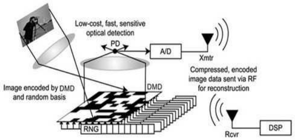
Fast Reconstruction Algorithms for Deterministic Sensing Matrices and Applications

Robert Calderbank *et al.*

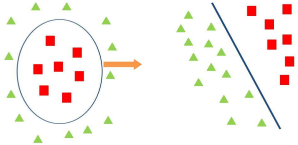
Program in Applied and Computational Mathematics
Princeton University
NJ 08544, USA.

- When sample by sample measurement is expensive and redundant:

- Compressive Sensing:
  - Transform to low dimensional measurement domain



- Machine Learning:
  - Filtering in the measurement domain



Take-Home Message

Compressed Sensing is a **Credit Card!**

We want one with no hidden charges

Geometry of Sparse Reconstruction

- **Restricted Isometry Property (RIP)**: An $N \times C$ matrix A satisfies (k, ϵ)-RIP if for any k-sparse signal $\boldsymbol{x}$:

$$(1-\epsilon)\|\boldsymbol{x}\|_2 \leq \|A\boldsymbol{x}\|_2 \leq (1+\epsilon)\|x\|_2.$$

- **Theorem [Candes,Tao2006]:**
  If the entries of $\sqrt{N}A$ are iid sampled from
  - N$(0,1)$ Gaussian
  - **U**$(-1,1)$ Bernoulli

  distribution, and $N = \Omega\left(k \log(\frac{C}{k})\right)$, then with probability $1 - e^{-cN}$, A has (k, ϵ)-RIP.

- **Reconstruction Algorithm [Candes,Tao 2006 and Donoho 2006]:**
  If A satisfies $(3k, \epsilon)$-RIP for $\epsilon \leq 0.4$, then given any k-sparse solution $\boldsymbol{x}$ to $A\boldsymbol{x} = \boldsymbol{b}$, the linear program

$$\text{minimize } \|\boldsymbol{z}\|_1 \text{ such that } A\boldsymbol{z} = \boldsymbol{b}$$

  recovers $\boldsymbol{x}$ successfully, and is robust to noise.

A: Adjacency matrix of a $(2k, \epsilon)$ expander graph

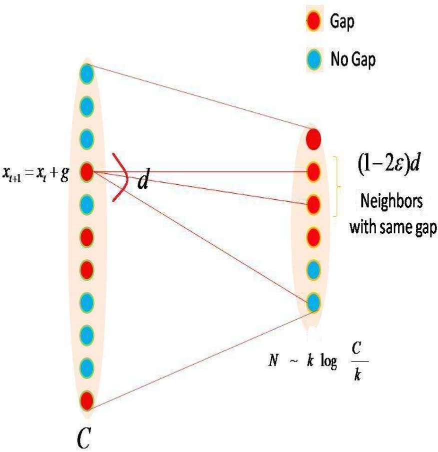
- No $2k$-sparse vector in the null space of A

Theorem [Jafarpour, Xu, Hassibi, Calderbank 2008]: If $\epsilon \leq 1/4$, then for any k-sparse solution $\boldsymbol{x}$ to $A\boldsymbol{x} = \boldsymbol{b}$, the solution can be recovered successfully in at most $2k$ rounds.

Gap: $\boldsymbol{g}_t = \boldsymbol{b} - A\boldsymbol{x}_t$

RHS proxy for difference between $\boldsymbol{x}_t$ and $\boldsymbol{x}$.

ALGORITHM. Greedy reduction of gap.



- Let A: adjacency matrix of an expander graph
- x^*: sparse
- Noisy compressed sensing measurements y in Poisson model
- $\hat{x} = \arg\min \sum_{j=1}^{N} ((Ax)_j - y_j \log(Ax)_j) + \gamma pen(x)$
- Optimization over the simplex (positive values)
- pen: a well chosen penalty function.
- Then $\hat{x} \approx x^*$

k-Sparse Reconstruction with Random Sensing Matrices

Approach	Measurements N	Complexity	Noise Resilience	RIP
Basis Pursuit (BP) [CRT]	$k \log\left(\frac{C}{k}\right)$	C^3	Yes	Yes
Orthogonal Matching Pursuit (OMP) [GSTV]	$k \log^{\alpha}(C)$	$k^2 \log^{\alpha}(C)$	No	Yes
Group Testing [CM]	$k \log^{\alpha}(C)$	$k \log^{\alpha}(C)$	No	No
Greedy Expander Recovery[JXHC]	$k \log\left(\frac{C}{k}\right)$	$C \log\left(\frac{C}{k}\right)$	No	RIP-1
Expanders (BP) [BGIKS]	$k \log\left(\frac{C}{k}\right)$	C^3	Yes	RIP-1
Expander Matching Pursuit(EMP) [IR]	$k \log\left(\frac{C}{k}\right)$	$C \log\left(\frac{C}{k}\right)$	Yes	RIP-1
CoSaMP [NT]	$k \log\left(\frac{C}{k}\right)$	$Ck \log\left(\frac{C}{k}\right)$	Yes	Yes
SSMP [DM]	$k \log\left(\frac{C}{k}\right)$	$Ck \log\left(\frac{C}{k}\right)$	Yes	Yes

Random Signals or Random Filters?

- Random Sensing
  1. Outside the mainstream of signal processing: Worst Case Signal Processing
  2. Less efficient recovery time
  3. No explicit constructions
  4. Larger storage
  5. Looser recovery bounds
- Deterministic Sensing
  1. Aligned with the mainstream of signal processing : Average Case Signal Processing
  2. More efficient recovery time
  3. Explicit constructions
  4. Efficient storage
  5. Tighter recovery bounds

k-Sparse Reconstruction with Deterministic Sensing Matrices

Approach	**Measurements** N	**Complexity**	**Noise Resilience**	**RIP**
LDPC Codes [BBS]	$k \log C$	$C \log C$	Yes	No
Reed-Solomon codes [AT]	k	k^2	No	No
Embedding ℓ_2 spaces into ℓ_1 (BP) [GLR]	$k(\log C)^\alpha$	C^3	No	No
Extractors [Ind]	$kC^{o(1)}$	$kC^{o(1)} \log(C)$	No	No
Discrete chirps [AHSC]	$\sqrt{C}$	$kN \log N$	Yes	StRIP
Delsarte-Goethals codes [CHS]	$2^{\sqrt{\log C}}$	$kN \log^2 N$	Yes	StRIP

StRIP is Simple to Design

A: $N \times C$ matrix satisfying

- columns form a group under pointwise multiplication
- rows are orthogonal and all row sums are zero

α: k-sparse signal where positions of the k nonzero entries are equiprobable

THEOREM: Given δ with $1 > \delta > \frac{k-1}{C-1}$, then with high probability

$$(1-\delta)\|\alpha\|_2 \leq \|A\alpha\|_2 \leq (1+\delta)\|\alpha\|_2$$

PROOF: Linearity of expectation

- $\mathbb{E}\left[\|A\alpha\|^2\right] \approx \|\alpha\|^2$
- VAR $\left[\|A\alpha\|^2\right] \to 0$ as $N \to \infty$

- **McDiarmid's inequality:** Given a function f for which $\forall\ x_1, \cdots, x_k, x_i'$:

$$\left| f(x_i, \cdots, x_i, \cdots, x_k) - f(x_i, \cdots, x_i', \cdots, x_k) \right| \leq c_i,$$

and given $X_1, \cdots, X_k$ independent random variables. Then

$$\Pr\left[f(X_1, \cdots, X_k) \geq \mathbf{E}[f(X_1, \cdots, X_k)] + \eta\right] \leq \exp\left(\frac{-2\eta^2}{\sum c_i^2}\right).$$

- **Relaxed assumption:**

$$\forall\ i, j : \quad \left| |\sum_x \varphi^i(x)|^2 - |\sum_x \varphi^j(x)|^2 \right| \leq N^{2-\eta},$$

- Then:
  1. Uniqueness of sparse representation
  2. $\ell 1$ recovery of complex Steinhaus (random phase arbitrary magnitude) signals.

Kerdock Sets

Kerdock set K_m: 2^m binary symmetric $m \times m$ matrices

Tensor $C^0(x, y, a) : \mathbb{F}_{2^m} \times \mathbb{F}_{2^m} \times \mathbb{F}_{2^m} \to \mathbb{F}_2$ given by

$$\mathrm{Tr}[xya] = (x_0, \ldots, x_{m-1}) P^0(a) (y_0, \ldots, y_{m-1})^T$$

Theorem: The difference of any two matrices $P^0(a)$ in K_m is nonsingular

Proof: Non-degeneracy of the trace

Example: $m = 3$, primitive irreducible polynomial $g(x) = x^3 + x + 1$

$$P^0(100) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}, P^0(010) = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix}, P^0(001) = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix}$$

Delsarte-Goethals Sets

Tensor $C^t(x,y,a) : \mathbb{F}_{2^m} \times \mathbb{F}_{2^m} \times \mathbb{F}_{2^m} \to \mathbb{F}_2$ given by

$$\begin{aligned} C^t(x,y,a) &= \mathrm{Tr}[(xy^{2^t} + x^{2^t}y)a] \\ &= (x_0,\ldots,x_{m-1})P^t(a)(y_0,\ldots,y_{m-1})^T \end{aligned}$$

Delsarte-Goethals Set $DG(m,r)$: $2^{(r+1)m}$ binary symmetric $m \times m$ matrices

$$DG(m,r) = \left\{ \sum_{t=0}^{r} P^t(a_t) | a_0, \ldots, a_r \in \mathbb{F}_{2^m} \right\}$$

Framework for exploiting prior information about the signal

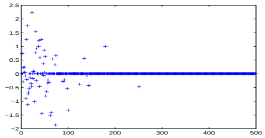
THEOREM: The difference of any two matrices in $DG(m,r)$ has rank at least $m - 2r$

PROOF: Non-degeneracy of the trace

- The Delsarte-Goethals structure imparts an order of preference on the columns of a Reed-Muller sensing matrix

$$K_m = DG(m,0) \subset DG(m,1) \subset \cdots \subset DG\left(m, \frac{m-1}{2}\right)$$

$$\text{Better inner products} \longleftrightarrow \text{Worse inner products}$$



- If a prior distribution on the positions of the sparse components is known, the DG structure provides a means to assign the best columns to the components most likely present

Reed-Muller Sensing Matrices

$$A = \left[\phi^{P,b}(x)\right] : \quad P \in DG(m,r),\ b \in \mathbb{Z}_2^m$$

A has $N = 2^m$ rows and $C = 2^{(r+2)m}$ columns

$$\phi^{P,b}(x) = i^{\mathrm{wt}(d_p)+2\mathrm{wt}(b)} i^{xPx^T + 2bx^T}$$

- Union of $2^{(r+1)m}$ orthonormal basis Γ_P
- Coherence between bases Γ_P and Γ_Q determined by $R = \mathrm{rank}(P+Q)$

THEOREM: Any vector in Γ_P has inner product $2^{-R/2}$ with 2^R vectors in Γ_Q and is orthogonal to the remaining vectors

PROOF: Exponential sums or properties of the symplectic group $Sp(2m,2)$

Quadratic Reconstruction Algorithm

$$f(x+a)\overline{f(x)} = \frac{1}{N}\sum_{j=1}^{k}|\alpha_j|^2(-1)^{aP_jx^T} + \frac{1}{N}\sum_{j\neq t}\alpha_j\overline{\alpha}_t\phi^{P_j,b_j}(x+a)\overline{\phi^{P_t,b_t}(x)}$$

$\frac{1}{N}\sum_{j=1}^{k}|\alpha_j|^2(-1)^{aP_jx^T}$: Concentrates energy at k Walsh-Hadamard tones.

$\frac{1}{N}\sum_{j=1}^{k}|\alpha_j|^4$: Signal energy in the Walsh-Hadamard tones

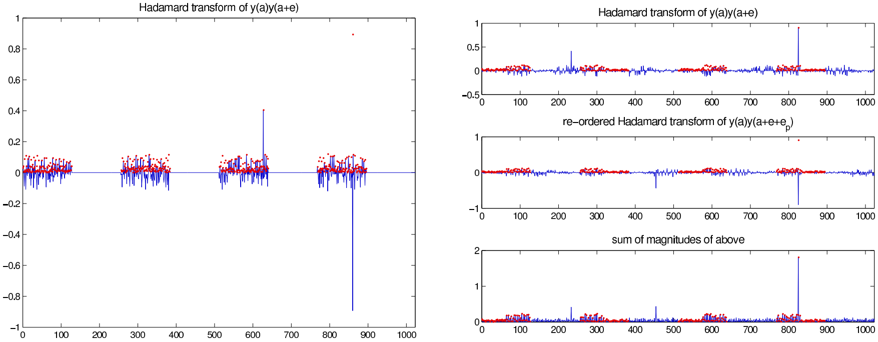
The second term distributes energy uniformly across all N tones – the l^{th} Fourier coefficient is

$$\Gamma_a^l = \frac{1}{N^{3/2}}\sum_{j\neq t}\alpha_j\overline{\alpha}_t\sum_{x}(-1)^{lx^T}\phi^{P_j,b_j}(x+a)\overline{\phi^{P_t,b_t}(x)}$$

THEOREM: $\lim_{N\to\infty}\mathbb{E}[N^2|\Gamma_a^l|^2] = \sum_{j\neq t}|\alpha_j|^2|\alpha_t|^2$

[Note: $\|f\|^4 = \left(\sum_{x,a}|f(x+a)\overline{f(x)}|^2\right)^2$]

Example: $N = 2^{10}$ and $C = 2^{55}$

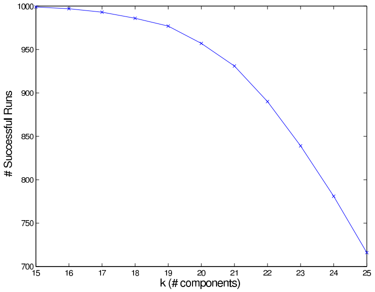


Fundamental Limits

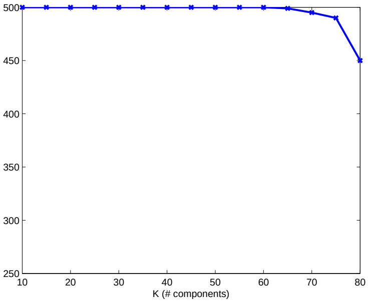
Information Theoretic Rule of Thumb: Number of measurements N required by Basis Pursuit satisfies

$$N > k \log_2 \left(1 + \frac{C}{k}\right)$$

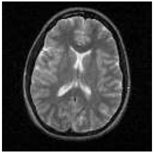
RM$(2, m): C = 2^{55}, k = 20$
$N = 1024$ versus 1014



Kerdock Sensing: $C = 2^{20}, k = 70$
$N = 1024$ versus 971



- Still images with controlled sparsity are reconstructed with good fidelity using compressive sensing with chirp matrices:
  - Original 128×128 image



  - Sparsified image
    - Daubechies-4 wavelet expansion
    - 10% of coefficients (1636) retained

  - Image reconstructed with deterministic algorithm from 4099 chirp measurements
    - About 4:1 compression
    - Essentially lossless reconstruction of sparsified image

Low Power Spread Spectrum Communication

- **Deutsche Telekom:** Energy cost of operations greater than people cost
- **Orthogonal CDMA:** RM(1,m)

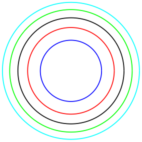
$$1 \text{ Walsh function } \leftrightarrow 1 \text{ bit to } 1 \text{ user}$$

- **Compressive CDMA:** RM(2,m)

$$\begin{aligned} 1 \text{ column} \quad &\leftrightarrow \quad 1 \text{ bit to } \binom{m}{2} \text{ users} \\ &\leftrightarrow \quad \binom{m}{2} \text{ bits to } 1 \text{ user} \end{aligned}$$

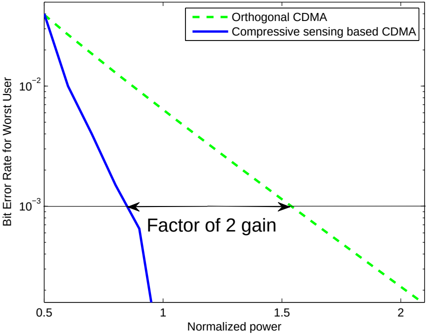
$$\Phi = \begin{bmatrix} & \phi_{P_i,b_i} & \end{bmatrix}_{2^m \times 2^{\binom{m+1}{2}}}$$

m=10:



$$X = \sum_{i=1}^{5} \sqrt{W_i}\phi_{P_i,b_i}$$

W_i = power of signal intended for i^{th} annulus

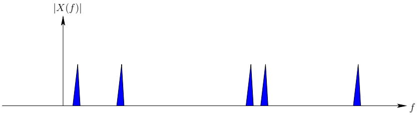


- **Orthogonal CDMA:** Many Sequences, No Interference, Low signal processing complexity
- **Compressive CDMA:** Few Sequences, Interference, More complex signal processing

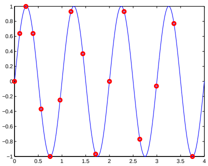
m=10, k=5 : MATLAB implementation of k-sparse signal reconstruction $\approx 1/20$ seconds

Using Chirps for A/D Conversion

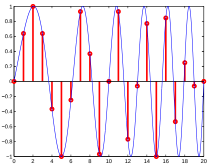
- Application: Sparse signals of tones over large band



- Idea: Non-uniform sampling to convert pure tones to chirps



$\longrightarrow$



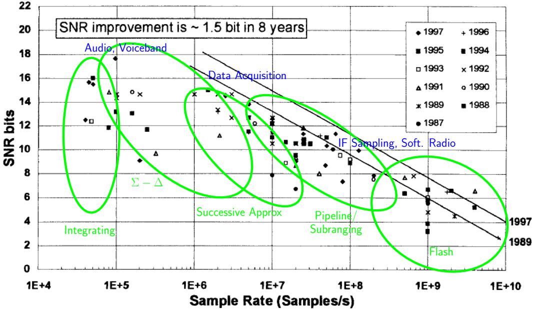
- Leverage RIP results from compressed sensing measurements

- Standard Performance metric

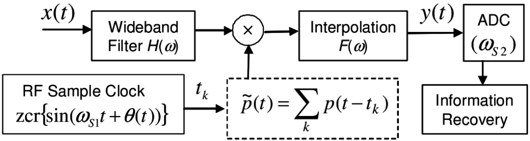
$$P = 2^{\text{SNR Bits}} f_{sampling}$$

  - Captures bandwidth/resolution trade-off. Eg: $\Delta\Sigma$ modulation

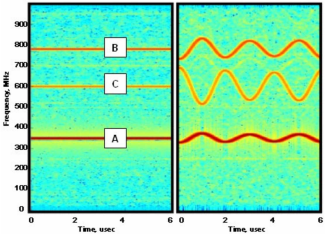
- [Walden (1999)] Slow rate of progress: 1.5 bit increase / 8 Years



Motivation: Nyquist Folding Analog-to-Information Receiver [G. Fudge *et al.*]



- Sampling at zero-crossing of a phase-modulated signal
- Undersampling aliases "Nyquist zones" together
- Stretching/reflection of phase-modulation resolves "Nyquist zone"
- Recovery visualized by spectrogram



Chirp Sampling

- Chirp Sample times: $t_n = nT + n^2 T\epsilon$
  - Converts pure tones to linear chirps

$$e^{j\omega t} \quad \rightarrow \quad e^{j\omega T n + j\omega T \epsilon n^2}$$

  - Can pick $T > \frac{1}{\text{bandwidth}}$ (under sampling)

- Discretized Model:
  - Pick P, Q with Q prime, $P \lessapprox Q$
  - Sample times $t_n = AQn + BPn^2$ for $A, B \in \mathbb{Z}$
    $n = 0, 1, \ldots, P-1$

$$e^{2\pi j \frac{k}{PQ} t} \quad \rightarrow \quad e^{2\pi j \frac{Ak}{P} n} e^{2\pi j \frac{Bk}{Q} n^2}$$

Recovery Conditions via Compressed Sensing

- Properties of Θ
  - Columns form a group
  - Rows are a tight frame

- Bound on column sum / inner-product

$$\left| \sum_{n=0}^{P-1} e^{2\pi j \frac{\alpha}{P} n} e^{2\pi j \frac{\beta}{Q} n^2} \right| \leq C P^{1-\frac{\delta}{2}} \quad \delta \approx 1$$

for constant C independent of P, Q.

Rife and Boorstyn Style Estimation Bounds

For P signal samples in noise given by

$$Z_n = be^{j\omega t_n} + W_n \qquad W_n \sim \mathcal{N}_C(0, 2\sigma^2)$$

Cramér-Rao lower bound on $\hat{\omega}$ with unknown b

$$\mathrm{var}\{\hat{\omega}\} \geq \frac{\sigma^2}{|b|^2(S_2 - S_1^2/P)}$$

where

$$S_1 = \sum_{n=0}^{P-1} t_n \quad , \quad S_2 = \sum_{n=0}^{P-1} t_n^2$$

Uniform sampling: $t_n = nT$

$$\mathrm{var}\{\hat{\omega}\} \geq \frac{\sigma^2}{|b|^2} \frac{12}{T^2 P(P^2-1)}$$

Chirp sampling: $t_n = nT + n^2 T\epsilon$

$$\mathrm{var}\{\hat{\omega}\} \geq \frac{\sigma^2}{|b|^2} \frac{1}{T^2 O(\epsilon^2 P^5)}$$

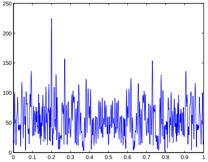
CRLB is Achievable with good SNR

- Can leverage efficiency of FFT after simple transform, converting chirps to tones

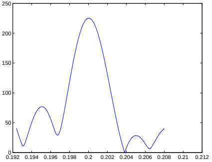
$$f[n] = \bar{y}[n]y[n+D]$$
$$= |b_1|e^{j\omega_1 D^2 T\epsilon + j\omega_1 DT} e^{j2\omega_1 DT\epsilon n} + |b_2|e^{j\omega_2 D^2 T\epsilon + j\omega_2 DT} e^{j2\omega_2 DT\epsilon n}$$
$$+ \cdots + \text{cross terms}$$

- FFT upon $f[n]$ gives initial estimates of ω_i from which we can narrow the search

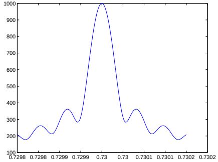
Initial FFT on $f[n]$



Rife Boorstyn refinement on ω_i from $f[n]$



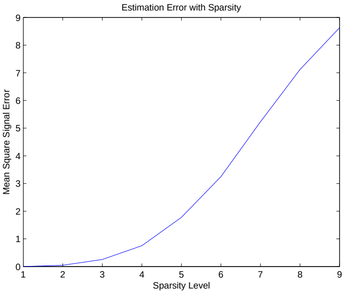
Final refinement on original samples $y[n]$



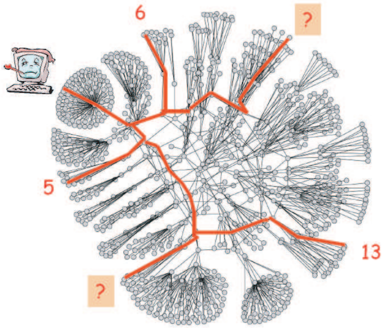
Simulation Results

	Parameter	Value
P	Samples	256
T	Sample Rate	$1/200$ s
ϵ	Relative Chirp rate	$1/10$
f_{max}	Observed BW	1000 Hz

$5\times$ undersampling (complex tones)



Passive Network Monitoring



Passive hop counts observed at multiple monitoring sites

Current methods

- fine grained analysis at a single node or flow
- collection of coarse statistics network wide

Limitations

- fail to leverage diverse detailed data from multiple vantage points
- too complex to extract knowledge from massive high-dimensional datasets

Challenge: Missing data

Recovery of low rank matrices: Keshavan, Montanari & Oh'09

M - $n \times m$ random matrix with rank r

$M = UV$ where U, V are independent random matrices with i.i.d. entries

M can be recovered up to precision δ from a random subset of $C(r,\delta)n$ observations. This can be accomplished efficiently via stochastic local search.

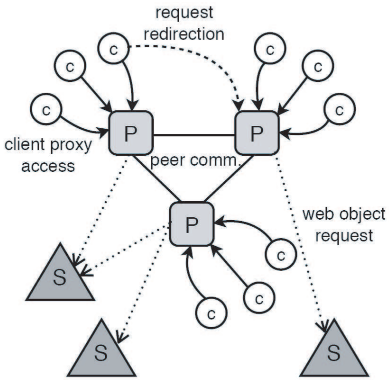
Verification of low rank: Rigidity Theory of matrices A. Singer

$$M \text{ is rank } r\text{, for example } M_{ij} = < c_i, c_j > \quad c_i \in \mathbb{R}^r$$

Given m entries, realization is rigid in r dimensions (completable) if for all observed (i,j)

$$< c_i, v_j > + < c_j, v_i > = 0 \quad \equiv \quad \dim(\text{null}(C)) \le r(r-1)/2$$

Content Distribution Network running on Planetlab



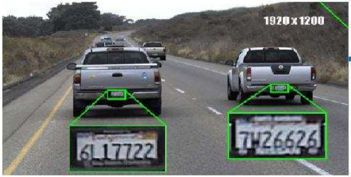
Monitors (columns) × End Hosts (rows)

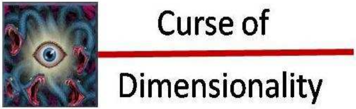
5		13	10	
	2	8		7
4	4			12
4			14	
		12	8	
	11	4		6
5	7			15
		8		9
	12			
		6	4	9
10	5	5	8	
	12	14		
			15	10
6	7	10		
7		10		11
		4		7
4				6
10	4			10
		13	10	5
		10	11	16

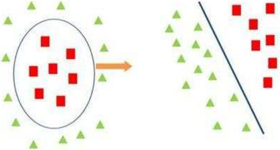
Monitors: Subset of Planetlab P nodes

End Hosts: Clients c, Sources S
Remaining Planetlab P nodes

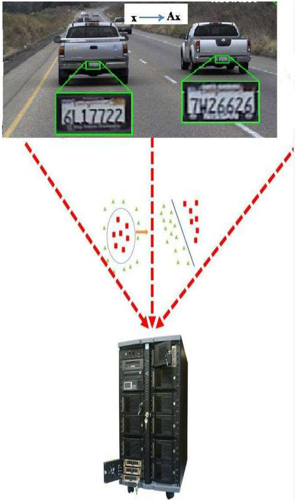
Compressive Learning

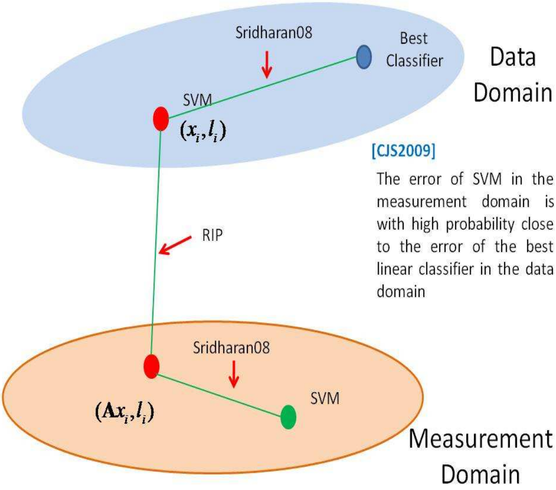






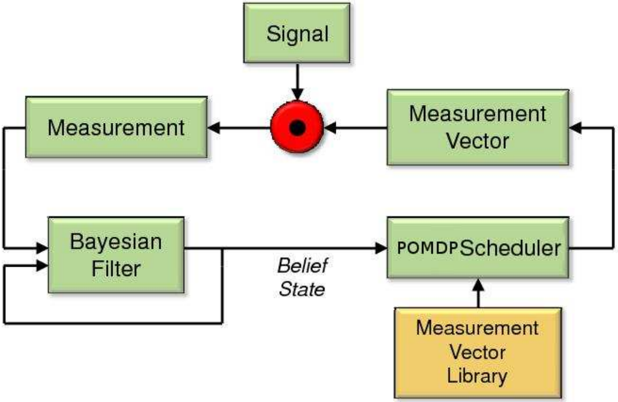
- Is it possible to find needles in compressively sampled haystack?
- If features can be learned in data domain, can they also be learned in the measurement domain?





[CJS2009]

The error of SVM in the measurement domain is with high probability close to the error of the best linear classifier in the data domain



[Castro, Haup, Nowak: AISTAT 09]: Distilled Sensing: Selective Sampling for Sparse Signal Recovery.