MATH 4280

Lecture Notes 3: Sparsity and compressed sensing

Compressed sensing

- Inherent structure in data implies the data admits a sparse representation in a suitable coordinate system
- Only a few parameters are required to characterize the data
- Instead of collecting high dimensional data, it is possible to acquire compressed measurement based on recent advent of compressed sensing
- Reconstruction of full signal using compressed or random measurements

Introduction

- Let $x \in \mathbb{R}^n$ be a signal that is K-sparse in the basis Ψ
- Assume y $\in \mathbb{R}^p$ be a compressed measurement, with $K < p \ll n$, given by

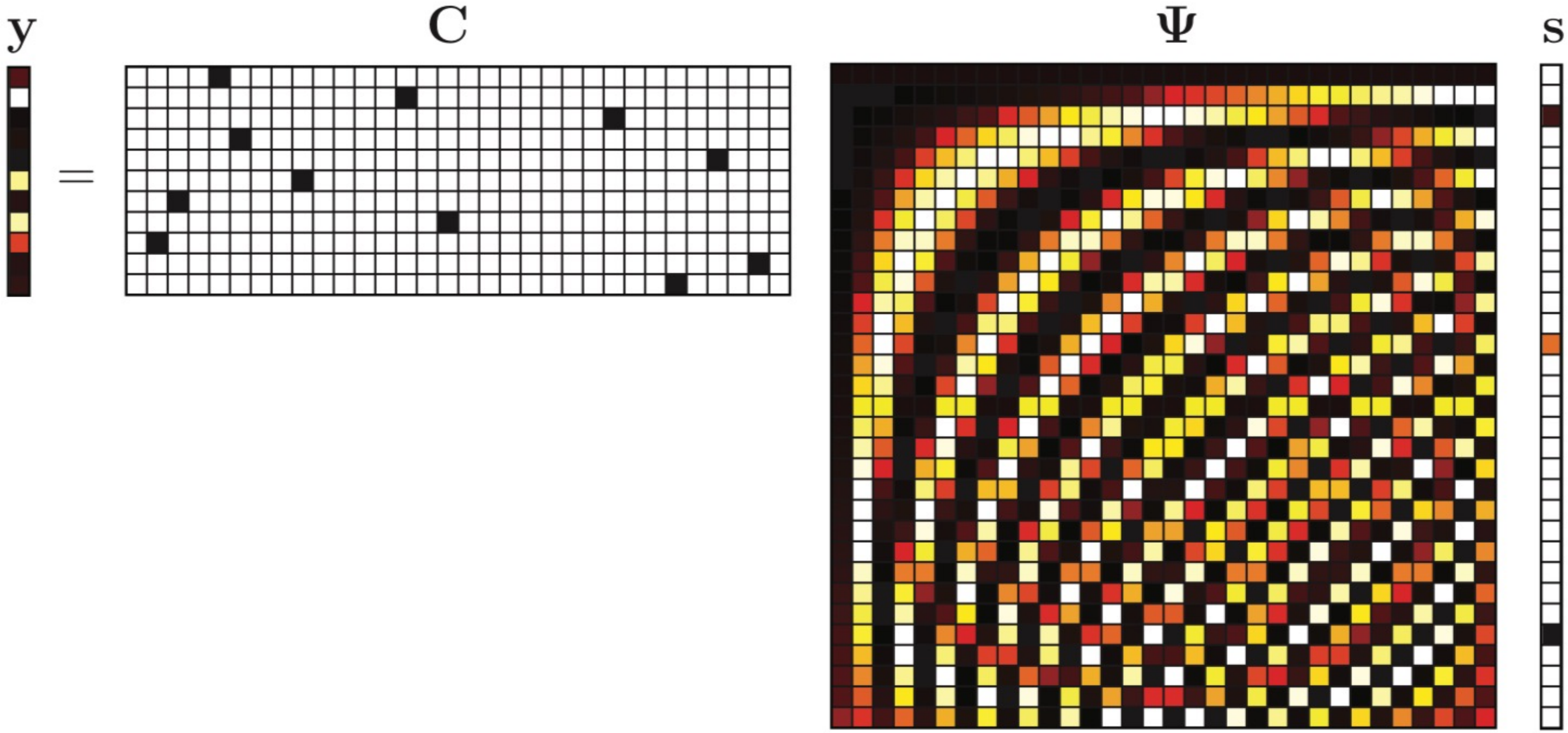
$$\mathbf{y} = \mathbf{C}\mathbf{x}$$

Spatial or time related. It maps the original signals s to the measured signal y.

- The $\boxed{\text{measurement matrix } \mathrm{C} \in \mathbb{R}^{p \times n}}$ represents a set of p linear measurements on the state x
- The choice of C is important (e.g. random projections)
- The goal is to find the sparest vector s that is consistent with the measurement

$$\mathbf{y} = \mathbf{C}\boldsymbol{\Psi}\mathbf{s} = \boldsymbol{\Theta}\mathbf{s}$$

$$\mathbf{y} = \mathbf{C}\boldsymbol{\Psi}\mathbf{s} = \boldsymbol{\Theta}\mathbf{s}$$



Mathematical formulation

- We find the sparest solution by solving the following

$$\hat{\mathbf{s}} = \underset{\mathbf{s}}{\operatorname{argmin}} \|\mathbf{s}\|_0 \quad \text{subject to} \quad \mathbf{y} = \mathbf{C}\boldsymbol{\Psi}\mathbf{s}$$

Non-convex

- The notation $\|\cdot\|_0$ denotes the ℓ_0 pseudo norm, defined as the number of nonzero elements
- The above is a combinatorial problem, which is expensive to solve
- Under some conditions on the matrix C, it is possible to formulate the problem as follows

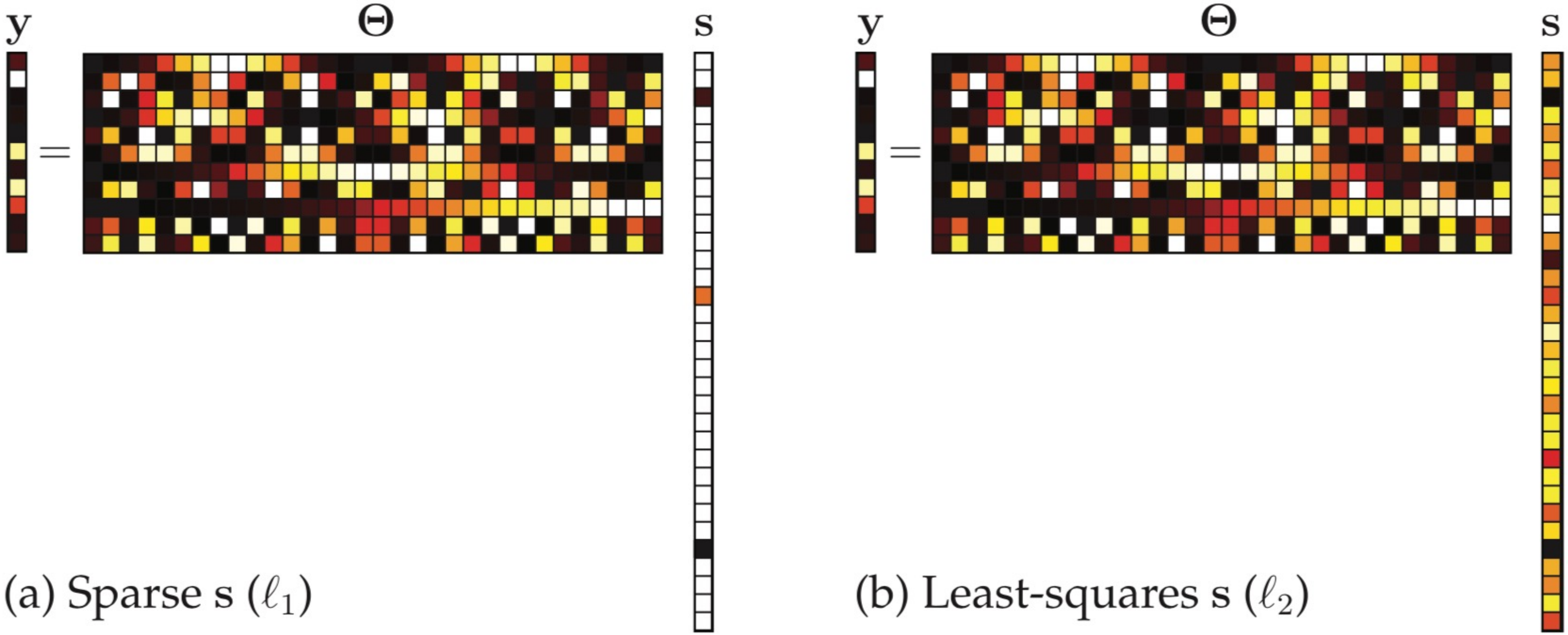
$$\hat{\mathbf{s}} = \underset{\mathbf{s}}{\operatorname{argmin}} \|\mathbf{s}\|_1 \quad \text{subject to} \quad \mathbf{y} = \mathbf{C}\boldsymbol{\Psi}\mathbf{s}$$

Convex

where

$$\|\mathbf{s}\|_1 = \sum_{k=1}^{n} |s_k|$$

L1 vs L2 minimization



Alternative formulation

- When the measurement y contains additive noise of magnitude ε, the problem can be formulated as

$$\hat{\mathbf{s}} = \underset{\mathbf{s}}{\operatorname{argmin}} \|\mathbf{s}\|_1, \quad \text{subject to} \quad \|\mathbf{C}\boldsymbol{\Psi}\mathbf{s} - \mathbf{y}\|_2 < \varepsilon$$

- Or

$$\hat{\mathbf{s}} = \underset{\mathbf{s}}{\operatorname{argmin}} \|\mathbf{C}\boldsymbol{\Psi}\mathbf{s} - \mathbf{y}\|_2 + \lambda\|\mathbf{s}\|_1$$

Example: underdetermined system

- We consider a linear system $\mathbf{y} = \mathbf{\Theta s}$ with $p = 200$ and $n = 1000$
- Use ℓ_1 minimization to obtain a sparse solution

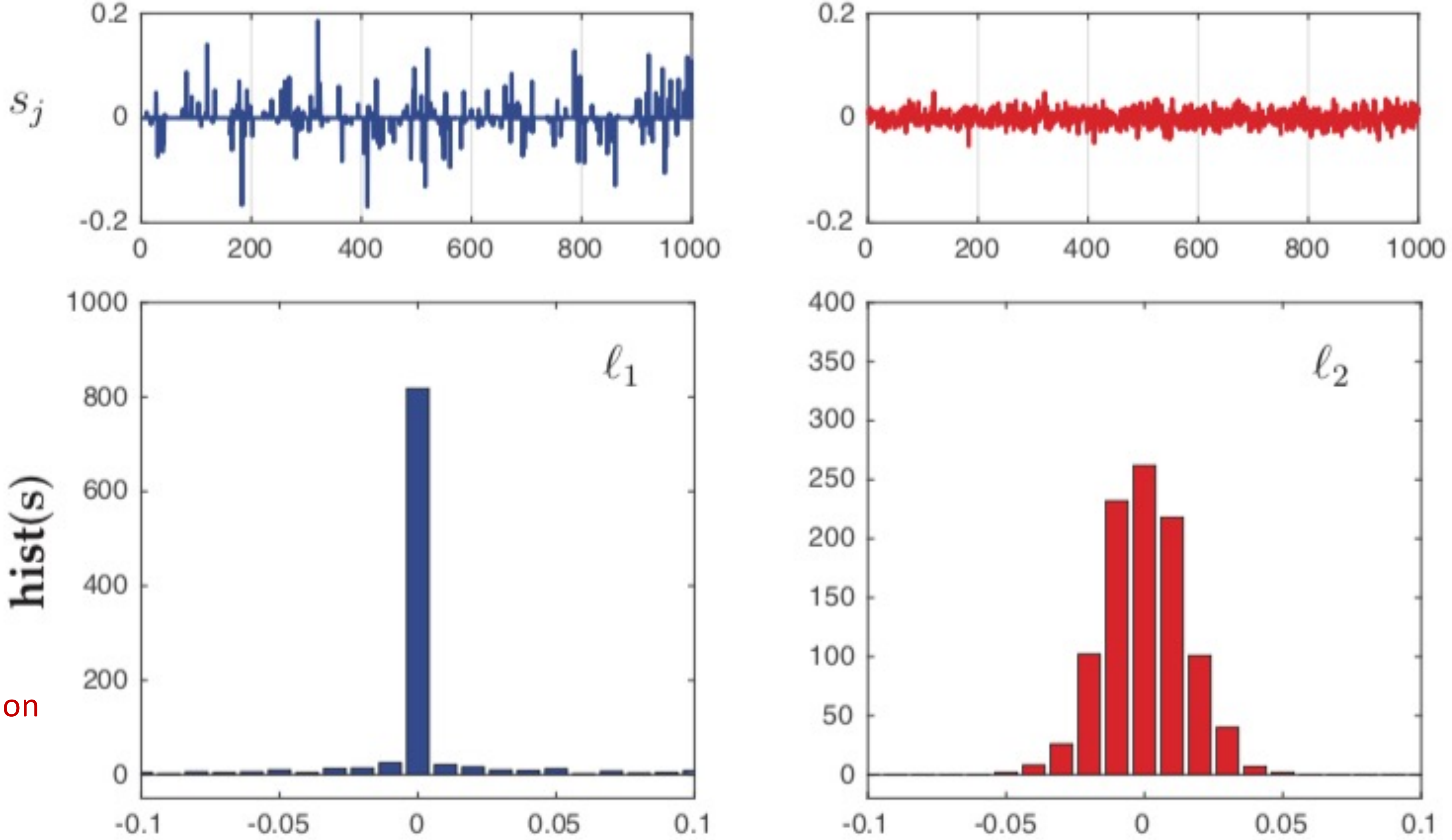
```
% Solve y = Theta * s for "s"
n = 1000;   % dimension of s
p = 200;    % number of measurements, dim(y)
Theta = randn(p,n);
y = randn(p,1);

% L1 minimum norm solution s_L1
cvx_begin;
    variable s_L1(n);
    minimize( norm(s_L1,1) );
    subject to
        Theta*s_L1 == y;
cvx_end;

s_L2 = pinv(Theta)*y;  % L2 minimum norm solution s_L2
```

cvx stands for convex optimization,
cvxpy is a python package

Use the CVX package



Sparse solution

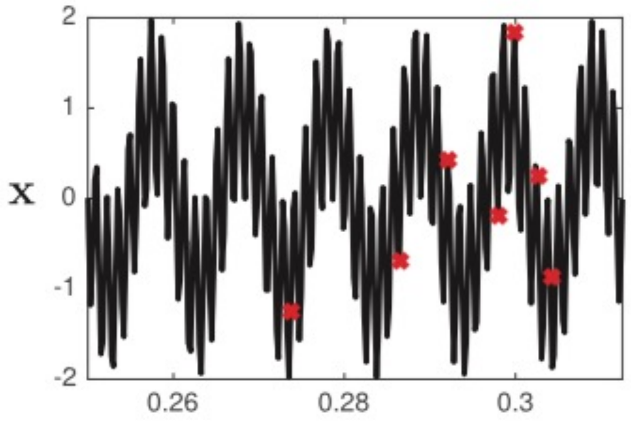
Example: recovering a signal

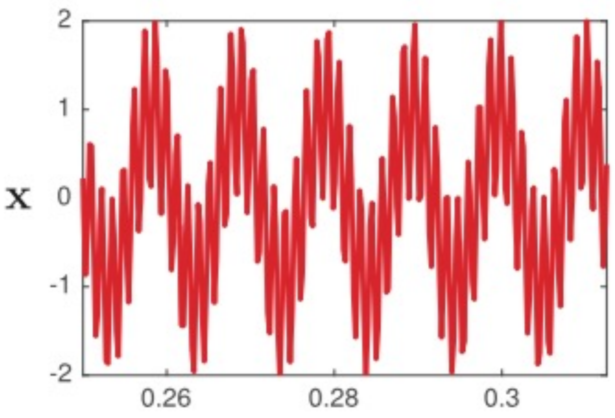
- Consider the following signal, and sample at some random locations

$$x(t) = \cos(2\pi \times 97t) + \cos(2\pi \times 777t)$$

- Sample random locations in time
- Note that the signal is sparse in cosine basis
- Formulation $\mathbf{y} = \mathbf{C}\boldsymbol{\Psi}\mathbf{s} = \boldsymbol{\Theta}\mathbf{s}$

Original signal





Recovered signal

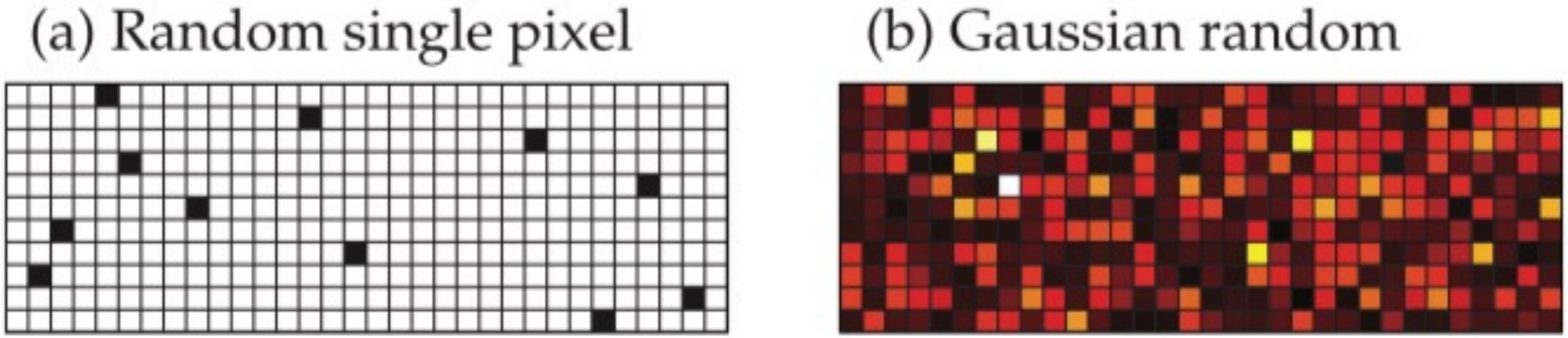
Some theoretical statements

- Recall the system $\mathbf{y} = \mathbf{C\Psi s} = \mathbf{\Theta s}$
- The matrix $\mathbf{C\Psi}$ should satisfy the restricted isometry property (RIP)

$$(1-\delta_K)\|\mathbf{s}\|_2^2 \le \|\mathbf{C\Psi s}\|_2^2 \le (1+\delta_K)\|\mathbf{s}\|_2^2$$

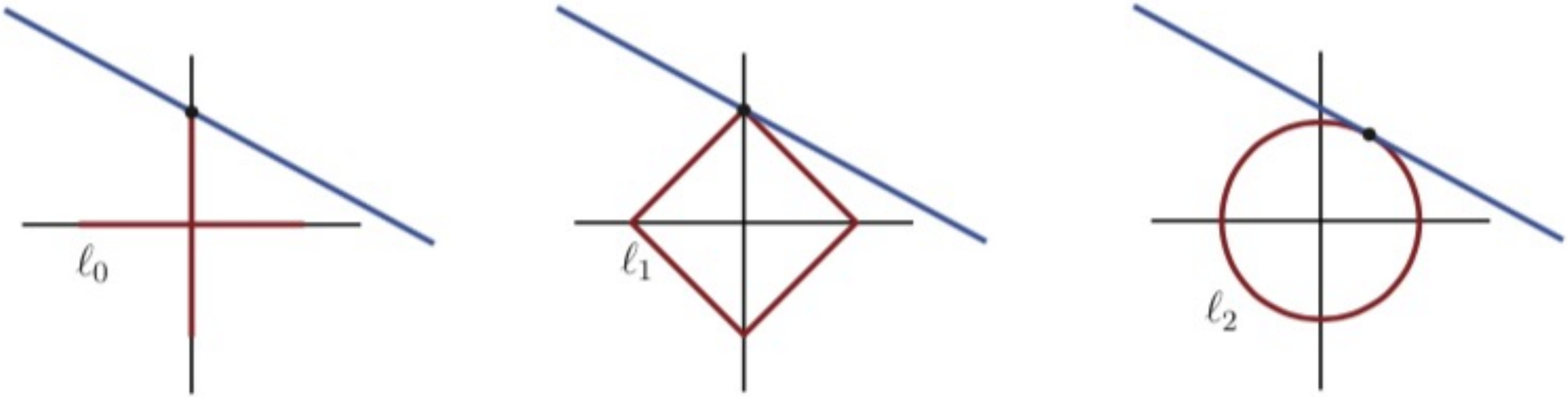
  That is, the matrix $\mathbf{C\Psi}$ is almost unitary

- The number of measurement scales as $p \sim \mathcal{O}(K \log(n/K))$
- Examples of measurement matrix



Geometric properties of norms

- Blue line represents feasible solution set
- Red line represents level set of minimum norm



Gives a sparse solution

Sparse representation for classification

- Sparse representation for classification (SRC) uses sparsity ideas
- Build an overcomplete library or dictionary (some collected or measured data)
- Given a new data, perform classification using L1 minimization

Build an overcomplete library

A given image/data with resolution 192 x 168

The corresponding downsampled image with resolution 12 x 10

Θ

Form the library matrix by putting each downsampled data in column

Person 7

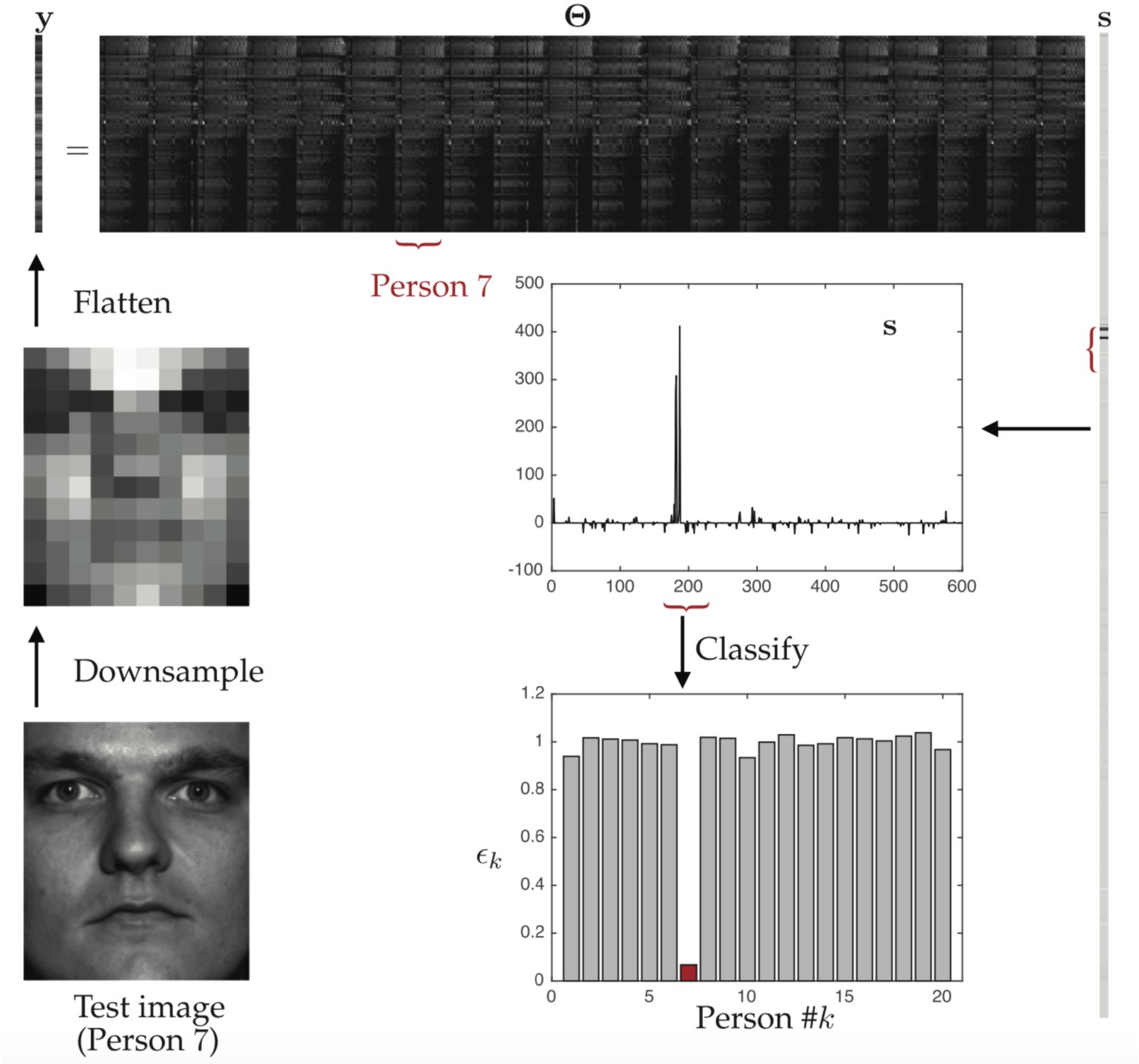
Each person can have several columns

Idea of SRC

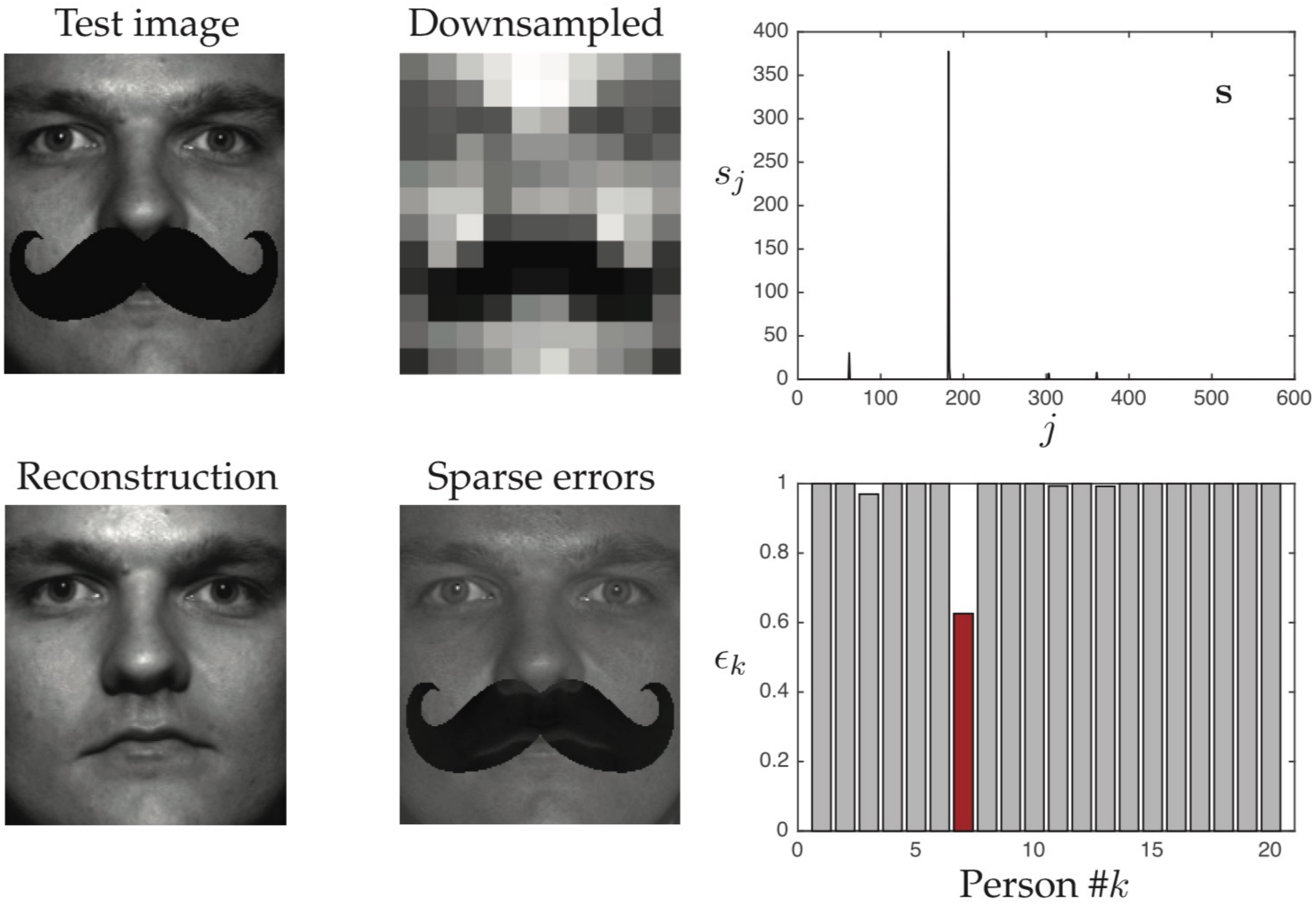
- Downsample the new data
- Solve a L1 minimization to obtain a sparse vector s
- The nonzero coefficients of s provide the required classification

$$\min \| s \|_1$$

$$subject\ to\ \ \| y - \Theta s \| \leq \varepsilon$$



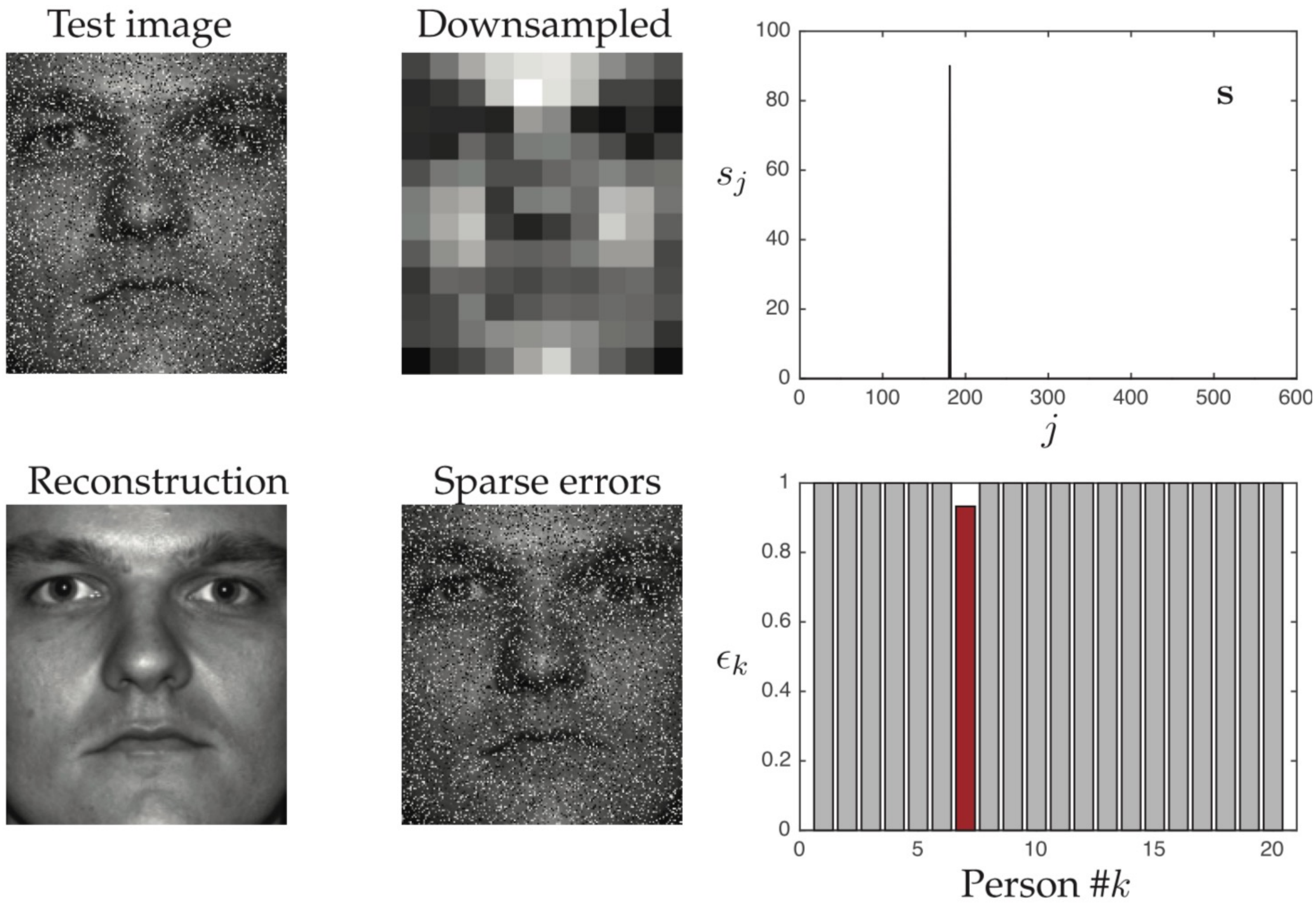
Ex 1



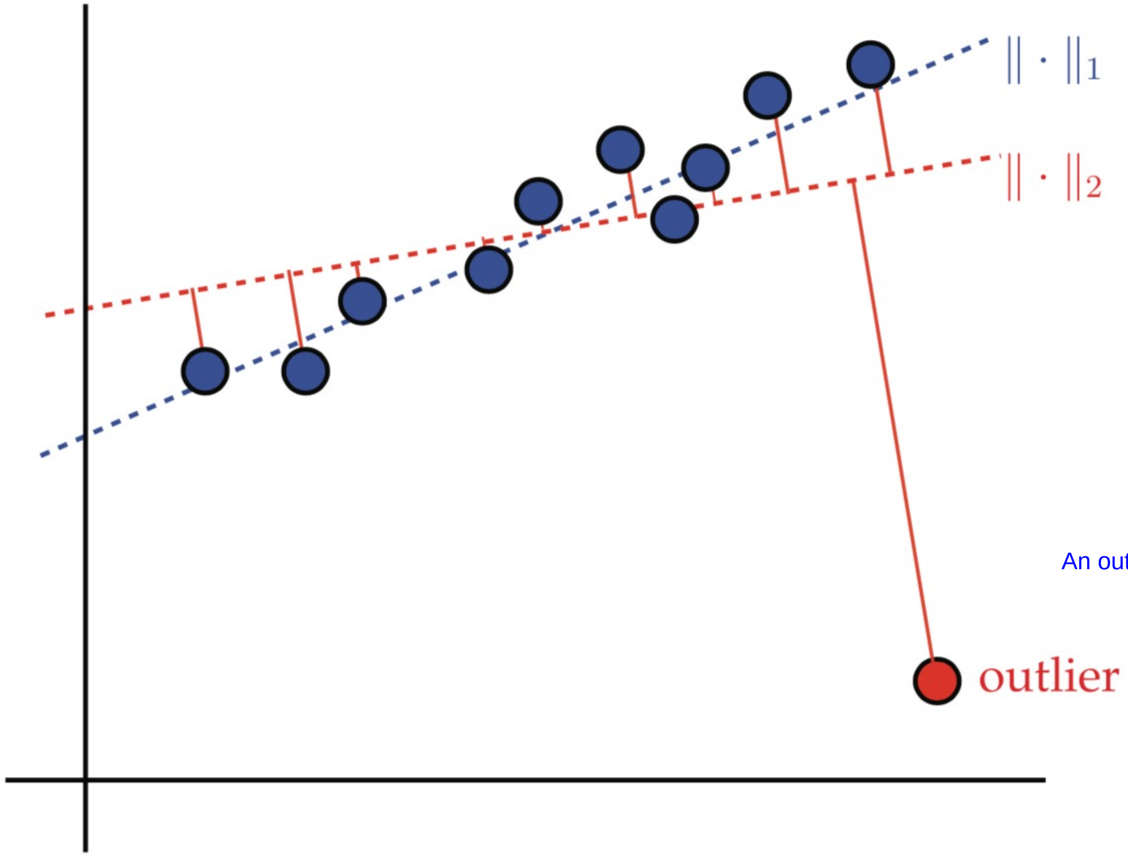
- The algorithm correctly identifies the person
- The reconstructed image removes the fake mushache

Ex 2

30% pixels are missing



Outliner removal using L1



$\| Ax - b \|_1$ Due to sparsity, L1 norm can effectively remove outliers

$\| Ax - b \|_2$ For L2 norm, the outlier has a larger weight

standard PCA would be affected by outliner a lot

An outliner can affect the regression a lot

L1 norm provide a more robust regression

Robust PCA

- Note that the PCA suffers from outliers and corrupted data
- The robust PCA (RPCA) aims to decompose the data matrix X as follows

L: clean image

$$\mathbf{X} = \mathbf{L} + \mathbf{S}$$

S: noise

where L is low-rank and S is sparse

- The principal component L is robust to outliers and corrupted data described by S
- Applications (e.g. video surveillance, face recognition, etc)

Mathematical formulation of RPCA

- We find L and S such that

$$\min_{\mathbf{L},\mathbf{S}} \operatorname{rank}(\mathbf{L}) + \|\mathbf{S}\|_0 \quad \text{subject to} \quad \mathbf{L} + \mathbf{S} = \mathbf{X}$$

- This problem is hard to solve
- We relax the above formulation as follows

$$\min_{\mathbf{L},\mathbf{S}} \|\mathbf{L}\|_* + \lambda\|\mathbf{S}\|_1 \quad \text{subject to} \quad \mathbf{L} + \mathbf{S} = \mathbf{X} \qquad \text{with} \qquad \lambda = 1/\sqrt{\max(n, m)}.$$

- This is called the Principal Component Pursuit (PCP), where $\|\cdot\|_*$ is the nuclear norm, given by the sum of singular values

Solving RPCA

- We define the shrinkage operator

$$\mathcal{S}_\tau(x) = \text{sign}(x)\max(|x| - \tau, 0)$$

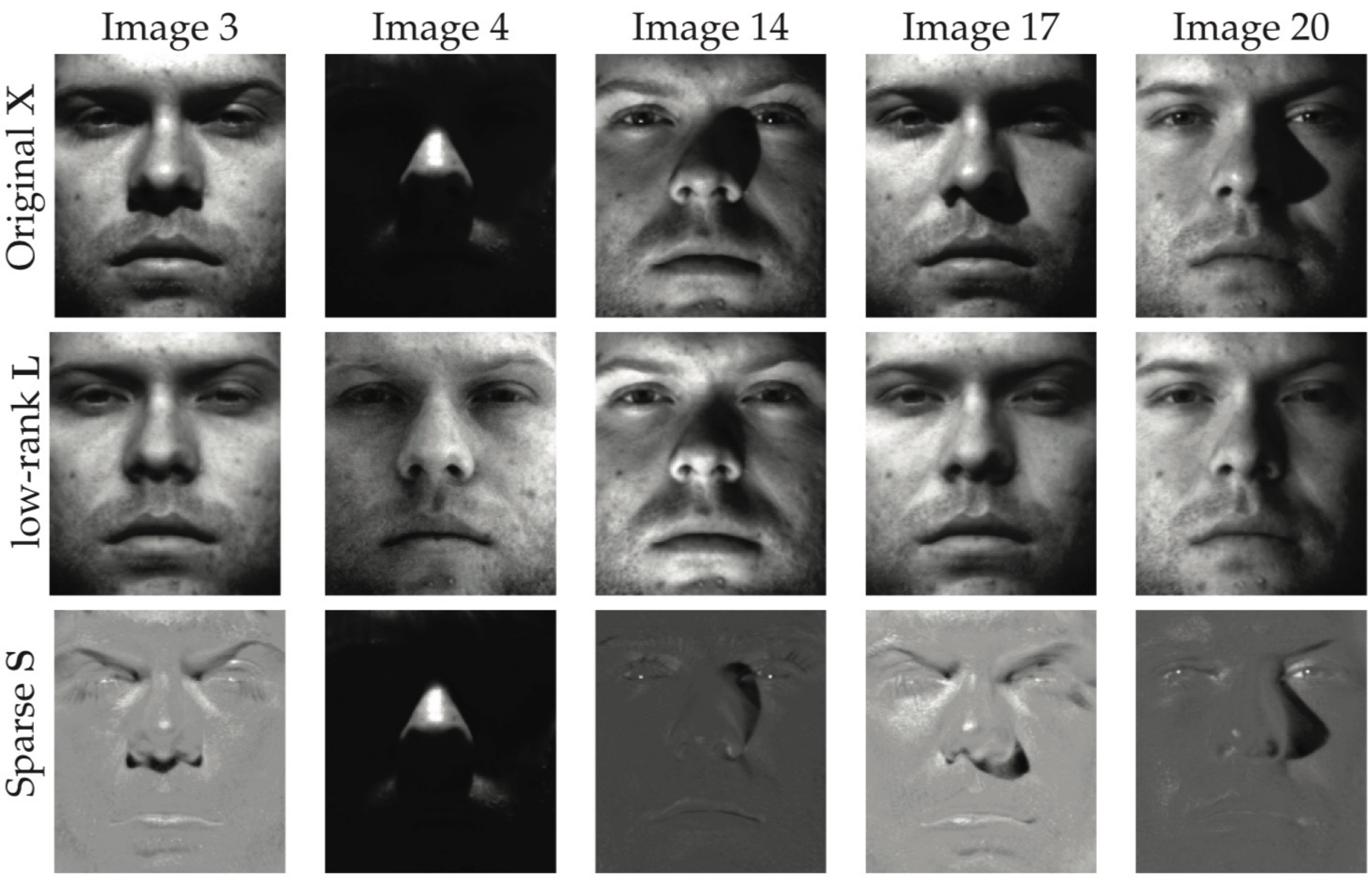
- We define a singular value threshold operator

$$\text{SVT}_\tau(\mathbf{X}) = \mathbf{U}\mathcal{S}_\tau(\mathbf{\Sigma})\mathbf{V}^*$$

- Perform the following iteration (alternating direction method)

$$\begin{aligned}
L_0 &= 0 \\
S_0 &= 0 \\
Y_0 &= 0 \\
L_{k+1} &= SVT_{1/\mu}(X - S_k + \mu^{-1}Y_k) \\
S_{k+1} &= Shrink_{\lambda/\mu}(X - L_k + \mu^{-1}Y_k) \\
Y_{k+1} &= Y_k + \mu(X - L_{k+1} - S_{k+1})
\end{aligned}$$

Yale B database example



Shadows are removed