# Fast algorithms for nonconvex compressive sensing

Rick Chartrand

Los Alamos National Laboratory

New Mexico Consortium

September 2, 2009

# Outline

Motivating Example

Nonconvex compressive sensing

Examples

Fast algorithm

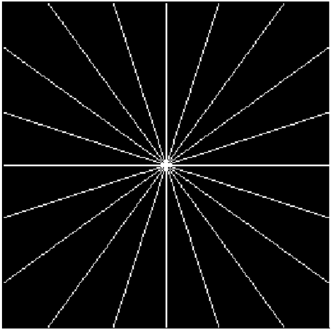
Summary

# Motivating example

Suppose we want to reconstruct an image from samples of its Fourier transform. How many samples do we need?

Consider radial sampling, such as in MRI or (roughly) CT.

Shepp-Logan phantom, $x$



$\Omega$

# Nonconvexity is better

Fewer measurements are needed with nonconvex minimization:

$$\min_{u} \|Du\|_p^p, \text{ subject to } (\mathcal{F}u)|_\Omega = (\mathcal{F}x)|_\Omega.$$

# Nonconvexity is better

Fewer measurements are needed with nonconvex minimization:

$$\min_u \|Du\|_p^p, \text{ subject to } (\mathcal{F}u)|_\Omega = (\mathcal{F}x)|_\Omega.$$

With $p = 1$, solution is $u = x$ with 18 lines ($\frac{|\Omega|}{|x|} = 6.9\%$).

backprojection, 18 lines

$p = 1$, 18 lines

# Nonconvexity is better

Fewer measurements are needed with nonconvex minimization:

$$\min_{\boldsymbol{u}} \|D\boldsymbol{u}\|_p^p, \text{ subject to } (\mathcal{F}\boldsymbol{u})|_\Omega = (\mathcal{F}\boldsymbol{x})|_\Omega.$$

With $p = 1$, solution is $u = x$ with 18 lines ($\frac{|\Omega|}{|x|} = 6.9\%$).

With $p = 1/2$, 10 lines suffice ($\frac{|\Omega|}{|x|} = 3.8\%$). (More than $10^{4500}$ local minima.)

backprojection, 18 lines

$p = 1$, 18 lines

$p = \frac{1}{2}$, 10 lines

$p = 1$, 10 lines

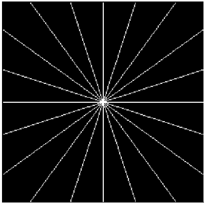
# New results

These are old results (Mar. 2006); what's new?

# New results

These are old results (Mar. 2006); what's new?

- Reconstruction (to 50 dB) in 13 seconds (in Matlab; versus literature-best 1–3 minutes).



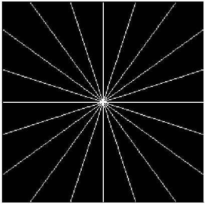
10 lines

fastest 10-line recovery

# New results

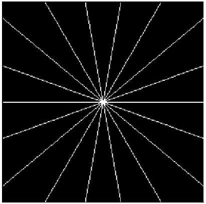
These are old results (Mar. 2006); what's new?

- Reconstruction (to 50 dB) in 13 seconds (in Matlab; versus literature-best 1–3 minutes).
- Exact reconstruction from 9 lines (3.5% of Fourier transform).



10 lines

fastest 10-line recovery

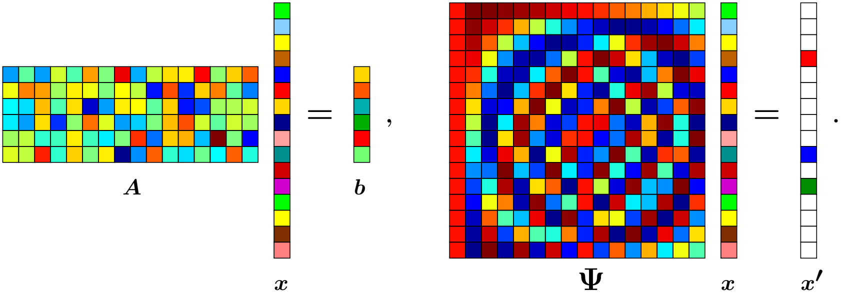


9 lines

recovery from fewest samples

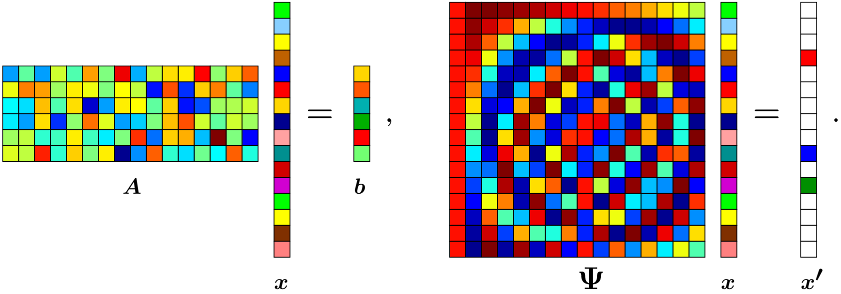
# Outline

# What is compressive sensing?



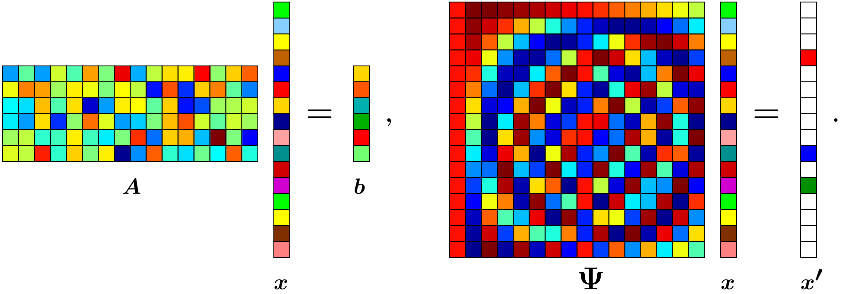

- Compressive sensing is the reconstruction of sparse signals $x$ from surprisingly few incoherent measurements $b = Ax$.

# What is compressive sensing?



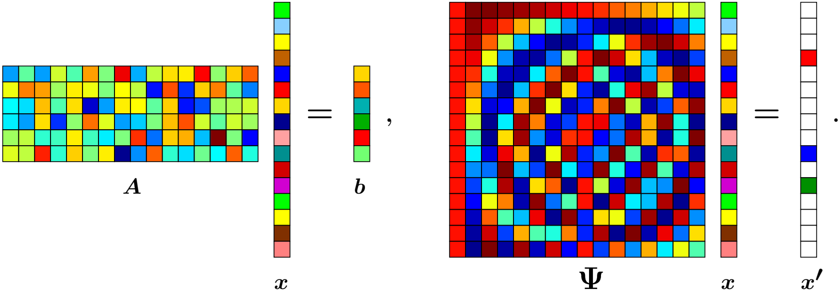

- Compressive sensing is the reconstruction of sparse signals $x$ from surprisingly few incoherent measurements $b = Ax$.
- We suppose the existence of an operator or dictionary $\Psi$ such that most of the components of $\Psi x$ are (nearly) zero.

# What is compressive sensing?




- An undersampled measurement $\boldsymbol{Ax}$ is tantamount to a compressed version of $\boldsymbol{x}$. If $\boldsymbol{x}$ is sufficiently sparse, it can be recovered perfectly.

# What is compressive sensing?




- An undersampled measurement $Ax$ is tantamount to a compressed version of $x$. If $x$ is sufficiently sparse, it can be recovered perfectly.
- We exploit the fact that sparsity is mathematically special, yet a general property of natural or human signals.

# Optimization for sparse recovery

- Let $x \in \mathbb{R}^N$ be sparse: $\|\Psi x\|_0 = K$, $K \ll N$.

# Optimization for sparse recovery

- Let $x \in \mathbb{R}^N$ be sparse: $\|\Psi x\|_0 = K$, $K \ll N$.
- Suppose $A$ is an $M \times N$ matrix, $M \ll N$, with $A$ and $\Psi$ *incoherent*. For example, $A = (a_{ij})$, i.i.d. $a_{ij} \sim N(0, \sigma^2)$. Let $b = Ax$.

# Optimization for sparse recovery

- Let $x \in \mathbb{R}^N$ be sparse: $\|\Psi x\|_0 = K$, $K \ll N$.
- Suppose $A$ is an $M \times N$ matrix, $M \ll N$, with $A$ and $\Psi$ *incoherent*. For example, $A = (a_{ij})$, i.i.d. $a_{ij} \sim N(0, \sigma^2)$. Let $b = Ax$.

$$\min_u \|\Psi u\|_0, \text{ s.t. } Au = b.$$

Unique solution is $u = x$ with optimally small $M$, but is NP-hard.

$M \geq 2K$ suffices with probability 1.

# Optimization for sparse recovery

- Let $x \in \mathbb{R}^N$ be sparse: $\|\Psi x\|_0 = K$, $K \ll N$.
- Suppose $A$ is an $M \times N$ matrix, $M \ll N$, with $A$ and $\Psi$ *incoherent*. For example, $A = (a_{ij})$, i.i.d. $a_{ij} \sim N(0, \sigma^2)$. Let $b = Ax$.

$$\min_u \|\Psi u\|_0, \text{ s.t. } Au = b.$$

Unique solution is $u = x$ with optimally small $M$, but is NP-hard.
$M \geq 2K$ suffices with probability 1.

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

Can be solved efficiently; requires more measurements for reconstruction. $M \geq CK \log(N/K)$

# Optimization for sparse recovery

- Let $x \in \mathbb{R}^N$ be sparse: $\|\Psi x\|_0 = K$, $K \ll N$.
- Suppose $A$ is an $M \times N$ matrix, $M \ll N$, with $A$ and $\Psi$ *incoherent*. For example, $A = (a_{ij})$, i.i.d. $a_{ij} \sim N(0, \sigma^2)$. Let $b = Ax$.

$$\min_u \|\Psi u\|_0, \text{ s.t. } Au = b.$$

Unique solution is $u = x$ with optimally small $M$, but is NP-hard. $M \geq 2K$ suffices with probability 1.

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

Can be solved efficiently; requires more measurements for reconstruction. $M \geq CK \log(N/K)$

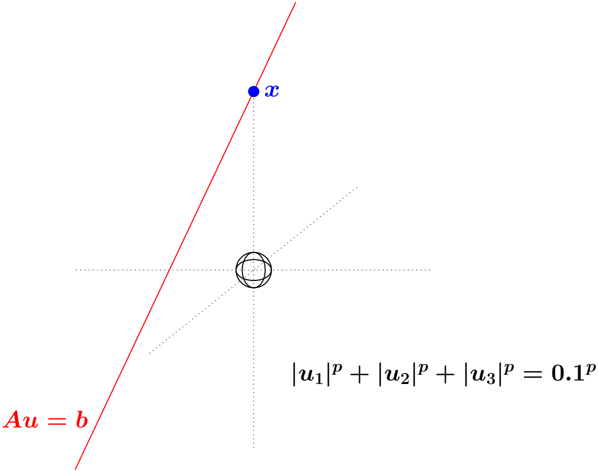
$$\min_u \|\Psi u\|_p^p, \text{ s.t. } Au = b,$$

where $0 < p < 1$. Solvable in practice; requires fewer measurements than $\ell^1$. $M \geq C_1(p)K + pC_2(p)K \log(N/K)$ (with V. Staneva)

# The geometry of $\ell^p$

$$\min_u \|u\|_p^p, \text{ subject to } Au = b$$

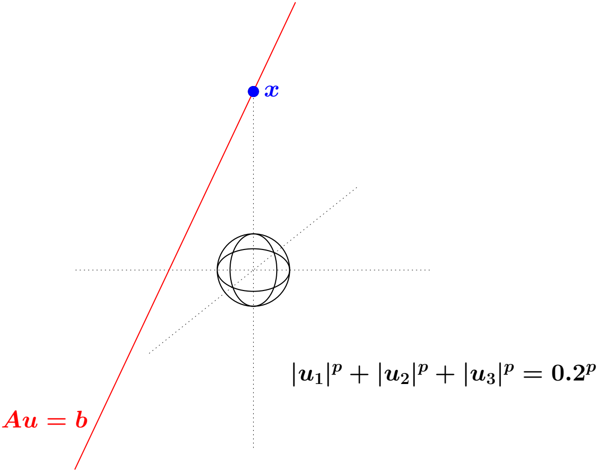
$p = 2$:



# The geometry of $\ell^p$

$$\min_u \|u\|_p^p, \text{ subject to } Au = b$$

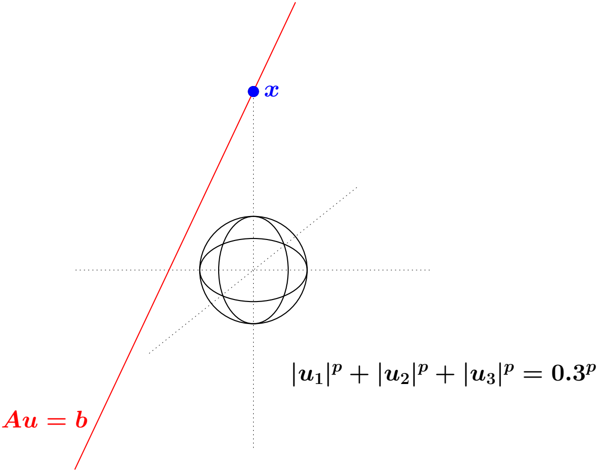
$p = 2$:



# The geometry of $\ell^p$

$$\min_u \|u\|_p^p, \text{ subject to } Au = b$$

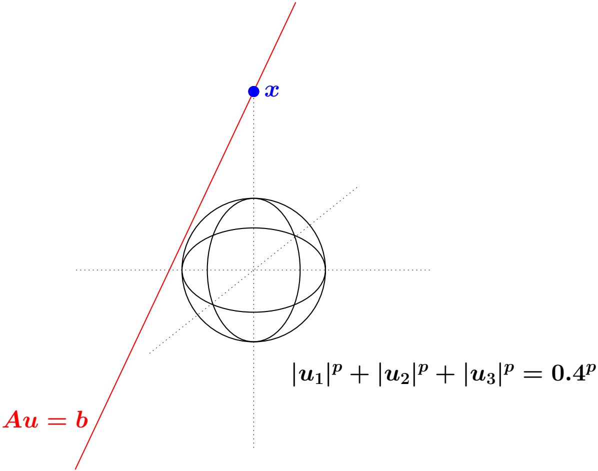
$p = 2$:



# The geometry of $\ell^p$

$$\min_u \|u\|_p^p, \text{ subject to } Au = b$$

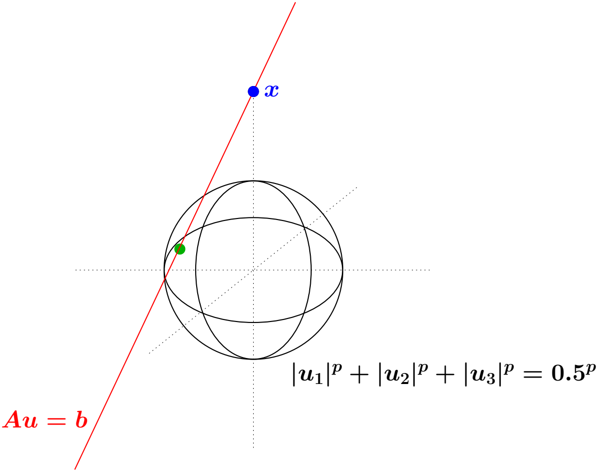
$p = 2$:



# The geometry of $\ell^p$

$$\min_u \|u\|_p^p, \text{ subject to } Au = b$$

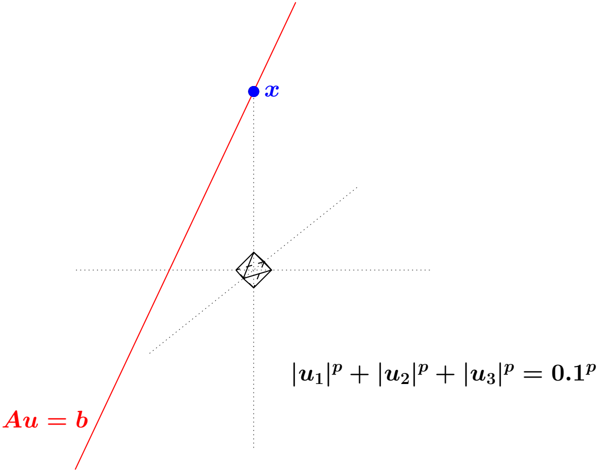
$p = 2$:



# The geometry of $\ell^p$

$$\min_u \|u\|_p^p, \text{ subject to } Au = b$$

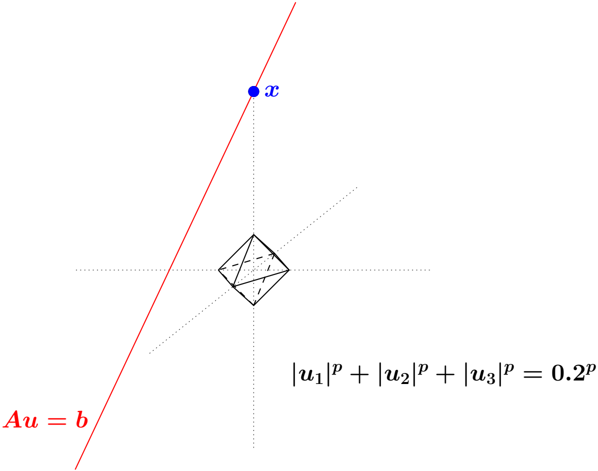
$p = 1$:



# The geometry of $\ell^p$

$$\min_u \|u\|_p^p, \text{ subject to } Au = b$$

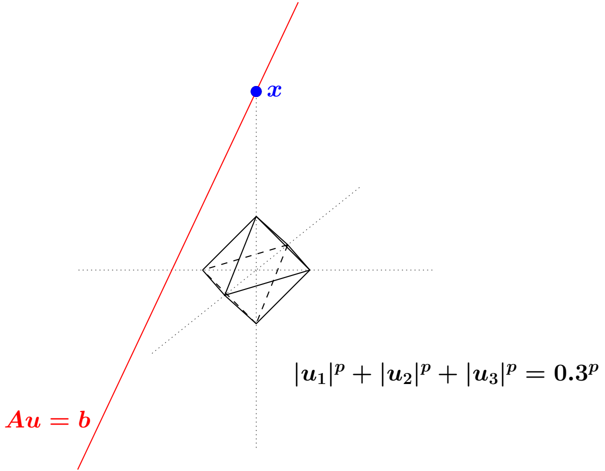
$p = 1$:



# The geometry of $\ell^p$

$$\min_u \|u\|_p^p, \text{ subject to } Au = b$$

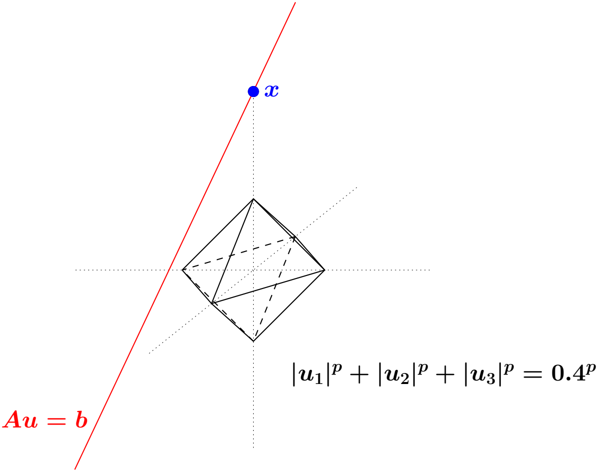
$p = 1$:



# The geometry of $\ell^p$

$$\min_u \|u\|_p^p, \text{ subject to } Au = b$$

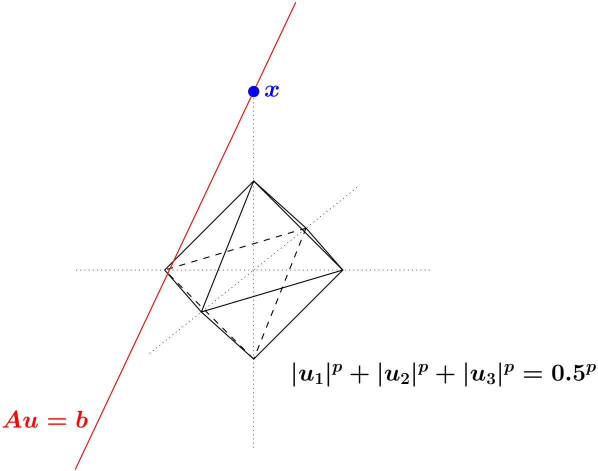
$p = 1$:



# The geometry of $\ell^p$

$$\min_u \|u\|_p^p, \text{ subject to } Au = b$$

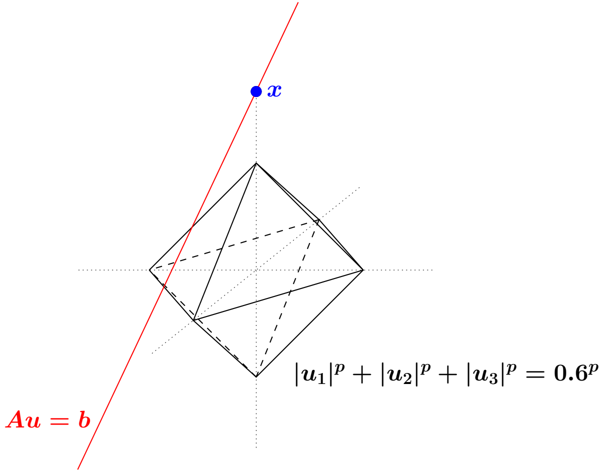
$p = 1$:



# The geometry of $\ell^p$

$$\min_u \|u\|_p^p, \text{ subject to } Au = b$$

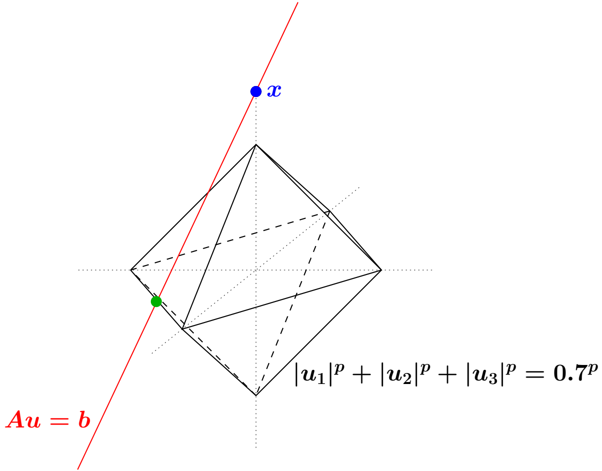
$p = 1$:



# The geometry of $\ell^p$

$$\min_u \|u\|_p^p, \text{ subject to } Au = b$$

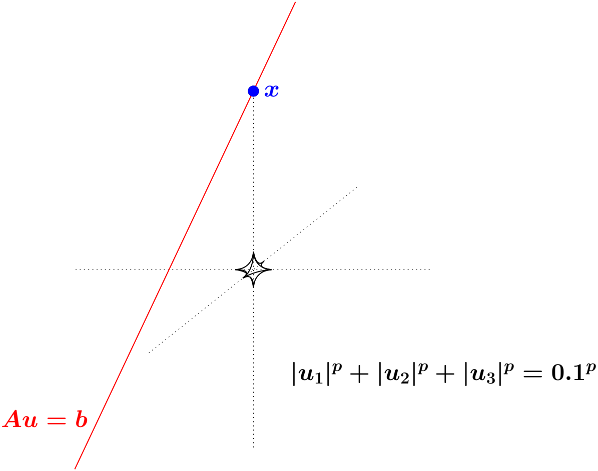
$p = 1$:



# The geometry of $\ell^p$

$$\min_u \|u\|_p^p, \text{ subject to } Au = b$$

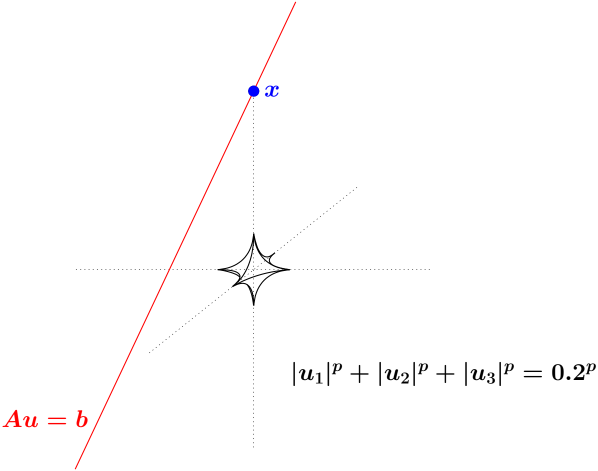
$p = 1/2$:



# The geometry of $\ell^p$

$$\min_u \|u\|_p^p, \text{ subject to } Au = b$$

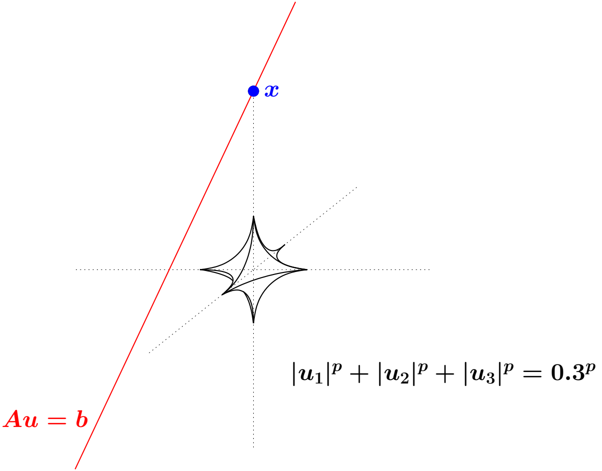
$p = 1/2$:



# The geometry of $\ell^p$

$$\min_u \|u\|_p^p, \text{ subject to } Au = b$$

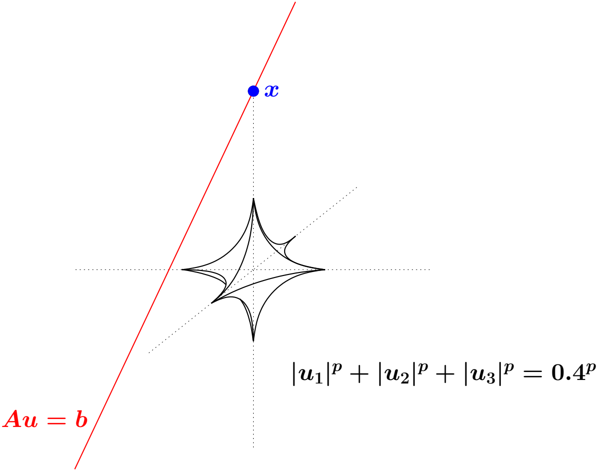
$p = 1/2$:



# The geometry of $\ell^p$

$$\min_u \|u\|_p^p, \text{ subject to } Au = b$$

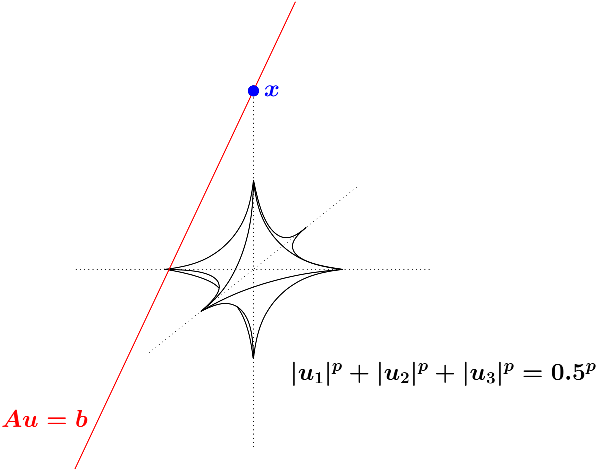
$p = 1/2$:



# The geometry of $\ell^p$

$$\min_u \|u\|_p^p, \text{ subject to } Au = b$$

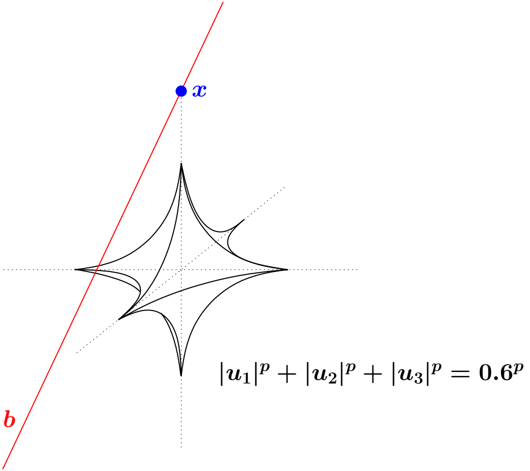
$p = 1/2$:



# The geometry of $\ell^p$

$$\min_u \|u\|_p^p, \text{ subject to } Au = b$$

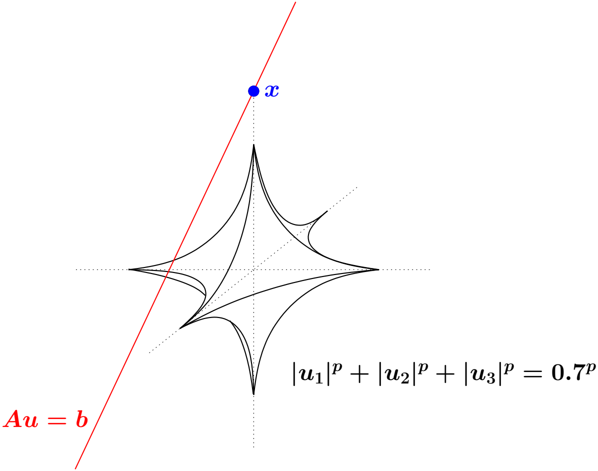
$p = 1/2$:



# The geometry of $\ell^p$

$$\min_u \|u\|_p^p, \text{ subject to } Au = b$$

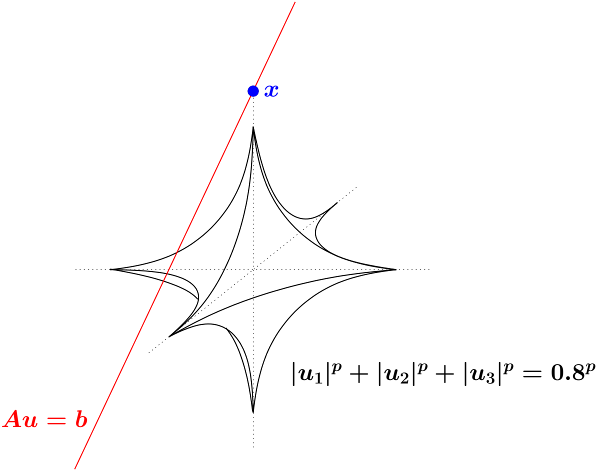
$p = 1/2$:



# The geometry of $\ell^p$

$$\min_u \|u\|_p^p, \text{ subject to } Au = b$$

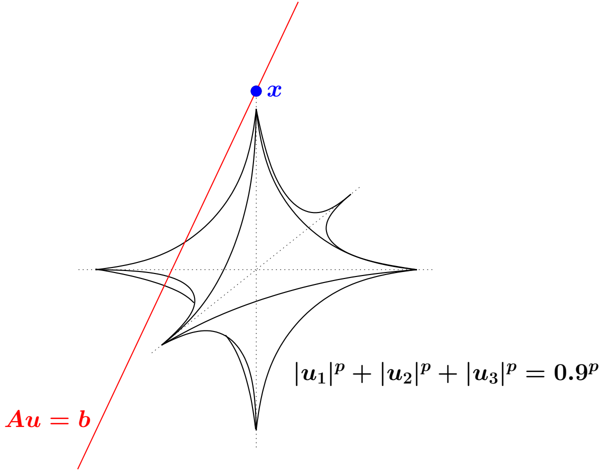
$p = 1/2$:



# The geometry of $\ell^p$

$$\min_u \|u\|_p^p, \text{ subject to } Au = b$$

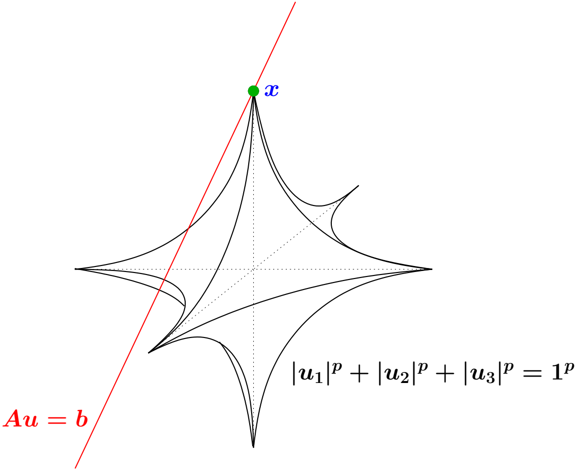
$p = 1/2$:



# The geometry of $\ell^p$

$$\min_u \|u\|_p^p, \text{ subject to } Au = b$$

$p = 1/2$:



# Why might global minimization be possible?

Consider an $\epsilon$-regularized objective, restricted to the feasible plane:

$$\sum_{i=1}^{N} \left(u_i^2 + \epsilon\right)^{p/2}.$$

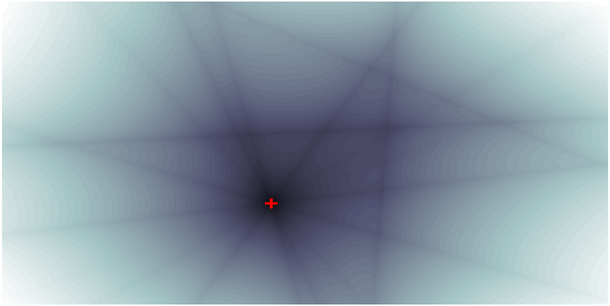
A moderate $\epsilon$ fills in the local minima.

# Why might global minimization be possible?

Consider an $\epsilon$-regularized objective, restricted to the feasible plane:

$$\sum_{i=1}^{N} (u_i^2 + \epsilon)^{p/2}.$$

A moderate $\epsilon$ fills in the local minima.



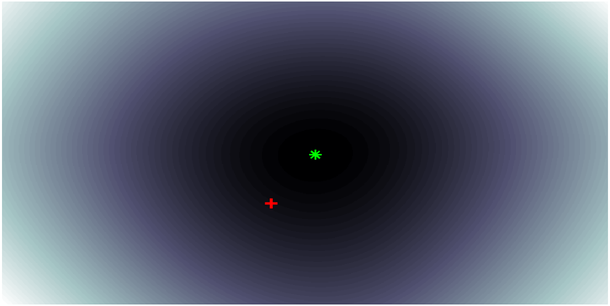
$\epsilon = 0$

# Why might global minimization be possible?

Consider an $\epsilon$-regularized objective, restricted to the feasible plane:

$$\sum_{i=1}^{N} (u_i^2 + \epsilon)^{p/2}.$$

A moderate $\epsilon$ fills in the local minima.



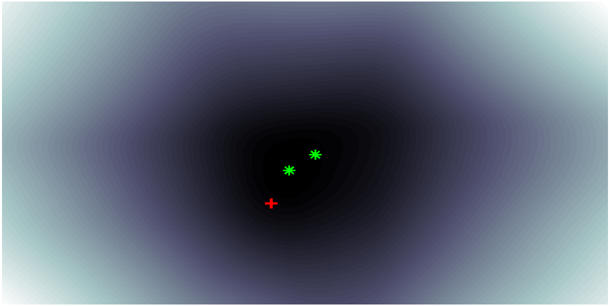
$\epsilon = 1$

# Why might global minimization be possible?

Consider an $\epsilon$-regularized objective, restricted to the feasible plane:

$$\sum_{i=1}^{N} (u_i^2 + \epsilon)^{p/2}.$$

A moderate $\epsilon$ fills in the local minima.



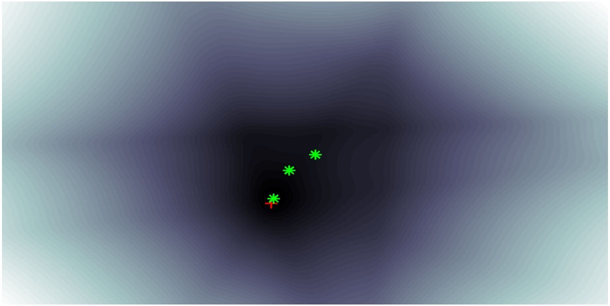
$\epsilon = 0.1$

# Why might global minimization be possible?

Consider an $\epsilon$-regularized objective, restricted to the feasible plane:

$$\sum_{i=1}^{N} (u_i^2 + \epsilon)^{p/2}.$$

A moderate $\epsilon$ fills in the local minima.



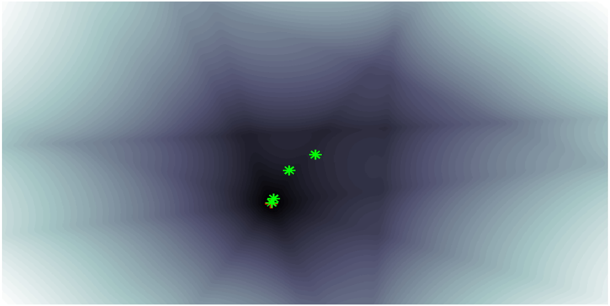
$\epsilon = 0.01$

# Why might global minimization be possible?

Consider an $\epsilon$-regularized objective, restricted to the feasible plane:

$$\sum_{i=1}^{N} (u_i^2 + \epsilon)^{p/2}.$$

A moderate $\epsilon$ fills in the local minima.



$$\epsilon = 0.001$$

# Outline

Motivating Example

Nonconvex compressive sensing

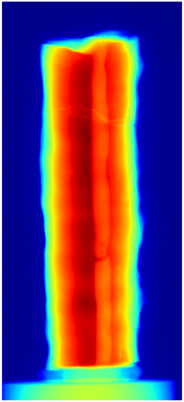
Examples

Fast algorithm

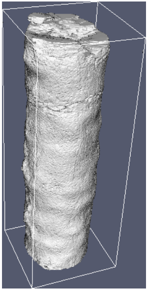
Summary

# 3-D tomography

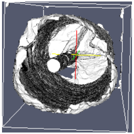
Six radiographs allow reconstruction of a stalagmite segment:



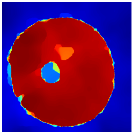
radiograph



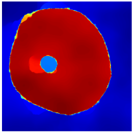
isosurface



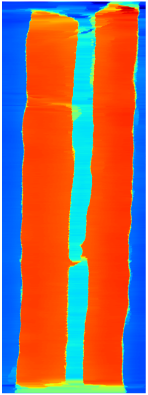
iso from end



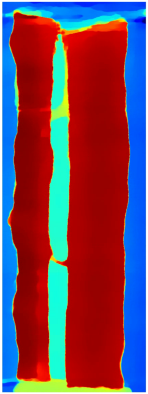
one $z$ slice



lower $z$ slice

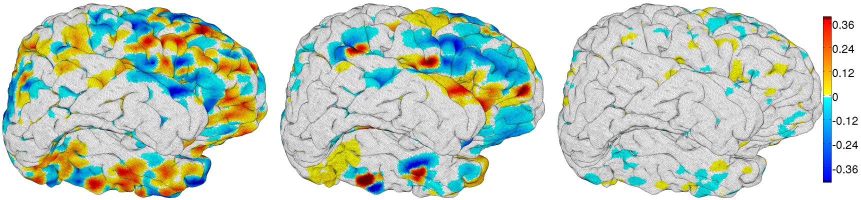


$x$ slice



$y$ slice

# Cortical activity reconstruction from EEG

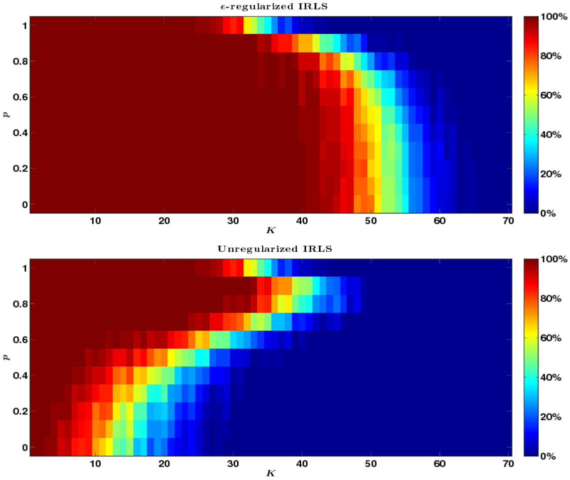



eigenfunction basis | wavelet basis | not sparse, noisy

Cortical activity is reconstructed perfectly from synthetic EEG data, consisting of 256 scalp potential measurements. The synthetic signals have 80 nonzero coefficients from graph-diffusion eigenfunction or wavelet bases. Making the signal only approximately sparse and adding noise results in very little reconstruction error.

# Numerical tests

Reconstruction frequency from 100 random measurements of 256-dimensional signals, using IRLS with and without regularization.



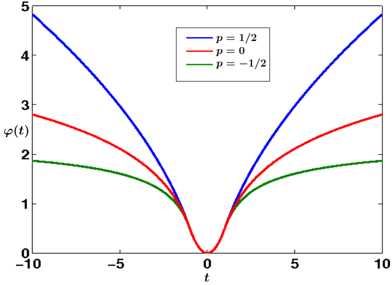
# Outline

# Semiconvex regularization

Now we generalize an approach of J. Yang, W. Yin, Y. Zhang, and Y. Wang. Consider a componentwise, mollified $\ell^p$ objective:

$$\varphi(t) = \begin{cases} \gamma|t|^2 & \text{if } |t| \le \alpha \\ |t|^p/p - \delta & \text{if } |t| > \alpha \end{cases}$$

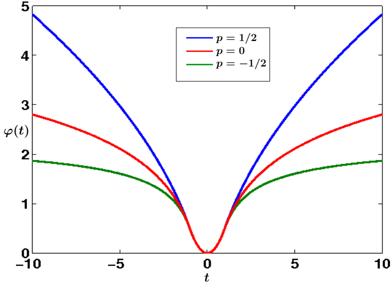
The parameters are chosen to make $\varphi \in C^1$.



# Semiconvex regularization

Now we generalize an approach of J. Yang, W. Yin, Y. Zhang, and Y. Wang. Consider a componentwise, mollified $\ell^p$ objective:

$$\varphi(t) = \begin{cases} \gamma|t|^2 & \text{if } |t| \leq \alpha \\ |t|^p/p - \delta & \text{if } |t| > \alpha \end{cases}$$




The parameters are chosen to make $\varphi \in C^1$.

Now we seek $\psi$ such that

$$\varphi(t) = \min_w \{\psi(w) + (\beta/2)|t - w|_2^2\}$$

This can be found by convex duality, as $|t|_2^2/2 - \varphi(t)/\beta$ is convex if $\beta = \alpha^{p-2}$.

# A splitting approach

Now we consider an unconstrained $\ell^p$ minimization problem, and replace

$$\min_{u} \sum_{i=1}^{N} \varphi((\Psi u)_i) + (\mu/2)\|Au - b\|_2^2$$

with the split version

$$\min_{u,w} \sum_{i=1}^{N} \psi(w_i) + (\beta/2)\|\Psi u - w\|_2^2 + (\mu/2)\|Au - b\|_2^2,$$

which we solve by alternate minimization.

# Easy iterations

Holding $u$ fixed, the $w$-subproblem is separable, and its solution comes from the convex duality:

$$w_i = \max\left\{0, |(\Psi u)_i| - \frac{|(\Psi u)_i|^{p-1}}{\beta}\right\}\frac{(\Psi u)_i}{|(\Psi u)_i|}.$$

This generalizes shrinkage ( or soft thresholding ) to $\ell^p$.

# Easy iterations

Holding $u$ fixed, the $w$-subproblem is separable, and its solution comes from the convex duality:

$$w_i = \max\left\{0, |(\Psi u)_i| - \frac{|(\Psi u)_i|^{p-1}}{\beta}\right\} \frac{(\Psi u)_i}{|(\Psi u)_i|}.$$

This generalizes shrinkage ( or soft thresholding ) to $\ell^p$.

Holding $w$ fixed, the $u$-problem is quadratic:

$$(\beta\Psi^*\Psi + \mu A^* A)u = \beta\Psi^* w + \mu A^* b.$$

# Easy iterations

Holding $u$ fixed, the $w$-subproblem is separable, and its solution comes from the convex duality:

$$w_i = \max\left\{0, |(\Psi u)_i| - \frac{|(\Psi u)_i|^{p-1}}{\beta}\right\} \frac{(\Psi u)_i}{|(\Psi u)_i|}.$$

This generalizes shrinkage ( or soft thresholding ) to $\ell^p$.

Holding $w$ fixed, the $u$-problem is quadratic:

$$(\beta\Psi^*\Psi + \mu A^* A)u = \beta\Psi^* w + \mu A^* b.$$

If A is a Fourier sampling operator and $\Psi$ is Fourier-diagonalizable (such as a derivative operator or orthogonal wavelet transform), we can solve this in the Fourier domain. This is very fast!

# Enforcing equality

$$\min_{u,w} \sum_{i=1}^{N} \psi(w_i) + (\beta/2)\|\Psi u - w\|_2^2 + (\mu/2)\|Au - b\|_2^2$$

Typically one enforces $w = \Psi u$ (and $Au = b$, if desired) by iteratively growing $\beta$ (and $\mu$). (continuation)

# Enforcing equality

$$\min_{u,w} \sum_{i=1}^{N} \psi(w_i) + (\beta/2)\|\Psi u - w\|_2^2 + (\mu/2)\|Au - b\|_2^2$$

Typically one enforces $w = \Psi u$ (and $Au = b$, if desired) by iteratively growing $\beta$ (and $\mu$). (continuation)

We get better results from an augmented Lagrangian (cf. *split Bregman*, T. Goldstein, S. Osher):

$$\min_{u,w} \sum_{i=1}^{N} \psi(w_i) + (\beta/2)\|Du - w - \lambda_1\|_2^2 + (\mu/2)\|Au - b - \lambda_2\|_2^2,$$

and update $\lambda_1^{n+1} = \lambda_1^n + w - \Psi u$, $\lambda_2^{n+1} = \lambda_2^n + b - Au$.

# Multiple penalty terms and other constraints

The same approach can easily handle two or more penalty terms in the objective:

$$\min_{u,w,v} \sum_{i=1}^{N} \psi_1(w_i) + (\beta_1/2)\|\Psi_1 u - w\|_2^2 + \sum_{i=1}^{N} \psi_2(v_i) + (\beta_2/2)\|\Psi_2 u - v\|_2^2 + (\mu/2)\|Au - b\|_2^2$$

# Multiple penalty terms and other constraints

The same approach can easily handle two or more penalty terms in the objective:

$$\min_{u,w,v} \sum_{i=1}^{N} \psi_1(w_i) + (\beta_1/2)\|\Psi_1 u - w\|_2^2 + \sum_{i=1}^{N} \psi_2(v_i) + (\beta_2/2)\|\Psi_2 u - v\|_2^2 + (\mu/2)\|Au - b\|_2^2$$

One can also use a similar splitting/augmented-Lagrangian approach to handle inequality noise constraints, nonnegativity constraints, etc. The method is very flexible.

# Summary

- Nonconvex compressive sensing allows compressible images to be recovered with even fewer measurements than "traditional" compressive sensing.
- Nonconvexity also improves robustness to noise and signal nonsparsity.
- Regularizing the objective appears to keep algorithms from converging to nonglobal minima.
- For Fourier-sampling measurements, the reconstruction can be done very fast.

`math.lanl.gov/~rick`