proof: **Quiz 9**

Condition for RIP and success of IHT

Theorem (RIP and success of IHT [FR, Th. 6.15 and Chap. 12.5]**)**

Let $M \in \mathbb{R}^{m,n}$ *s.t.* $M_{i,j} \sim \mathcal{N}(0, 1/m)$. *Let* $\epsilon > 0, \delta > 0$.
If $m \geq \frac{4}{\epsilon^2}\left(2s \ln \frac{en}{s} + 7s + 2 \ln \frac{2}{\delta}\right)$, *then*

$$\mathbb{P}_M\left(\sup_{x \in \Sigma_s} \left|\frac{||Mx||_2^2}{||x||_2^2} - 1\right| > \epsilon\right) \leq \delta \quad (10)$$

In particular: $\exists c_1, c_2, c_3 > 0$ s.t. if $m \geq c_1 s \ln \frac{n}{s} + c_2 s + c_3$,
then with probability at least $1 - \delta$

$$\hat{x}_{IHT} = x$$

Proof: **Quiz 10**

Part 4 - Compressive sensing - what it is good for?

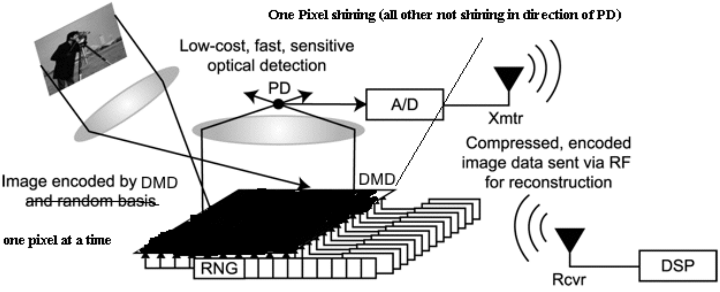
How to spot a compressive sensing system?

Case 1

- Think about systems that use a raster mode for sampling then think of physical ways to perform multiplexing instead
- Once you perform the multiplexing, use compressive sensing solvers to reconstruct signal
- Does it work better or as well with fewer measurements ?

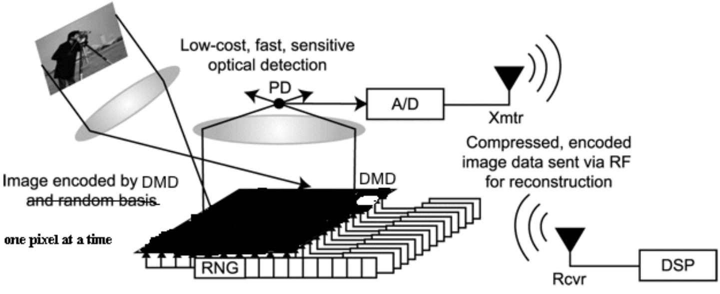
Example: single pixel camera

Classical i=1



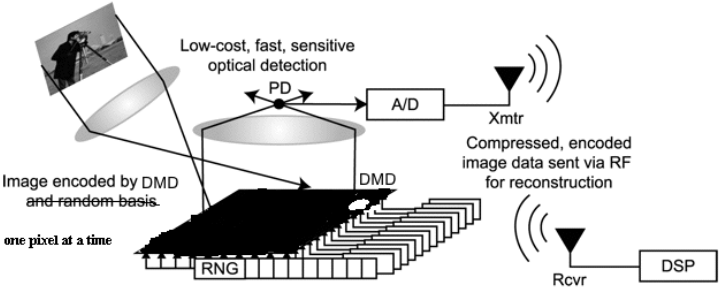
Example: single pixel camera

Classical i=2



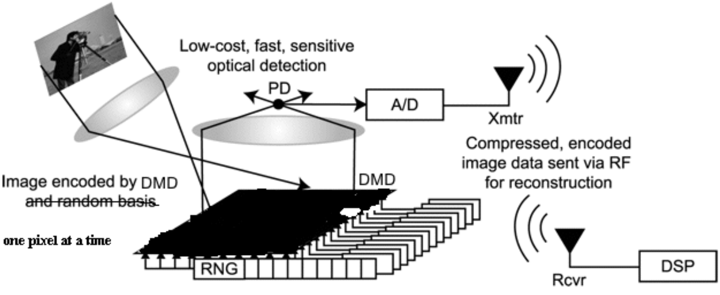
Example: single pixel camera

Classical i=3



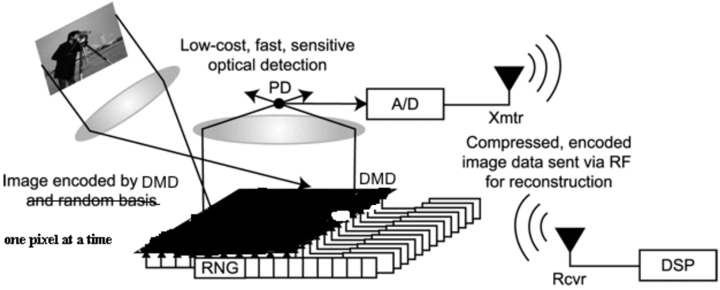
Example: single pixel camera

Classical i=4



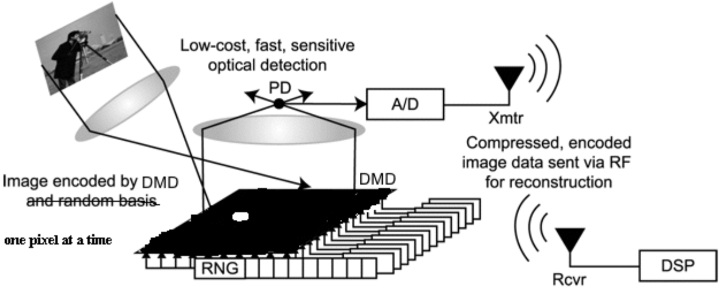
Example: single pixel camera

Classical i=5



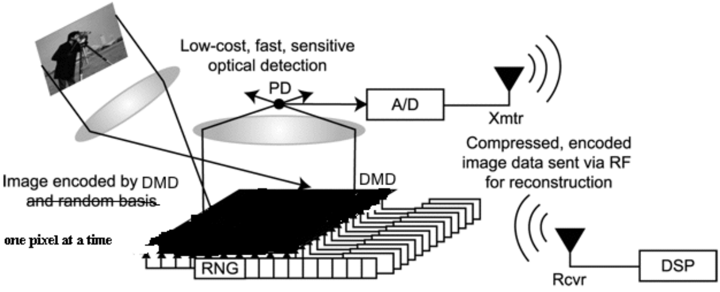
Example: single pixel camera

Classical i=1000



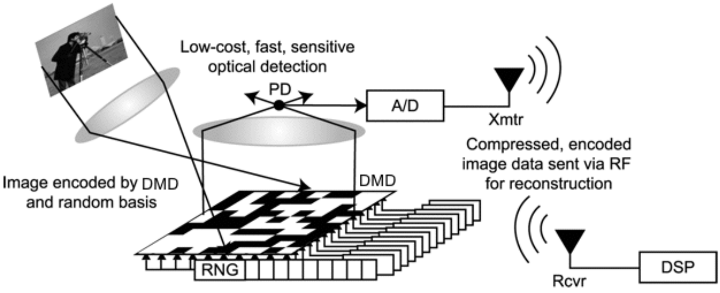
Example: single pixel camera

Classical i=10000000



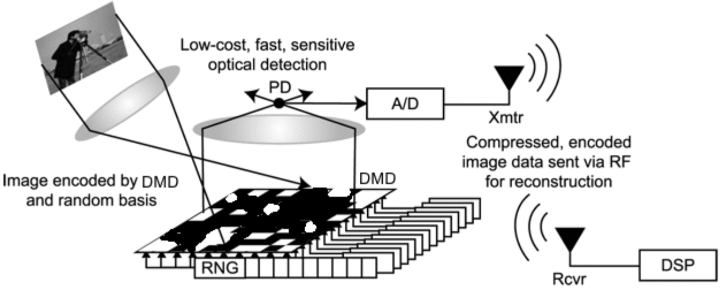
Example: single pixel camera

Compressive sensing i=1



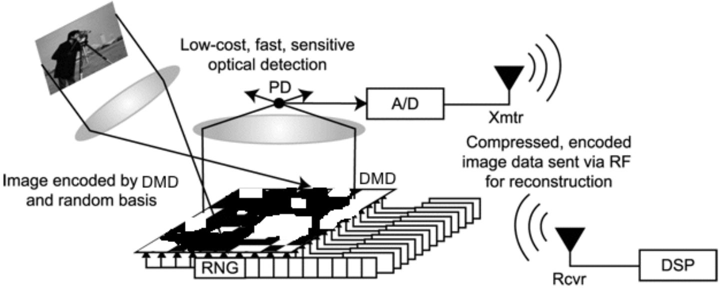
Example: single pixel camera

Compressive sensing i=2



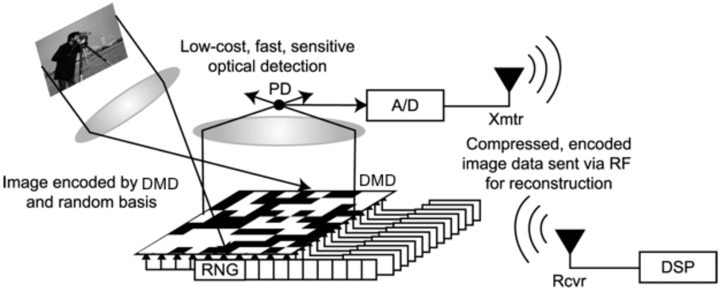
Example: single pixel camera

Compressive sensing i=3



Example: single pixel camera

Compressive sensing



if image is 3-sparse, the sufficient number of measurements scales with 3 and not the size of the image!!!!

How to spot a compressive sensing system?

Case 2

- Look for acquisition schemes that multiplexes a signal already
- Is the signal produced by this system sparse in some basis?
- If yes, subsample the acquisition,
  use compressive sensing solvers to reconstruct signal
- Does it work better or as well with fewer measurements ?

Part 5 - Compressive sensing - summary

Compressive sensing overview

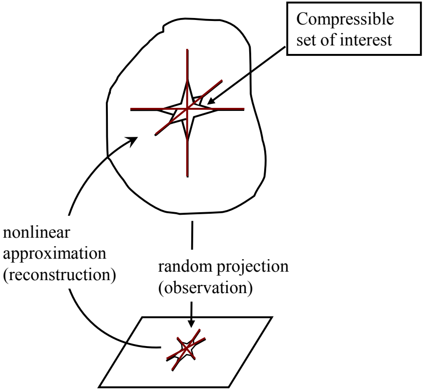
Observe $x \in \mathbb{R}^n$ via m measurements, with $m \ll n$
More precisely, $y = Mx$ where $y \in \mathbb{R}^m$

Assumptions:

- signal approximately s-sparse

- use $m \geq c\ s \log \frac{n}{s}$, c=constant, random linear measurements

- reconstruct by a non linear mapping



Mike Davies.