"Ca va être compliqué": Islands of knowledge, Mathematician-Pirates and the Great Convergence

Igor Carron,
https://www.linkedin.com/in/IgorCarron
http://nuit-blanche.blogspot.com

IFPEN presentation, March 30th, 2015

Outline

- Sensing
- Big Data
- What have we learned from compressive sensing, advanced matrix factorization ?
- Machine Learning
- Two words

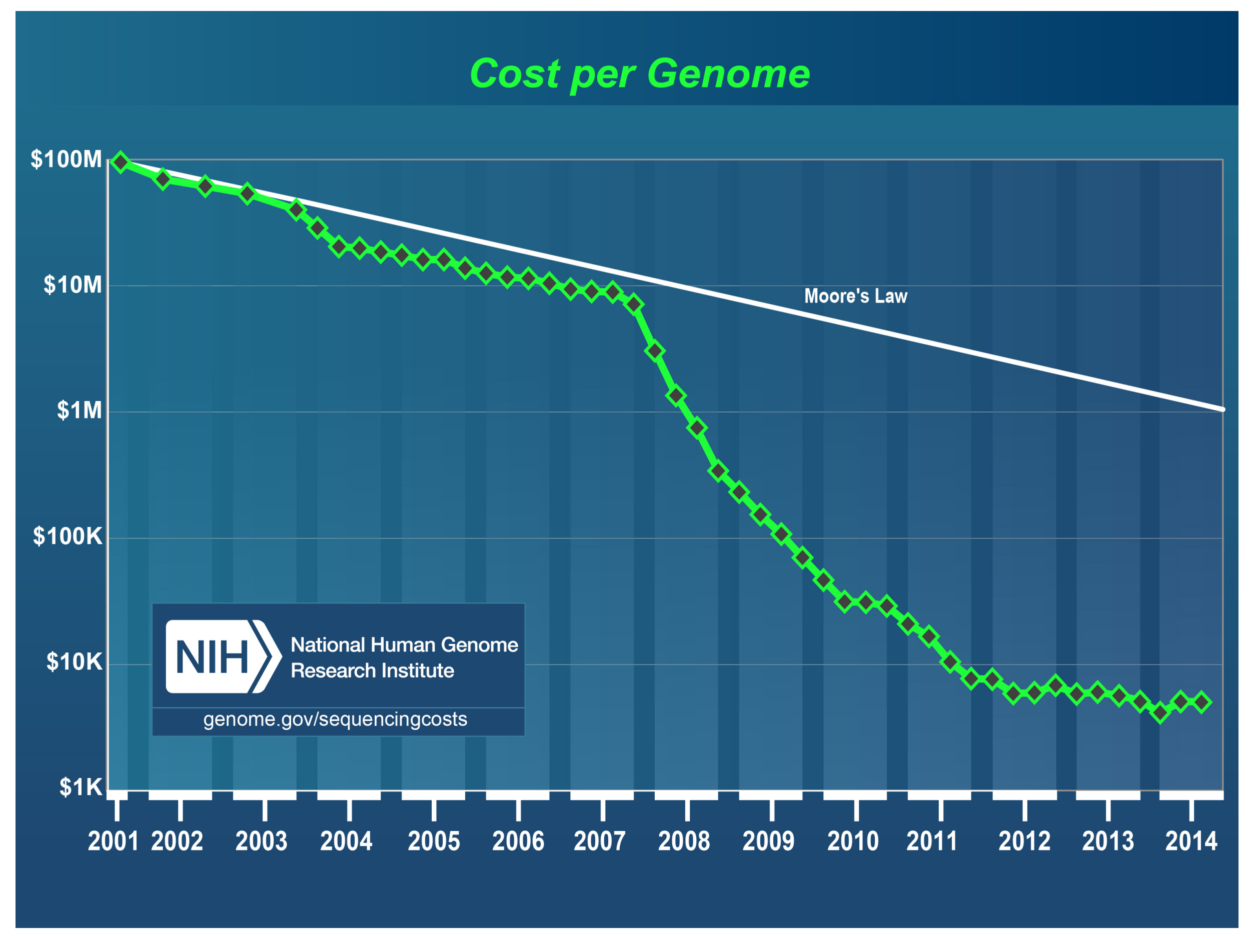
Sensing

Phenomena -> Sensor -> Making Sense of that Data

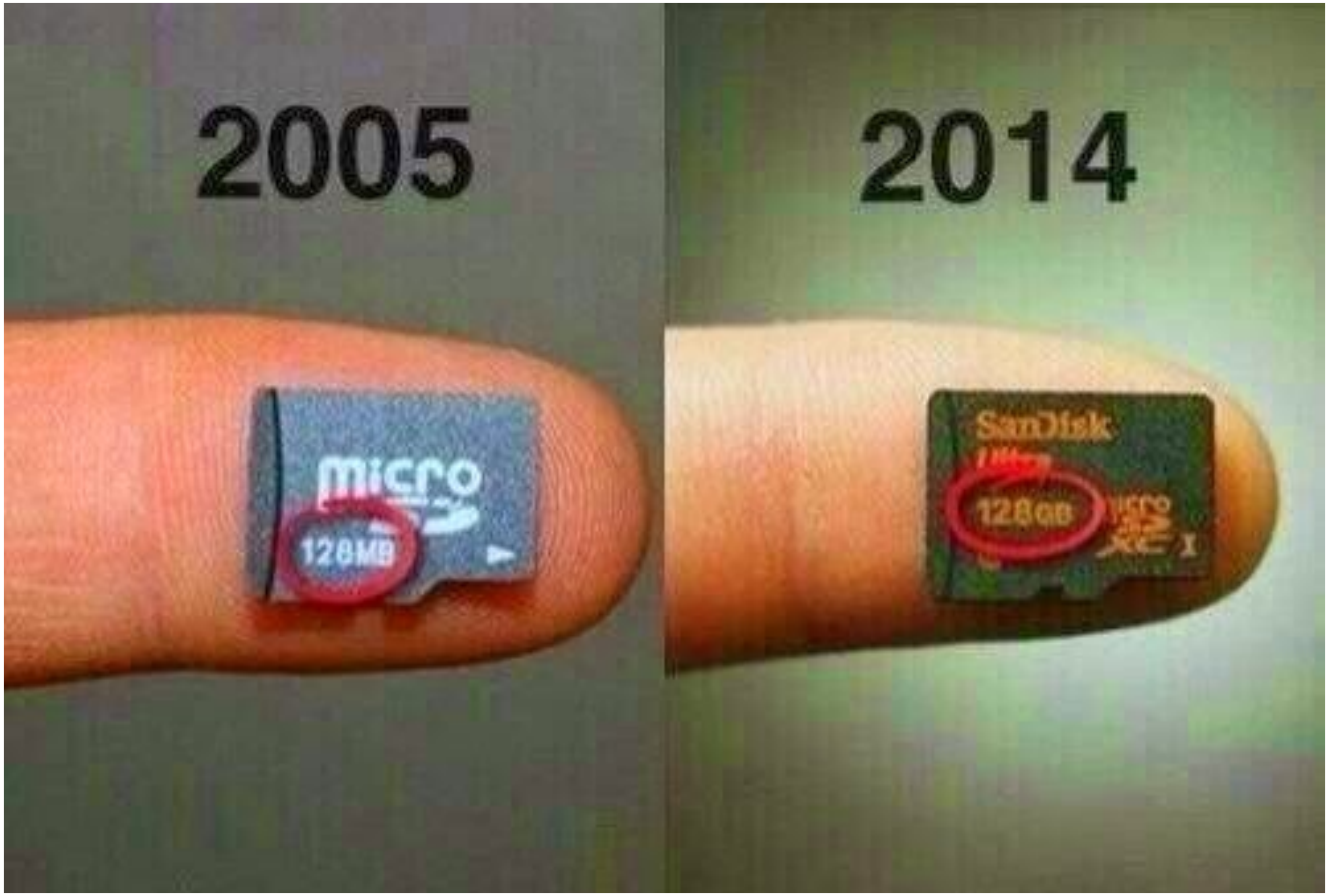
Phenomena -> Sensor -> Making Sense of that Data

Information rich and cheap sensors

- YouTube videos
    - 18/04/11. 35hrs uploaded per minute
    - 23/05/12. 60hrs uploaded per minute
    - 29/11/14. **100 hrs** uploaded **per minute**
- DNA sequencing cost
- Single cell sequencing
    - 2011: 1 cell
    - May 2012: 18 cells,
    - March 2015:~200,000 cells



Moore's law is not just for sensors



Algorithm-wise

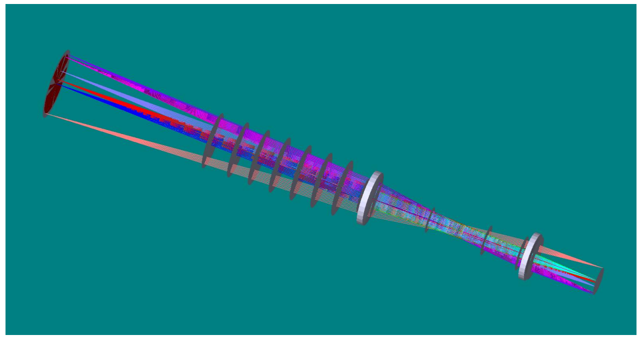
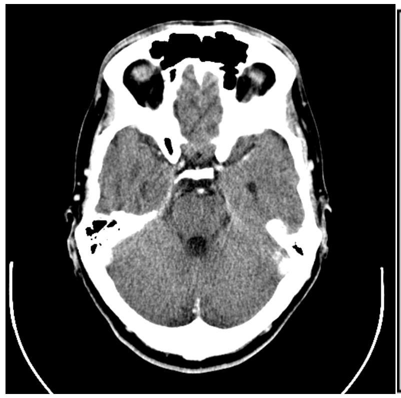
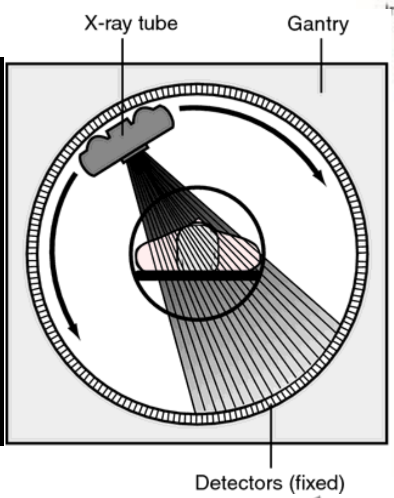
- Some problems used to be NP-Hard, relaxations have been found.

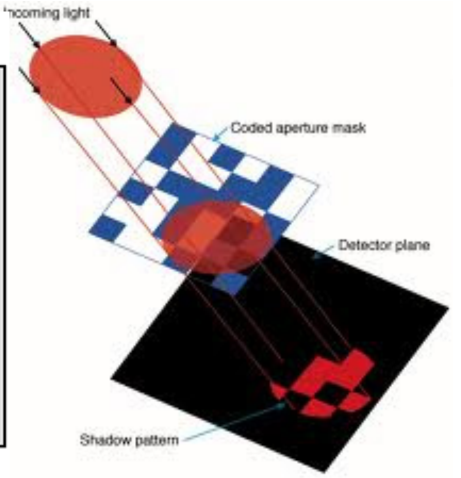
- Parallel to Moore's law, algorithms and sensors have changed the nature of the complexity of the problem

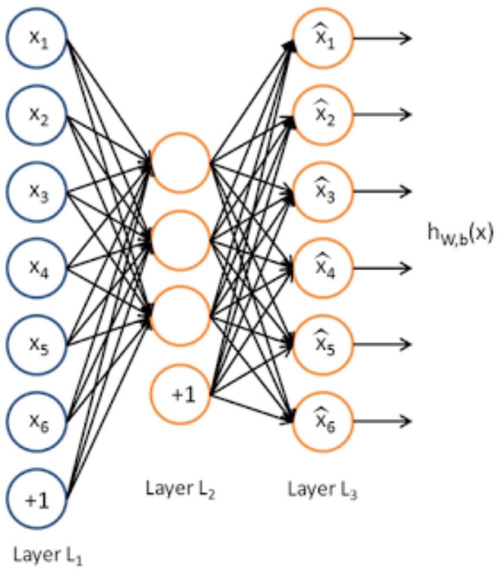
Phenomena -> Sensor -> Making Sense of that Data

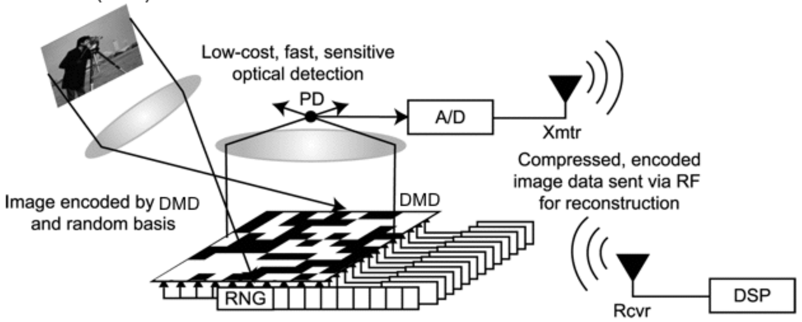
Sensing as the Identity

- x = I x for a perfect sensor
- x = (AB)x and AB=I, ex Camera
- x = L(Ax), ex. Coded aperture, CT
- x = **N**(Ax) or even x = **N**(A(Bx)), ex Compressive Sensing
- Hx = **N**(Ax), ex, classification in Compressive Sensing
- x = **N2**(**N1**(x)), ex autoencoders
- Hx = **N4**(**N3**(**N2**(**N1**(x)))), deep autoencoders

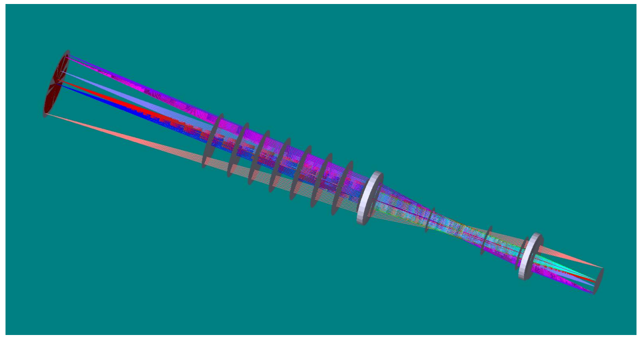


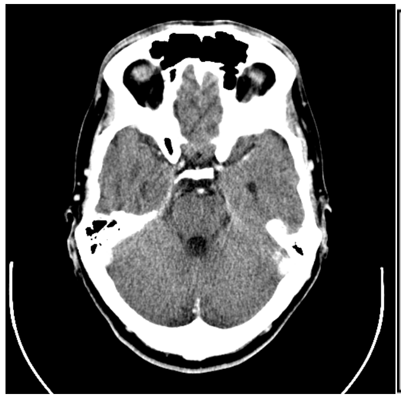


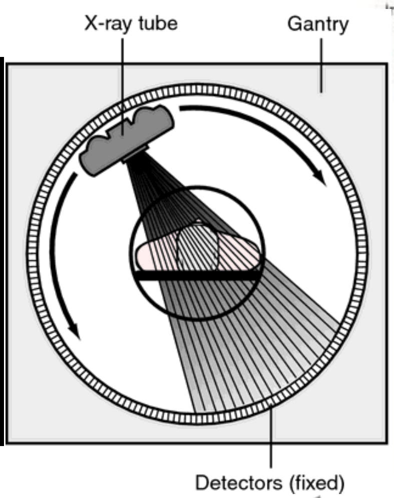


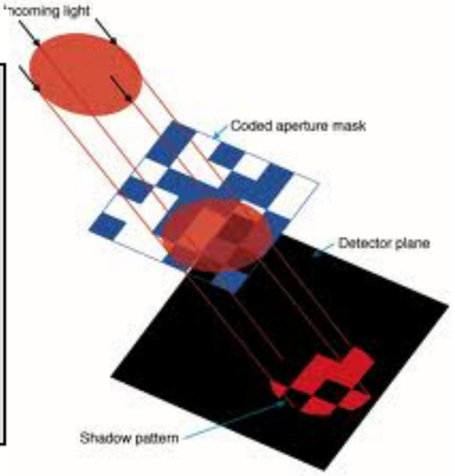

Sensing as the Identity

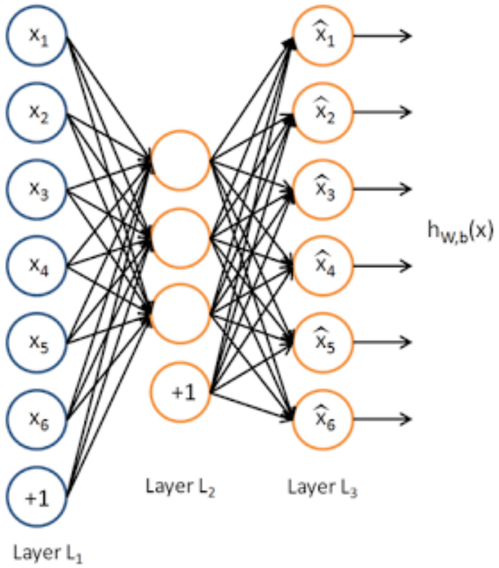
- $x = I\,x$ for a perfect sensor
- $x = (AB)x$ and $AB=I$, ex Camera
- $x = L(Ax)$, ex. Coded aperture, CT
- $x =$ **N**(Ax) or even $x =$ **N**$(A(Bx))$, ex Compressive Sensing
- $Hx =$ **N**(Ax), ex, classification in Compressive Sensing
- $x =$ **N2**(**N1**(x)), ex autoencoders
- $Hx =$ **N4**(**N3**(**N2**(**N1**(x)))), deep autoencoders

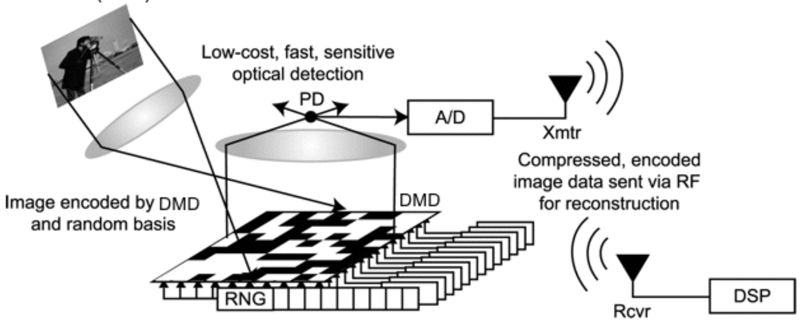




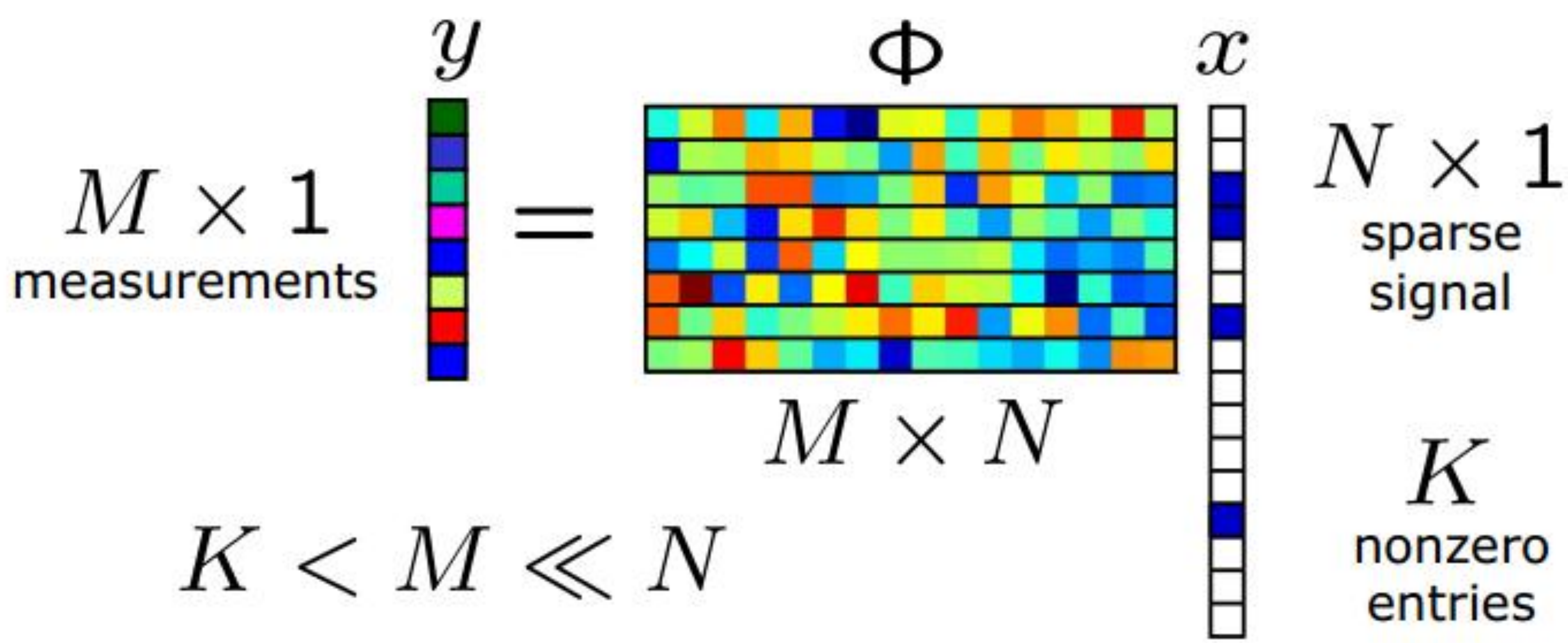


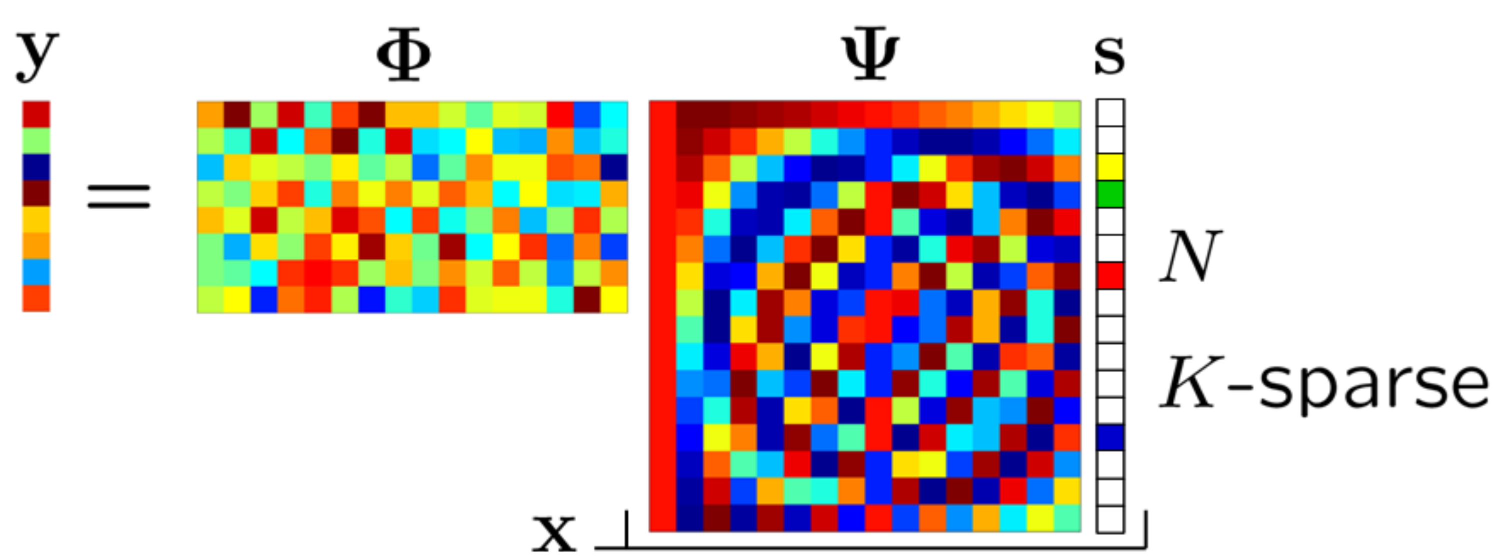




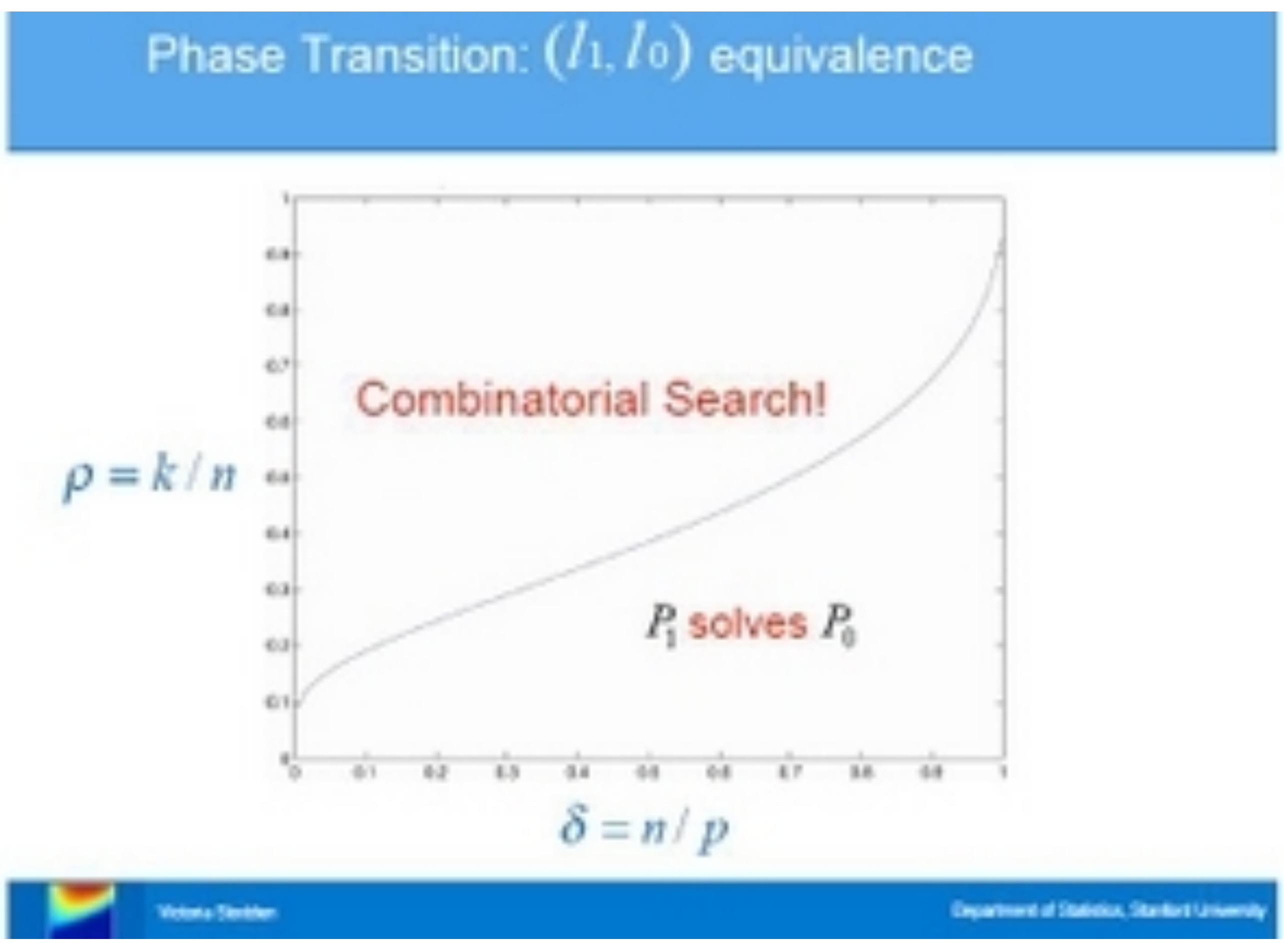


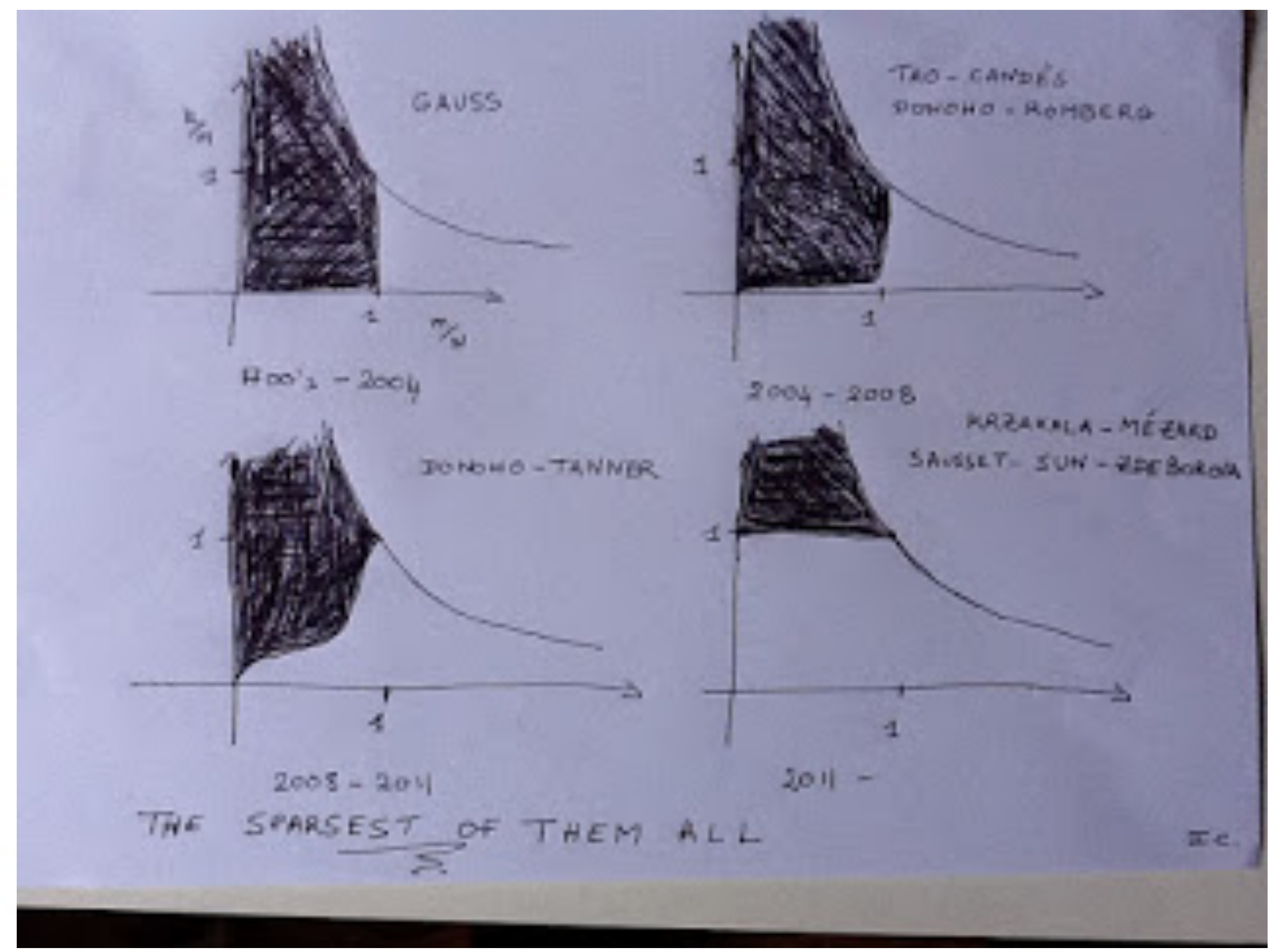
Compressive Sensing



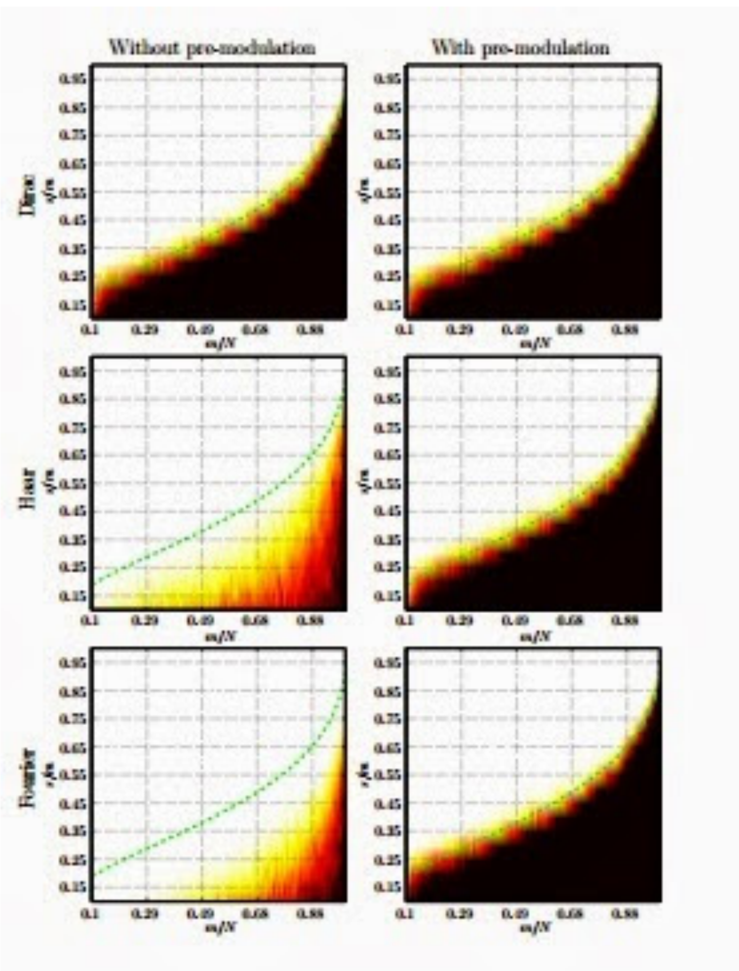


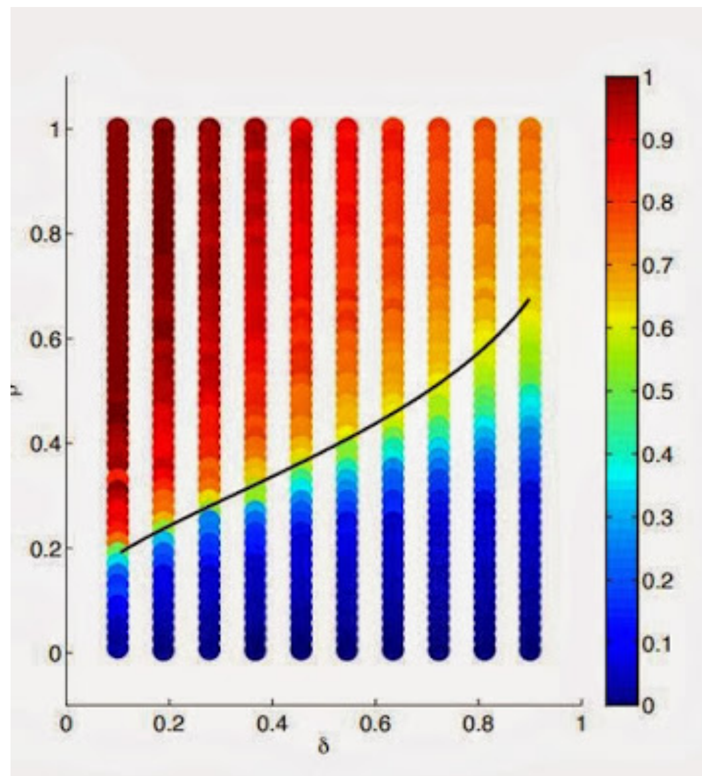
The relaxations and the bounds

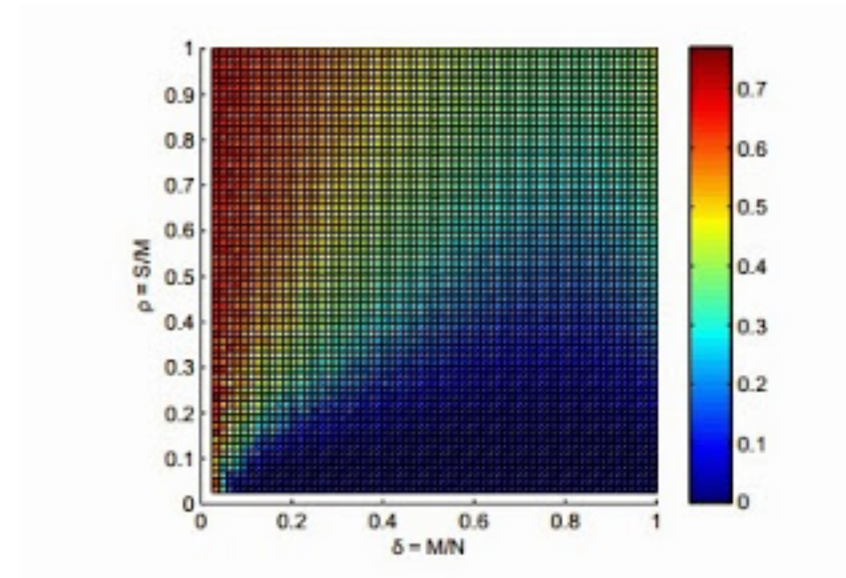




The bounds as sensor design limits







(a) SNR 10 dB

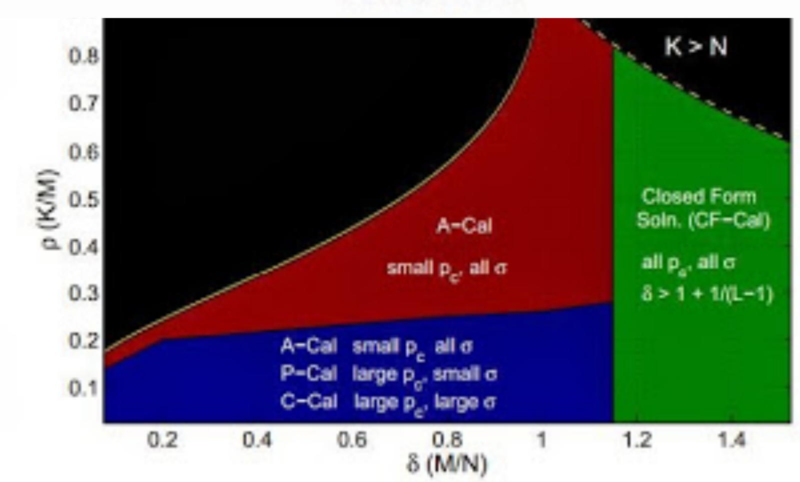


Fig. 7: A depiction of the algorithms to be used for different calibration scenarios. Note that for each case L must also be sufficiently high depending on σ and p_c. The solid yellow line indicates the Donoho-Tanner phase transition curve for fully calibrated compressed sensing recovery [20].

http://nuit-blanche.blogspot.fr/2013/11/sunday-morning-insight-map-makers.html

Convenience clouds the mind
ex: least squares

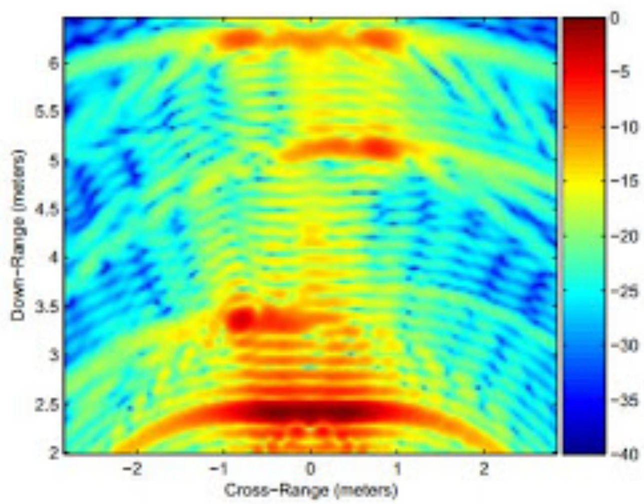


Fig. 9. Through-the-wall radar image using conventional backprojection with 100% data.

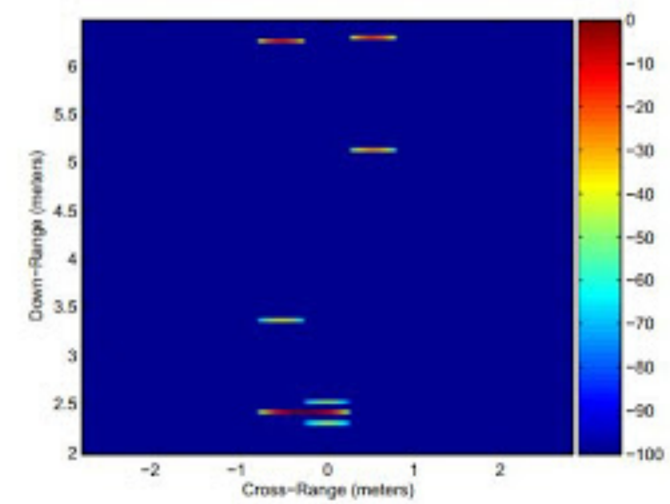
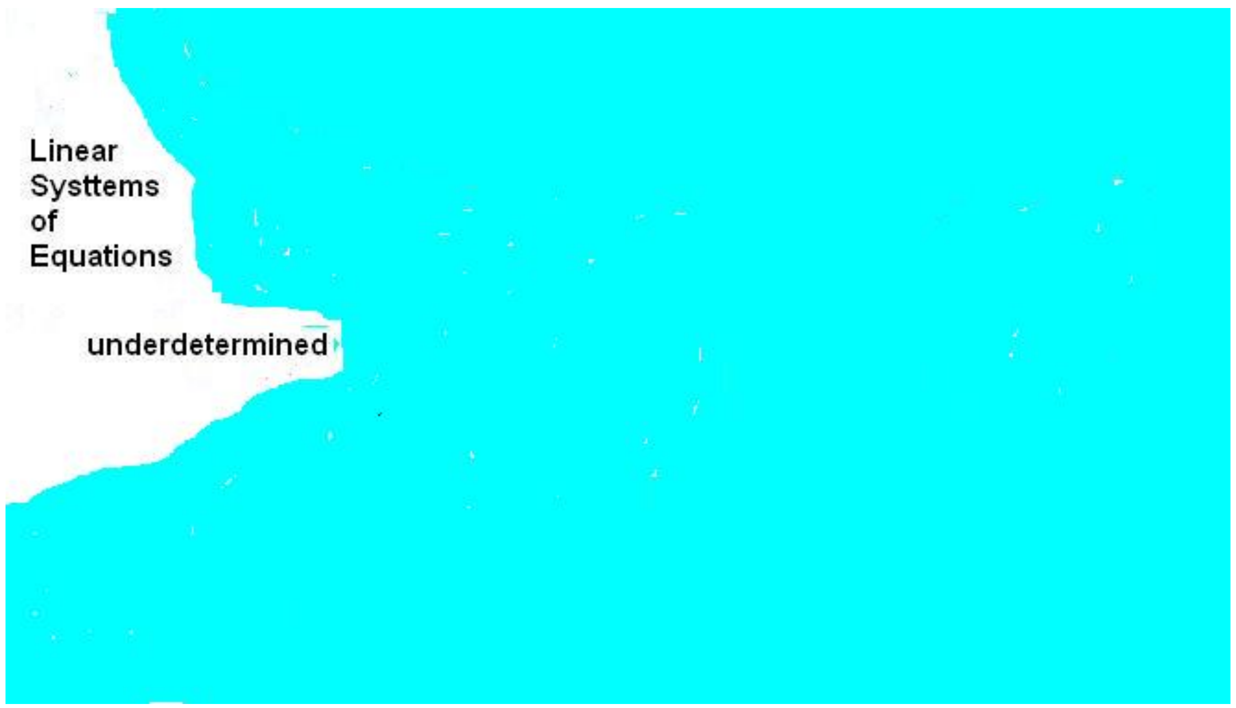
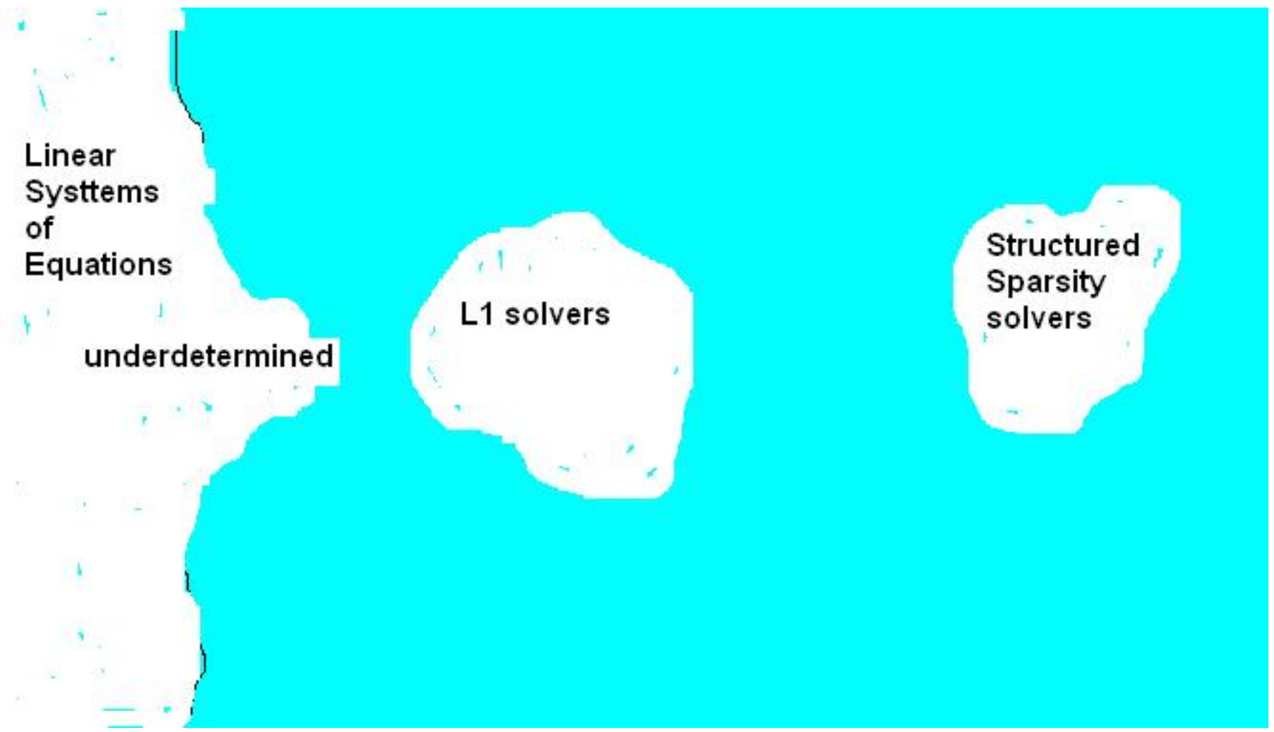
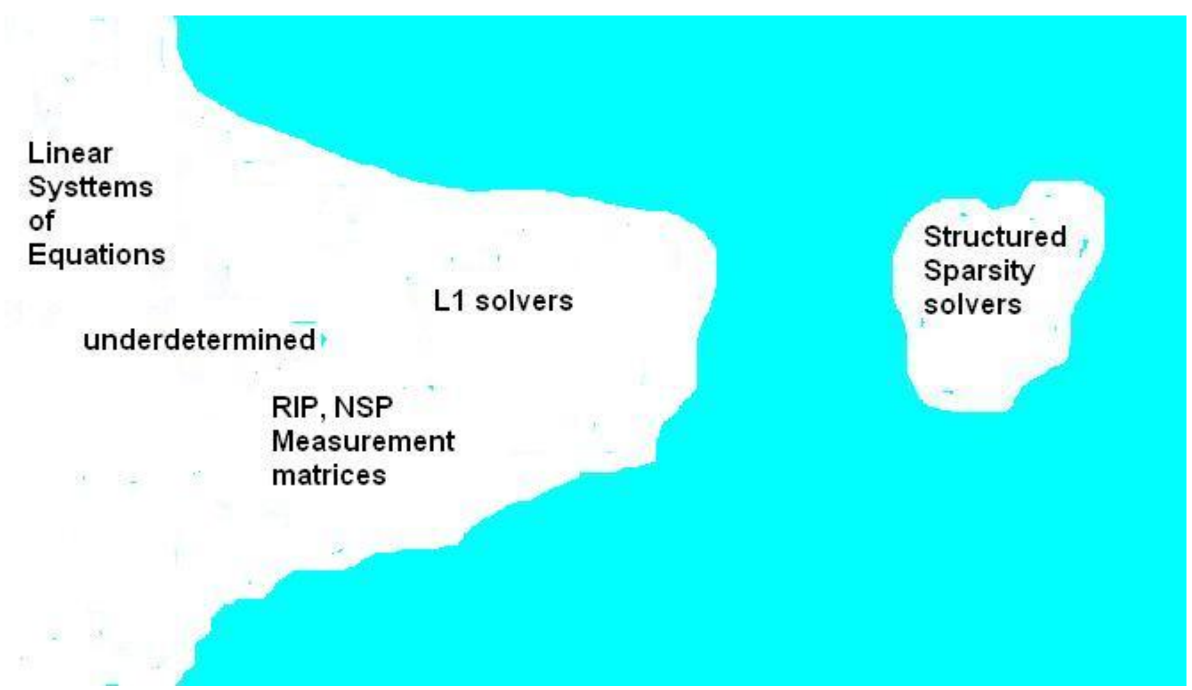


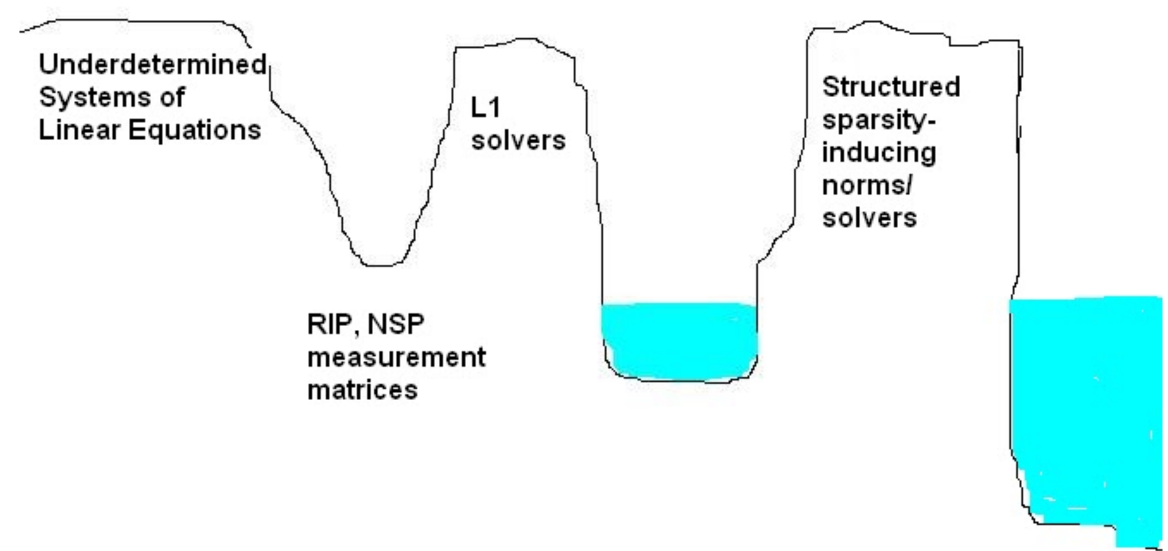
Fig. 11. Reconstructed image using BOMP and 6.4% data. The model used is $\breve{\mathbf{y}} = \mathbf{\Phi}\mathbf{\Psi}\mathbf{r}$.

Islands of knowledge









Islands of knowledge

Partial list of Contributions in Compressed Sensing

1991 Leahy & Jeffs : ℓ_p minimization $(0 < p < 1)$ for sparse beamforming array
1992 Rao et al. : iterative reweighted least squares for Magnetoencephalography (MEG)
1995 Rao et al. : FOCUSS – ℓ_p relaxation $(0 < p < 1)$ using reweighted least-squares with equality constraints
1996 J.J. Fuchs: Sensor array processing using ℓ_1.
1996 Feng & Bresler : spectrum blind sampling, guarantee of MUSIC for MMV, spark conditions
1996 Delaney & Bresler : iterative tomographic reconstruction from sparse samples with non-convex relaxation and reweighting algorithms
1997 Harikumar & Bresler : sparse solutions to linear inverse problems using convex and non-convex relaxations, and reweighting algorithms
1997 Feng : algebraic conditions for MMV
1998 Venkataramani & Bresler : algebraic conditions (spark) for uniqueness
1999 Bresler, Gastpar, & Venkataramani : "Image Compression on the Fly"
1999 Fuch et al. : ℓ_1 minimization
2000 Couvreur & Bresler : guarantee for backward greedy
2000 Gastpar & Bresler : information-theoretic analysis of compressive sensing with noise and with quantization

ℓ_1 regularization—known empirically

- Geology/geophysics
  - Claerbout and Muir (1973)
  - Taylor et al. (1979)
  - Levy and Fullager (1981)
  - Oldenburg et al. (1983)
  - Santosa and Symes (1988)
- Radio astronomy
  - Högbom (1974)
  - Schwarz (1978)
- Fourier transform spectroscopy
  - Kawata et al. (1983)
  - Mammone (1983)
  - Minami et al. (1985)
- NMR spectroscopy
  - Barkhuijsen (1985)
  - Newman (1988)
- Medical ultrasound
  - Papoulis and Chamzas (1979)

Common elements:

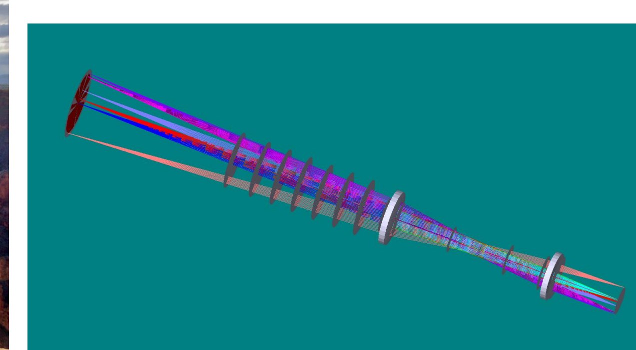
- Sparsity assumption enables improved resolution of estimate (beyond bandwidth of acquisition)
- Sparsity in space or gradient with respect to space
- ℓ_1 minimization to promote sparsity
- Sparse domain given by nature

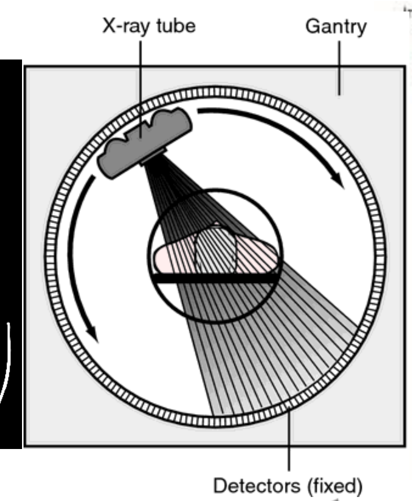
Mathematical limits explored in [Donoho (1992)]

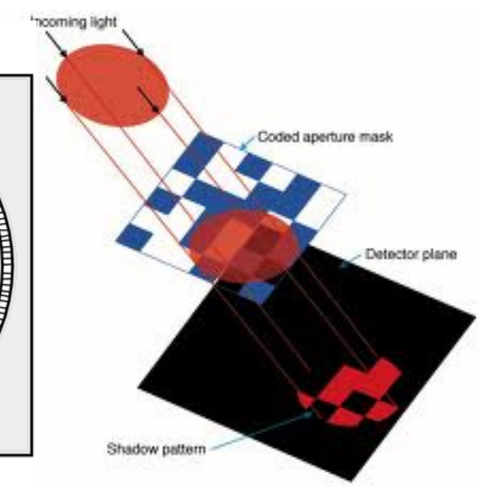
Beyond Compressive Sensing

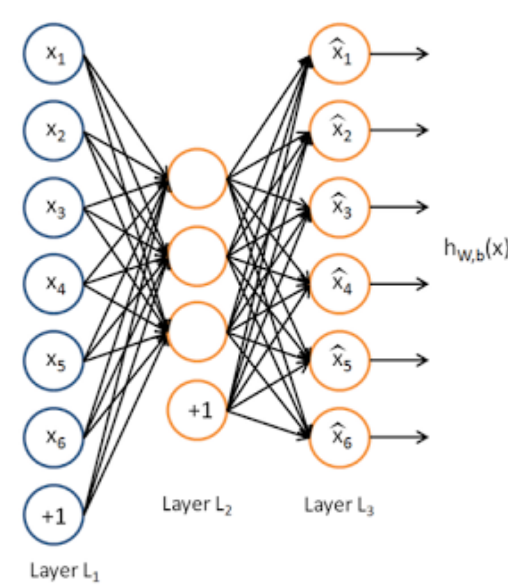
Sensing as the Identity

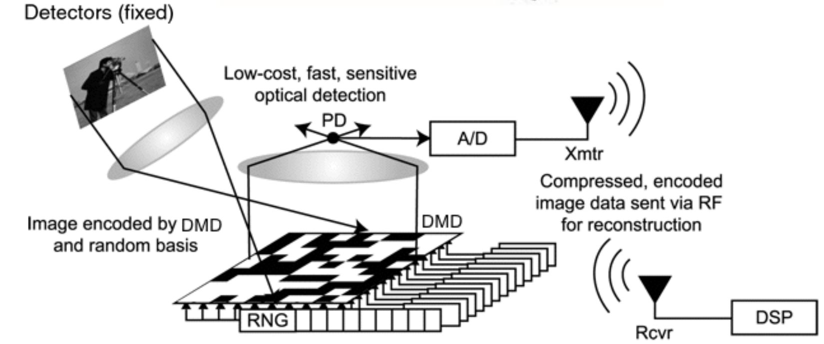
- x = I x for a perfect sensor
- x = (AB)x and AB=I, ex Camera
- x = L(Ax), ex. Coded aperture, CT
- x = **N**(Ax) or even x = **N**(A(Bx)), ex Compressive Sensing
- Hx = **N**(Ax), ex, classification in Compressive Sensing
- x = **N2**(**N1**(x)), ex autoencoders
- Hx = **N4**(**N3**(**N2**(**N1**(x)))), deep autoencoders









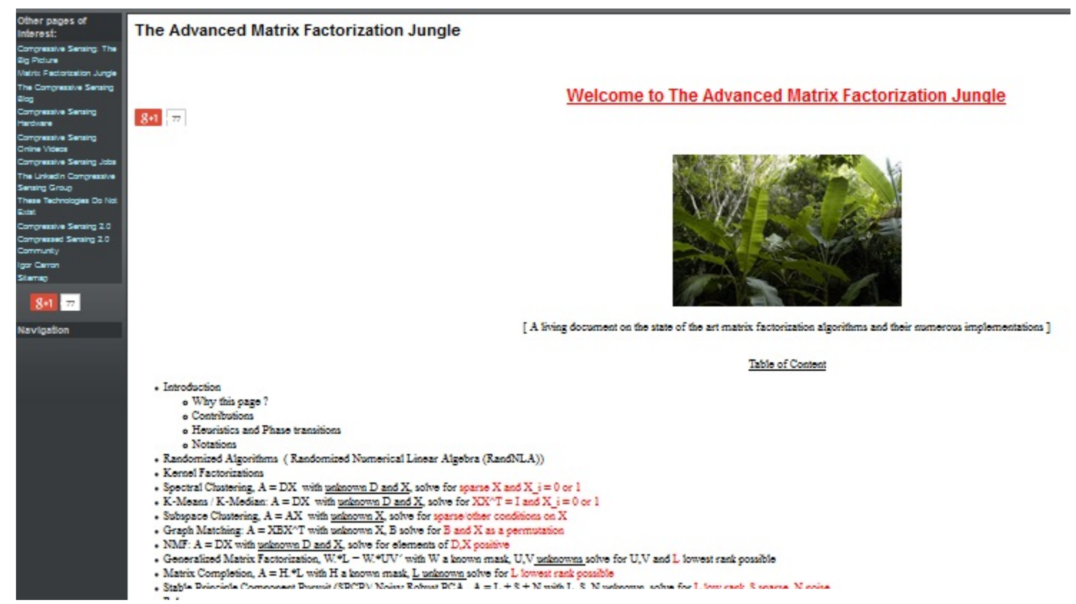


Advanced Matrix Factorizations

- Also Linear Autoencoders:
- A = B C s.t B or C or B and C have specific features
- Examples: NMF, SVD, Clustering, ...
- Use: hyperspectral unmixing,....

Advanced Matrix Factorizations

- Spectral Clustering, A = DX with unknown D and X, solve for sparse X and X_i = 0 or 1
- K-Means / K-Median: A = DX with unknown D and X, solve for XX^T = I and X_i = 0 or 1
- Subspace Clustering, A = AX with unknown X, solve for sparse/other conditions on X
- Graph Matching: A = XBX^T with unknown X, B solve for B and X as a permutation
- NMF: A = DX with unknown D and X, solve for elements of D,X positive
- Generalized Matrix Factorization, W.*L – W.*UV′ with W a known mask, U,V unknowns solve for U,V and L lowest rank possible
- Matrix Completion, A = H.*L with H a known mask, L unknown solve for L lowest rank possible
- Stable Principle Component Pursuit (SPCP)/ Noisy Robust PCA, A = L + S + N with L, S, N unknown, solve for L low rank, S sparse, N noise
- Robust PCA : A = L + S with L, S unknown, solve for L low rank, S sparse
- Sparse PCA: A = DX with unknown D and X, solve for sparse D
- Dictionary Learning: A = DX with unknown D and X, solve for spars



Method	$\mathcal{D}(\mathbf{X};\mathbf{BC})$	$\mathcal{B}$	$\mathcal{C}$
SVD	$\|\mathbf{X}-\mathbf{BC}\|_F^2$	$\mathbf{B}^T\mathbf{B}=\mathbf{I}$	$\mathbf{CC}^T=\Lambda$
k-means	$\|\mathbf{X}-\mathbf{BC}\|_F^2$	-	$\mathbf{CC}^T=\mathbf{I}$, $c_{ij}=\{0,1\}$
k-medians	$\|\mathbf{X}-\mathbf{BC}\|_1$	-	$\mathbf{CC}^T=\mathbf{I}$, $c_{ij}=\{0,1\}$
pLSI [27]	$KL(\mathbf{X};\mathbf{BC})$	$\mathbf{1}^T\mathbf{B}\mathbf{1}=1$, $b_{ij}\geq 0$	$\mathbf{1}^T\mathbf{C}=\mathbf{1}$, $c_{ij}\geq 0$
NMF [23]	$KL(\mathbf{X};\mathbf{BC})$	$b_{ij}\geq 0$	$c_{ij}\geq 0$

Table 1. Criteria which satisfy the EBD conditions (1)(2)(3).

	$f(Z)$	Ω
SSC	$\|Z\|_0$ or $\|Z\|_1$	$\{Z\|X=XZ,\ \mathrm{diag}(Z)=0\}$
LRR	$\|Z\|_*$	$\{Z\|X=XZ\}$
SSQP	$\|Z^TZ\|_1$	$\{Z\|X=XZ, Z\geq 0, \mathrm{diag}(Z)=0\}$
MSR	$\|Z\|_1+\delta\|Z\|_*$	$\{Z\|X=XZ, \mathrm{diag}(Z)=0\}$
Other choices	$(\sum_{i=1}^n\sum_{j=1}^n \lambda_{ij}\|Z_{ij}\|^{p_{ij}})^s$ $\lambda_{ij}>0, p_{ij}>0, s>0$	$\{Z\|X=XZ, \mathrm{diag}(Z)=0\}$

Bounds on Advanced Matrix Factorizations

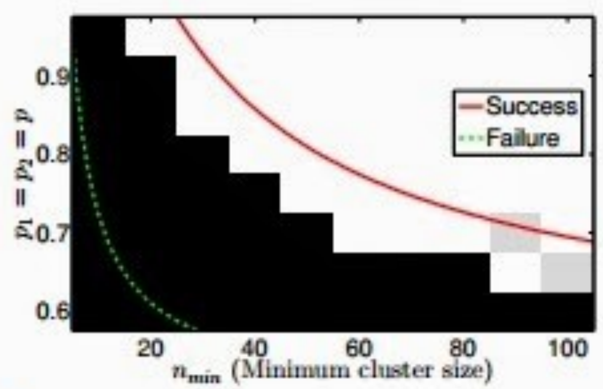


Fig. 2: Simulation results showing the region of success (white region) and failure (black region) of Program 1 with $\lambda = 0.99\Lambda_{succ}$. Also depicted are the thresholds for success (solid red curve on the top-right) and failure (dashed green curve on the bottom-left) predicted by Theorem 1.

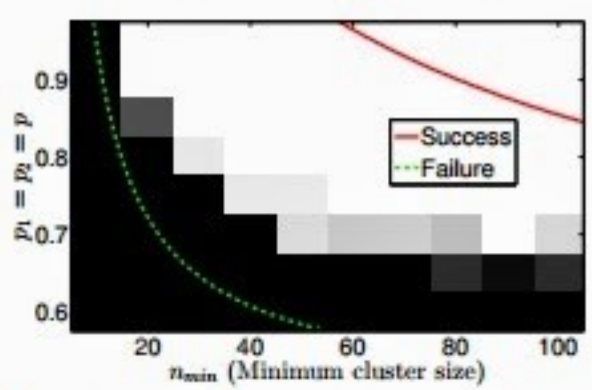


Fig. 3: Simulation results showing the region of success (white region) and failure (black region) of Program 1 with $\lambda = 2ED_{min}^{-1}$. Also depicted are the thresholds for success (solid red curve on the top-right) and failure (dashed green curve on the bottom-left) predicted by Theorem 2.

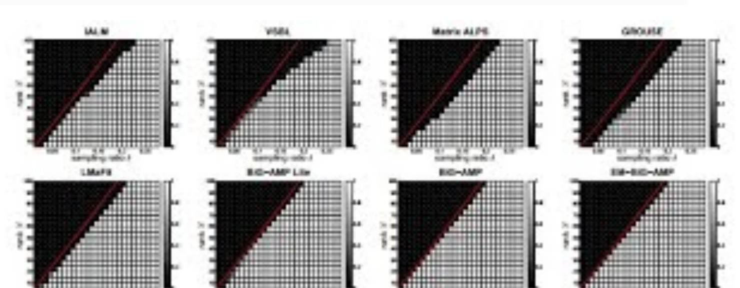


Fig. 2. Empirical success rates for noiseless completion of an $M \times L$ matrix sampled uniformly at random, as a function of sampling ratio $\delta = \frac{|\Omega|}{ML}$ and rank N. Here, "success" is defined as NMSE < -100 dB, success rates were computed from 10 random realizations, and $M = L = 1000$. Points above the red curve are infeasible, as described in the text.

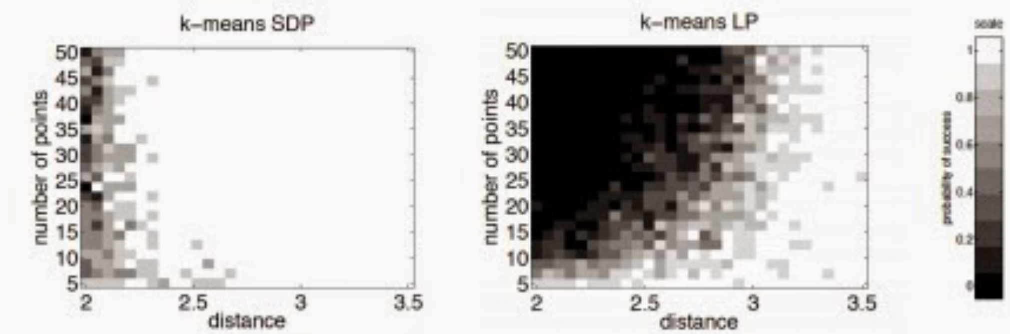


FIGURE 1. Empirical probability of integrality of convex relaxation-based clustering. Lighter color corresponds to higher probability of success. We consider 2 clusters in $\mathbb{R}^3$, $4 \leq N \leq 50$, $2 \leq \Delta \leq 3.5$. The k-median LP always provided an integral solution. These numerical results suggest superiority of the k-median LP vs k-means SDP, and of k-means SDP with respect to k-means LP, when $k = 2$.

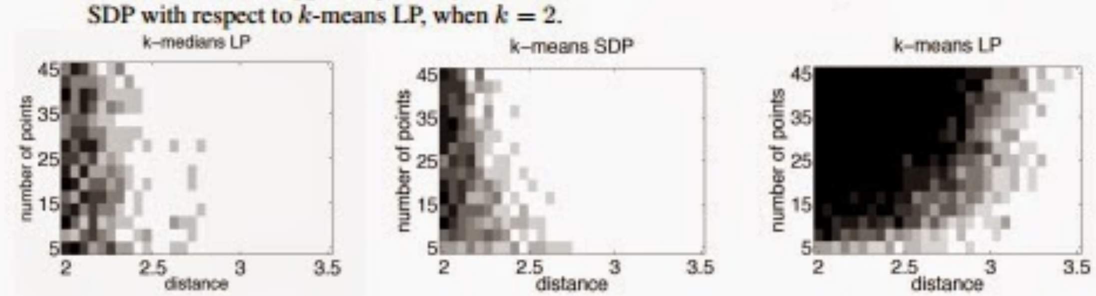


FIGURE 2. When we consider 3 clusters in $\mathbb{R}^3$, $6 \leq N \leq 42$, $2 \leq \Delta \leq 3.5$, the k-median and k-means SDP show a very similar behavior. These numerical results suggest the performance of the k-median LP degrades with k.

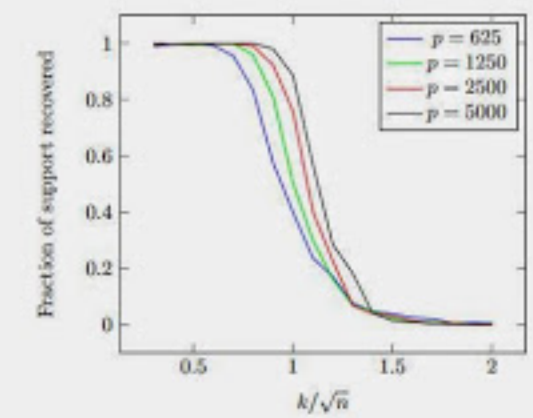


Figure 2: The support recovery phase transition for the 'data-driven' algorithm described in Section 3, which is agnostic of the support size k. This indicates that the 'data-driven' algorithm also works when $k \leq c\sqrt{n}$ independent of p. Note that these results are very similar to the one of the Covariance Thresholding algorithm of Figure 1. The problem does not appear to be significantly harder when k is unknown.

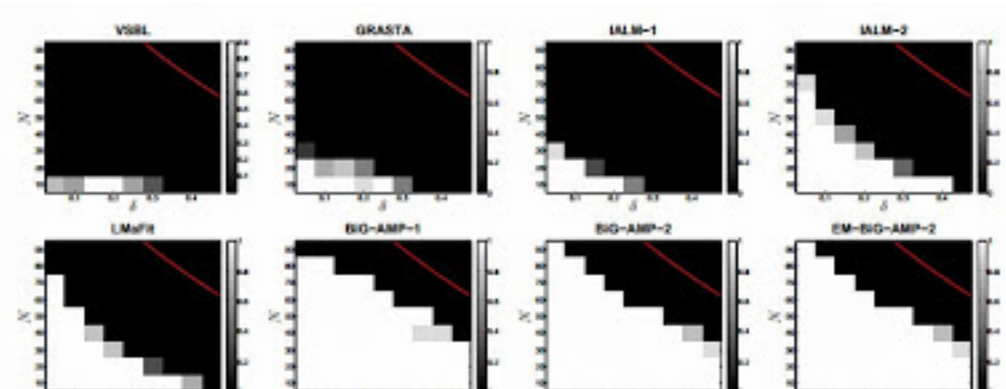


Fig. 8. Empirical success rates for RPCA with a 200×200 matrix of rank N corrupted by a fraction δ of outliers with amplitudes uniformly distributed on $[-10, 10]$. Here, "success" is defined as NMSE < -80 dB, and success rates were averaged over 10 problem realizations. Points above the red curve are infeasible, as described in the text.

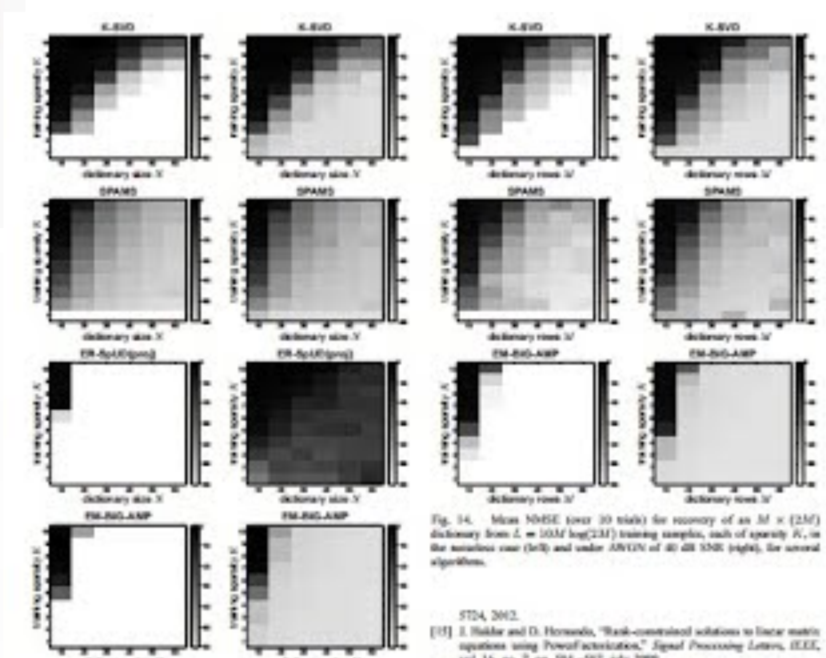


Fig. 12. Mean NMSE (over 10 trials) for recovery of an $N \times N$ dictionary from $L = 5N \log N$ training samples, each of sparsity K, in the noiseless case (left) and under AWGN of 40 dB SNR (right), for several algorithms.

Fig. 14. Mean NMSE (over 10 trials) for recovery of an $M \times (2M)$ dictionary from $L = 10M \log(2M)$ training samples, each of sparsity K, in the noiseless case (left) and under AWGN of 40 dB SNR (right), for several algorithms.

5724, 2012.

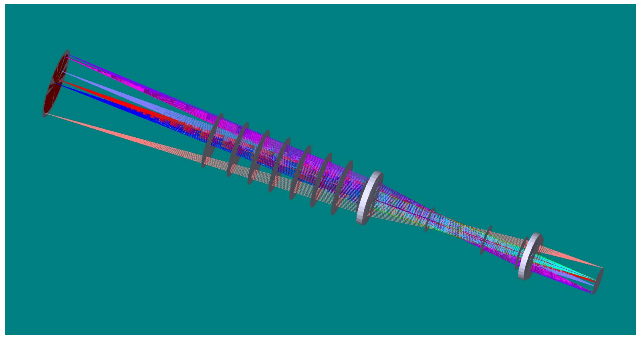
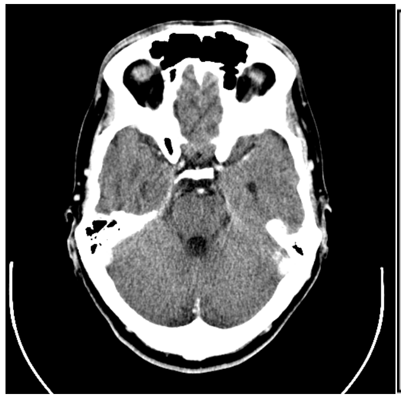
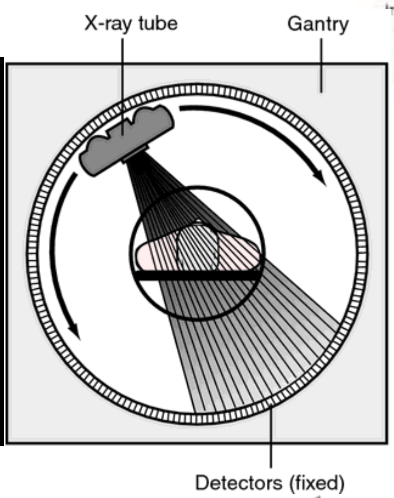
[15] J. Haldar and D. Hernando, "Rank-constrained solutions to linear matrix equations using PowerFactorization," Signal Processing Letters, IEEE, vol. 16, no. 7, pp. 584–587, july 2009.

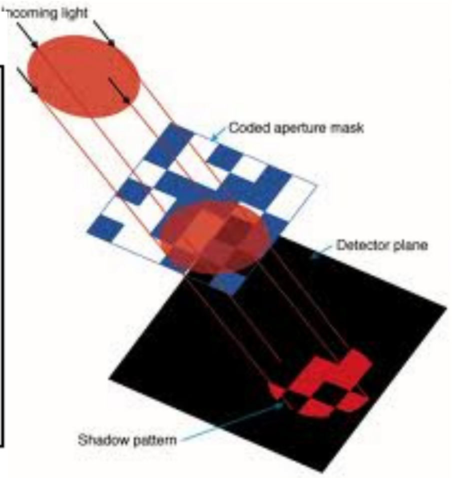
[16] L. Balzano, R. Nowak, and B. Recht, "Online identification and tracking of subspaces from highly incomplete information," Arxiv preprint arXiv:1006.4046, 2010.

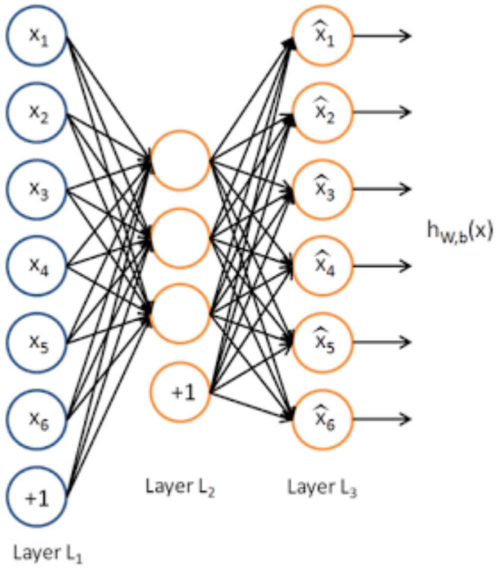
[17] R. Keshavan, A. Montanari, and S. Oh, "Matrix completion from a few entries," Information Theory, IEEE Transactions on, vol. 56, no. 6, pp. 2980–2998, june 2010.

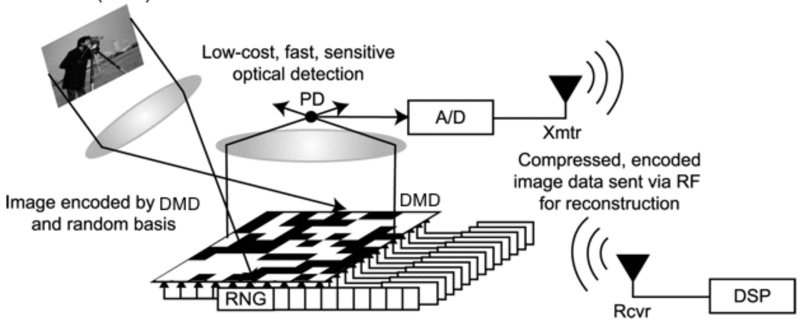
Sensing as the Identity

- x = I x for a perfect sensor
- x = (AB)x and AB=I, ex Camera
- x = L(Ax), ex. Coded aperture, CT
- x = **N**(Ax) or even x = **N**(A(Bx)), ex Compressive Sensing
- Hx = **N**(Ax), ex, classification in Compressive Sensing
- x = **N2**(**N1**(x)), ex autoencoders
- Hx = **N4**(**N3**(**N2**(**N1**(x)))), deep autoencoders and more

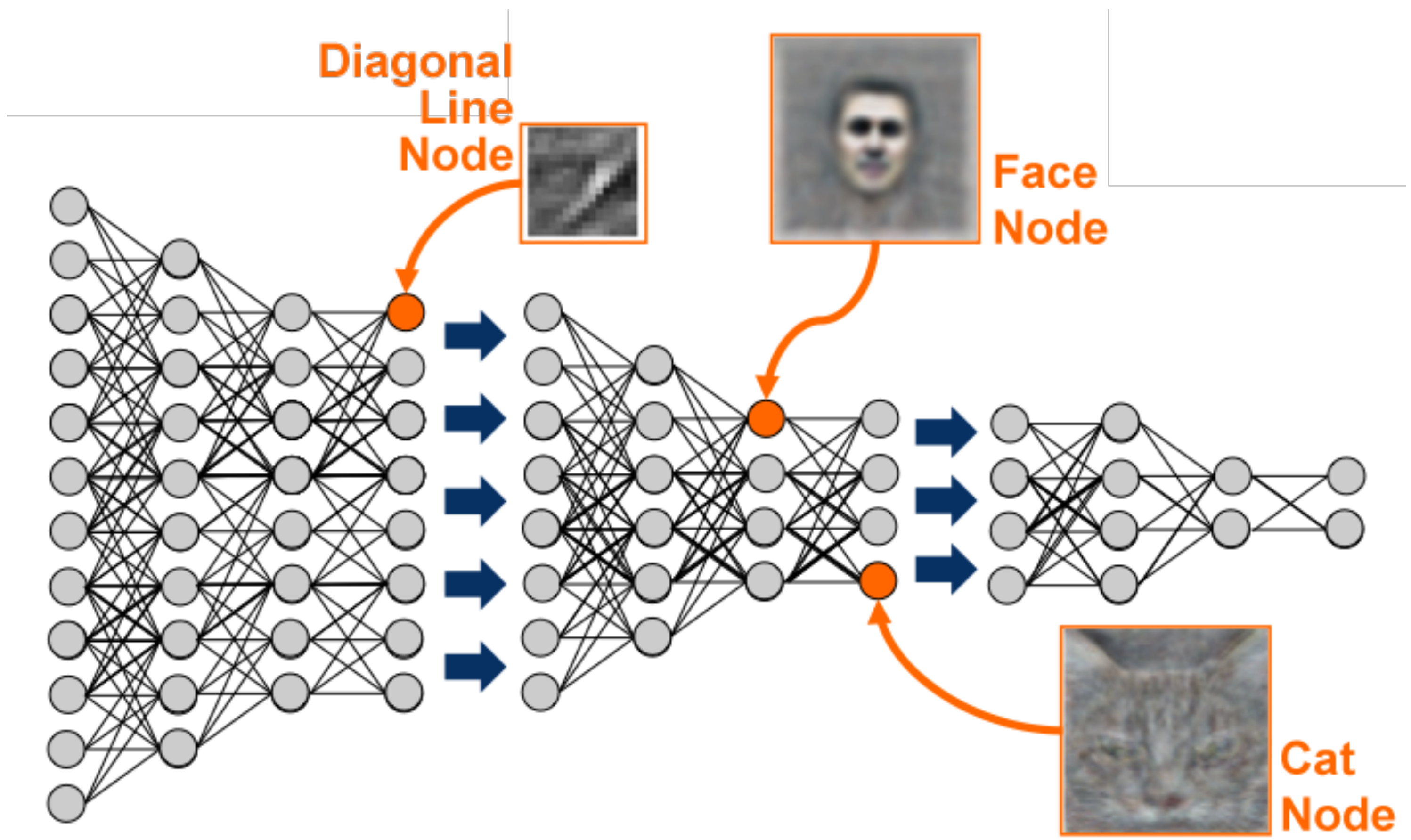





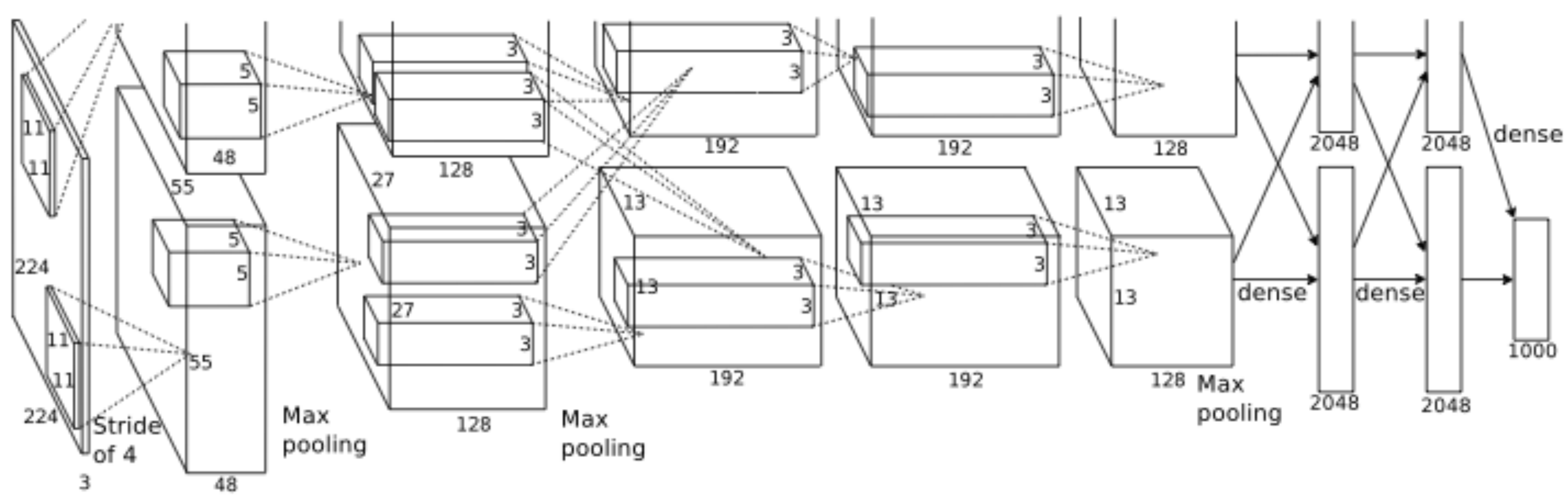







Machine Learning / Deep Neural Networks





Bounds and Limits DNNs

- Currently unknown.
- DNNs could even be complicated regularization schemes of simpler approach (but we have not found which)

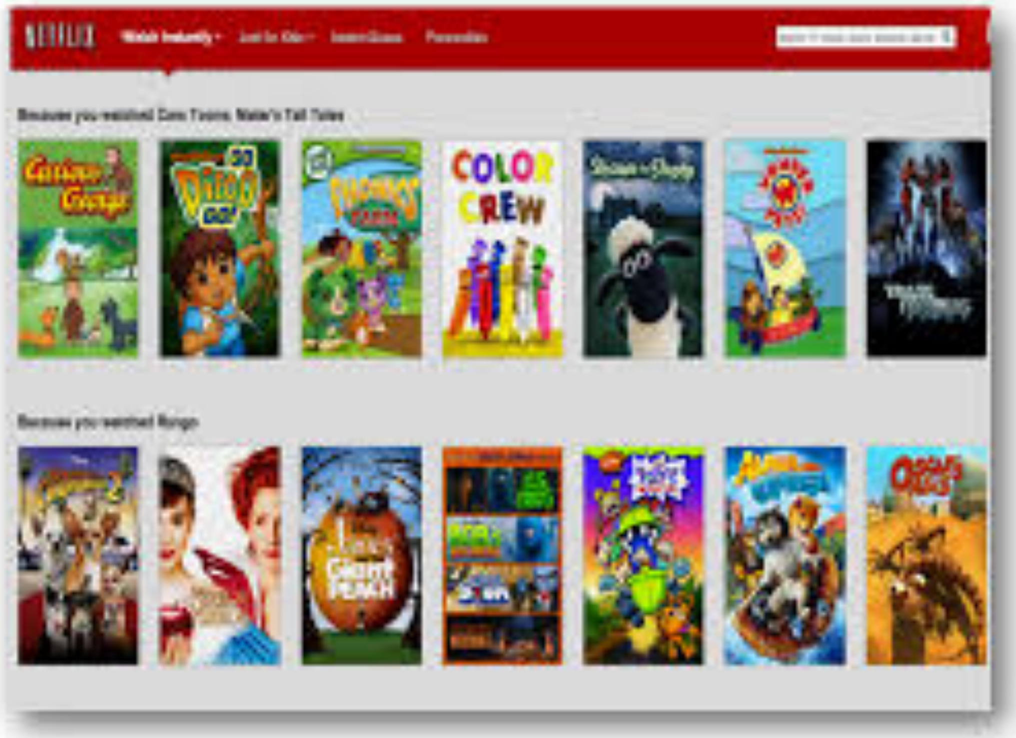
The Great Convergence ?

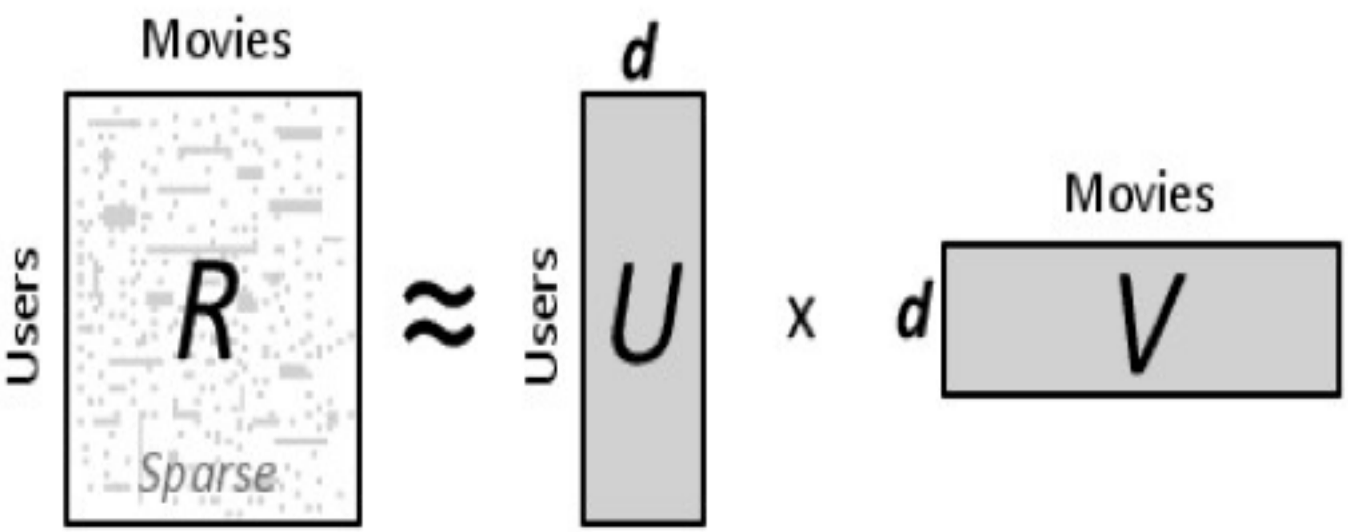
- Recent use of Deep Neural Networks structure to perform MRI reconstruction, Error Correcting Coding, Blind Source Separation.....

Two more words

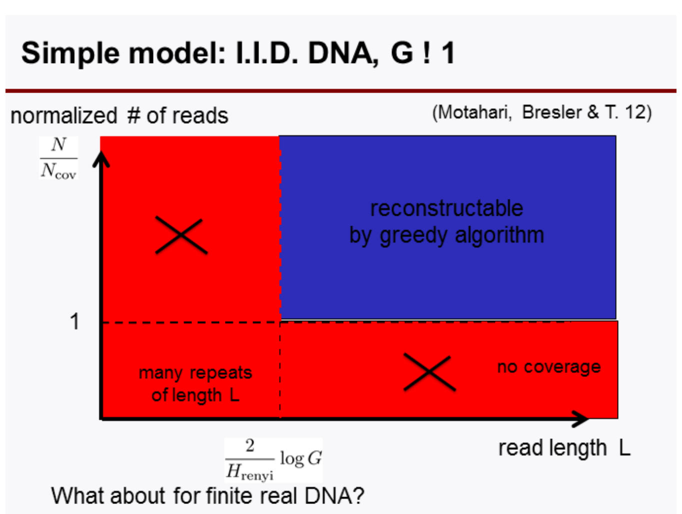
Advanced Matrix Factorization

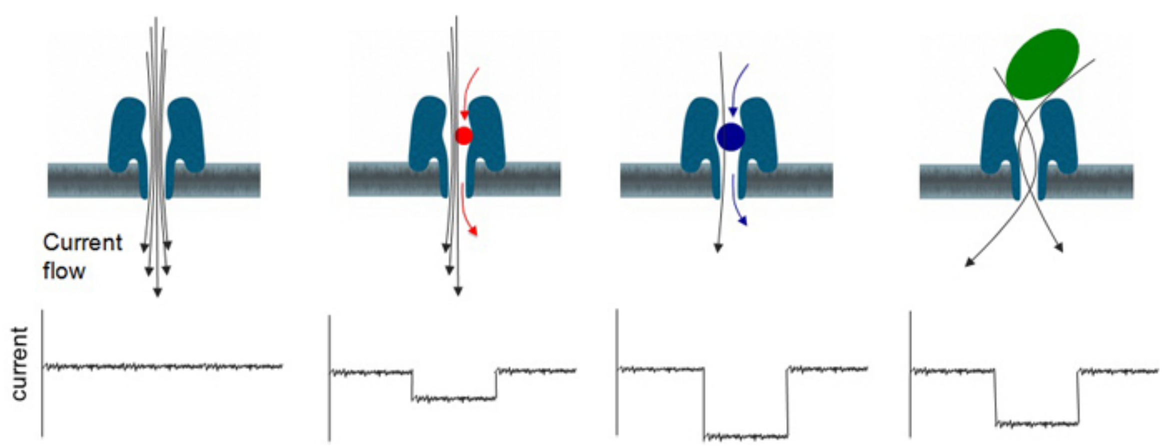
- Recommender systems

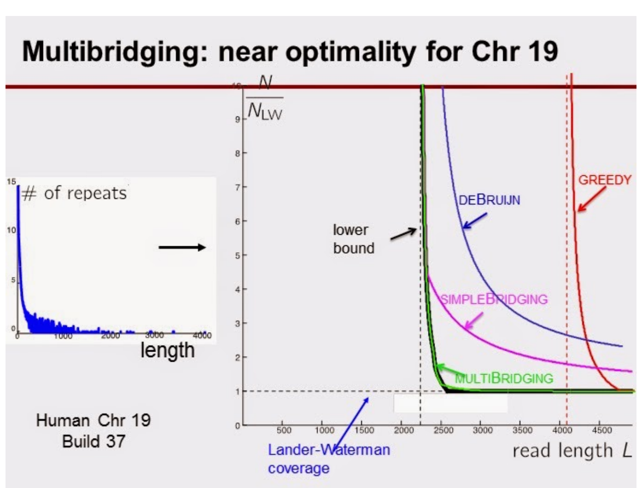




What happens when the sensor makes the problem not NP-hard anymore ?







More infos

- http://nuit-blanche.blogspot.com
- Paris Machine Learning meetup, http://nuit-blanche.blogspot.com/p/paris-based-meetups-on-machine-learning.html