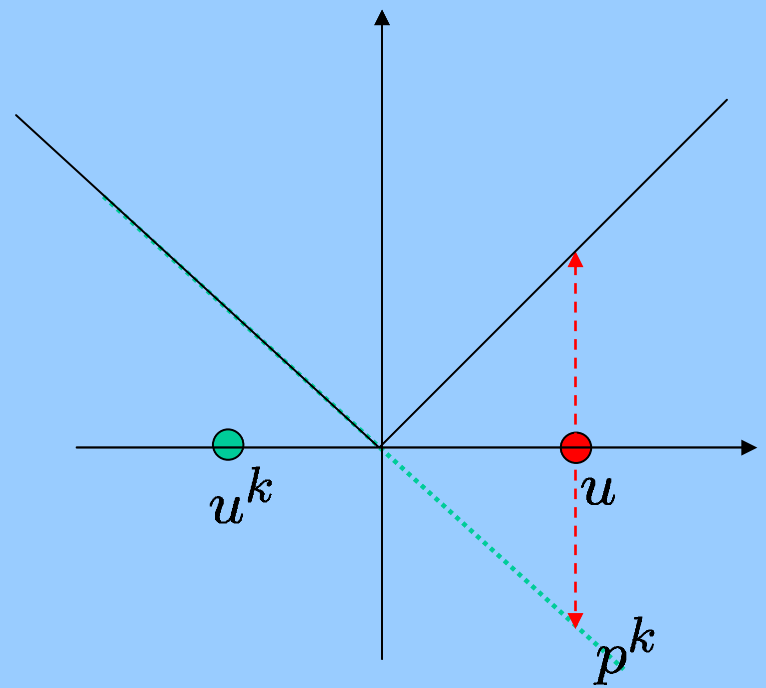
Compressed Sensing and L^1-Related Minimization



Yin Wotao

Computational and Applied Mathematics

Rice University

Jan 4, 2008 – Chinese Academy of Sciences – Inst. Comp. Math

The Problems of Interest

- Unconstrained

$$\min_u \ \mu\|u\|_1 + \frac{1}{2}\|Au - b\|_2^2$$

- Constrained

$$\min_u \ \|u\|_1, \ \text{s.t.} \ Au = b$$

- They found applications in compressed sensing, statistics, ...

Compressed Sensing

- We want to “compress” a signal
- Classical compression algorithms must know the signal first, but here the compression is done before signal is physically acquired
- The goal is to reduce the number of measurements

u: a signal with lots of null entries

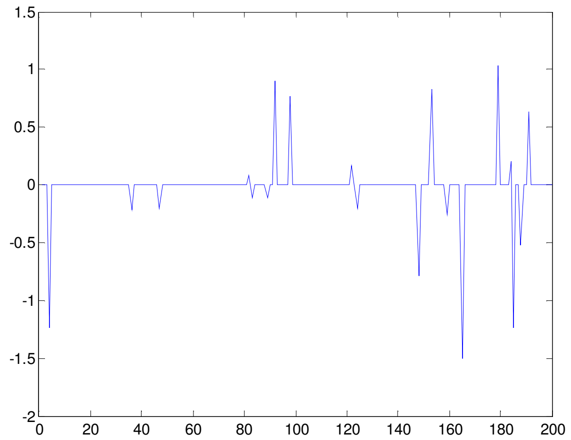
- Apply a linear transform A to u, #row(A)<#col(A)
- Physically acquire $b = Au$
- How do we get u from b when $b = Au$ is underdetermined?
- When u is sufficiently sparse and b is long enough, exact reconstruction is possible (both in theory and in practice)

A test in MATLAB

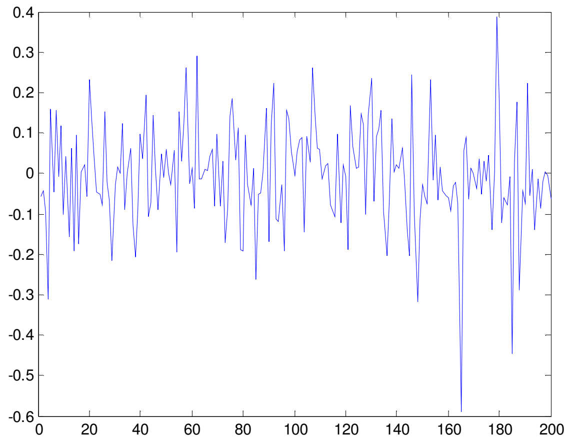
```
% Construct the compressed sensing problem
n = 200;
m = 100;
A = randn(m,n);
u = sprandn(n,1,.1);
b = A*u;

% least squares reconstruction
% min ||u||_2, subject to A*u=b
u_recon = pinv(A)*b;
```

A test in MATLAB



u



u_recon

A test in MATLAB via YALMIP and CVX

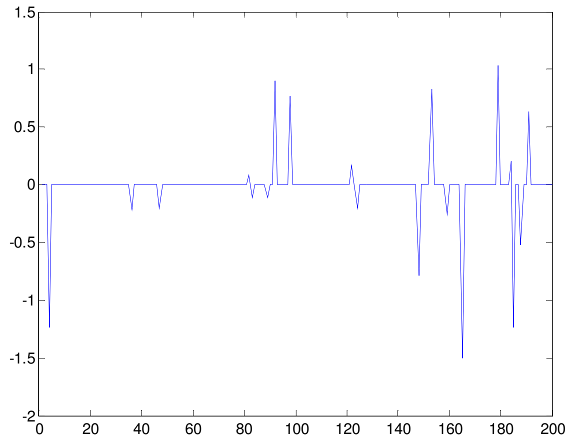
- Candes and Tao, Donoho suggest

$$\min_u \quad \|u\|_1 \qquad \text{st} \quad Au = b$$

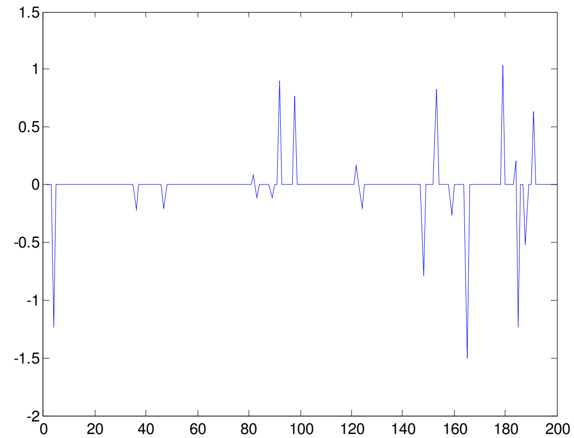
```
>> x=sdpvar(n,1);
>> solvesdp([A*x==b],norm(x,1));
...
>> u_recon=double(x);
```

```
cvx_begin
  variable u_recon(n)
  minimize(norm(u_recon,1))
  subject to
      A*u_recon==b
cvx_end
```

A test in MATLAB



u



u_recon

A test in MATLAB: noisy case

```
% Construct the compressed sensing problem
>> n = 200;
>> m = 100;
>> A = randn(m,n);
>> u = full(sprandn(n,1,.1));
>> b = A*u + 0.01*randn(m,1); % add noise
```

$$\begin{array}{ll} \min_{u} & \|u\|_1 \\ \text{st} & \|Au - b\|_2 \leq \epsilon \end{array}$$

```
>> epsilon = 0.02;
```

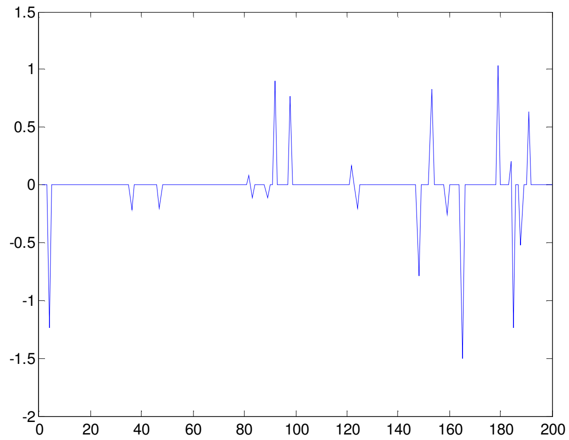
YALMIP and CVX

$$\begin{array}{ll} \min\limits_{u} & \|u\|_1 \\ \text{st} & \|Au - b\|_2 \le \epsilon \end{array}$$

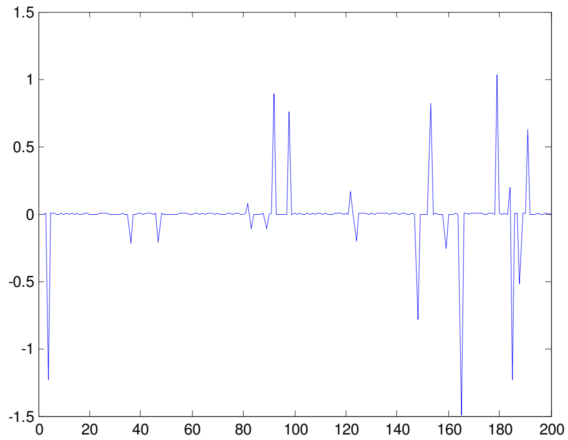
```
>> x=sdpvar(n,1);
>> solvesdp(norm(A*x-b)<=epsilon,norm(x,1));
...
>> u_recon=double(x);
```

```
cvx_begin
  variable u_recon(n)
  minimize(norm(u_recon,1))
  subject to
      norm(A*x-b,2)<=epsilon
cvx_end
```

A test in MATLAB



u



u_recon

When is l_1-reconstruction successful?

Def:

- n=dim(u) signal size
- k=$||u||_0$ signal sparsity
- m=#.row(A) sample size

- When A is subgaussian (e.g. Gaussian, Bernoulli, etc.) (Mendelson et al)
  - $m \approx O(k$ log $(n/k))$
- When A is Fourier submatrix (Candes, Romberg and Tao, 2004)
  - $m \approx O(k$ log $(n))$
- m can be even less using $||u||_p$ for p<1 minimization (Candes-Wakin-Boyd, Chartrand, and Chartrand-Y)

What if u is not sparse?

- u is not sparse in most cases, but it needs to be sparse in some sense

- Sparse under basis (e.g., Fourier, wavelet, curvelet)
- Sparse under linear transforms (overcomplete basis, not invertible)
- Sparse under nonlinear transforms (e.g., total variation)

$$TV(u) = \sum_{ij} \|(\nabla u)_{ij}\|_2$$

- Has a low rank
- Has a low dimension embedding

Compressive Imaging

- Natural images u are sparse in the wavelet domain
- Φu is approximately sparse

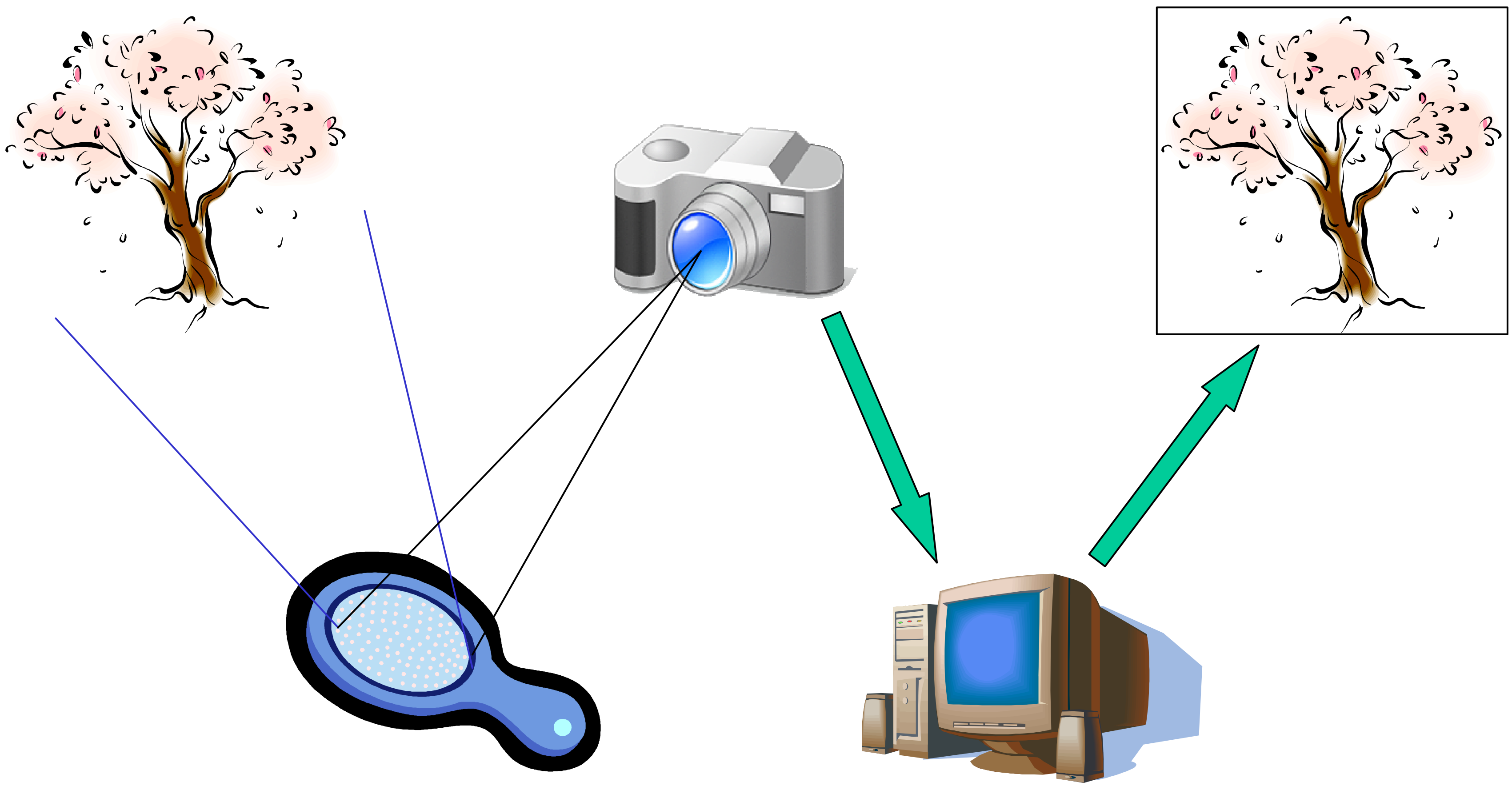
- To recover u from $Au=b$, solve

$$\begin{array}{rl} \min\limits_{u} & \|\Phi u\|_1 \\ \text{st} & \|Au - b\|_2 \leq \epsilon \end{array}$$

or

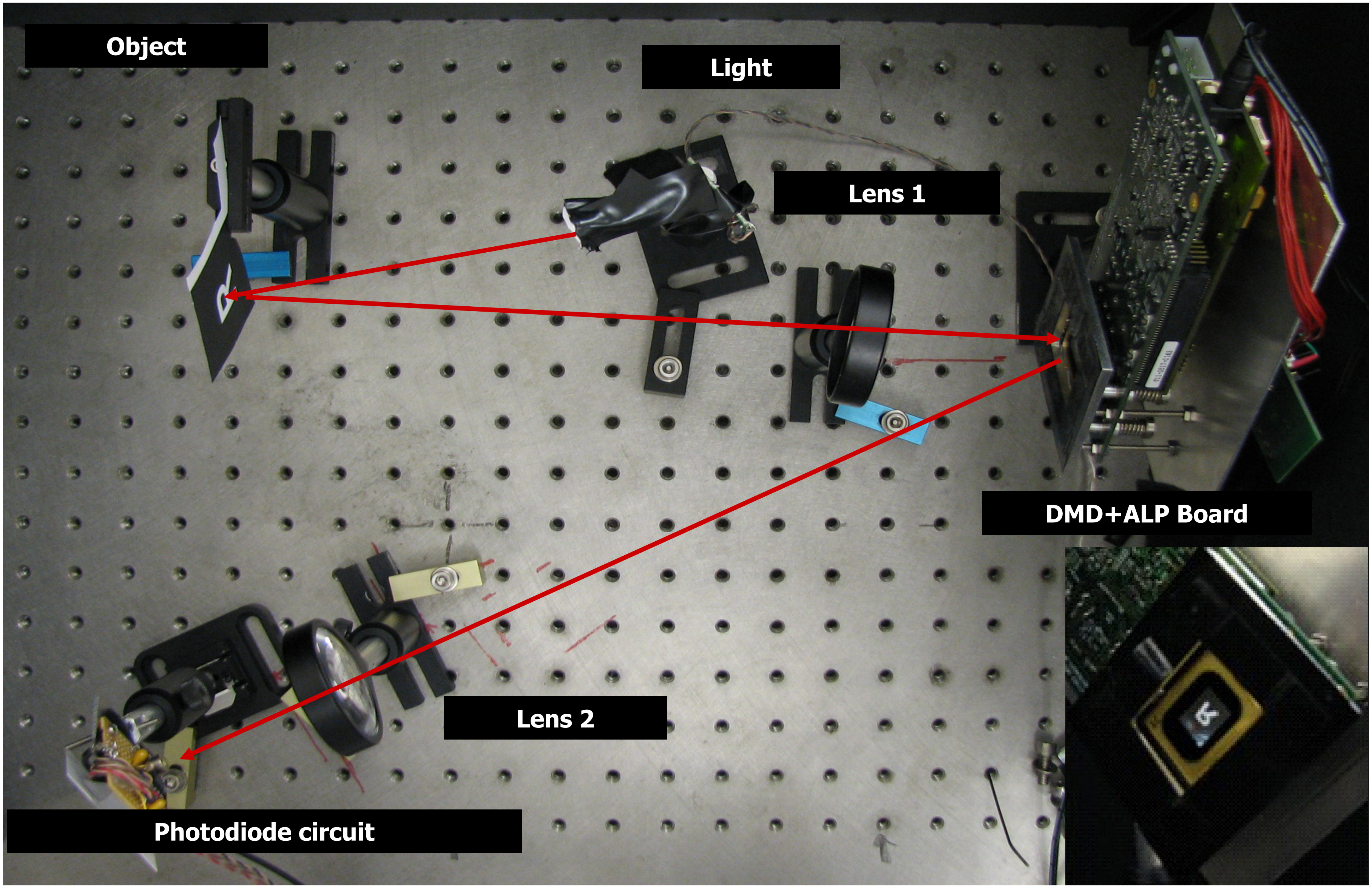
$$\begin{array}{rl} \min\limits_{x} & \|x\|_1 \\ \text{st} & \|(A\Phi^{-1})x - b\|_2 \leq \epsilon \end{array}$$

Introduction to *Compressed Sensing*

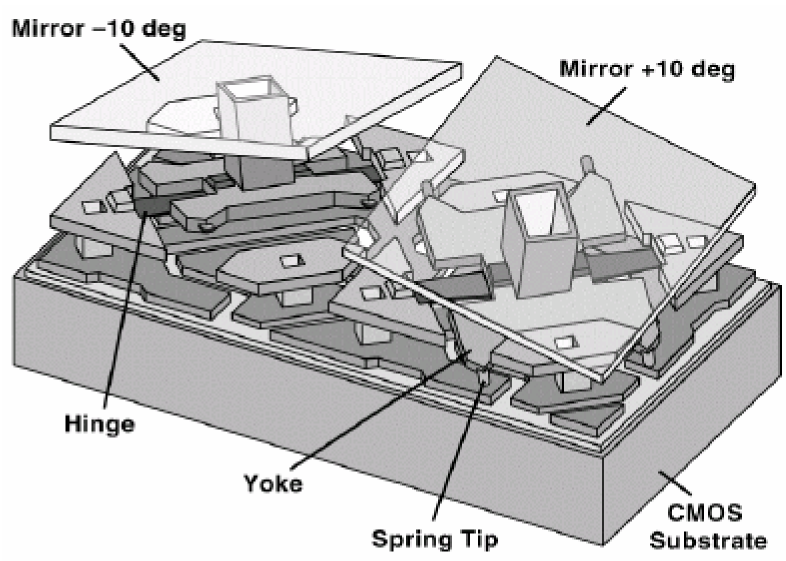


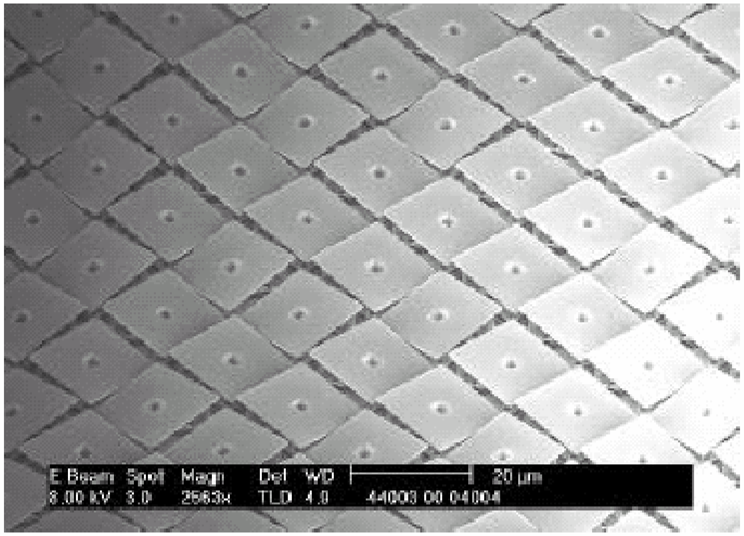
Input	Linear encoding	Signal acquisition	Signal reconstruction	Signal representation
u	Au	$b = Au$	$x = \Phi u$	$\Phi^{-1}x$

Rice Single-Pixel Camera (Wakin *et al*)



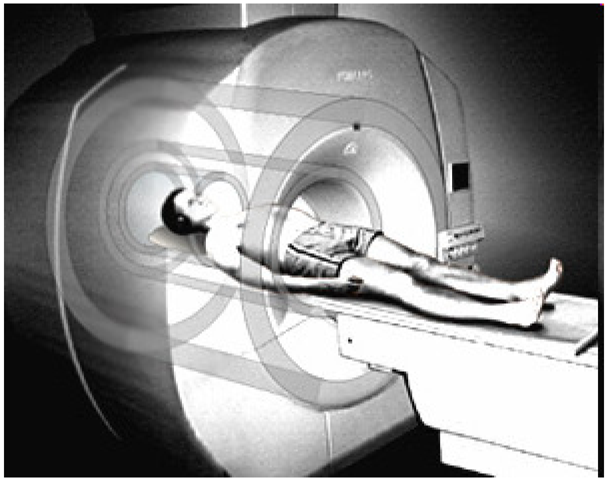
TI Digital Micromirror Device (DMD)



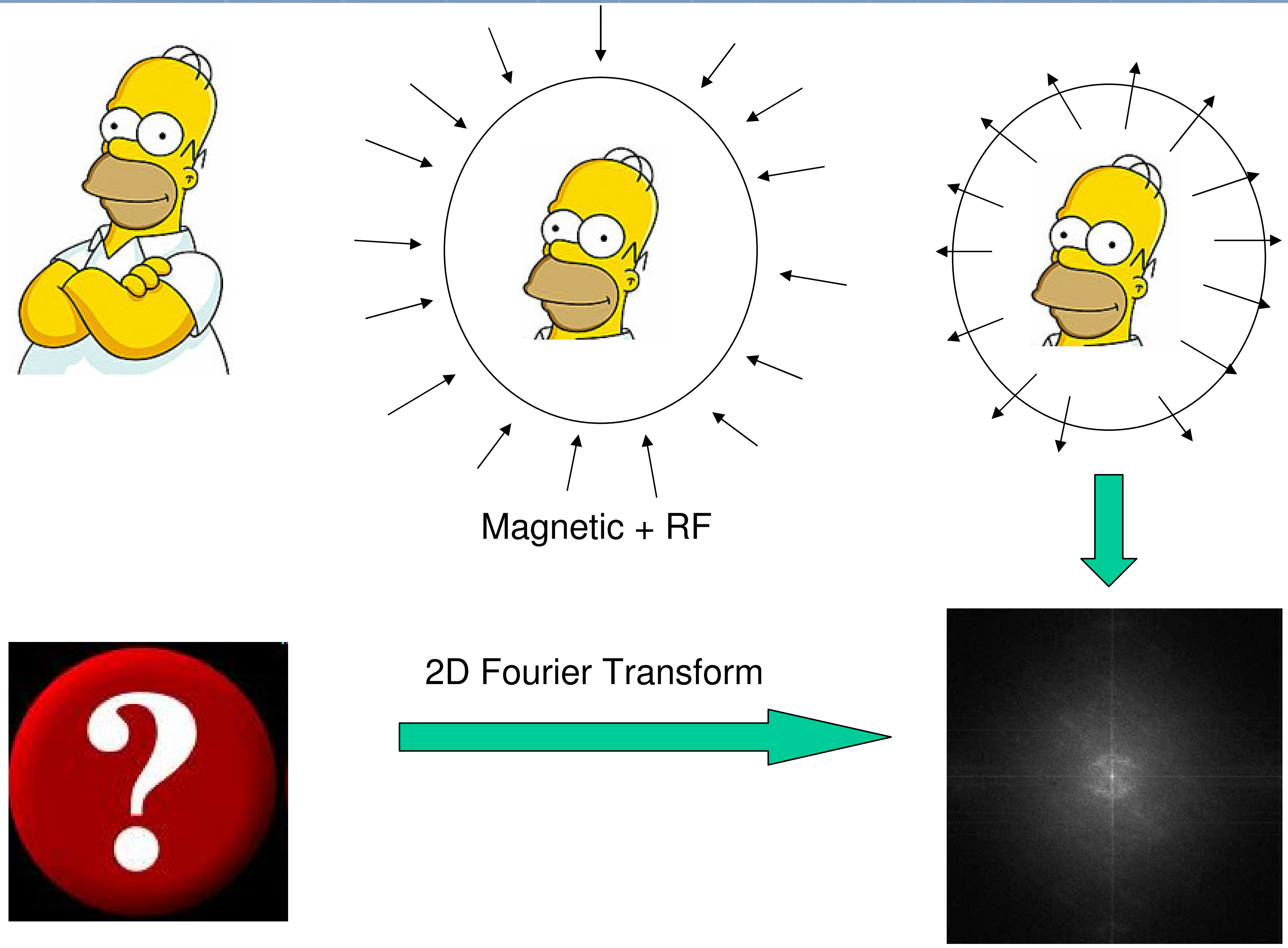


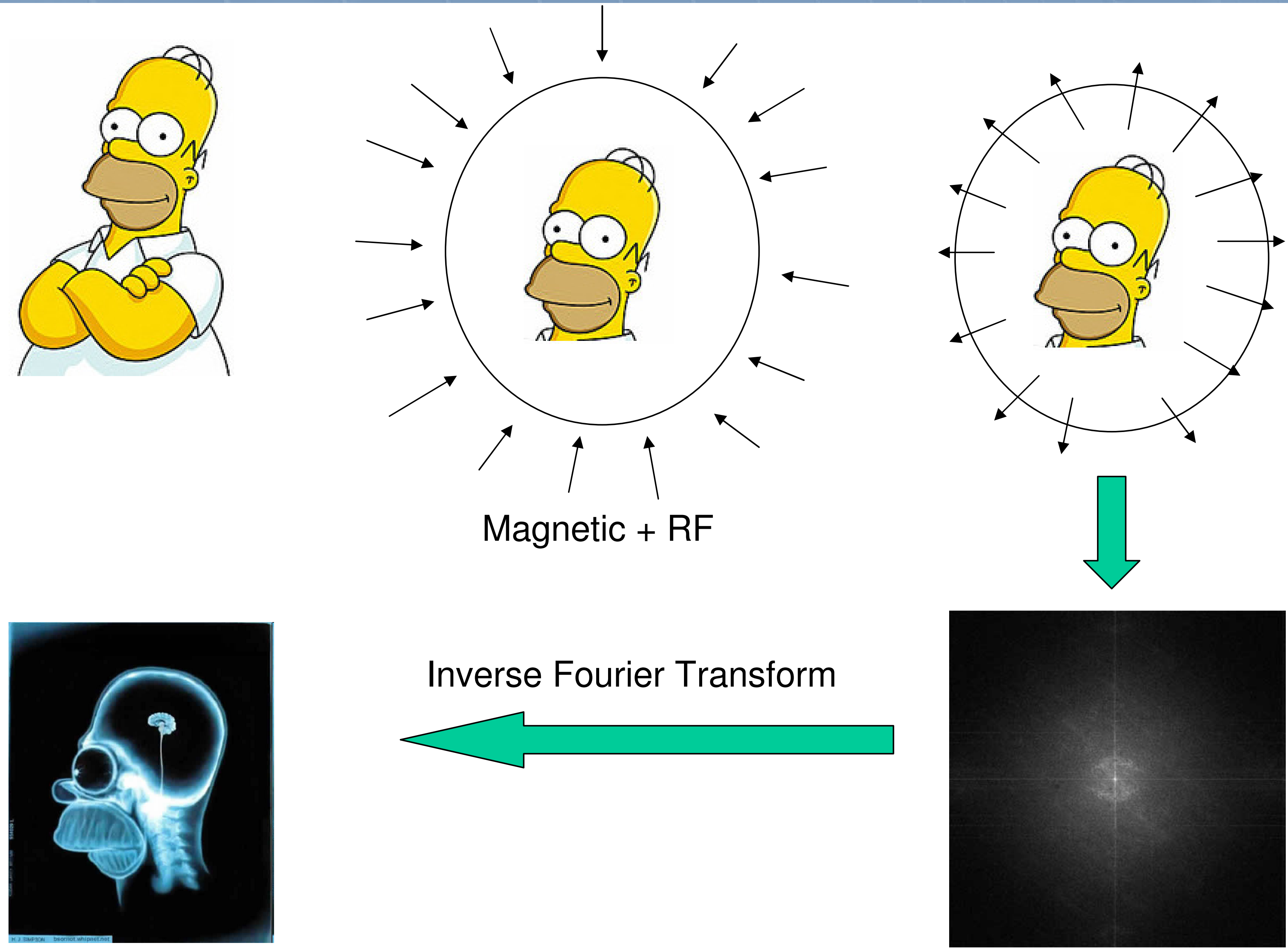
DLP 1080p --> 1920 x 1080 resolution

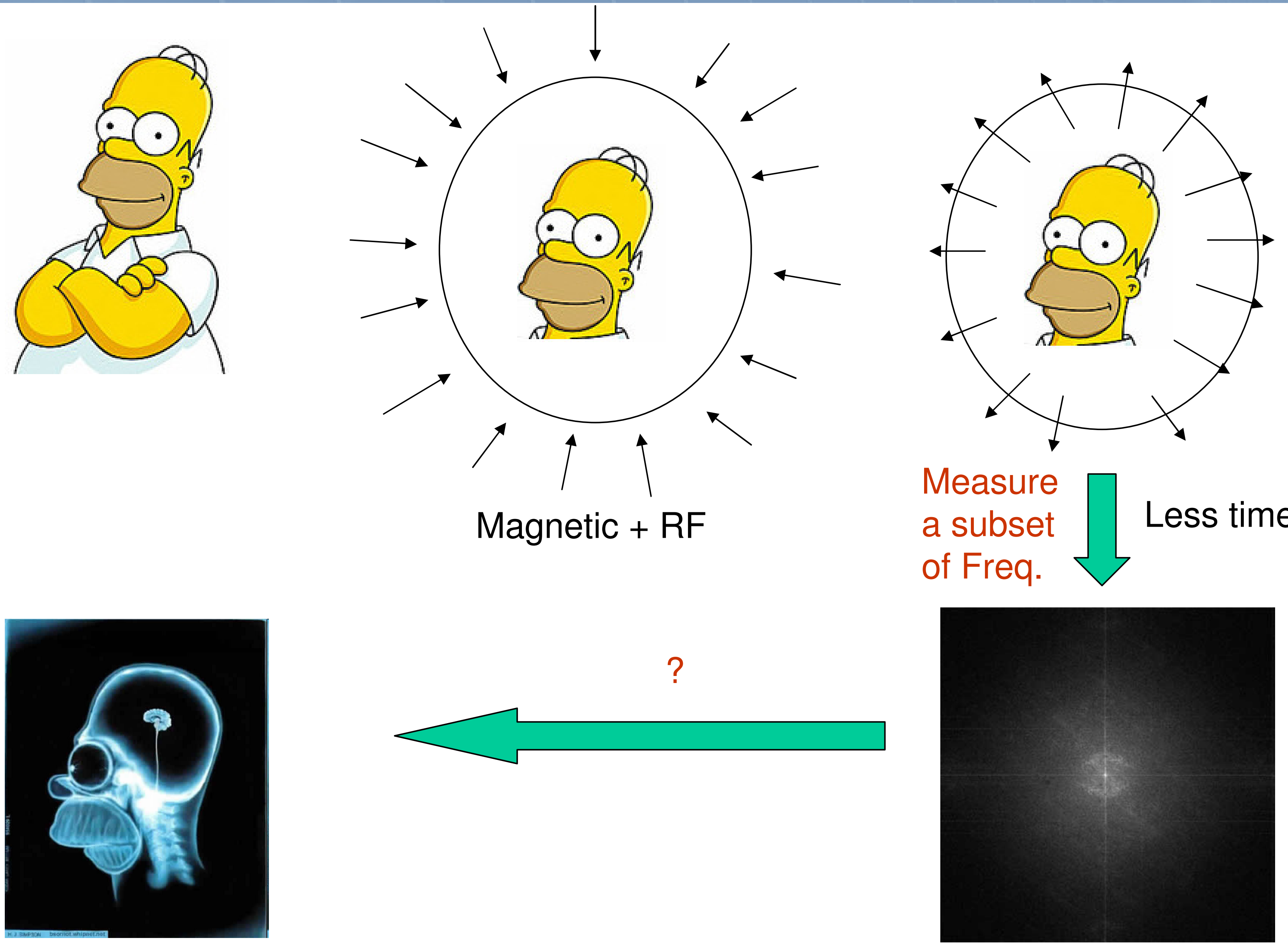
Compressed MR Imaging



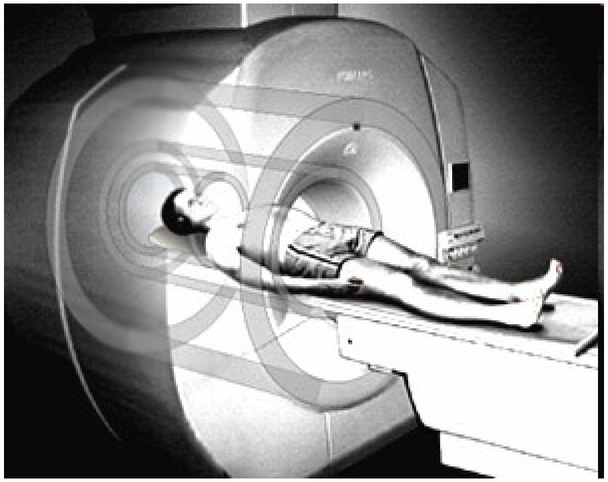
Compressed Sensing and L1-Related Minimization




Magnetic + RF
Inverse Fourier Transform


Magnetic + RF
Measure
a subset
of Freq.
Less time
?

Compressed MR Imaging



- Data Model: b=Ru; R is Fourier
- Input: frequency response b
- Reconstruction:
  - R is full: solve b=Ru, solution is u=$R^{-1}b$
  - R is partial: b=Ru is underdetermined
    - u is sparse in the wavelet domain
    - u has a small total variation (the l_1–norm of ∇u)

$$\begin{array}{ll} \min_{u} & \alpha\|\Phi u\|_1 + TV(u) \\ \text{s.t.} & Ru = b. \end{array}$$

Numerical Examples

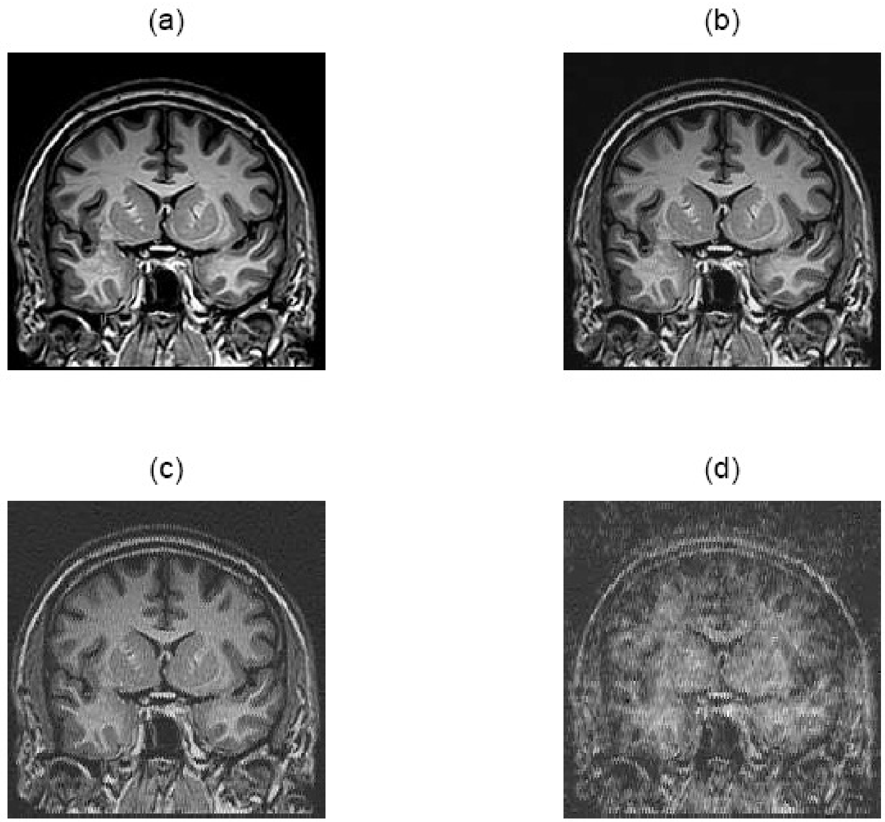


Figure 2. (a) is the original Brain image. (b), (c) and (d) are the recovered images at the sampling ratios of 38.65%, 21.67% and 8.66%, respectively.

Compressed Sensing

- Utilize the *sparsity* of a signal to reconstruct it from *a smaller number of measurements* than what is usually needed
- Procedure: encoding, sensing, decoding
  - Encoding: **nonadaptive**, linear, “random”
  - Sensing: acquiring measurements
  - Decoding: using optimization to recover the signal
- Advantages: improve the capacities of physical devices
  - Infra-red Imaging: higher resolution
  - CT: less radiation dosage
  - MRI: less time
  - multi-sensor distributive network: higher throughput, longer battery time
  - DNA microarrays: less cost
  - Low-light imaging, microcopy, video acquisition, hyper-spectral image classification

The Challenges

Such optimization problems

- Have large-scale and dense data (though not always)
- Are non-smooth

However,

- It is the solution that is sparse

Classical l_1 solvers

- Most of them solve

$$\min_y \|By - c\|_1$$

  using simplex-type methods, interior-point, or Huber-norm approximation.
- Applied to geophysics and economics problems

- Invert or factorize a matrix involving B

Next, recent solvers ...

Different Formulations

- Unconstrained

$$\min_u \ \mu\|u\|_1 + \frac{1}{2}\|Au - b\|_2^2$$

- Basis pursuit

$$\min_u \ \|u\|_1, \ \text{s.t.} \ \|Au - b\|_2 \leq \delta$$

- LASSO

$$\min_u \ \|Au - b\|_2, \ \text{s.t.} \ \|u\|_1 \leq \gamma$$

- Constrained

$$\min_u \ \|u\|_1, \ \text{s.t.} \ Au = b$$

(A Subset of) Recent Algorithm Types

- Path-following: *LARS, etc.*
  - Start from an easy problem
  - Gradually transform the easy problem to the original one
  - Solution path is piece-wise linear
  - Solve one (small) system of linear equations for each breakpoint
- Specialized interior-point method: l1_ls
  - More accurate than first-order methods
  - Use truncated Newton's method and preconditioned conjugate gradient
- Operator splitting:
  - Cheap per-iteration cost, more iterations
  - Accelerated by line search and continuation
  - Obtain optimal support quickly
- Gradient projection: GPSR, similar to operator splitting
- Bregman method:
  - Originally for the constrained problem, finite convergence

Operator Splitting

- Observation:
  - $||u||_1$ is nonsmooth but is separable
  - $||Au\text{-}b||_2^2$ is smooth but non-separable
- Solution: apply different operations

$$T_1 = \partial_u(\mu\|u\|_1) \qquad\qquad T_2 = \partial_u \tfrac{1}{2}\|Au - b\|_2^2$$

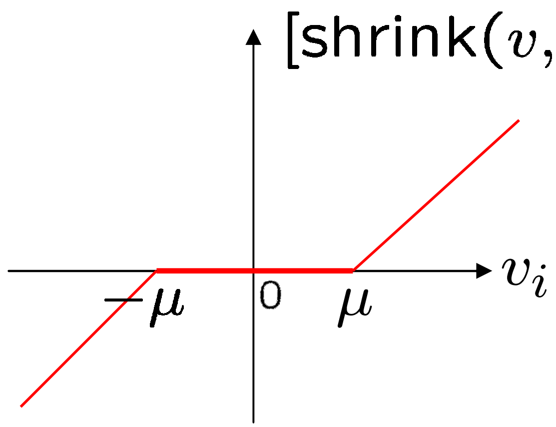
$$\begin{aligned} u^* = \arg\min \quad &\Leftrightarrow \quad \vec{0} \in (T_1 + T_2)(u^*) \\ &\Leftrightarrow \quad \vec{0} \in (I + \tau T_1 - I + \tau T_2)(u^*) \\ &\Leftrightarrow \quad (I - \tau T_2)(u^*) \in (I + \tau T_1)(u^*) \\ &\Leftrightarrow \quad u^* = (I + \tau T_1)^{-1}(I - \tau T_2)(u^*) \end{aligned}$$

$I + \tau T_1$: gradient descent with stepsize τ

$(I - \tau T_1)^{-1}$: shrinkage, a closed-form operator

Shrinkage, a component-wise separable operator:

- Shrinkage (soft thresholding) previously used by, as far as I know,
  - Chambolle, DeVore, Lee and Lucier
  - Figueiredo, Nowak, and Wright
  - Daubechies, De Frise and DeMul
  - Elad, Matalon and Zibulevsky
  - Hale, Yin and Zhang
  - Darbon and Osher
  - Combettes and Pesquet



MATLAB command: sign(v).*max(0,|v|-mu)

FPC (Hale, Y, and Zhang, 07)

$$u^{k+1} \leftarrow (I + \tau T_1)^{-1}(I - \tau T_2)(u^k)$$

- Method is not new, theorems are
- Theorem:
  Under mild conditions, the algorithm converges linearly.
- Theorem:
  The algorithm obtains the optimal support in a finite number of iterations.

For the constrained problem: $\min_u \ \|u\|_1, \text{ s.t. } Au = b$

- Let μ go to 0, and solve the unconstrained problem
- Bregman Iterative Method (Y, Osher, Goldfarb, & Darbon, 2007)

$$\text{Let } u^0 = 0 \text{ and } b^0 = 0. \text{ For } k = 0, 1, \ldots$$

$$b^{k+1} \leftarrow b + (b^k - Au^k)$$

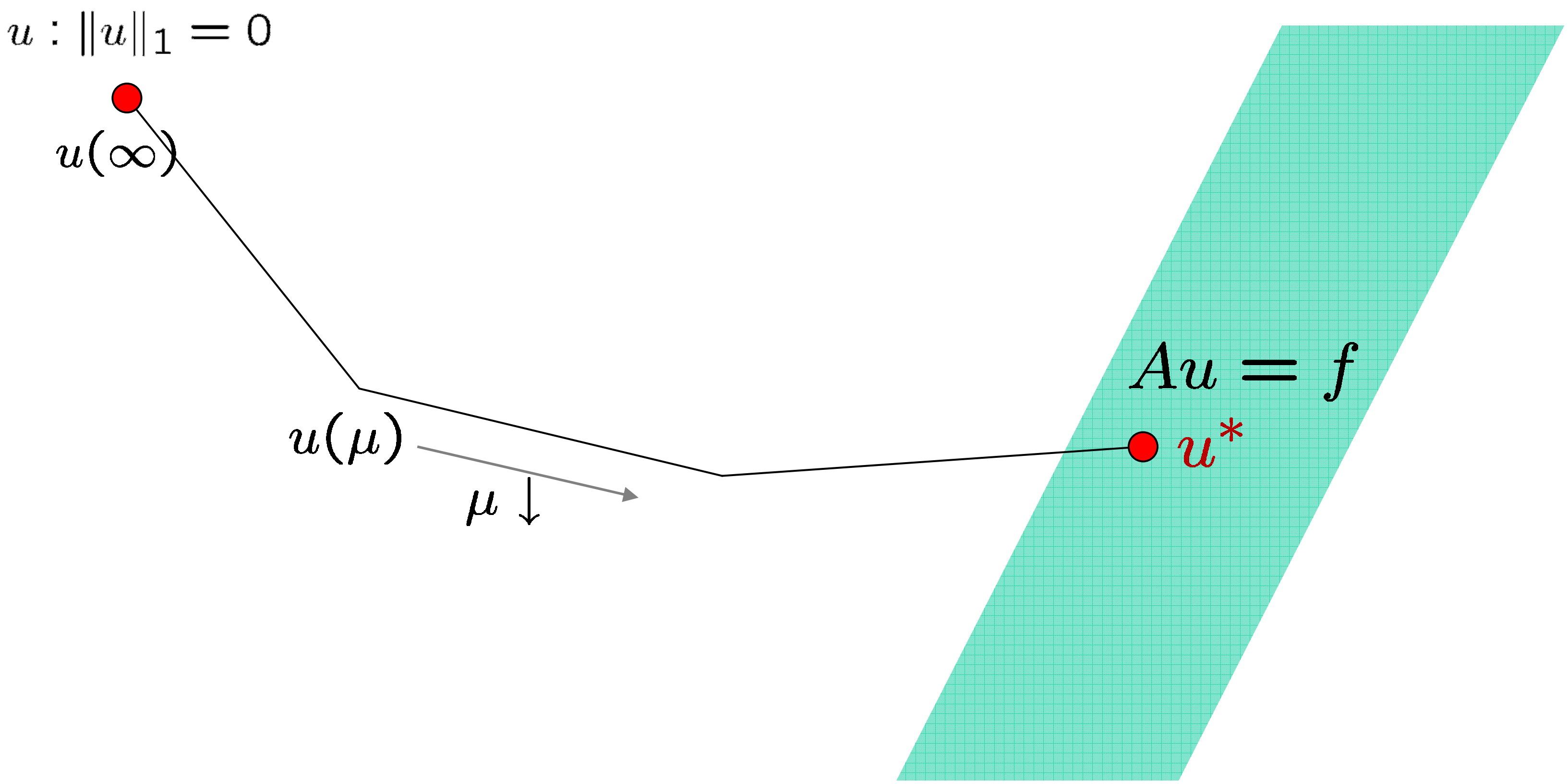
$$u^{k+1} \leftarrow \arg\min_u \ \mu\|u\|_1 + \tfrac{1}{2}\|Au - b^k\|^2$$

- Theorem:

  The Bregman method converges to an exact solution in a finite number of steps for any μ>0.

Understanding the Unconstrained Problem

$$u(\mu) := \arg\min_u \ \mu\|u\|_1 + \frac{1}{2}\|Au - f\|^2$$

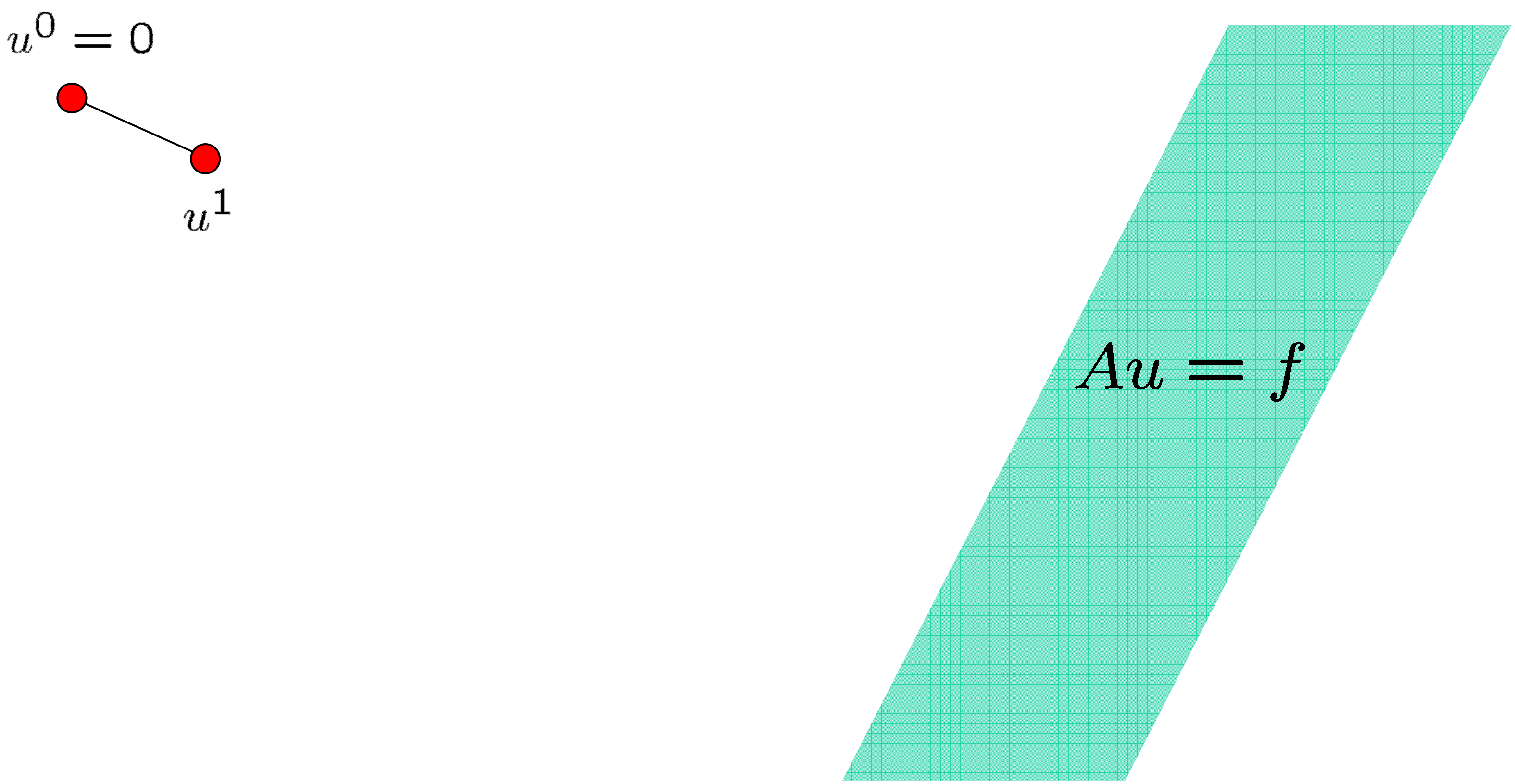


Understanding the Nonlinear Bregman Iterations

Start from $u^0 = 0$ and $p^0 = 0$

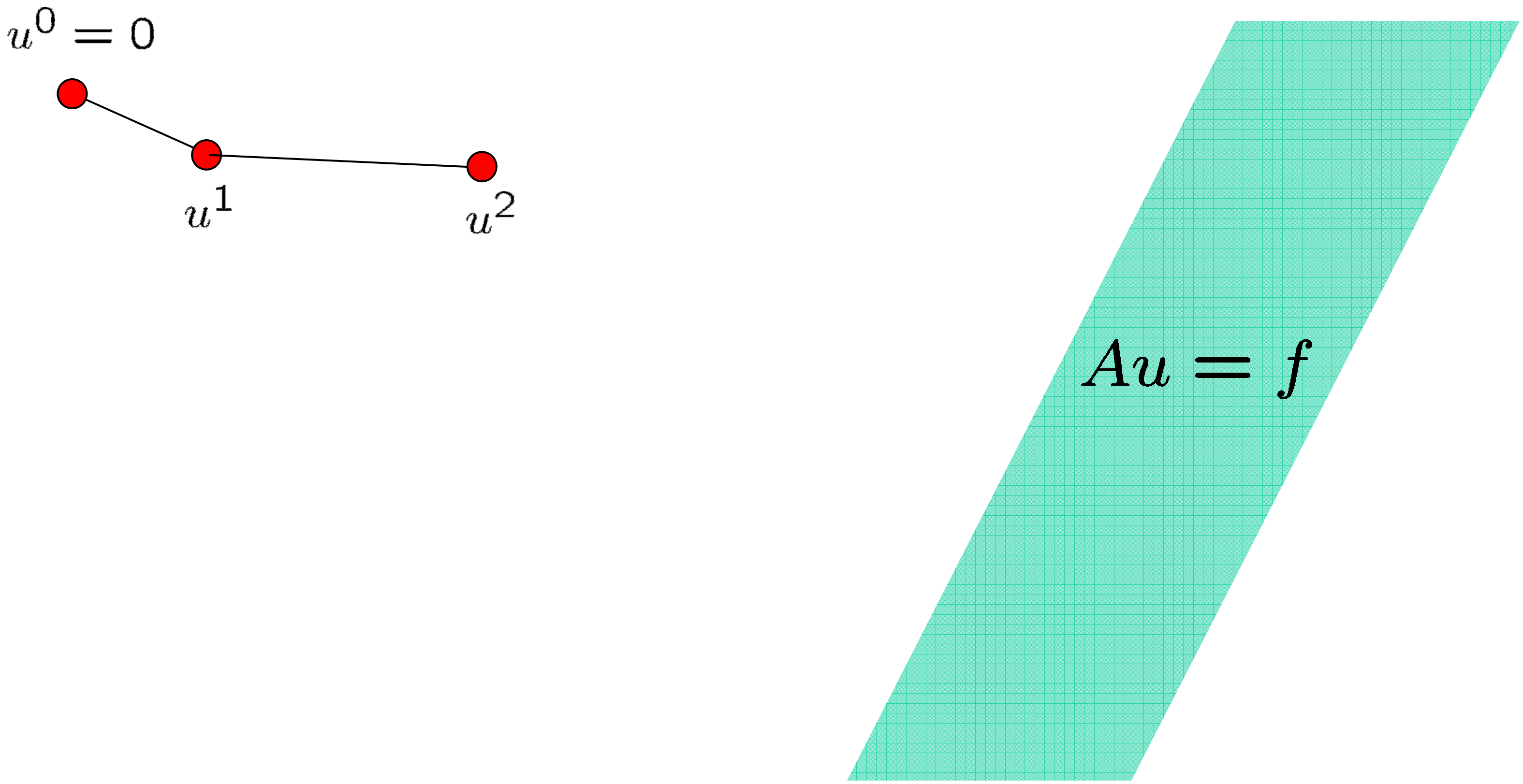
Understanding the Nonlinear Bregman Iterations

$$\text{Step 1: } u^1 = \min_u \ \mu D(u; u^0, p^0) + \frac{1}{2}\|Au - f\|^2$$



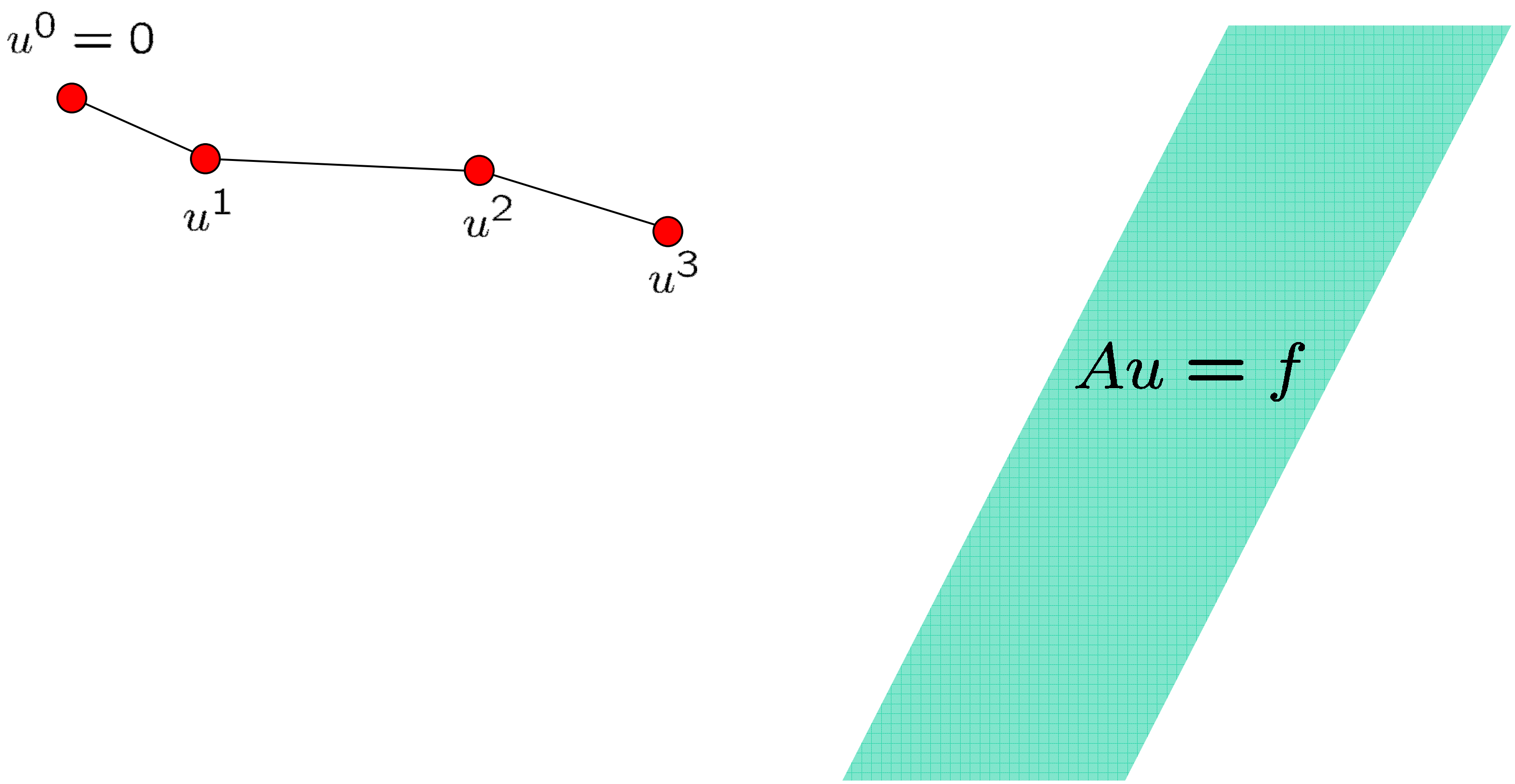
Understanding the Nonlinear Bregman Iterations

$$\text{Step 2: } u^2 = \min_u \ \mu D(u; u^1, p^1) + \tfrac{1}{2}\|Au - f\|^2$$



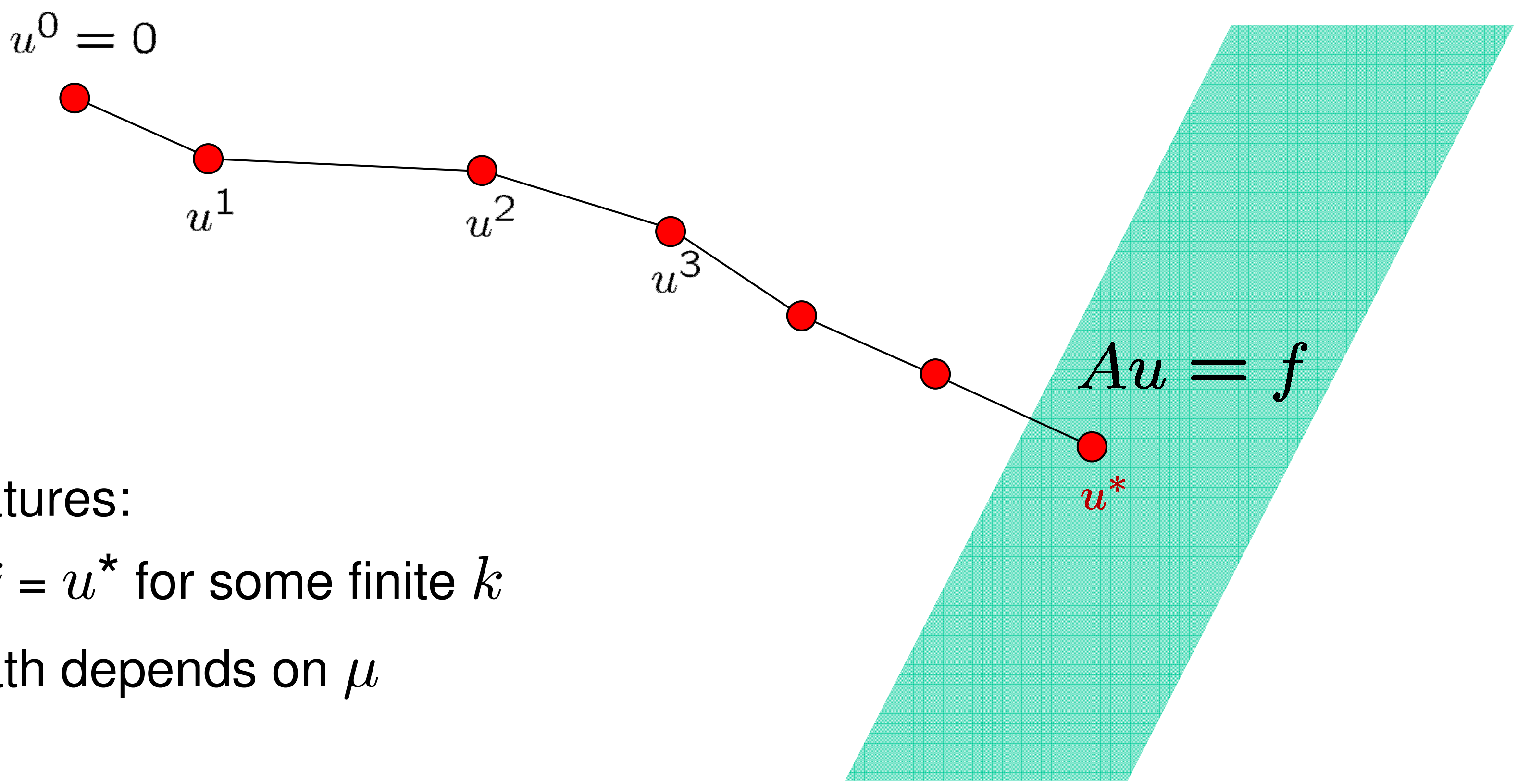
Understanding the Nonlinear Bregman Iterations

Step 3: $u^3 = \min_u \ \mu D(u; u^2, p^2) + \frac{1}{2}\|Au - f\|^2$



Understanding the Nonlinear Bregman Iterations

$$\text{Step k:} \quad u^k = \min_u \; \mu D(u; u^{k-1}, p^{k-1}) + \frac{1}{2}\|Au - f\|^2$$



Features:

- $u^k = u^*$ for some finite k
- path depends on μ

Linearized Bregman Iterations

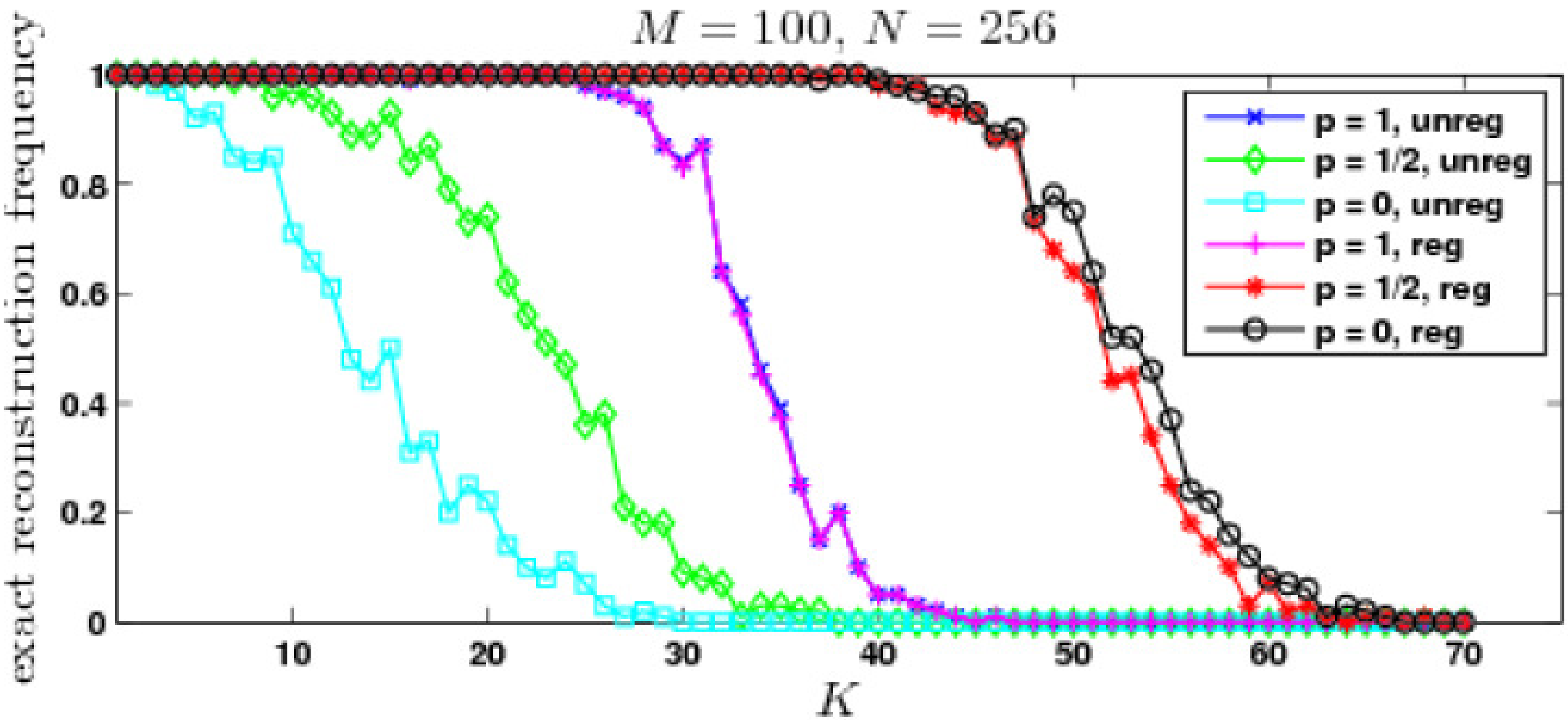
Not converges finitely, but

- Code is extremely simple (MATLAB: 2 lines + stop. criteria)
- Larger components come out first, but code still fast to recover small components

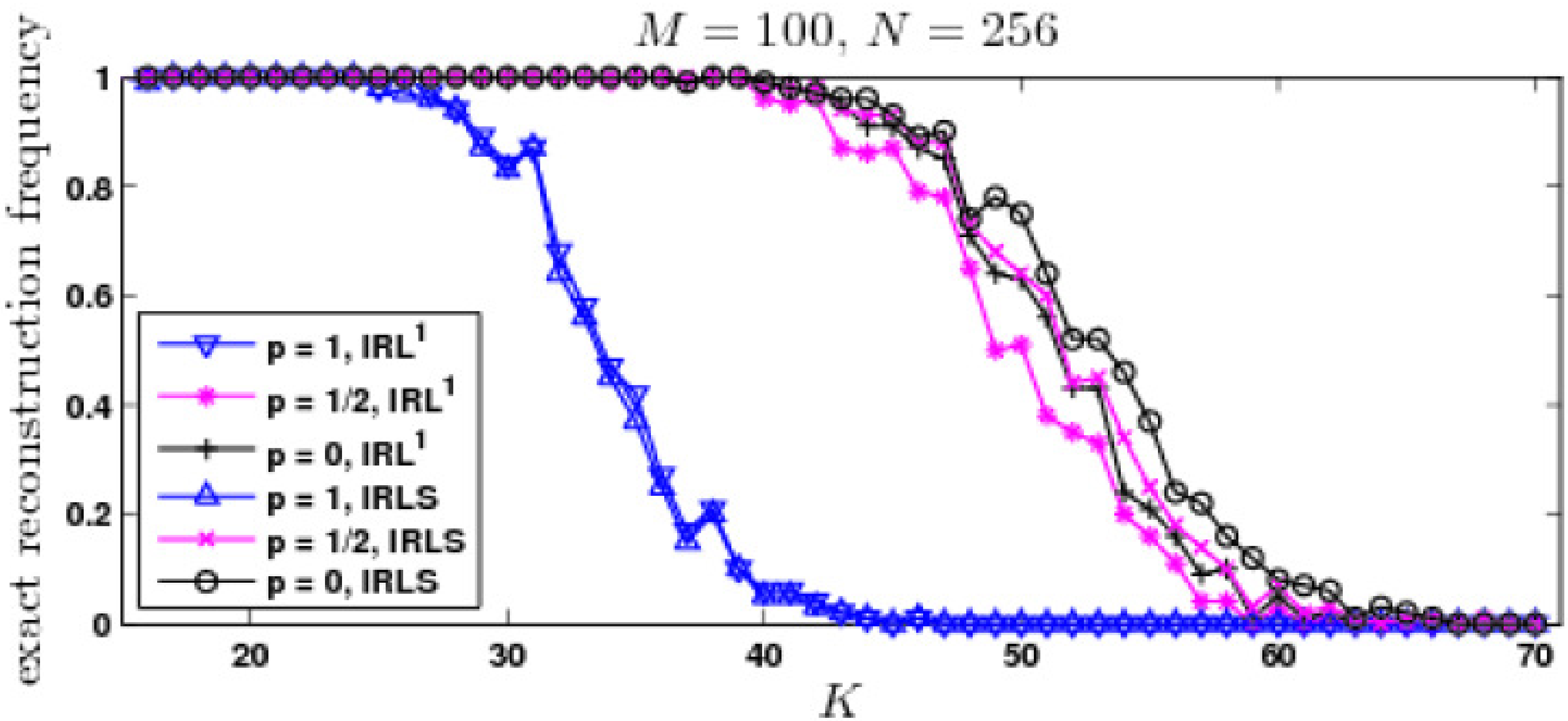
- For 2^{10}x2^{20} problems, more like 40 seconds.

l_p–minimization versus l_1–minimization

- Rao-Kreutz-Delgado, FOCUSS'99, Chartrand'06, Candes-Wakin-Boyd'07
- Chartrand-Y'07:



Iteratively Reweighted *LS* seems to be better than l_1

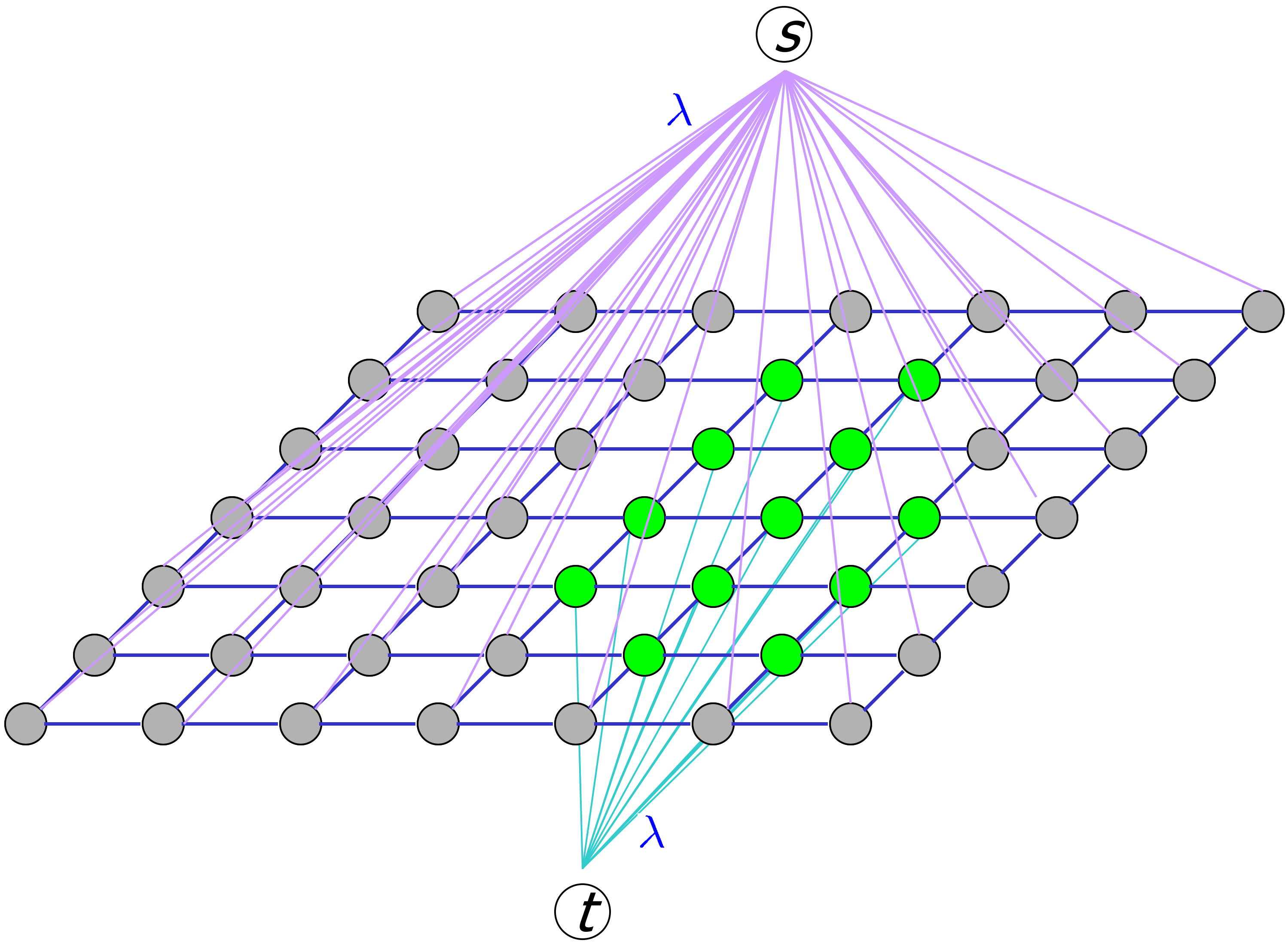


Sparsity under Nonlinear Transforms

For total variation: $TV(u) = \sum_{ij} \|(\nabla u)_{ij}\|_2$

Long story short, 3 existing choices:

- Operator splitting (Darbon-Osher, Wang-Y.-Zhang, 07)
  - Solve $\min_u \mu TV(u) + \frac{1}{2}\|u - u^{k+1/2}\|_2^2$
    using max-flow (Darbon-Sigelle, Chambolle, Goldfarb-Y)



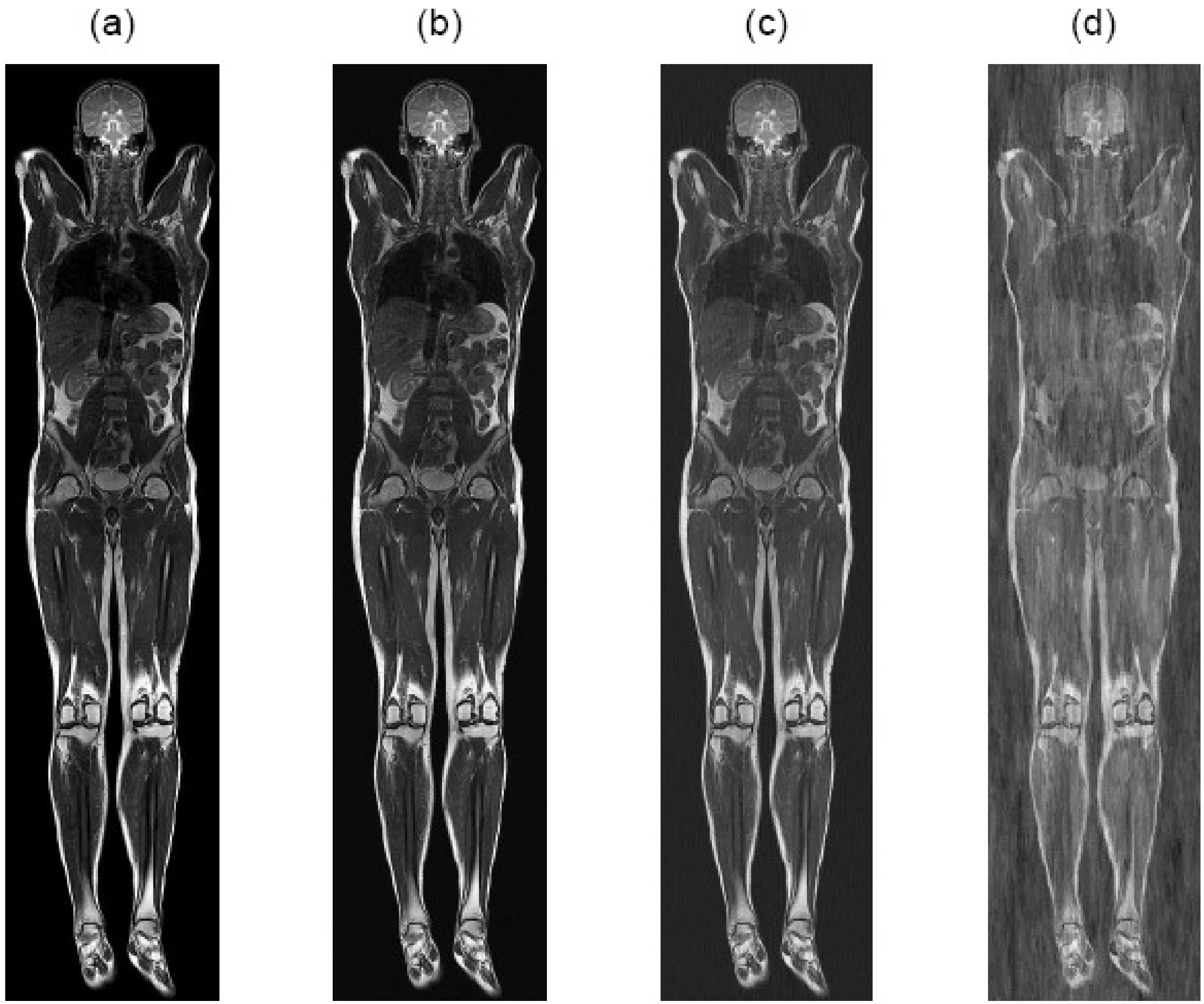
s
λ
λ
t

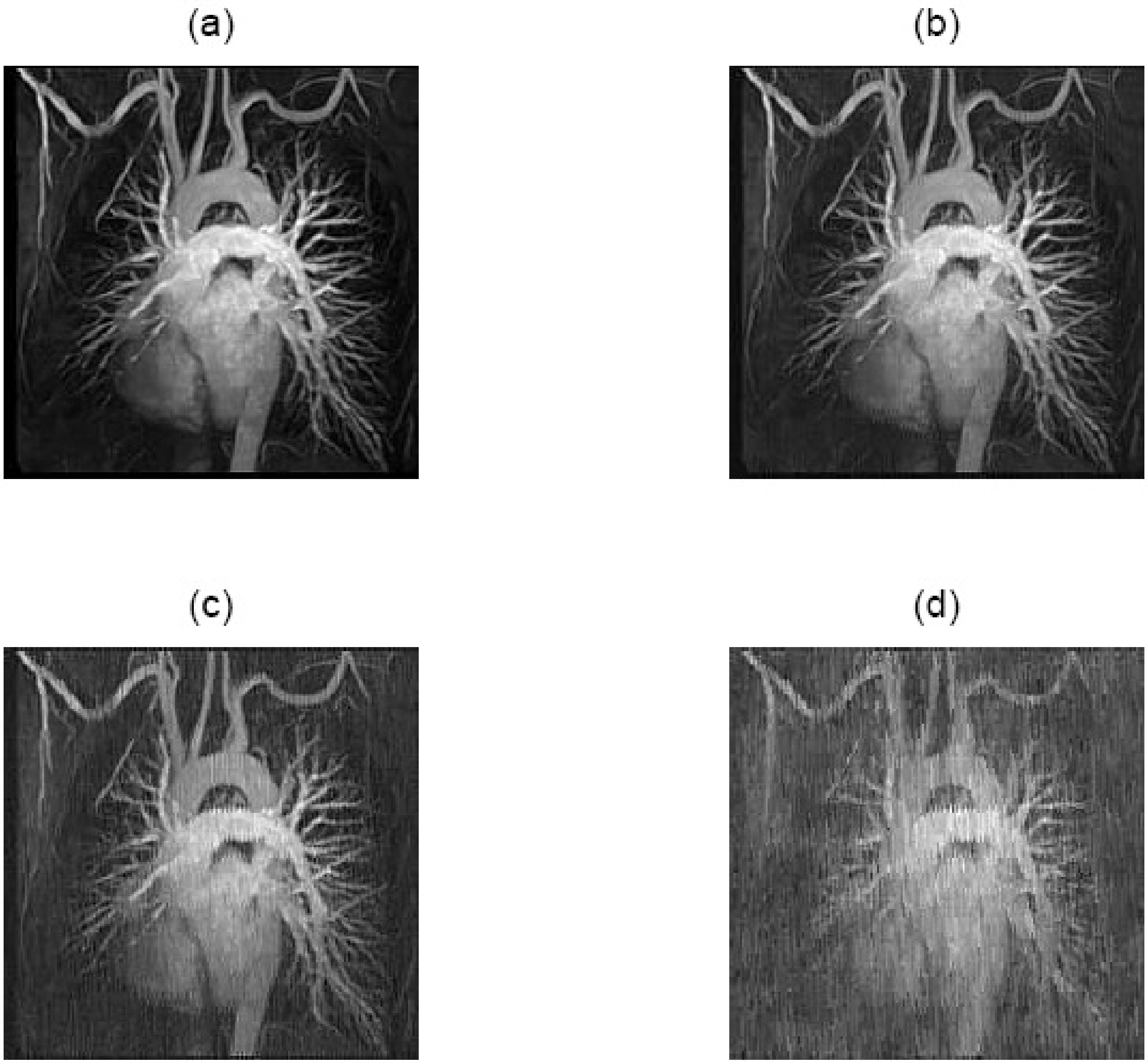
Sparsity under Nonlinear Transforms

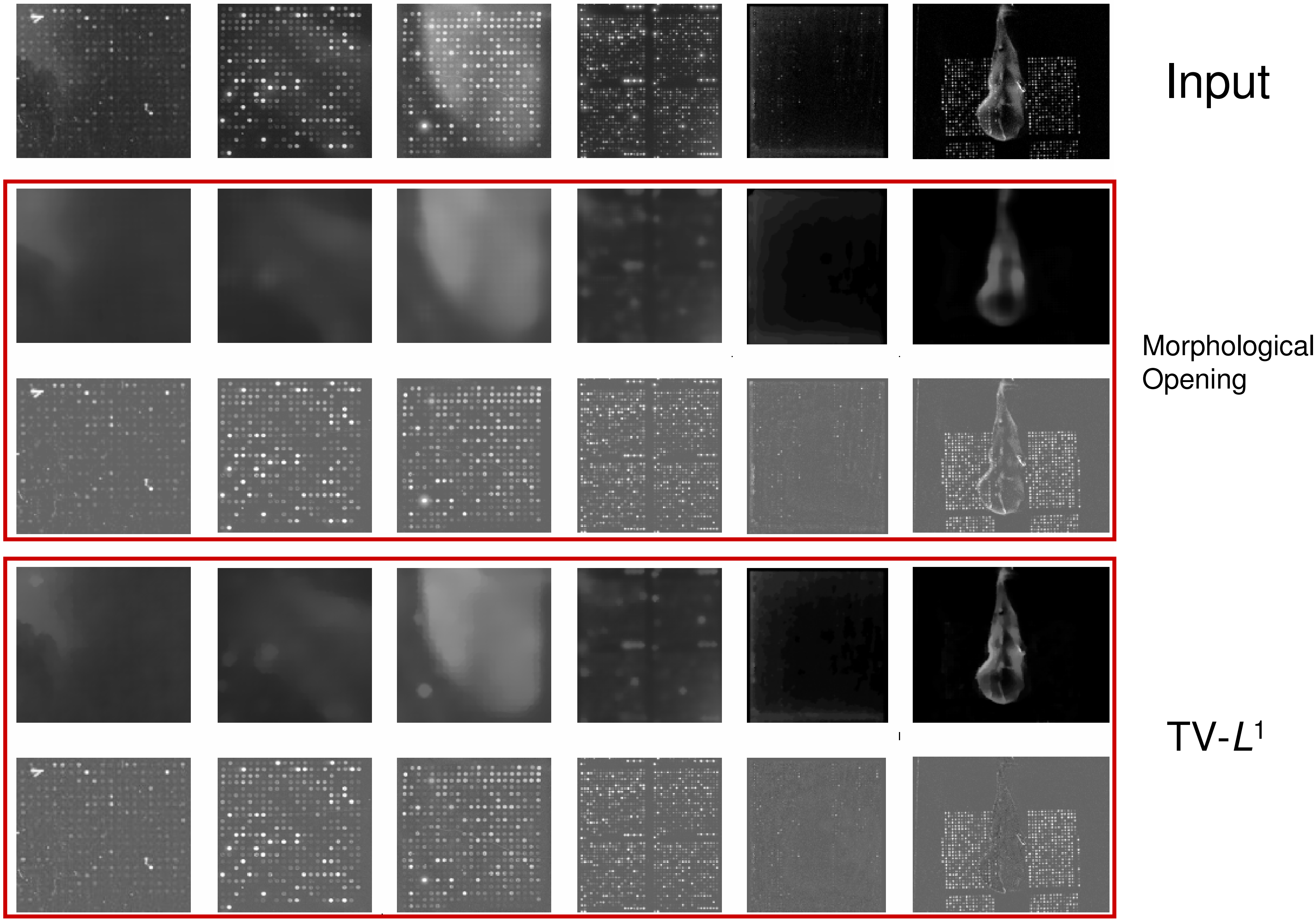
For total variation: $TV(u) = \sum_{ij} \|(\nabla u)_{ij}\|_2$

Long story short, 3 existing choices:

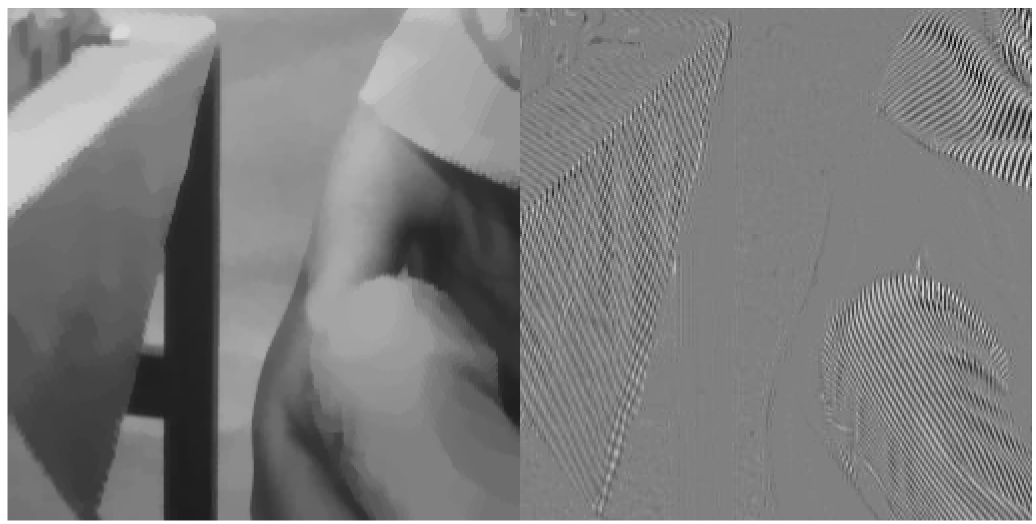
- Operator splitting (Darbon-Osher, Wang-Y.-Zhang, 07)
  - Solve $\min_u \mu TV(u) + \frac{1}{2}\|u - u^{k+1/2}\|_2^2$
    using max-flow (Darbon-Sigelle, Chambolle, Goldfarb-Y)
- A different splitting (Wang-Y.-Zhang, 07)
  $$\min_{\vec{w},u} \mu\| \, |\vec{w}_{ij}| \, \|_1 + \beta\|\vec{w} - \nabla u\|_2^2 + \frac{1}{2}\|Au - b\|_2^2$$
- Legrender-Fenchel Transform / Duality (Y., 07)

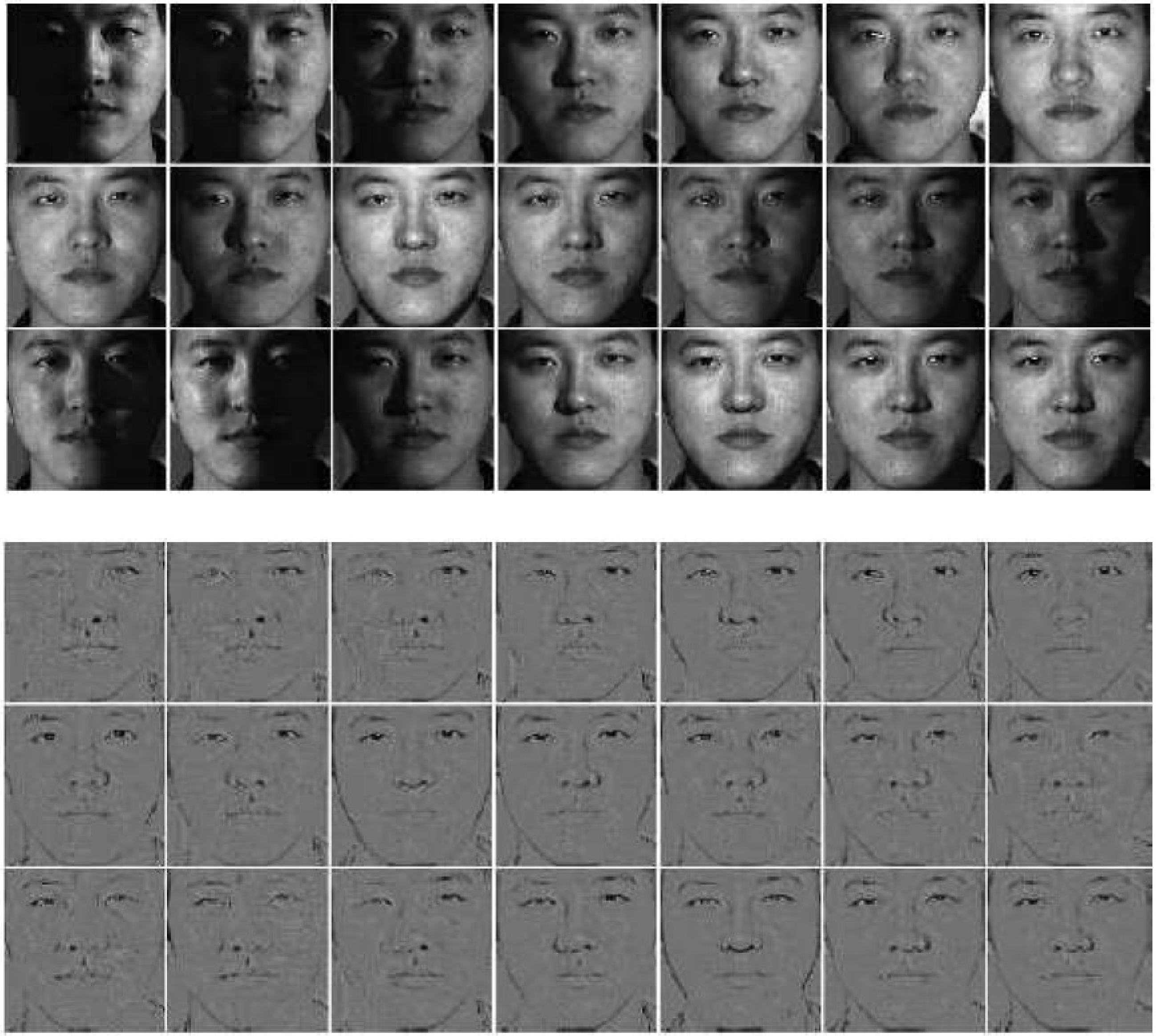

(a)
(b)
(c)
(d)


(a)
(b)
(c)
(d)


Input
Morphological
Opening
TV-L^1

Compressed Sensing and L1-Related Minimization





Compressed Sensing and L1-Related Minimization

Acknowledgements

Special Thanks:

袁老师 and 戴老师

for their invitation and

刘昕

for his arrangement.

Collaborators on L^1-related minimization:

- Columbia: Don Goldfarb, Jianing Shi
- Rice CAAM: Yin Zhang, Elaine Hale, Yilun Wang
- Rice ECE: Rich Baraniuk, Kevin Kelly, etc.
- UCLA: Stan Osher, Jerome Darbon, Bin Dong, Yu Mao