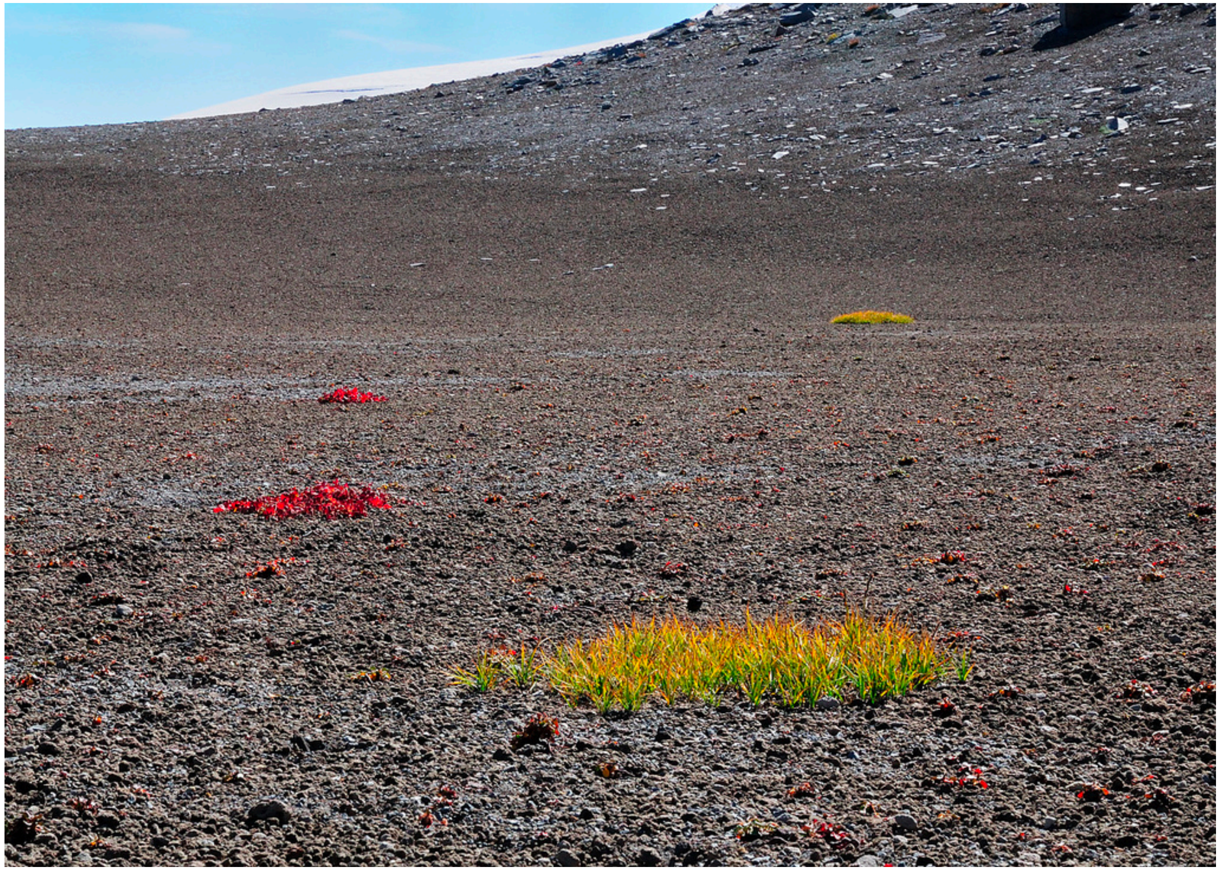


EE225A Lecture 21

Compressed Sensing Potpourri

Jon Tamir

“Sparse plants on pumice plain, Mt Rainier”

https://www.flickr.com/photos/jeffbottman/8088689063/

Logistics

1. Problem Set 5 – Now Due Friday

1. Problem Set 6 Out Soon

Problem on board

Particle Filtering: Example 7.4.2

Outline for Today

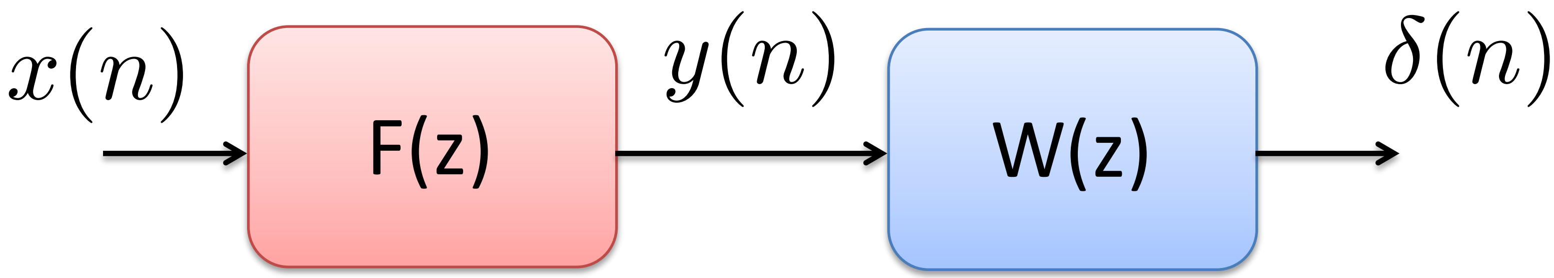
1. Application of LMS – System Identification

2. Introduction to Modern Compressed Sensing Topics
   » Phase Transition Curves
   » Low Rank Matrix Recovery
   » Low Rank + Sparse
   » Multi-scale Low Rank
   » Dictionary Learning

LMS and System ID

- **LMS:** Stochastic gradient descent for Wiener-Hopf Equations

- **Special Case:** Wiener Filter Deconvolution



$$\hat{d}(n) = y(n) * w(n) = x(n)$$

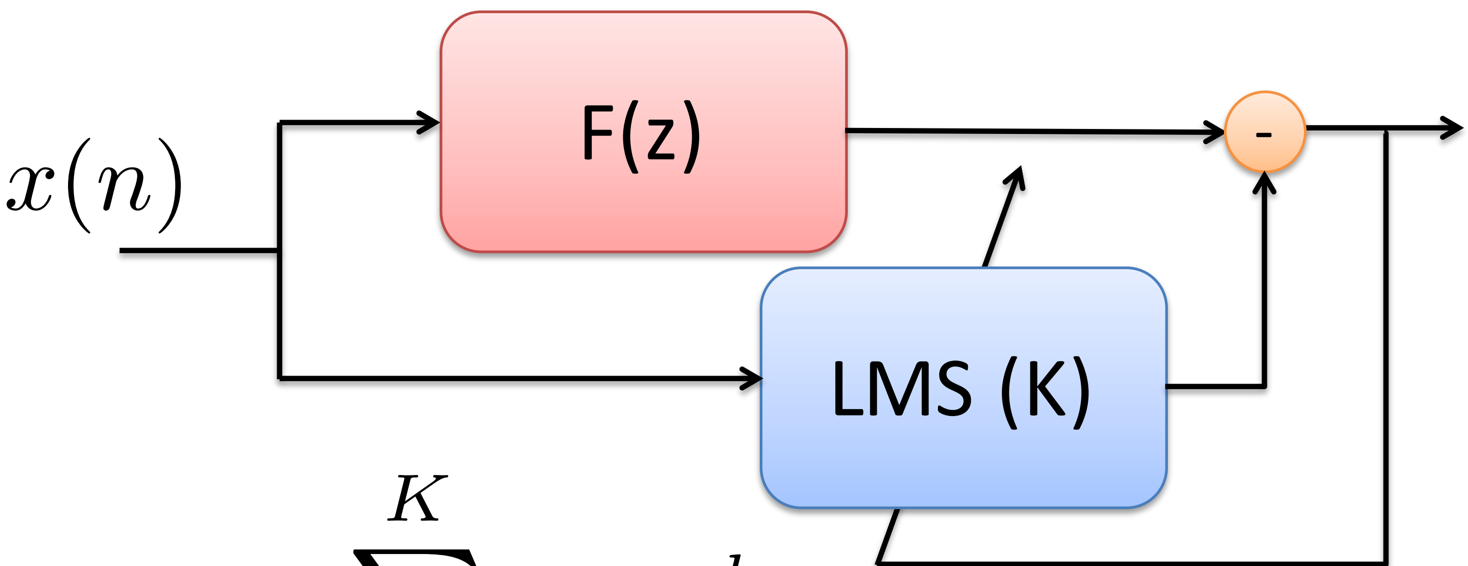
- **Need to know F(z)…**

LMS and System ID

» If we don't know F(z), we can use LMS to learn it

» **Case 1:** F(z) is FIR(K), have access to input

- » At convergence, error should be zero

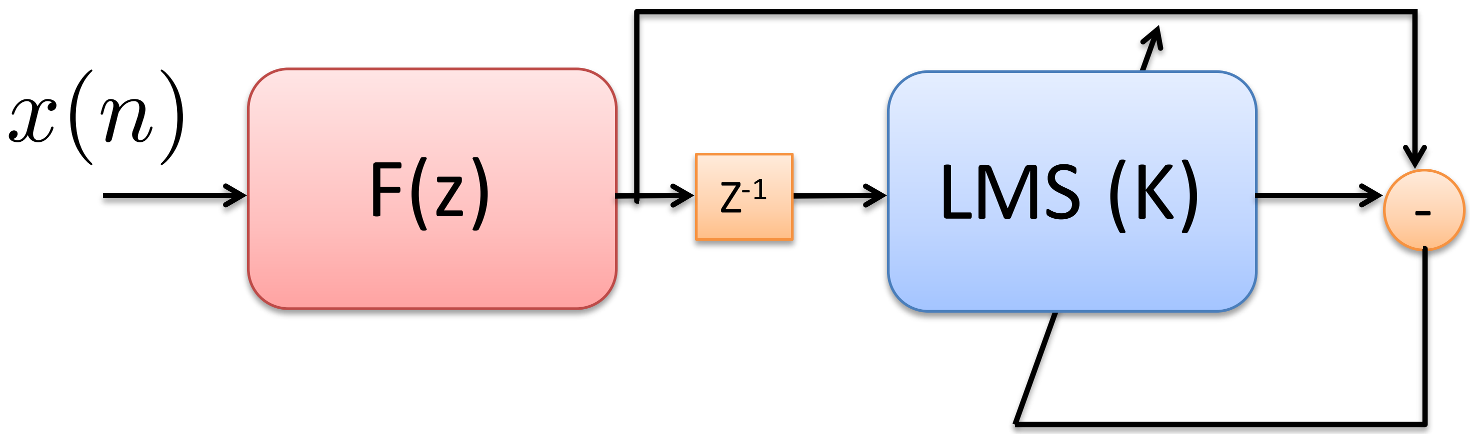


$$F(z) = \sum_{k=1}^{K} \alpha_k z^{-k}$$

LMS and System ID

- If we don't know F(z), we can use LMS to learn it

- **Case 2:** F(z) is an AR(K) process, do not have access to input
  - At convergence, error should equal the innovation



$$F(z) = \frac{1}{1 - \sum_{k=1}^{K} \alpha_k z^{-k}}$$

Phase Transition Curves

» **Recall:** Compressed Sensing Recipe

1. **Signal is k-sparse**
2. **Incoherent and sufficient sampling**
3. **Non-linear reconstruction**

$$y = Ax$$
$$y \in \mathbb{R}^n$$
$$x \in \mathbb{R}^N$$
$$||x||_0 \leq k$$

» **Goal:** Recover x **exactly** from measurements y

» **Setup:**

$$A \sim N(0, 1/n)$$

$n/N \to \delta$ (undersampling fraction)

$k/n \to \rho$ (measure of sparsity)

Donoho, Maleki, Montanari, "Message-passing algorithms for compressed sensing"

Phase Transition Curves

- **Goal:** Recover x **exactly** from the measurements y
- **Question:** When will compressed sensing fail?
  - Depends on ρ , δ , and the specific algorithm

- For l1-minimization (as in homework),

(P1)
$$\begin{aligned} \underset{x}{\text{minimize}} \quad & ||x||_1 \\ \text{subject to} \quad & Ax = y \end{aligned}$$

(convex relaxation of true problem of interest)

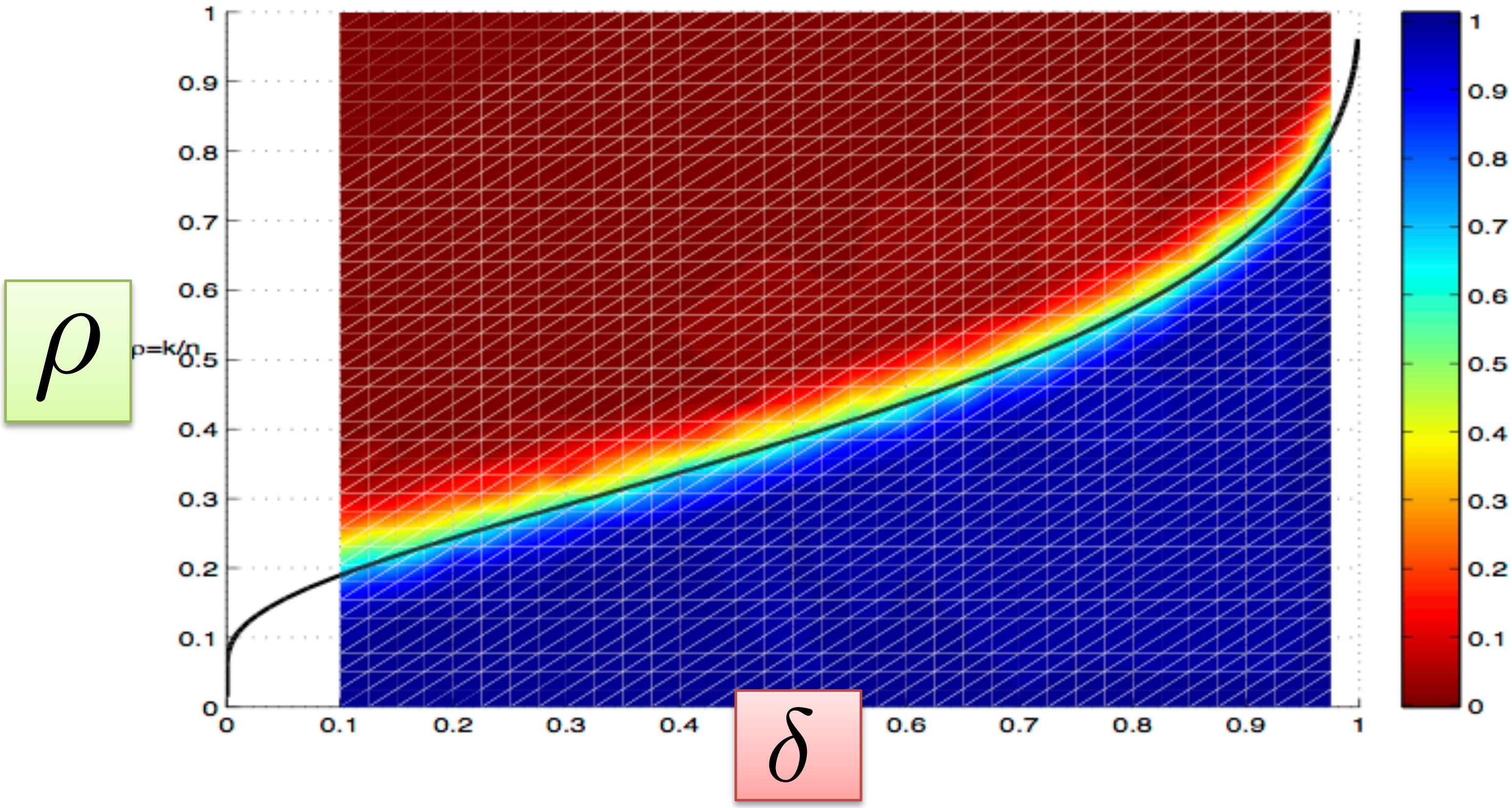


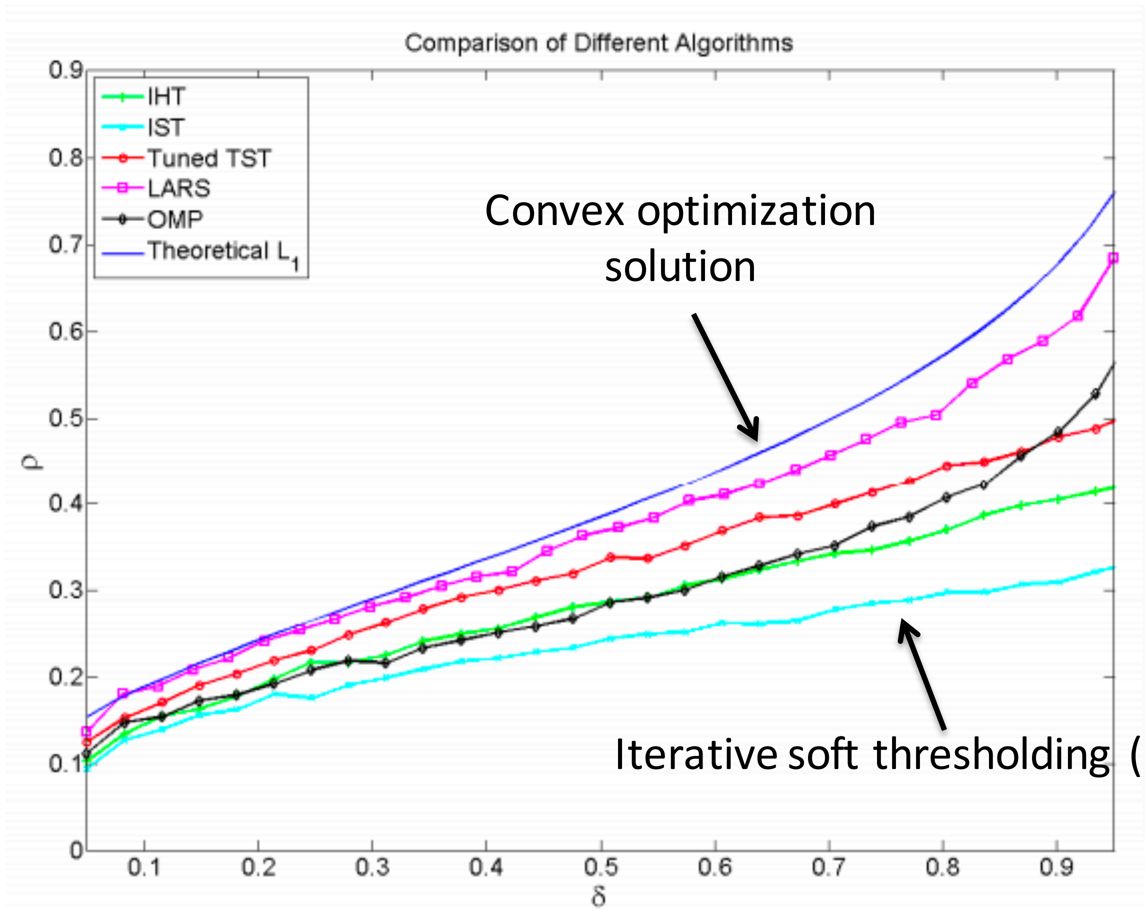
FIGURE 3. *Compressed Sensing from random Fourier measurements.* Shaded attribute: fraction of realizations in which $\ell_1$minimization (1.2) reconstructs an image accurate to within six digits. Horizontal axis: undersampling fraction $\delta = n/N$. Vertical axis: sparsity fraction $\rho = k/n$.

Donoho, Tanner, "Observed universality of phase transitions in high-dimensional geometry…"

Best Phase Transition

- Each reconstruction algorithm has a sharp phase transition curve
  - Below the curve is **exact recovery!** Above the curve is **total failure!**

ρ



δ

Beating IST: Approximate Message Passing

- Compressed Sensing LASSO: l1 minimization

(P2) $$\underset{x}{\text{minimize}} \quad \frac{1}{2}||y - Ax||_2^2 + \lambda||x||_1$$

- Simple Algorithm: **Iterative Soft Thresholding**

$$x^{(i+1)} = \eta_i \left(A^T z^{(i)} + x^{(i)} \right)$$
$$z^{(i)} = y - Ax^{(i)}$$

  - Converges to (P2), but only to (P1) in certain cases…
  - Computationally cheap

Beating IST: Approximate Message Passing

- Approximate Message Passing (AMP)
  - Add a "correction factor" to error vector
  - Correction term makes noise statistics look Gaussian at each iteration
  - Sparse signal + Gaussian noise fits the sparse denoising assumptions

$$x^{(i+1)} = \eta_i \left(A^T z^{(i)} + x^{(i)} \right)$$

$$z^{(i)} = y - Ax^{(i)} + \frac{1}{\delta} z^{(i-1)} \left\langle \eta'_{(i-1)} \left(A^T z^{(i-1)} + x^{(i-1)} \right) \right\rangle$$

  - Converges to (P1)!
  - Still computationally cheap

$$\eta'_i(x) = \frac{\partial}{\partial s} \eta_i(x)$$

$$\langle u \rangle = \frac{1}{N} \sum_{j=1}^{N} u(j)$$

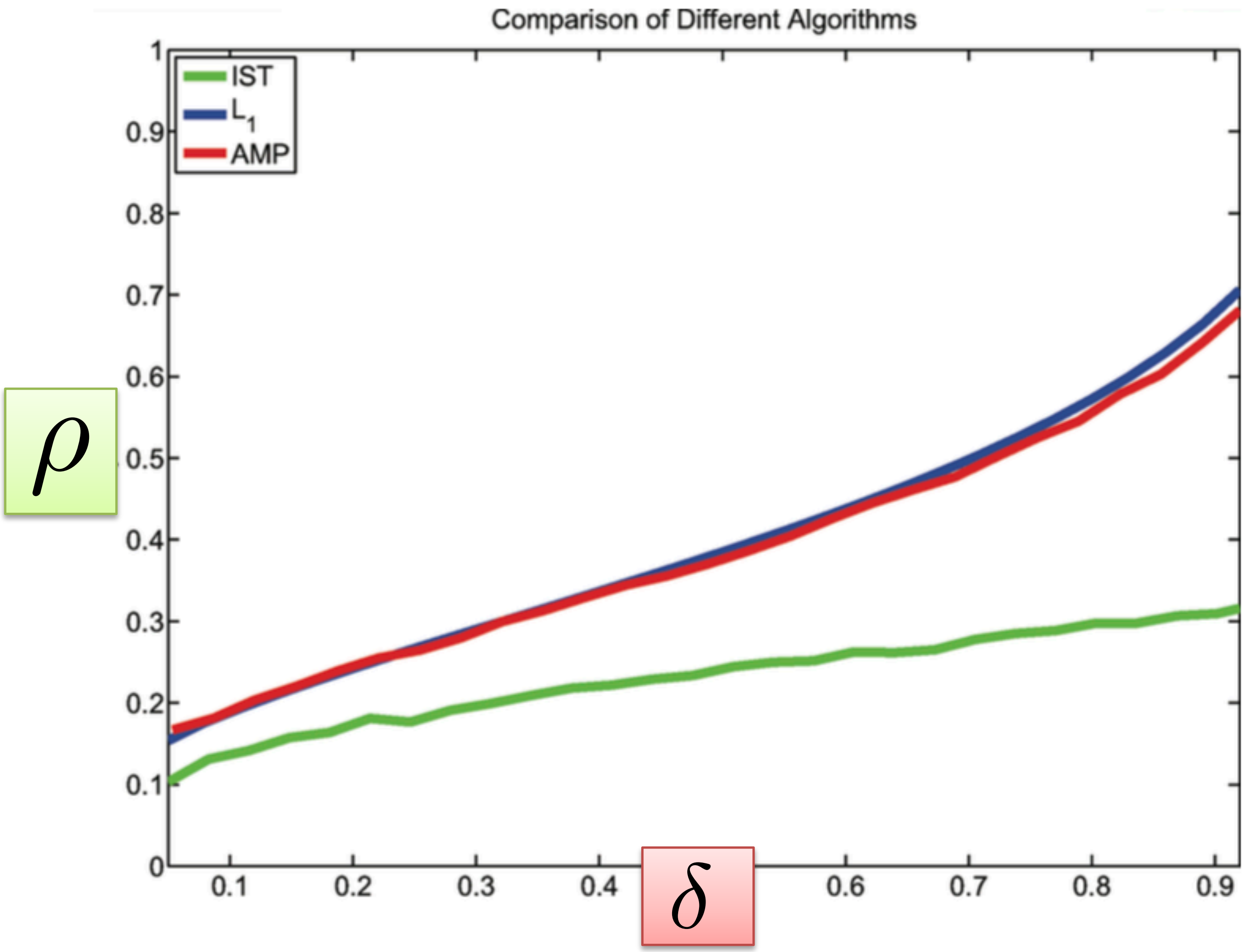
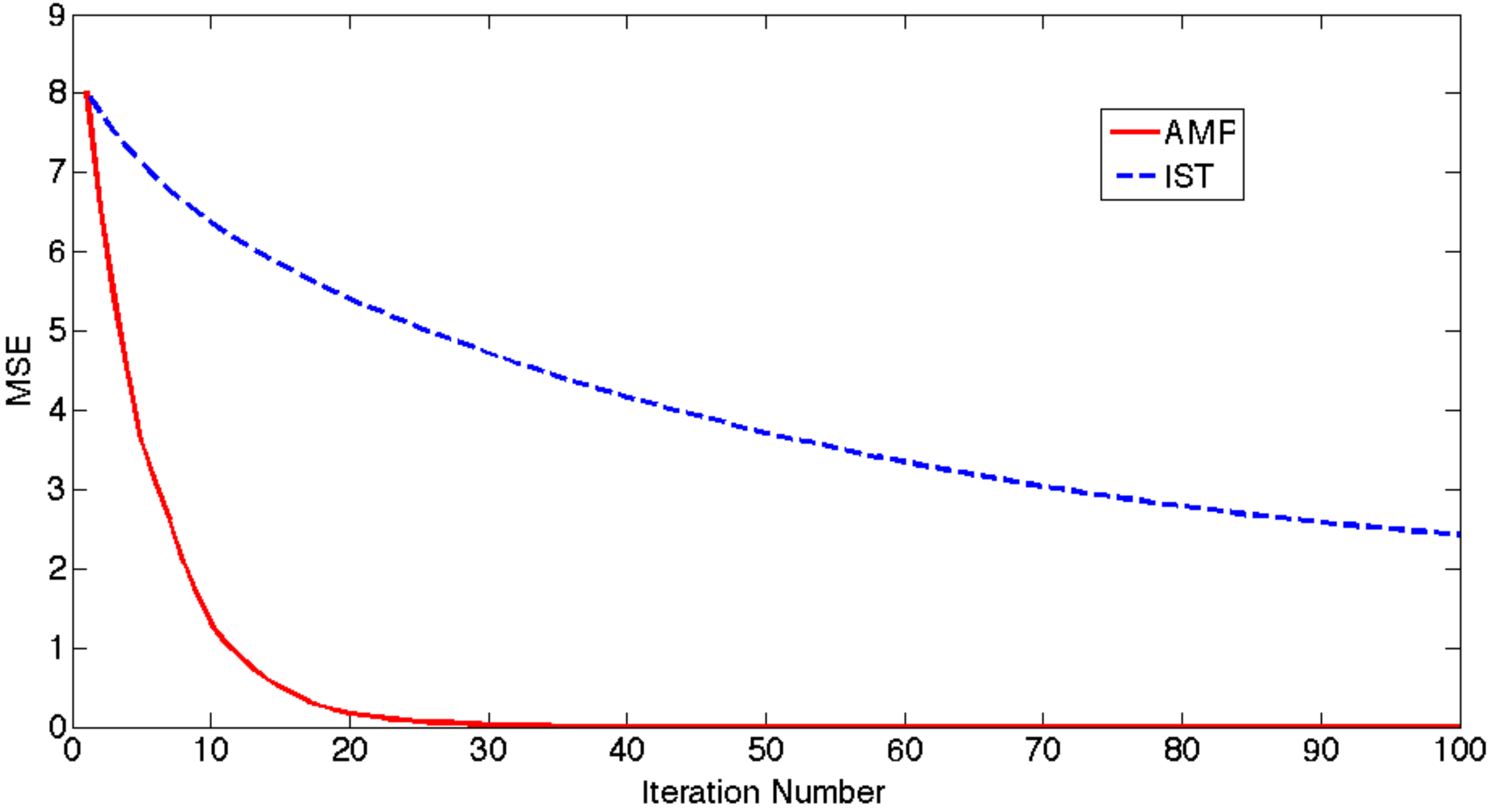


Fig. 3. Observed phase transitions of reconstruction algorithms. Red curve, AMP; green curve, iterative soft thresholding (IST); blue curve, theoretical ℓ_1 transition. Parameters of IST tuned for best possible phase transition (3). Reconstruction signal length $N = 1,000$. $T = 1,000$ iterations. Empirical phase transition is value of ρ at which success rate is 50%. Details are in *SI Appendix*.

AMP in Practice

```
% data gen
N = 1000;
n = 500;
delta = n/N;
k = 50;

% sensing matrix
A = 1/sqrt(n)*randn(n,N);

% signal
x = [zeros(N-k, 1); randn(k,1)]; x = x(randperm(N));
y = A*x;

for i=2:nitr
  % IST
  xts_ista(:,i) = SoftThresh(xts_ista(:,i-1) + mu_ist*A'*(y - A*xts_ista(:,i-1)), mu_ist*lambda_ist);
  mse_ista(i) = norm(xts_ista(:,i) - x);

  % AMP
  % calculate x_{t+1}
  gamma_amp = xts_amp(:,i-1) + A'*zts_amp(:,i-1);
  tmp = sort(abs(gamma_amp), 'descend');
  lambdas_amp(i) = tmp(n);
  xts_amp(:,i) = SoftThresh(gamma_amp, lambdas_amp(i));
  % calculate z_{t+1}
  zts_amp(:,i) = (y - A*xts_amp(:,i)) + ...
      zts_amp(:,i-1)/n*sum(abs(gamma_amp) > lambdas_amp(i));
  mse_amp(i) = norm(xts_amp(:,i) - x);
```

AMP in Practice



Phase Transition Curves Summary

» For a particular matrix A and reconstruction algorithm:

1. Choose ρ, δ
2. Monte Carlo simulation of many problem instances
3. Count ratio of successes vs. total problem instances
4. Repeat

» Extensive theory for phase transitions

- » Deep connections between compressed sensing and denoising
- » Provable optimality of Approximate Message Passing
- » (i.e. it's not all simulation)

» Framework for testing other sparse models/algorithms

» This slide intentionally left blank.

Extending Sparsity

- Original Compressed Sensing formulation
  - Recover a sparse signal x from incoherent measurements y

- Are there other notions of sparsity?
  - Block-sparse vector
  - **Low rank matrix**
  - Permutation matrix
  - Sign vector

General Notion of Sparsity

» "Low-dimensional subspace embedded in high-dimensional signal"

» **Atomic Norm Decomposition**

  » Signal can be decomposed as a linear combination of building blocks

» **Sparse Vectors:** linear combination of standard basis

$$\mathbf{x} = \sum_i c_i \mathbf{e_i}$$

» **Low Rank Matrices:** linear combination of rank-1 matrices

$$\mathbf{X} = \sum_i c_i \mathbf{u_i} \mathbf{v_i}^T$$

"The Convex Geometry of Linear Inverse Problems," Venkat Chandrasekaran, Benjamin Recht, Pablo A. Parrilo, and Alan Willsky

“Netflix Problem”

- After you watch a cool movie or TV show, you rate the it on a scale of 1 to 5
  - House of Cards: 5
  - Orange is the New Black: 3
  - Real World: 1
  - …

- Based on your (and others’) ratings, can I predict how you (and others) will rate a movie/show you haven’t seen?

Collaborative Filtering

» Predict ratings for movies across all users

Movies

$$R \in \mathbb{R}^{m \times n} \qquad \begin{bmatrix} 2 & 3 & ? & ? & 5 & ? \\ 1 & ? & ? & 4 & ? & 3 \\ ? & ? & 3 & 2 & ? & 5 \\ 4 & ? & 3 & ? & 2 & 4 \end{bmatrix}$$

Users

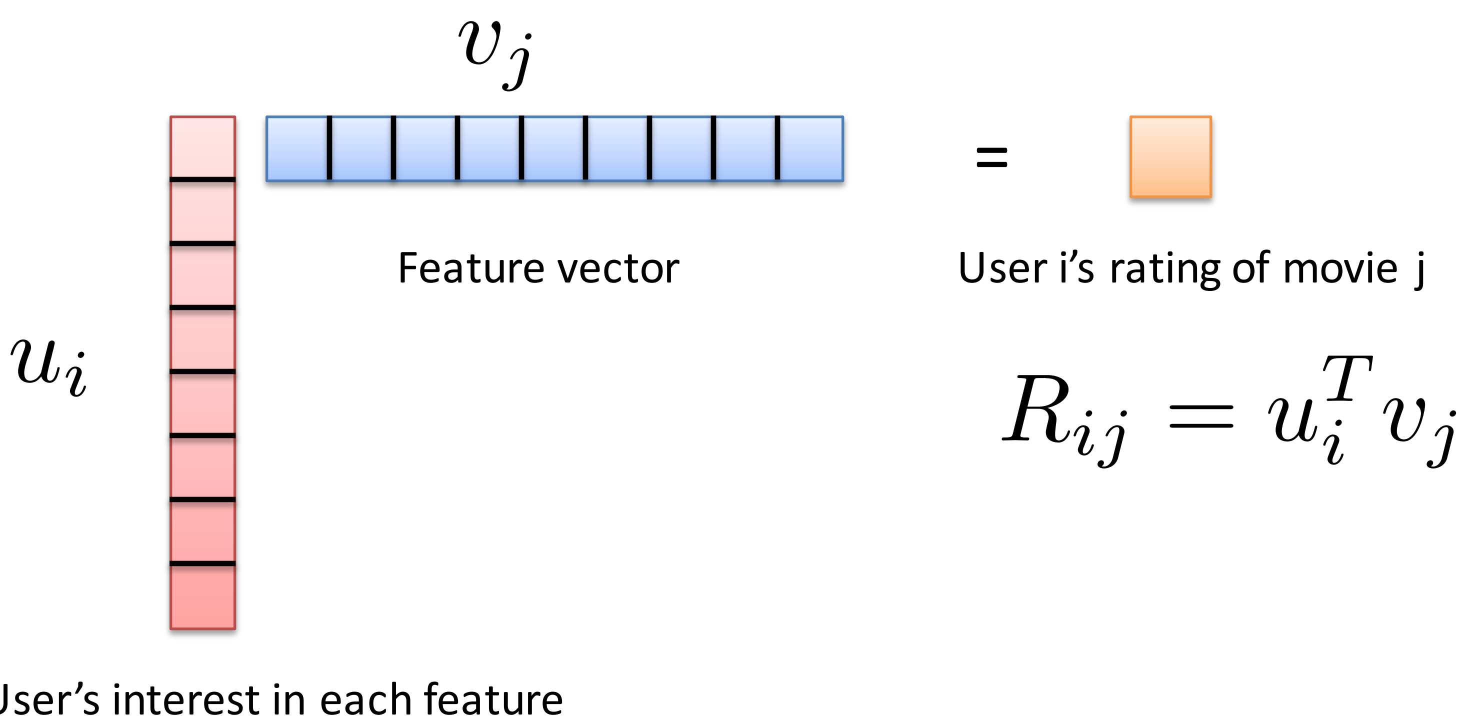
» Only know R_{ij} for $(i, j) \in \Omega$

  » Don't have ratings for every movie from every user

http://www.ece.ust.hk/~palomar/ELEC5470_lectures/17/slides_low_rank_matrix_optim.pdf

Low Rank Matrix Decomposition

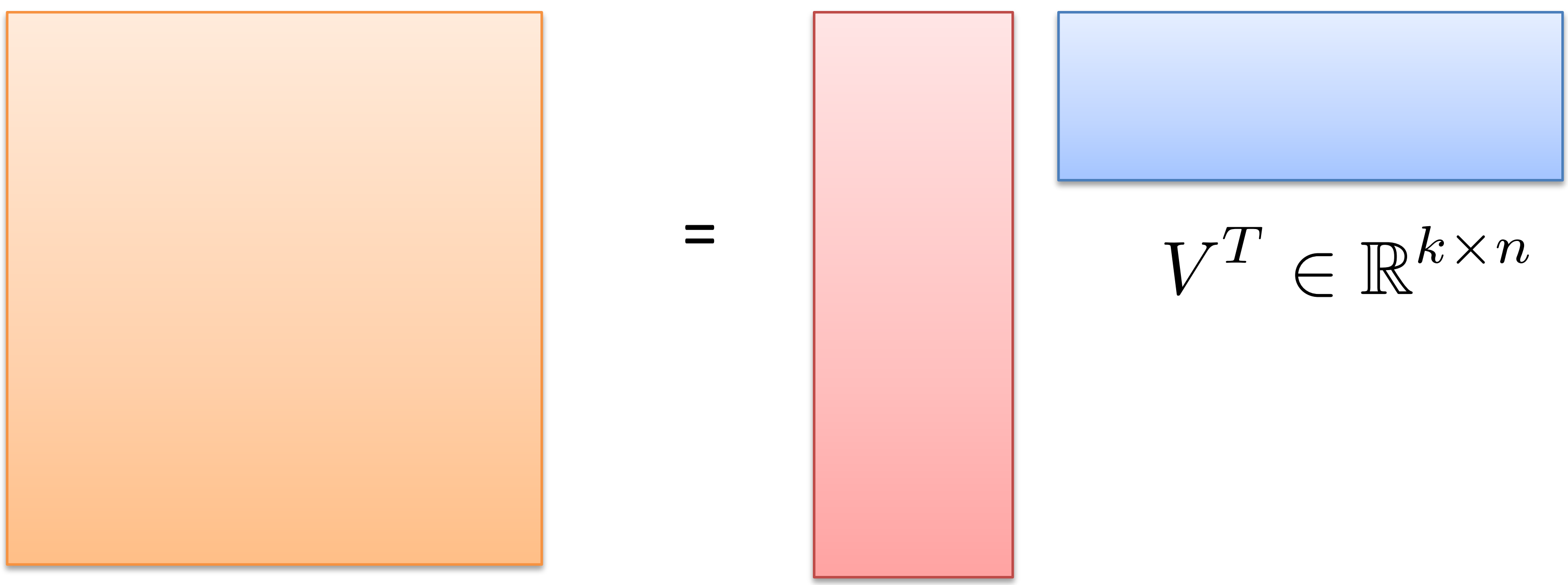
- Can "explain" a movie rating by a small (k) number of features
  - Actors, genre, storyline, length, year, …
- Each user has a preference for the features



$$R_{ij} = u_i^T v_j$$

Low Rank Matrix Decomposition

- Can "explain" a movie rating by a small (k) number of features
  - Actors, genre, storyline, length, year, …
- Each user has a preference for the features
- Matrix R is low rank, with rank k << m, k << n



$$V^T \in \mathbb{R}^{k \times n}$$

$$R = UV^T$$

$$U \in \mathbb{R}^{m \times k}$$

Low Rank Matrix Completion

» Given that the true R is low rank, find a matrix X that is low rank and agrees with R at the observed entries:

$$\begin{aligned} \underset{X}{\text{minimize}} \quad & \text{rank } X \\ \text{subject to} \quad & \mathcal{A}(X) = y \end{aligned}$$

$$(X_{ij} = R_{ij} \quad \forall (i,j) \in \Omega)$$

$$y = \mathcal{A}(R)$$

$$\mathcal{A}(X) = \{X_{ij} \;:\; (i,j) \in \Omega\}$$

Low Rank Matrix Completion

» Given that the true R is low rank, find a matrix X that is low rank and agrees with R at the observed entries:

(P3)
$$\begin{aligned} \underset{X}{\text{minimize}} \quad & ||X||_* \\ \text{subject to} \quad & \mathcal{A}(X) = y \end{aligned}$$

Convex Relaxation

$$||X||_* = \sum_i \sigma_i$$

(sum of singular values)

Low Rank Matrix Completion

- How can we solve the low rank matrix completion problem?

- Intuition:
  - A low rank matrix has a small number of non-zero singular values
  - We see a linear mixture of these singular values (through SVD)
  - How about we apply IST on the singular values of X?

- POCS Algorithm:
  1. Take the SVD: $X = P\Sigma Q^T$ ← Not low rank
  2. Soft Threshold: $\hat{\Sigma} = S_\lambda(\Sigma)$
  3. Form Matrix: $\hat{X} = P\hat{\Sigma}Q^T$ ← Low rank but inconsistent with data
  4. Enforce data consistency:

$$X_{ij} = R_{ij} \quad \forall (i,j) \in \Omega$$

When Does the Algorithm Work?

- Matrix Restricted Isometry Property (RIP):
  - An operator A satisfies Matrix-RIP if, for all X with rank X at most r,

$$(1-\epsilon)||X||_F^2 \leq ||\mathcal{A}(X)||_2^2 \leq (1+\epsilon)||X||_F^2$$

  - (describes the incoherence of the measurement operator A)

- **Theorem:**
  - If X is rank-r, A satisfies Matrix-RIP, and y = A(x),
  - Then X will be the optimum of (P3)

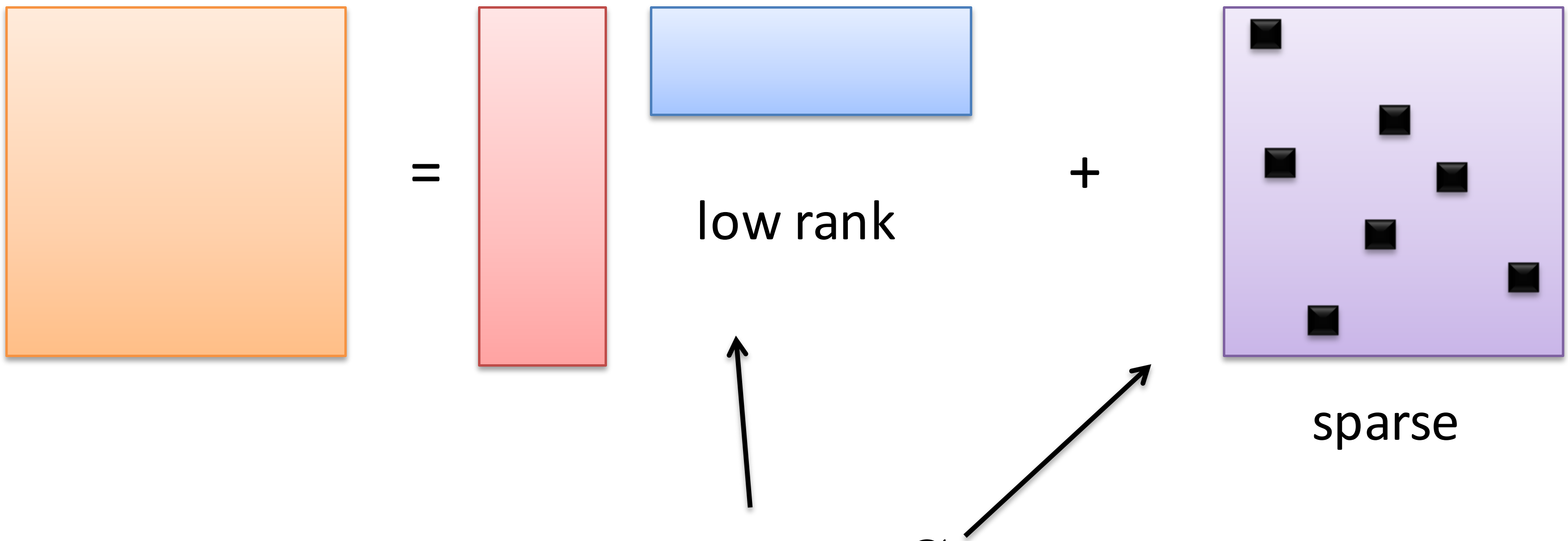
- **Uniqueness:**
  - There are multiple matrices that satisfy the problem (P3), but only one will be low rank

- Other Assumptions? (yes – incoherence on rows/cols of R)

(Recht, Fazel, Parrilo)

Sparse + Low Rank Matrix Decomposition

» Low rank matrix corrupted by sparse errors



$$R = L + S$$

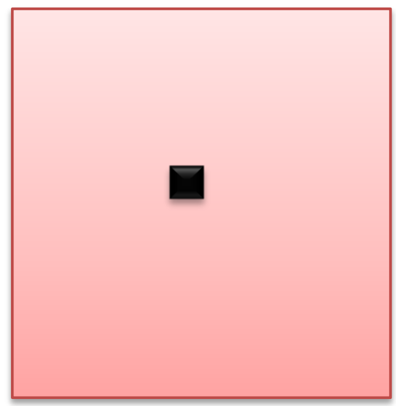
» **Task:** Given R, find L and S exactly

» (not well-defined in general)

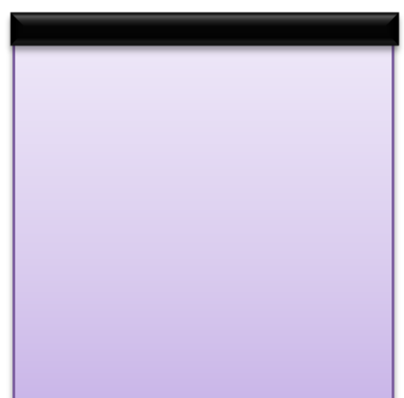
"Sparse and Low-Rank Decompositions," Chandrasekaran, Sanghavi, Parrilo, Willsky

Sparse + Low Rank Matrix Decomposition

» Base Cases

L = 

No hope of separating from S

S = 

No hope of recovering first row of L

Sparse + Low Rank Matrix Decomposition

» Under certain assumptions, we can recover L and S through

$$\begin{aligned} \underset{L,S}{\text{minimize}} \quad & ||L||_* + \lambda ||S||_1 \\ \text{subject to} \quad & R = L + S \end{aligned}$$

Candès E, Li X, Ma Y, Wright J. "Robust principal component analysis? "

Input Video

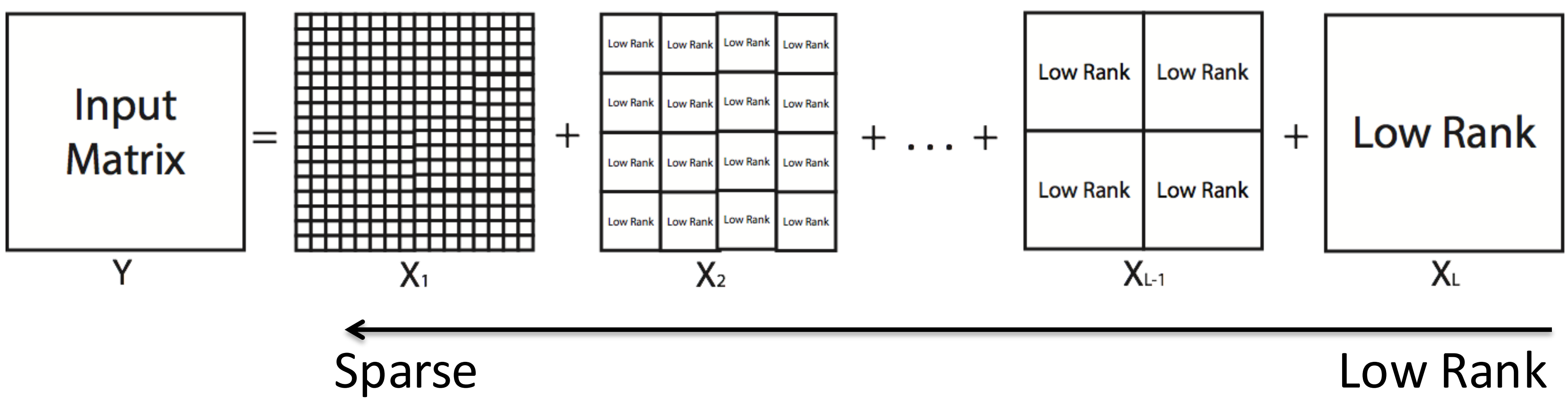
Low Rank

+

Sparse

Beyond Low Rank + Sparse: Multi-scale Low Rank Modeling

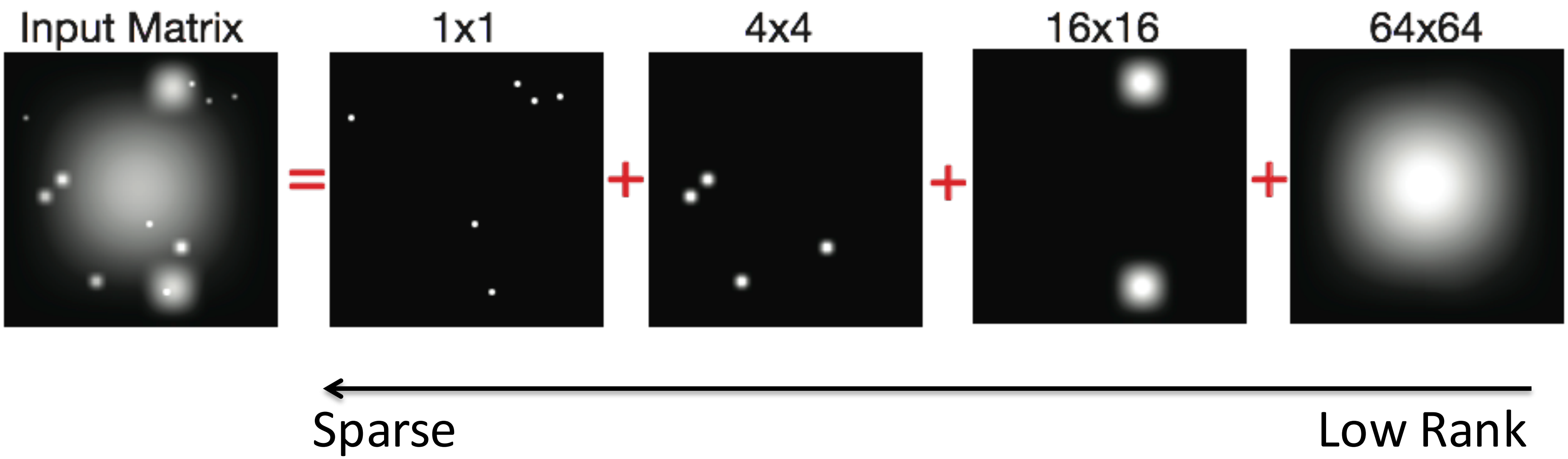
- Sum of matrices with increasing scales of correlation
  - Exploit **all** scales of correlation



Work by Frank Ong

Beyond Low Rank + Sparse: Multi-scale Low Rank Modeling

- Sum of matrices with increasing scales of correlation
  - Exploit **all** scales of correlation



Work by Frank Ong

Input Video

Multi-scale Low Rank

Work by Frank Ong

Multi-scale Low Rank Decomposition

$$\underset{X_i}{\text{minimize}} \quad \sum_{i=0}^{L-1} \lambda_i \|X_i\|_{(i)}$$

$$\text{subject to} \quad Y = \sum_{i=0}^{L-1} X_i$$

» POCS Algorithm:

1. Enforce block low rank for each X_i (Block-wise SVD + IST)
2. Enforce data consistency

Work by Frank Ong

Application: Face Shadow Removal

» Given: face images with different illuminations

» **Want to remove shadows**

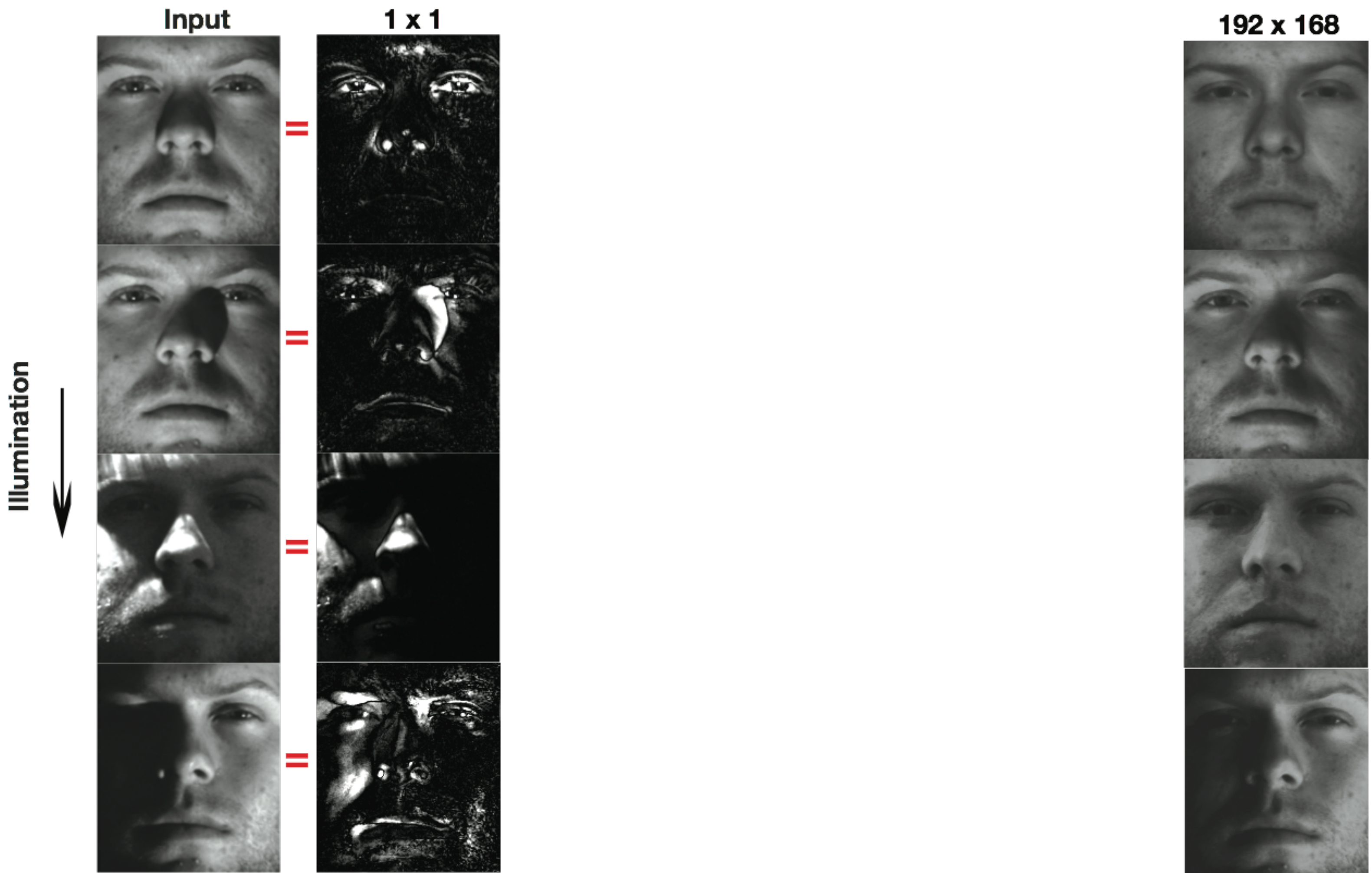
» **Faces are low rank**

» **Shadows are not**

Work by Frank Ong

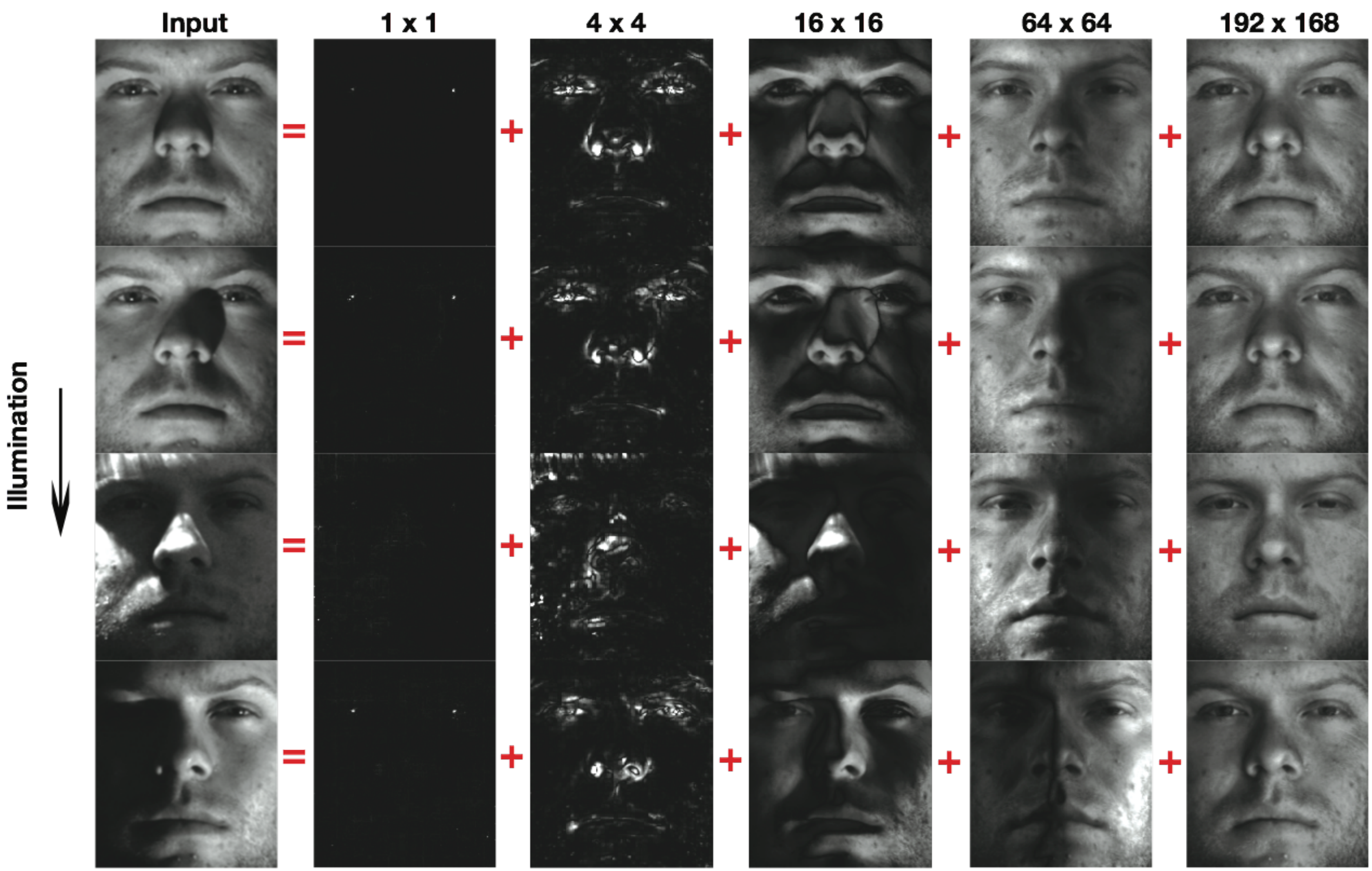
Application: Face Shadow Removal

Low Rank + Sparse



Application: Face Shadow Removal

Multi-scale Low Rank

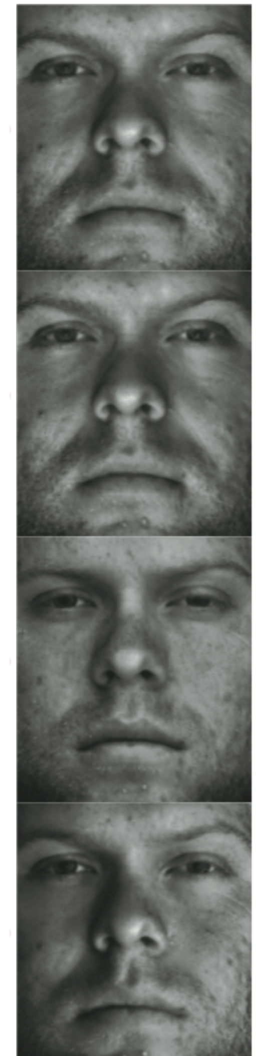


Application: Face Shadow Removal

Input

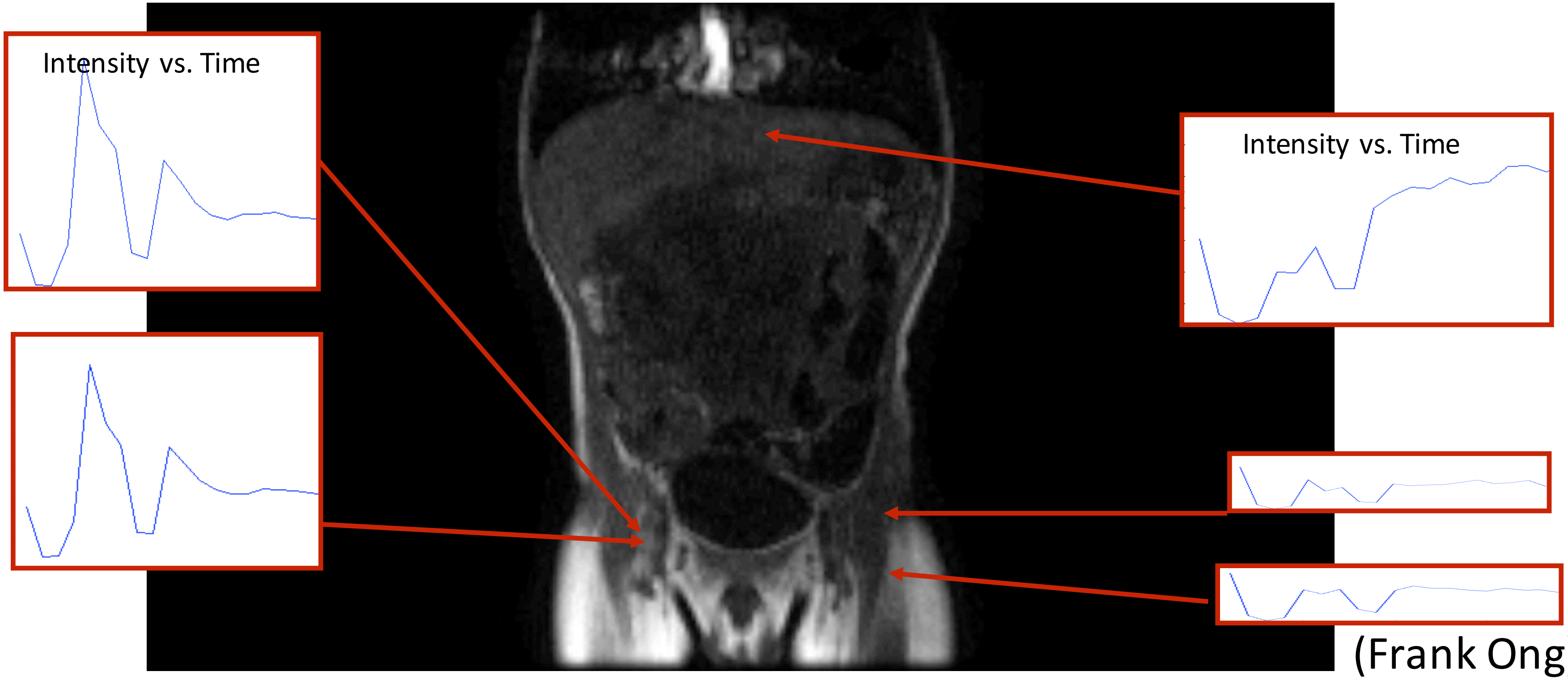
Low Rank + Sparse

Multi-scale Low Rank



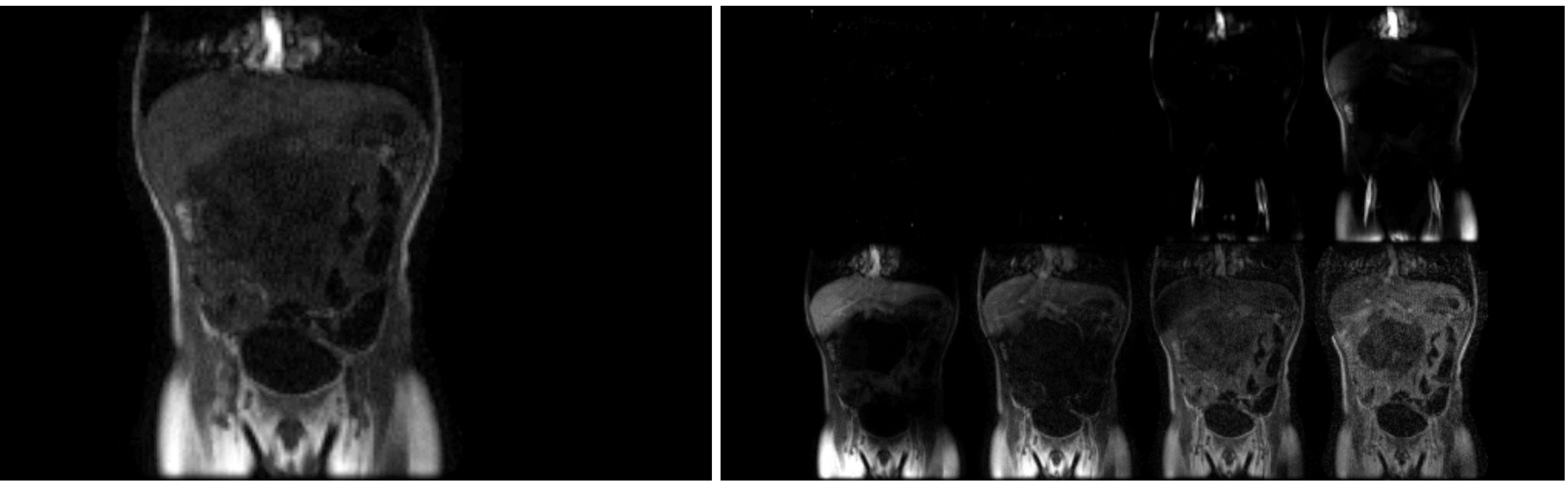
Application: Dynamic Contrast Enhanced MRI

- » Contrast agent injected into patient
- » A series of images are acquired over time
- » Different blood permeability gives different signature signal



(Frank Ong)

Application: Multi-scale Low Rank MRI



(Frank Ong)

Dictionary Learning

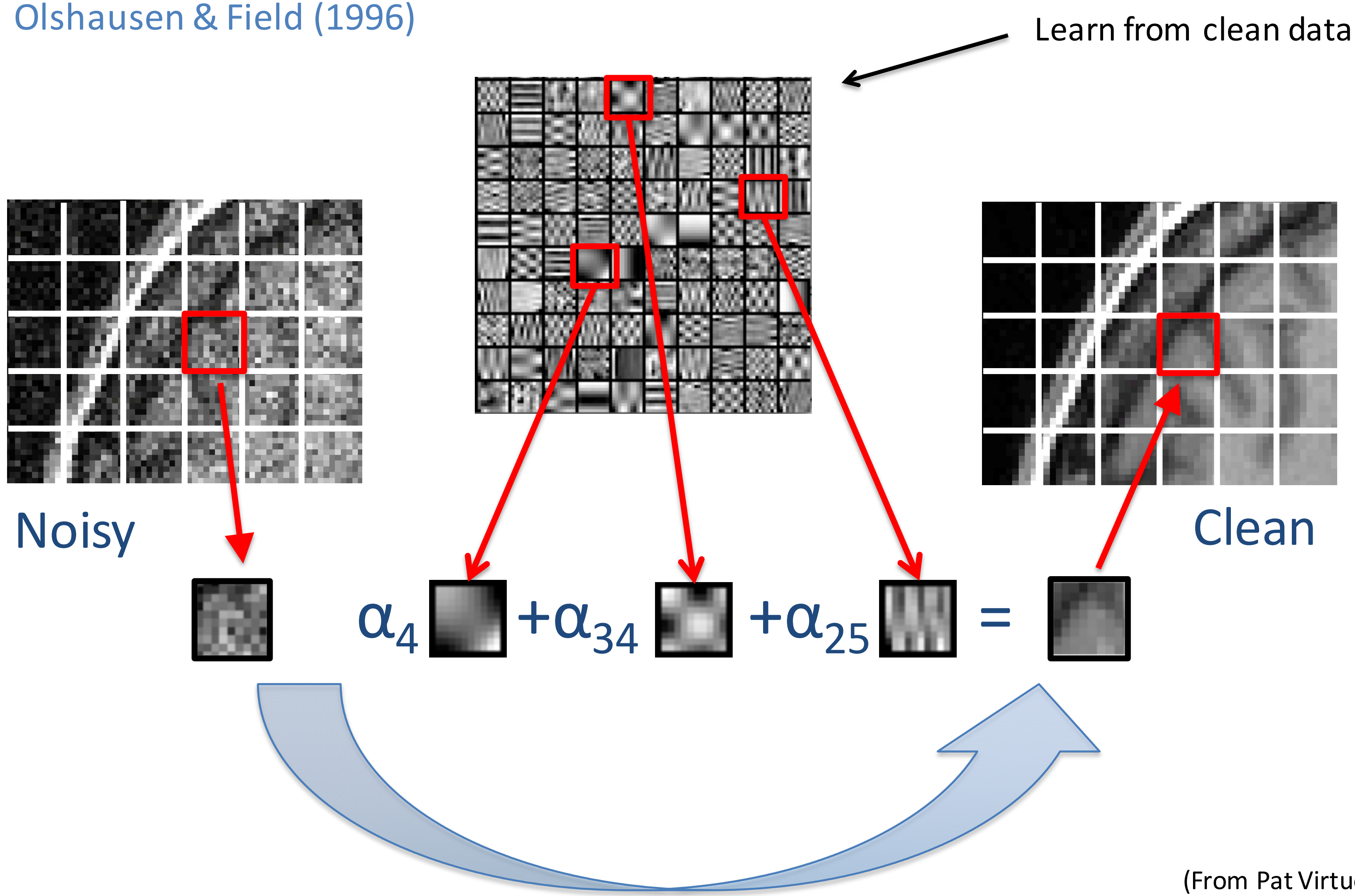
- So far, we always know how our signal is sparse
  - Reverberation/communication systems: time domain
  - Natural/medical images – wavelet domain
  - Low rank matrices – singular values

- What if we don't know how our signal is sparse?

- **Dictionary Learning –** *learn* a sparse representation
  - Feed in many examples of signals from a class of interest
  - Learn a sparse representation for all the signals in the class
  - Often *overcomplete representation* – more elements than unknowns

Dictionary Learning

Olshausen & Field (1996)



(From Pat Virtue)

Dictionary Learning Formulation

- Given training signals

$$\mathbf{Y} = \begin{bmatrix} \mathbf{y_1} & \cdots & \mathbf{y}_N \end{bmatrix}$$

- Find (jointly) a dictionary D and coefficients X such that

$$\mathbf{Y} = \mathbf{DX}$$

- Where x_i has at most k non-zero elements

- Many ways to solve: MLE, MAP, alternating minimization, …

K-SVD: State of the Art Dictionary Learning

- Extension of K-means
  - K-means: each signal is represented by a single cluster centroid
  - Vector Quantization: each signal is represented by a single cluster centroid, times a weight
  - K-SVD: each signal is represented by a linear combination of a small number of cluster centroids (atoms)

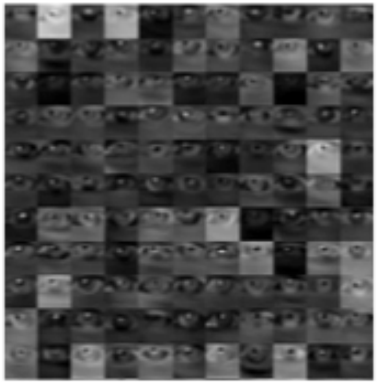
Michal Aharon, Michael Elad, Alfred Bruckstein, "K-SVD: An Algorithm for Designing Overcomplete Dictionaries for Sparse Representation"

Leveraging Non-stationary Statistics

Registration & local dictionary learning

Face compression

Bryt and Elad. JVCIR, 19(4):270 – 282, 2008.



1. Register faces
2. Train different dictionaries at different locations
3. (e.g. eye)

(From Pat Virtue)

Compressed Sensing Potpourri Summary

- Compressed sensing theory is reaching maturity
  - Large focus on specialized applications
- Compressed sensing algorithms in active development
  - AMP and generalizations
  - Phase transition theory
- More general notions of sparsity are still blossoming
  - Deep connections to high-dimensional statistics
  - Manifold learning, subspace clustering