There are special spare signals that vanish nearly everywhere in the Ψ domain. $1-\delta$ accept some probability for failure.

Sensing matrices whose entries are i.i.d. Gaussian/Rademacher are incoherent with any given orthonormal basis Ψ with a very high probability.

Fourier Transform

$$F(u)=\frac{1}{\sqrt{N}}\sum_{n=0}^{N-1} f(n)e^{\frac{-i2\pi un}{N}}$$

$$f(n)=\frac{1}{\sqrt{N}}\sum_{u=0}^{N-1} F(u)e^{\frac{i2\pi un}{N}}$$

$$f=HF=\sum_{u=0}^{N-1} H_u F(u)$$

$$F=H^Tf \qquad HH^T=I$$

2D

$$F(u,v)=\frac{1}{\sqrt{MN}}\sum_{x=0}^{M-1}\sum_{y=0}^{N-1} f(x,y)e^{-2\pi\left(\frac{ux}{M}+\frac{vy}{N}\right)i}$$

$$f(x,y)=\frac{1}{\sqrt{MN}}\sum_{u=0}^{M-1}\sum_{v=0}^{N-1} F(u,v)e^{2\pi\left(\frac{xu}{M}+\frac{yv}{N}\right)i}$$

$$f=H_{MN\times MN}F_{MN\times 1}$$

DCT in 2D

$$F(u,v)=\sum_{n=0}^{N-1}\sum_{m=0}^{M-1} f(n,m)H^*_{nmuv}$$

$$f(n,m)=\sum_{u=0}^{M-1}\sum_{v=0}^{N-1} F(u,v)H_{nmuv}$$

$$H_{nmuv}=\alpha(u)\alpha(v)\cos\left(\frac{\pi(2n+1)u}{2N}\right)\cos\left(\frac{\pi(2m+1)}{2M}\right)$$

$$u=0,\dots,N-1 \qquad v=0,\dots,M-1$$

$$\alpha(u)=\begin{cases}\sqrt{1/N} & u=0\\ \sqrt{2/N}\end{cases}$$

$$\alpha(v)=\begin{cases}\sqrt{1/M} & v=0\\ \sqrt{2/M}\end{cases}$$

Shannon Band-Limited signal can be reconstructed if $f_s>2f_{max}$

CS $\vec{y}=\Phi\vec{f}$, measurement matrix

No unique sol^n as under-determined.

Unique if $m<n$

1. $\vec{f}$ is sparse 2. Φ incoherent

Many signals have sparse representations in standard ortho normal bases

$\vec{f}=\Psi\vec{\theta}$, $\|\theta\|_0<<n$, $\Psi^T\Psi=I$

$\|\theta\|_0$ – no. of non zero elements

Coherence. $\mu(\Psi,\Phi)=\sqrt{n}\max_{\substack{1\le j\le m\\1\le i\le n}}|\langle\Phi^{(j)},\Psi_i\rangle|$ (j^{th} row, i^{th} column), $\in[1,\sqrt{n}]$

- Sensing matrices whose entries are Rademacher R.V. are incoherent with any given orthonormal basis Ψ with very high probability.

$\vec{y}=\Phi\Psi\vec{\theta}$

P0 $\min\|\theta\|_0$ such that $\vec{y}=\Phi\Psi\vec{\theta}$

BP P1 $\min\|\theta\|_1$ s.t. $\vec{y}=\Phi\Psi\vec{\theta}$

If any $2S$ columns of an $m\times n$ Φ are linearly independent. Then any S-sparse signal f can be uniquely constructed from measurements $y=\Phi f$

Probabilistic **Theorem 1** $f=\Psi\theta$, sparse representation basis $\|\theta\|_0<<n$

If $m\ge C\log\left(\frac{m}{\delta}\right)\|\theta\|_0\,\mu^2(\Psi,\Phi)$ the sol^n to following is exact with probability ~~atle~~ $1-\delta$

$\min\|\theta\|_1$ s.t. $y=\Phi\Psi\theta$ $\quad y=\Phi f=\Phi\Psi\theta$

Intuition Behind Coherence Measurement f^n should be non-sparse linear combination of DCT basis vectors = having measurement which are non-sparse linear combination of all DCT C.E.

Restricted Isometric Property (RIP) RIC-δ_S of $\underset{m\times n}{A}=\Phi\Psi$ is smallest number such that for any S-sparse vector θ, $(1-\delta_S)\|\theta\|^2\le\|A\theta\|^2\le(1+\delta_S)\|\theta\|^2$

Bad Event: $A\theta^{(1)}=A\theta^{(2)}$ for $\theta^{(1)}\ne\theta^{(2)}$, sol^n design A such that $A\theta^{(1)}\approx A\theta^{(2)}\leftrightarrow\theta^{(1)}\approx\theta^{(2)}$

Deterministic **Theorem 2** (stronger conclusions than theorem 1. If $\theta=\theta_S$ then exact recovery)

L1 If $A=\Phi\Psi$ has RIP of order $2S$, (with $\delta_{2S}<0.41$)

Let θ^* be sol^n of P1, $\theta_S=\theta$ with S Largest magnitude others C.E 0

Then $\|\theta^*-\theta\|_2\le C_0/\sqrt{S}\,\|\theta-\theta_S\|_1$ $\quad C_0=f(\delta_{2S})$, $f'(\cdot)>0$

$\|\theta^*-\theta\|_1\le C_0\|\theta-\theta_S\|_1$

C_0 ind. of n; it is an inc f^n of δ_{2S} in $[0,1]$

→ Reconstruction for S-sparse signal ALWAYS exact if A has RIP

→ Many signals are compressible, sharp decay in |CE| in standard orthonormal bases

Compressive Sensing Under Noise. $y=\Phi f+\eta=\Phi\Psi\theta+\eta$

$\min\|\theta\|_1$ such that $\|y-\Phi\Psi\theta\|_2^2\le\varepsilon$

A has RIP 2S **Theorem 3** L1: does not require Ψ to be orthonormal; error due to noiseless data; here also for $\delta_{2S}<\sqrt{2}-1$

$$\|\theta^*-\theta\|_2\le\frac{C_0}{\sqrt{S}}\|\theta-\theta_S\|_1+C_1\varepsilon$$

$C_1\varepsilon$: error due to noise. C_0, C_1 increasing f^n of δ_{2S}

If $\|\eta\|_2^2\le K$ choose $\varepsilon\ge K$ with high probability.

Randomness is super-cool.

For a fixed orthonormal basis Ψ, sensing matrix Φ $A=\Phi\Psi$ will obey RIP of order S with h.p. provided no. of rows $m\ge CS\log\left(\frac{n}{S}\right)$ if

1. Φ contains entries from $N(0,\frac{1}{m})$ 2. Rademacher $\left(\pm\frac{1}{\sqrt{m}}\right)$

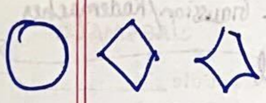
3. Columns of Φ sampled uniformly randomly from unit-sphere in m-d cube.

A sensing matrix Φ containing a random subset of rows of DFT also obeys RIP. → Provided $m\ge O(S\log^4 n)$

Noiseless Case: Theorem 3 holds whenever A obeys RIP

Noisy Case: Theorem 3 holds with some prob < 1 whenever A obeys RIP

Theorem 3 Choosing ε: if η_i: uni$(-r,r)$ then $\varepsilon\ge r^2m$; if $\eta_i\sim N(0,\sigma^2)$ $\eta_i^2:\chi$ sq. with h.p. $\|\vec{\eta}\|$ will lie within 3 std dev from mean. Set $\varepsilon\ge 9m\sigma^2$



CS for piecewise const signals.

Cartoon

Gradients are sparse (except along edges)

min $TV(f)$ s.t. $y = \Phi\Psi^T f$

$f_x + jf_y$

$TV(f) \leftarrow \iint \sqrt{f_x^2 + f_y^2}\, dx\, dy$

= L1-norm of 2d gradient vector.

Theorem 4 (like Theorem 1)

Consider piecewise const signal $\vec{f}_{n\times 1}$

If $m \geq C \log\frac{n}{\delta} \|\theta\|_0 \mu^2(\Psi, \Phi)$ then sol^n to PTV is exact atleast with probability $1-\delta$.

Relevant Theorem behind Candes' Experiment

Consider 1d signal f.

Mutual Coherence.

$$\mu(A) = \max_{i,j,i\neq j} \frac{|A_i^t A_j|}{\|A_i\|_2 \|A_j\|_2}$$

Theorem 5 like theorem 3 handles noise as well as compressible signals but uses μ & not δ (unlike theorem 3)

does not required Ψ to be orthonormal

$y = \Phi x + \eta = \Phi\Psi\theta + \eta = A\theta + \eta$

Estimate x via θ

Let θ_S be a sub-vector of θ with S largest elements in correct location.

if $S < 0.5\left(1 + \frac{1}{\mu(A)}\right)$ then

P2 $\min \|\theta\|_1$, s.t. $\|y - A\theta\|_2 \leq \varepsilon'$, $\varepsilon \leq \varepsilon'$

sol^n to P2 yields following error bounds

$\|\theta - \theta^*\|_2 \leq C_0(\varepsilon + \varepsilon') + C_1\|\theta - \theta_S\|_1$

$\|\eta\|_2 \leq \varepsilon$

δ_S cannot be computed efficiently.

μ can be computed in $O(mn^2)$ time.

Consider for S-sparse θ:

$$(1-\delta_S)\|\theta\|^2 \leq \|A\theta\|^2 \leq (1+\delta_S)\|\theta\|^2$$

$\delta_S = \max\{1 - \lambda_{min}, \lambda_{max} - 1\}$

$$\lambda_{max} = \max_{\substack{\theta_T \in R^S,\\ |T| \leq S}} \frac{\|A_T\theta_T\|^2}{\|\theta_T\|^2}$$

λ_{max} & λ_{min} indicate Max & Min. EV of $(A_T)^T(A_T)$

Gershgorin Circle Theorem

Let A be a complex $n\times n$ matrix with entries a_{ij}

$R_i = \sum_{j\neq i} |a_{ij}|$. $D(a_{ii}, R_i) \subseteq \mathbb{C}$ closed disc centered at a_{ii}

Theorem: Every eigenvalue of A lies within atleast one of the Gershgorin Discs $D(a_{ii}, R_i)$

Reconstruction Problem

P0 : $\min \|\theta\|_0$ s.t. $\|y - A\theta\|_2^2 \leq \varepsilon$, $A = \Phi\Psi$

P1 : $\min \|\theta\|_1$ s.t. $\|y - A\theta\|_2^2 \leq \varepsilon$

Cat 1| P0 is NP-hard

Run an approximation algo

Algo: Matching Pursuit, Orthogonal MP, Iterative hard Thresholding, Co-SAMP

Cat 2| Solve P1 using algo like.

Iterative Shrinkage Thresholding algo (ISTA)

Matching Pursuit.

$r^{(0)} = y, i = 0$

while ($\|r^{(i)}\|^2 > \varepsilon$)

{ $j = \arg\max_l \left|\frac{r^{(i)T} a_l}{\|a_l\|_2}\right|$

$\theta_j = \frac{r^{(i)T} a_j}{\|a_j\|^2}$, $r^{(i+1)} = r^{(i)} - \theta_j a_j$

$i = i+1$

}

Output: $\{\theta_j\}$

decomposes y to

$\vec{y} = \frac{y^T a_k}{a_k^T a_k} a_k + r^{(-k)}$

$\|y\|^2 = (y^T a_k)^2 + \|r^{-k}\|^2$

. May pick same dictionary column multiple times.

Orthogonal Matching Pursuit

$r^{(0)} = y$, $\theta = 0$; $T^{(0)} = \phi$; $i = 0$

while ($\|r^{(i)}\|^2 > \varepsilon$)

{ (1) $j = \arg\max_k \left|\frac{r^{(i)T} a_k}{\|a_k\|_2}\right|$

(2) $T^{(i+1)} = T^{(i)} \cup j$; $i = i+1$

(3) $\theta_{T^{(i)}} = \arg\min_w \|y - A_{T^{(i)}} w\|^2$

(4) $r^{(i)} = y - A_{T^{(i)}}\theta_{T^{(i)}}$

The process of massive data acquisition followed by compression is extremely wasteful.

ISTA (Iterative Shrinkage & Thresholding Algo)

PU $\min_\theta J(\theta) = \|y - A\theta\|_2^2 + \lambda\|\theta\|_1$

$y \in \mathbb{R}^m$ $\theta \in \mathbb{R}^n$ $A \in \mathbb{R}^{m\times n}$, $m << n$, n large.

Iterative approach $J(\vec{\theta}_{k+1}) < J(\vec{\theta}_k)$

$\vec{\theta}_{k+1}$ chosen to minimize Majorizer f^n $M_k(\vec{\theta})$

such that $M_k(\vec{\theta}) >= J(\vec{\theta})$ $\forall\vec{\theta}$ & $M_k(\vec{\theta}_k) = J(\vec{\theta}_k)$

1. $k=0$, init θ_0
2. Choose $M_k(\vec{\theta})$ such that $M_k(\vec{\theta}) >= J(\vec{\theta})$ $\forall\vec{\theta}$ & $M_k(\vec{\theta}_k) = J(\vec{\theta}_k)$
3. Select $\vec{\theta}_{k+1}$ to minimize $M_k(\vec{\theta})$
4. $k = k+1$

One possible Majorizer

$M_k(\vec{\theta}) = \|\vec{y} - A\vec{\theta}\|_2^2 +$ non-ve f^n of $\theta + \lambda\|\theta\|_1$

$= \|\vec{y} - A\vec{\theta}\|_2^2 + (\vec{\theta} - \vec{\theta}_k)^T(\alpha I - A^TA)(\vec{\theta} - \vec{\theta}_k) + \lambda\|\theta\|_1$

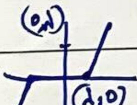
for $\alpha >$ max eigen value of A^TA

Initialize θ_0 randomly, $k=0$

$\alpha = EV_{Max}(A^TA)$

Repeat till convergence

$\vec{\theta}_{k+1} = \text{soft}\left(\vec{\theta}_k + \frac{1}{\alpha}A^T(\vec{y} - A\vec{\theta}_k), \frac{\lambda}{2\alpha}\right)$



$x = \text{soft}(y;\lambda) = \begin{cases} 0 & |y| < \lambda \\ y - \lambda & y \ge \lambda \\ & y \le -\lambda \end{cases}$

$y = x + \lambda\,\text{sign}(x) \Leftrightarrow x = \text{soft}(y;\lambda)$

$\vec{\theta}_{k+1} + \frac{\lambda}{2\alpha}\text{sign}(\vec{\theta}_{k+1}) = \vec{\theta}_k + \frac{1}{\alpha}A^T(\vec{y} - A\vec{\theta}_k)$

Compressed sensing for pool testing.

$\vec{y} = A\vec{x} + \vec{\eta}$

$\min \|x\|_1$ s.t. $\|y - Ax\|_2 \le \varepsilon$

$J_{Lasso}(x; y, A) := \|y - Ax\|_2^2 + \lambda\|x\|_1$

· Smaller CT (cycle Threshold) = Greater Viral Load

$z_i(1+q)^{\tau_i} = F$ $\quad \tau_i = t_i + e_i$, $e_i \sim N(0, \sigma^2)$ $\sigma^2 << 1$ $\quad \sim 16\text{-}40$

$A^i x(1+q)^{t_i} = F \Rightarrow z_i(1+q)^{\tau_i} = A^i x(1+q)^{t_i}$

$\log z_i = \log A^i x - e_i \log(1+q)$

$y_i = A\omega$

$\omega \sim \text{Poisson}(x)$

Removing -ve samples ~ Combinatorial group testing ~~problem~~ algo called COMP

[Combinatorial Orthogonal MP]

Sensitivity = $\frac{\text{Correctly detected +ve}}{\text{True +ve}}$ $\quad$ Specificity = $\frac{\text{Correctly detected -ve}}{\text{True -ve}}$

→ Dorfman Pooling:

First round groups: 2nd round individual testing of each person in +ve pool.

$$G(f)=\int_{-\infty}^{\infty} f(t)e^{-j2\pi ft}dt$$

$$f(t)=\int_{-\infty}^{\infty} G(f)e^{j2\pi ft}df$$

$$I = I_0 \exp\left(-\int_L f(x,y,z)\,dL\right)$$

$$g(\rho,\theta) = \log\frac{I_0}{I} = \int_L f(x,y,z)\,dL$$

Radon Transform of f

fixed θ varying

$$R_\theta(f) = g(\rho,\theta) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\,\delta(x\cos\theta + y\sin\theta-\rho)\,dx\,dy$$

$$= \sum_{x=0}^{M-1}\sum_{y=0}^{N-1} f(x,y)\,\delta(x\cos\theta+y\sin\theta-\rho)$$

Sinogram a radon transform plotted as an image in a (ρ,θ) grid.

Let $\hat{f}_{\theta_k}(x,y) = g(\rho,\theta_k)$

Back Projection

$$\hat{f}(x,y)=\int_0^{\pi} g(x\cos\theta+y\sin\theta,\theta)\,d\theta = \int_0^{\pi}\hat{f}_\theta(x,y)\,d\theta$$

gives back $f(x,y)$ blurred with $\frac{1}{\sqrt{x^2+y^2}}$

FT of RT — Projection/Fourier Slice Theorem

$$G(\mu,\theta) = \int_{-\infty}^{\infty} g(\rho,\theta)\,e^{-j2\pi\mu\rho}\,d\rho$$

$$= \iint f(x,y)\,e^{-j2\mu\pi(x\cos\theta+y\sin\theta)}\,dx\,dy$$

$$= \iint f(x,y)\,e^{-j2\pi(xu+yv)}\,dx\,dv$$

$$\therefore G(\mu,\theta) = [F(u,v)]_{u=\mu\cos\theta,\ v=\mu\sin\theta}$$

it states that FT of a 2D object along some direction θ $(G(\mu,\theta))$ = Slice of 2D F.T. of object along same direction θ (in f-plane) passing through origin.

Consider 2D-iFT of

$$f(x,y)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} F(u,v)\,e^{j2\pi(xu+yv)}\,du\,dv$$

$$f(x,y)=\int_0^{2\pi}\int_0^{\infty} F(\mu\cos\theta,\mu\sin\theta)\,e^{j2\pi\mu(x\cos\theta+y\sin\theta)}\,\mu\,d\mu\,d\theta$$

$$=\int_0^{2\pi}\int_0^{\infty} G(\mu,\theta)\,e^{j2\pi\mu(x\cos\theta+y\sin\theta)}\,\mu\,d\mu\,d\theta$$

$$=\int_0^{\pi}\int_{-\infty}^{\infty} G(\mu,\theta)\,e^{j2\pi\mu(x\cos\theta+y\sin\theta)}\,|\mu|\,d\mu\,d\theta$$

$$\frac{1}{\sqrt{x^2+y^2}} \longleftrightarrow \frac{1}{\sqrt{u^2+v^2}}$$

$ git log --oneline --all

HEAD → main, origin/main, origin/HEAD

Github HEAD is currently on remote main.

Github main is currently in sync with local main

Your local repo is currently on the main branch

Dictionary Learning: columns/atoms

$\vec{y}_i = D\vec{S}_i$ D: Dictionary $\vec{S}_i$: dictionary coefficients

Algorithms: 1. PCA 2. Non-ve sparse coding 3. Method of Optimal Directions 4. Union of orthonormal bases 5. K-SVD

PCA Extracting k features from $\vec{x}_{i\,d\times1}$ $k \ll d$

1 Mean $\vec{\bar{x}} = \frac{1}{N}\sum_{i=1}^{N} \vec{x}_i$,

$\bar{x}_i = x_i - \bar{x}$

Covariance Matrix $C = \frac{1}{N-1}\sum_{i=1}^{N} \vec{\bar{x}}_i \vec{\bar{x}}_i^T$

$d\times d$ Symmetric, Positive Definite.

4 Find $\vec{EV}$ $CV = V\Lambda$ Λ: $D(\lambda_i)$

Extract k $\vec{EV}$ corresponding to k largest EV.

$\hat{V}_k = V(:, 1:k)$ columns corresponding to k-largest $\vec{EV}$

$d\times k$

6 Project each point onto the Eigenspace, giving a vector of k eigen-coefficients for that point.

$\alpha_{ik} = \hat{V}_k^T \bar{x}_i$ $\alpha_{ik} \in R^k$; $\alpha_i = V^T \bar{x}_i$, $\alpha_i \in R^d$.

As V is orthonormal we have.

$\bar{x}_i = V\alpha_i = V(:,1)\alpha_i(1) + \cdots + V(:,d)\alpha_i(d)$

$\approx \hat{V}_k \alpha_{ik} = \hat{V}_k(:,1)\alpha_{ik}(1) + \hat{V}_k(:,2)\alpha_{ik}(2) + \cdots + \hat{V}_k(:,k)\alpha_{ik}(k)$

α_{ik} is a vector of EigenCoefficients of i^{th} sample point, and it has k element, j^{th} element of this vector is denoted to $\alpha_{ik}(j)$

Non-negative Matrix Factorisation

$Y \in R^{n\times N}$ $Y \approx WH$, $W, H > 0$

W: columns basis   H: values of columns are C.E.

$W_{n\times r}$ $H_{r\times N}$

minimise $E(W,H) = \|Y - WH\|_F^2$

s.t. $W \geq 0, H \geq 0$

→ Method of alternating optimization using projected Gradient Descent with adaptive step size

→ Multiplicative Update

extension with sparseness penalty imposed on C.E.

$E(W,H) = \|Y - WH\|_F^2 + \lambda\|H\|_1$

→ Well posedness unit norm constraint on columns of W.

Basis need not be orthonormal ex. Full rank matrix

→ with complete bases, representation always unique.

→ advantage: greater compactness

~~Fix basis~~

Learning the bases: Method 1 Fix basis A & obtain sparse CE then gradient descent on A

→ Learning Bases: Method of Optimal Directions Method 2

$E(A) = \sum_{i=1}^{n} \|y_i - AS_i\|^2 = \|Y - AS\|_F^2$

assume sparse codes are computed for every signal using OMP, MP.

· we want to find A that minimizes this error assuming fixed sparse codes

→ Method 3 Union of orthonormal Bases

$X = AS + E$

$(A,S) = \min_{A,S} \|X - AS\|^2 + \lambda\|S\|_1$

A overcomplete assume union of orthonormal bases.

$A = [A_1\ A_2 \cdots A_M]$ $A_i A_i^T = I$ $\vec{S} = \begin{pmatrix} \vec{S}_1 \\ \vec{S}_2 \\ \vdots \\ \vec{S}_M \end{pmatrix}$

C.E. in S can be estimated using block-coordinate descent

for $m = 1:M$ {

$X_m = X - \sum_{j\neq m} A_j S_j$

Update S_m as: $S_m = \arg\min_{S^*} \|X_m - A_m S^*\|^2 + \lambda_t \|S^*\|_1$

Blind Compressed Sensing

knowledge of basis not available

→ inferring signal basis as well as signal C.E. directly from compressed data.

$\vec{y}_i = \Phi_i \vec{x}_i + \eta_i$ $\vec{x}_i = D\vec{\theta}_i$

$E(D, \{\theta_i\}) = \sum_{i=1}^{N} \|y_i - \Phi_i D\theta_i\|^2$ s.t. $\|\theta_i\|_0 \leq T$, $d_k^t d_k = 1$

Alternating

1. Sparse coding assuming a fixed dictionary.
2. Dictionary updates using sparse codes

Dictionary Update

$E(d_k) = \sum_{i=1}^{n} \|y_i - \Phi_i \sum_{l=1}^{K} \theta_{il} \vec{d}_l\|^2$

$= \sum_{i=1}^{N} \|y_i - \Phi_i \sum_{l=1, l\neq k}^{K} d_l \theta_{il} - \Phi_i d_k \theta_{ik}\|^2$

$= \sum_{i=1}^{n} \|y_i^{-k} - \Phi_i d_k \theta_{ik}\|^2$

$(y_i^{-k} - \Phi_i d_k \theta_{ik})^T (y_i^{-k} - \Phi_i d_k \theta_{ik})$

$-2 y_i^{-k} \Phi_i d_k + \theta_{ik}^T d_k^T \Phi_i^T \Phi_i d_k \theta_{ik}$

$-2 y_i^{-k} \Phi_i + \theta_{ik}^2\, 2 d_k^T \Phi_i^T \Phi_i$

$-2 \Phi_i^T y_i^{-k} + \theta_{ik}^2\, 2\Phi_i^T \Phi_i d_k$

$-2\Phi_i^T (y_i^{-k} - \cdot$

M T W T F S S
Page No.:
Date:
YOUVA

Lambertian Model

$I = L\rho\cos\theta$ — ρ: surface reflectivity, L: Lightning Intensity

$\cos\theta < 0 \Rightarrow$ surface doesn't face light source. $I = \max(0, L\rho\cos\theta)$

Theorem 1 $M_{n_1 \times n_2}$ rank $r < \min(n_1, n_2)$

Observed only fraction of entries $T(i,j) = M(i,j)$ $\forall (i,j) \in \Omega$

If (1) M has row & column spaces that are sufficiently incoherent with canonical basis
(2) r is 'sufficiently' small.
(3) Ω is " Large, then we can accurately recover M from T by solving $M^* = \min_M \text{rank}(M)$ subject to $M(i,j) = T(i,j)$ $\forall (i,j) \in \Omega$.

Theorem 2

Q0 $M^* = \min_M \|M\|_*$ s.t $M(i,j) = T(i,j) \; \forall (i,j) \in \Omega$

Trace Norm/Nuclear Norm sum of its singular values (Softened version of rank)

gives same M^* as theorem 1

Rank-Nullity Theorem. $M_{m\times n}$

$\text{rank}(M) + \text{nullity}(M) = n$

$\text{rank}(A_{m\times n}) \leq \min(m,n)$

$\text{rank}(AB) \leq \min(\text{rank}(A), \text{rank}(B))$

$M_{n_1 \times n_2} = \underset{\text{Low-rank}}{L} + \underset{\text{Sparse (with unknown support)}}{S}$

Robust PCA

$M = L + S$ — support is uniformly randomly distributed in the entries of M.

Principal Component Pursuit yields L,S with h.p.

$$E(L', S') = \min_{(L,S)} \|L\|_* + \frac{1}{\sqrt{\max(n_1, n_2)}}\|S\|_1$$

subject to $L + S = M$

$\|S\|_1 = \sum_{i\,1 \text{ to } n_1} \sum_{j\,1 \text{ to } n_2} |S_{ij}|$

$\text{rank}(A+B) \leq \text{rank}(A) + \text{rank}(B)$

$\text{rank}(A^T A) = \text{rank}(AA^T) = \text{rank}(A) = \text{rank}(A^T)$

2D-DCT matrix is orthogonal.

F

Compressive RPCA

$M = L + S$

$y = A(L+S)$

retrieves L, S given A, y

Objective Function

(P1) $\min \|y - A(L+S)\|_2$

subject to $\text{rank}(L) \leq r$

$\|\text{vec}(S)\|_0 \leq k$

CCTV footage, hyperspectral image.

(Compressive) Low Rank Matrix Recovery.

$y = A(M)$

$y_i = \text{vec}(A_i) \cdot \text{vec}(M)$

$\{A_i\}$: Set of matrices constituting linear map A

$\vec{y} = A\,\text{vec}(M)$

RIP for linear maps

$(1-\delta_r)\|M\|_F \leq \|A(M)\|_2 \leq (1+\delta_r)\|M\|_F$

$\forall$ matrices M of rank atmost r.

$\vec{y} = H\vec{x} + \vec{\varepsilon}$

$P(x|y) = \dfrac{P(y|x)P(x)}{P(y)} \propto P(y|x)P(x)$

$P(y|x)$: Likelihood

$P(x)$: Prior Model on Signal.

If $\vec{\varepsilon} \sim N$ then $p(y|x) \propto \exp\left(-\dfrac{\|y-Hx\|_2^2}{2\sigma^2}\right)$

MAP $\hat{x} = \text{argmax}_x\, p(x|y)$

$= \text{argmax}_x\, P(y|x)p(x)$

MLE $\hat{x}_{MLE} = \text{argmax}_x\, p(y|x)$

MMSE $\hat{x} = \text{argmin}_z \int \|x-z\|^2 p(x|y)\, dx$

$= E[X|Y]$

Prior in IP image x has DCT C.E which are Laplacian Distributed.

$y = h * x + n = Hx + n,\ n \sim N(0, \sigma)$

Equivalently Represented as a circulant matrix H derived from h.

→ Statistical Compressed Sensing.

$\vec{y} = \Phi \vec{x} + \vec{\eta}$

$\eta \sim N(0, \sigma^2 I_{m\times m}),\ x \in N(0, \Sigma)$

→ If Likelihood & Prior are both Gaussian, then MAP = MMSE

→ The decay of the EigenValues of the covariance matrix is the equivalent of signal sparsity or compressibility in an appropriate orthonormal basis,

→ Gaussian Mixture Models are known to be universal PDF estimators.

∂