Fast algorithms for nonconvex compressive sensing

Rick Chartrand

Los Alamos National Laboratory

New Mexico Consortium

February 25, 2009

Outline

Motivating example

Nonconvex compressive sensing

Algorithms

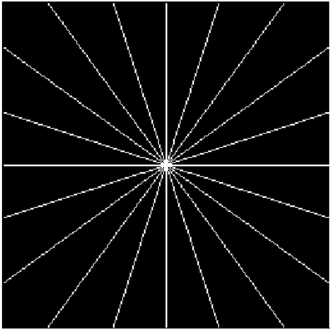
Summary

Sparse tomography

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



Ω

Nonconvexity

Fewer measurements are needed with nonconvex minimization:

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

Nonconvexity

Fewer measurements are needed with nonconvex minimization:

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

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

backprojection, 18 lines

$p = 1$, 18 lines

Nonconvexity

Fewer measurements are needed with nonconvex minimization:

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

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

With $p = 1/2$, 10 lines suffice ($\frac{|\Omega|}{|x|} = 3.8\%$).

backprojection, 18 lines

$p = 1$, 18 lines

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

$p = 1$, 10 lines

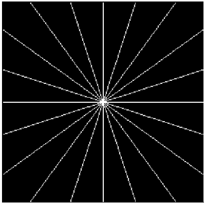
New results

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 (versus literature-best 1–3 minutes).



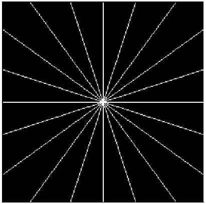
10 lines

fastest 10-line recovery

New results

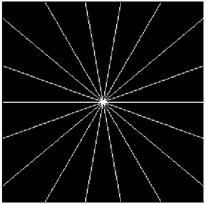
These are old results (Mar. 2006); what's new?

- Reconstruction (to 50 dB) in 13 seconds (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

Optimization for sparse recovery

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

$$\min_u \|\Psi u\|_0, \text{ s.t. } \Phi u = \Phi x.$$

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

$M \geq 2K$ suffices w.h.p.

$$\min_u \|\Psi u\|_1, \text{ s.t. } \Phi u = \Phi x.$$

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 Φ is an $M \times N$ matrix, $M \ll N$, with Φ and Ψ incoherent. For example, $\Phi = (\varphi_{ij})$, i.i.d. $\varphi_{ij} \sim N(0, \sigma^2)$.

$$\min_u \|\Psi u\|_0, \text{ s.t. } \Phi u = \Phi x.$$

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

$$\min_u \|\Psi u\|_1, \text{ s.t. } \Phi u = \Phi x.$$

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

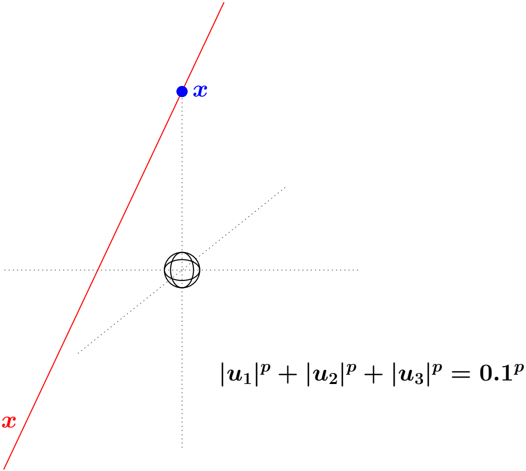
$$\min_u \|\Psi u\|_p^p, \text{ s.t. } \Phi u = \Phi x,$$

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

The geometry of ℓ^p

$$\min_u \|u\|_p^p, \text{ subject to } \Phi u = \Phi x$$

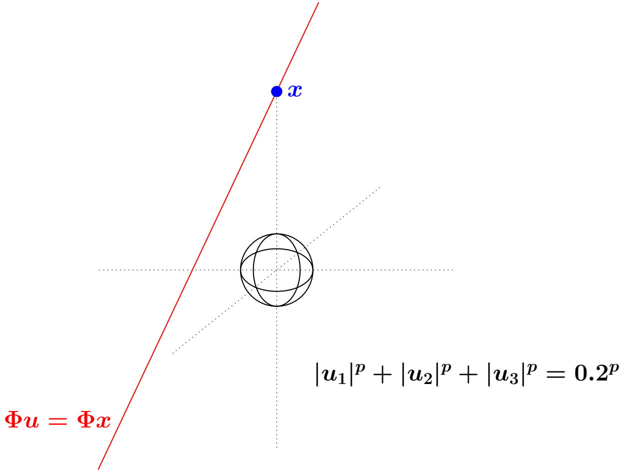
$p = 2$:



The geometry of ℓ^p

$$\min_u \|u\|_p^p, \text{ subject to } \Phi u = \Phi x$$

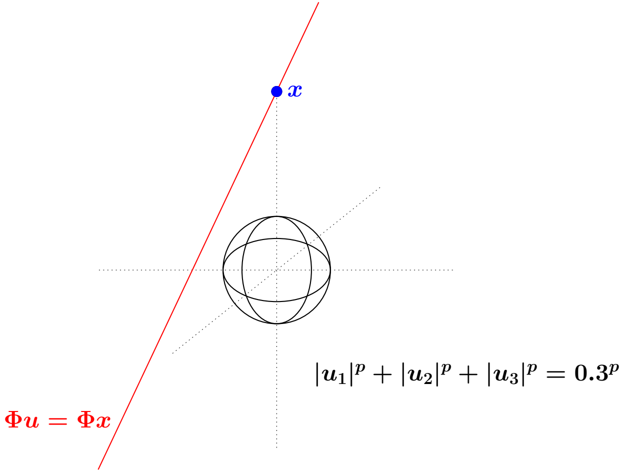
$p = 2$:



The geometry of ℓ^p

$$\min_u \|u\|_p^p, \text{ subject to } \Phi u = \Phi x$$

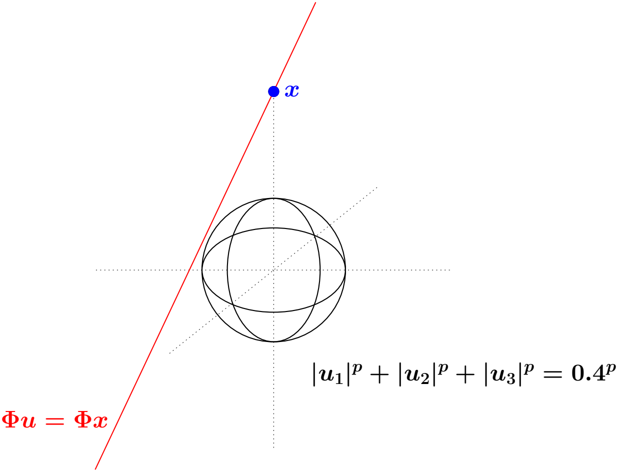
$p = 2$:



The geometry of ℓ^p

$$\min_u \|u\|_p^p, \text{ subject to } \Phi u = \Phi x$$

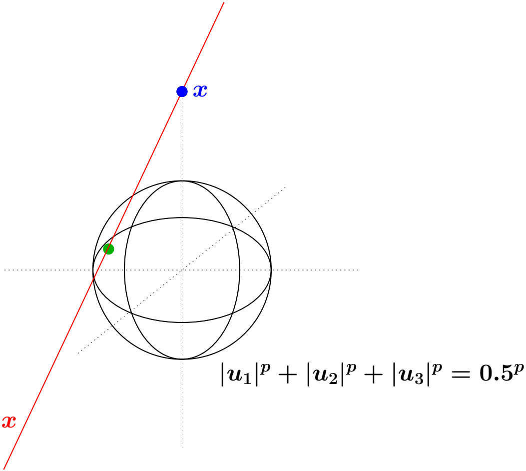
$p = 2$:



The geometry of ℓ^p

$$\min_u \|u\|_p^p, \text{ subject to } \Phi u = \Phi x$$

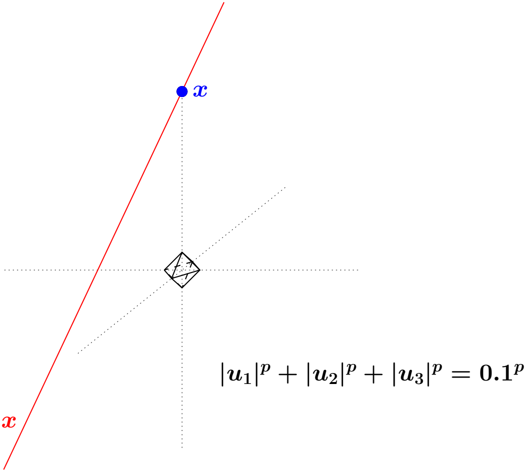
$p = 2$:



The geometry of ℓ^p

$$\min_u \|u\|_p^p, \text{ subject to } \Phi u = \Phi x$$

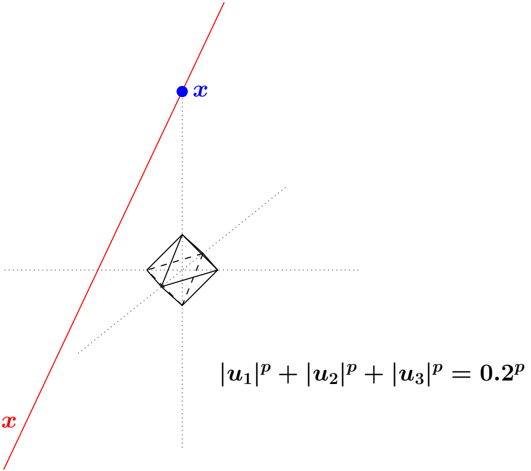
$p = 1$:



The geometry of ℓ^p

$$\min_u \|u\|_p^p, \text{ subject to } \Phi u = \Phi x$$

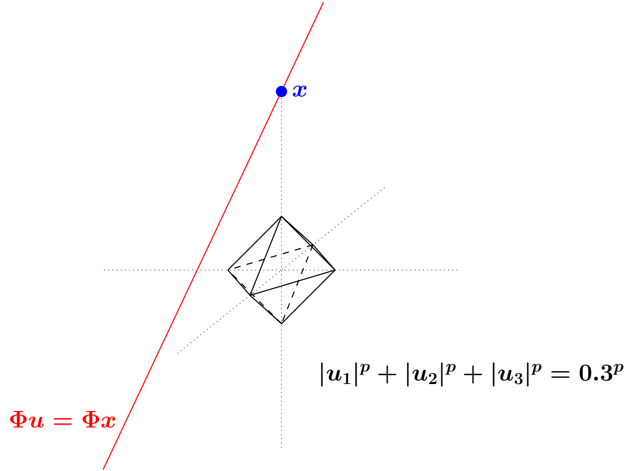
$p = 1$:



The geometry of ℓ^p

$$\min_u \|u\|_p^p, \text{ subject to } \Phi u = \Phi x$$

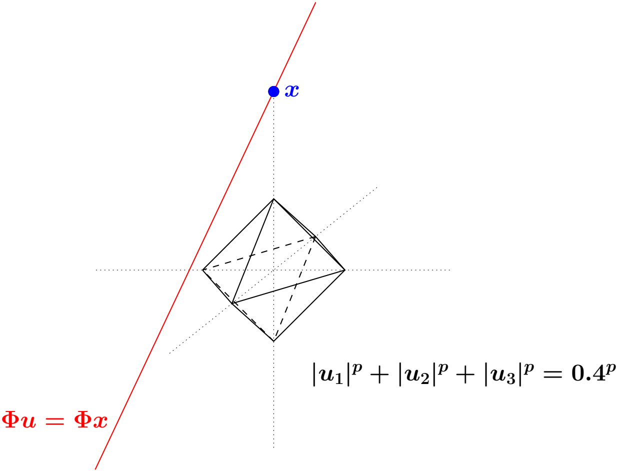
$p = 1$:



The geometry of ℓ^p

$$\min_u \|u\|_p^p, \text{ subject to } \Phi u = \Phi x$$

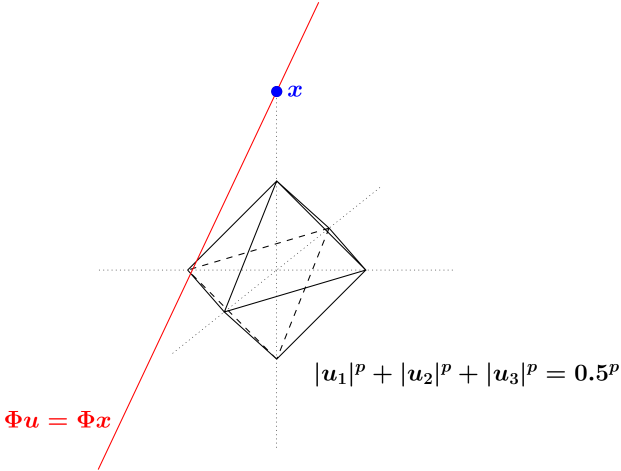
$p = 1$:



The geometry of ℓ^p

$$\min_u \|u\|_p^p, \text{ subject to } \Phi u = \Phi x$$

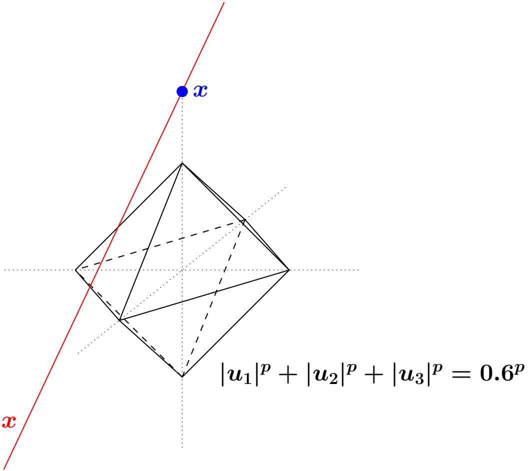
$p = 1$:



The geometry of ℓ^p

$$\min_u \|u\|_p^p, \text{ subject to } \Phi u = \Phi x$$

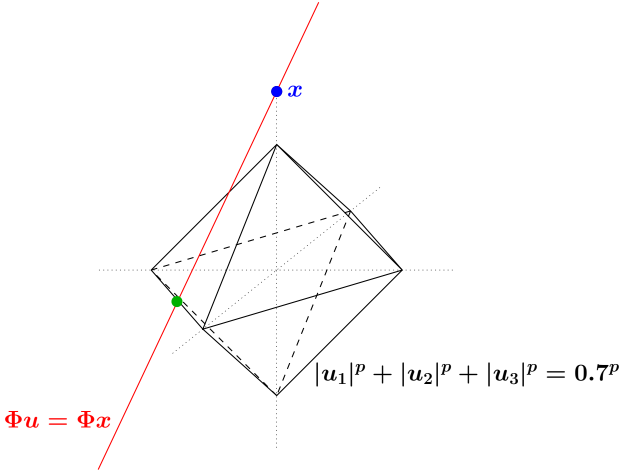
$p = 1$:



The geometry of ℓ^p

$$\min_u \|u\|_p^p, \text{ subject to } \Phi u = \Phi x$$

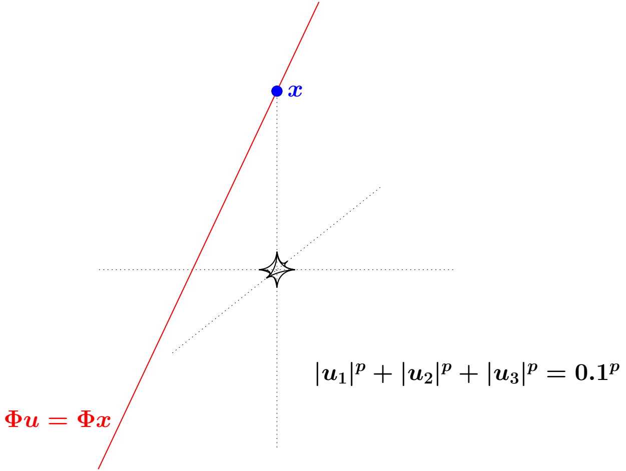
$p = 1$:



The geometry of ℓ^p

$$\min_u \|u\|_p^p, \text{ subject to } \Phi u = \Phi x$$

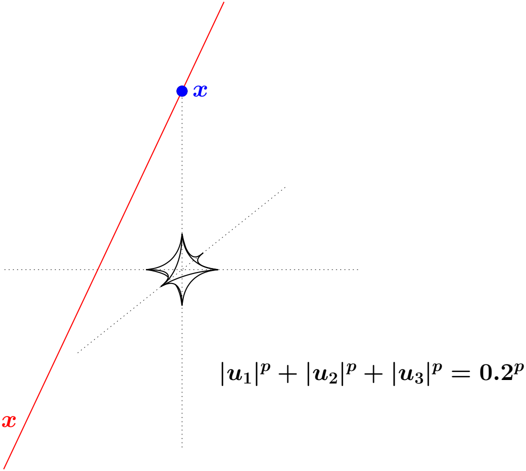
$p = 1/2$:



The geometry of ℓ^p

$$\min_u \|u\|_p^p, \text{ subject to } \Phi u = \Phi x$$

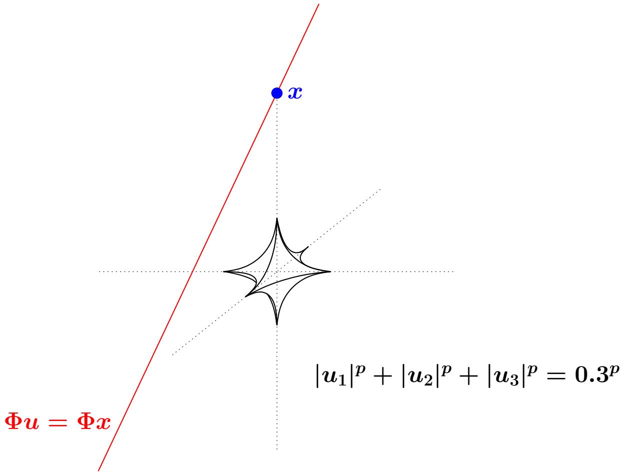
$p = 1/2$:



The geometry of ℓ^p

$$\min_u \|u\|_p^p, \text{ subject to } \Phi u = \Phi x$$

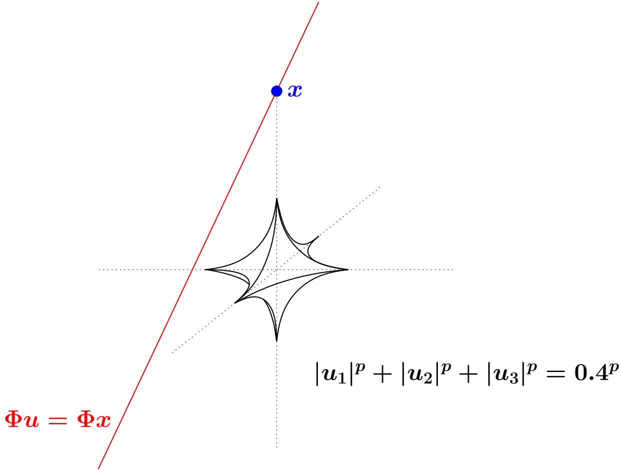
$p = 1/2$:



The geometry of ℓ^p

$$\min_u \|u\|_p^p, \text{ subject to } \Phi u = \Phi x$$

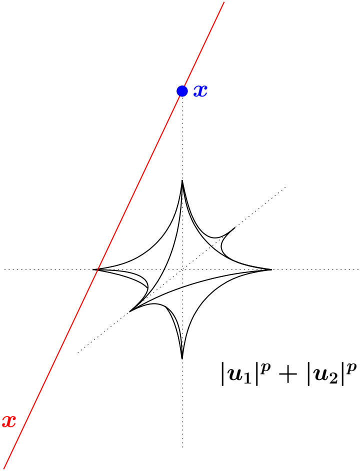
$p = 1/2$:



The geometry of ℓ^p

$$\min_u \|u\|_p^p, \text{ subject to } \Phi u = \Phi x$$

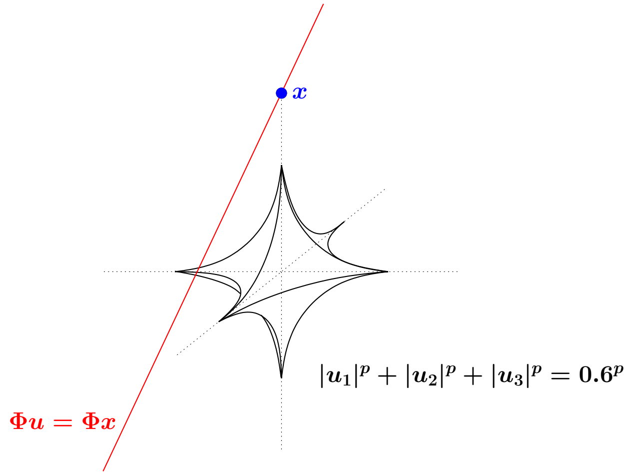
$p = 1/2$:



The geometry of ℓ^p

$$\min_u \|u\|_p^p, \text{ subject to } \Phi u = \Phi x$$

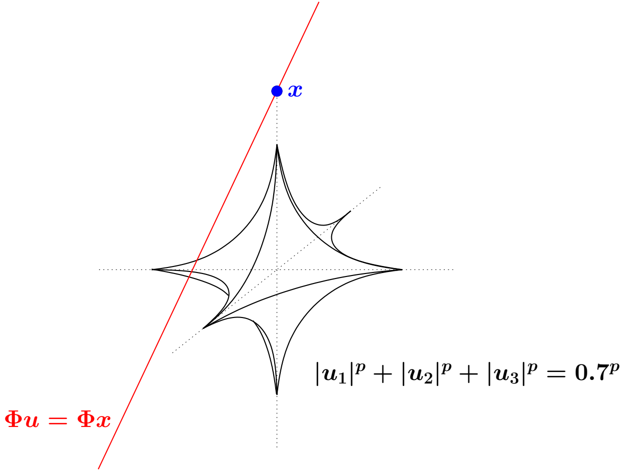
$p = 1/2$:



The geometry of ℓ^p

$$\min_u \|u\|_p^p, \text{ subject to } \Phi u = \Phi x$$

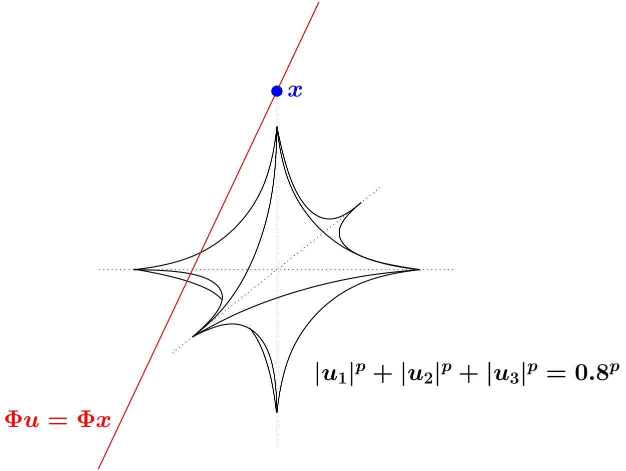
$p = 1/2$:



The geometry of ℓ^p

$$\min_u \|u\|_p^p, \text{ subject to } \Phi u = \Phi x$$

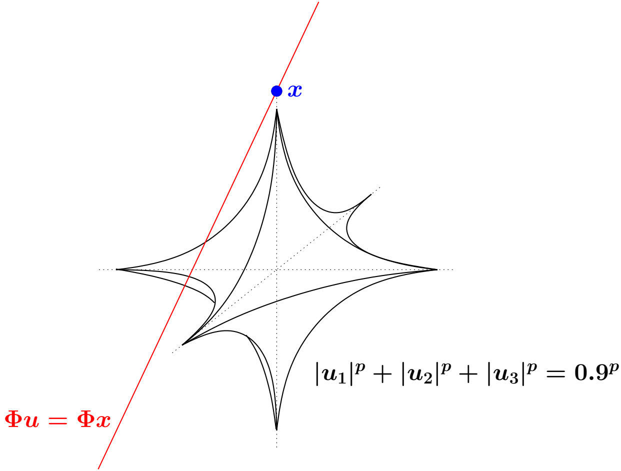
$p = 1/2$:



The geometry of ℓ^p

$$\min_u \|u\|_p^p, \text{ subject to } \Phi u = \Phi x$$

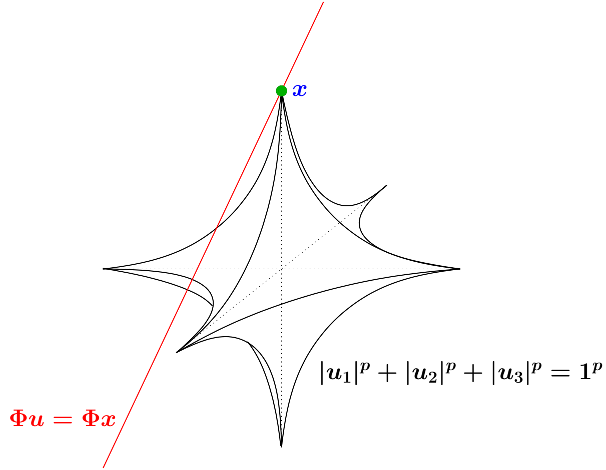
$p = 1/2$:



The geometry of ℓ^p

$$\min_u \|u\|_p^p, \text{ subject to } \Phi u = \Phi x$$

$p = 1/2$:



Why might global minimization be possible?

Consider the ϵ-regularized, constraint-eliminated objective:

$$F_\epsilon(t) = \sum_{i=1}^{N} \left\{ \left[x_i + (Vt)_i \right]^2 + \epsilon \right\}^{p/2},$$

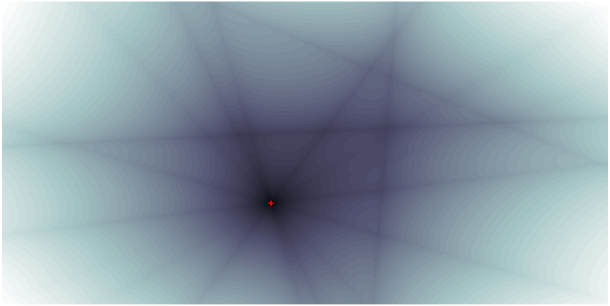
where $\mathcal{R}(V) = \mathcal{N}(\Phi)$. A moderate ϵ fills in the local minima.

Why might global minimization be possible?

Consider the ϵ-regularized, constraint-eliminated objective:

$$F_\epsilon(t) = \sum_{i=1}^{N} \left\{ \left[x_i + (Vt)_i \right]^2 + \epsilon \right\}^{p/2},$$

where $\mathcal{R}(V) = \mathcal{N}(\Phi)$. A moderate ϵ fills in the local minima.



$\epsilon = 0$

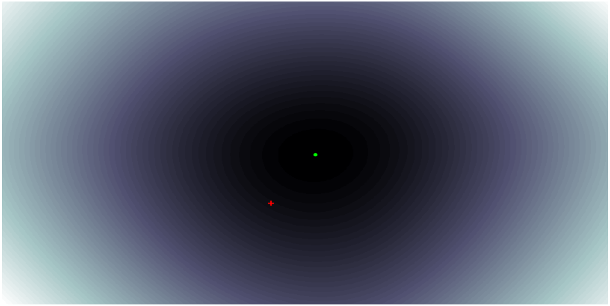
Los Alamos NATIONAL LABORATORY EST. 1943

Why might global minimization be possible?

Consider the ϵ-regularized, constraint-eliminated objective:

$$F_\epsilon(t) = \sum_{i=1}^{N} \left\{ \left[x_i + (Vt)_i \right]^2 + \epsilon \right\}^{p/2},$$

where $\mathcal{R}(V) = \mathcal{N}(\Phi)$. A moderate ϵ fills in the local minima.



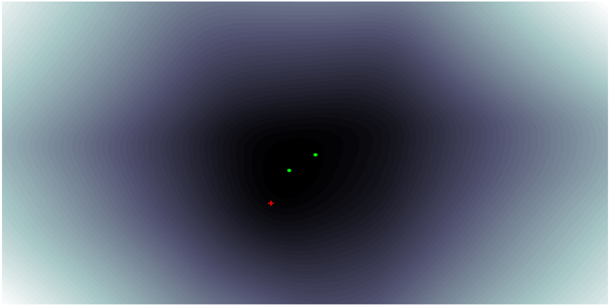
$\epsilon = 1$

Why might global minimization be possible?

Consider the ϵ-regularized, constraint-eliminated objective:

$$F_\epsilon(t) = \sum_{i=1}^{N} \left\{ \left[x_i + (Vt)_i\right]^2 + \epsilon \right\}^{p/2},$$

where $\mathcal{R}(V) = \mathcal{N}(\Phi)$. A moderate ϵ fills in the local minima.



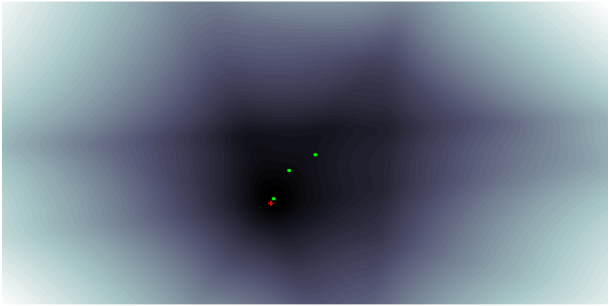
$\epsilon = 0.1$

Why might global minimization be possible?

Consider the ϵ-regularized, constraint-eliminated objective:

$$F_\epsilon(t) = \sum_{i=1}^{N} \left\{ \left[x_i + (Vt)_i \right]^2 + \epsilon \right\}^{p/2},$$

where $\mathcal{R}(V) = \mathcal{N}(\Phi)$. A moderate ϵ fills in the local minima.



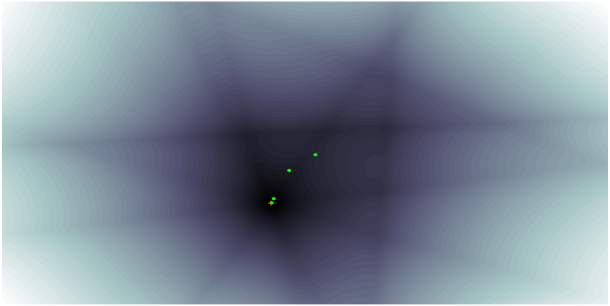
$\epsilon = 0.01$

Why might global minimization be possible?

Consider the ϵ-regularized, constraint-eliminated objective:

$$F_\epsilon(t) = \sum_{i=1}^{N} \left\{ \left[x_i + (Vt)_i \right]^2 + \epsilon \right\}^{p/2},$$

where $\mathcal{R}(V) = \mathcal{N}(\Phi)$. A moderate ϵ fills in the local minima.

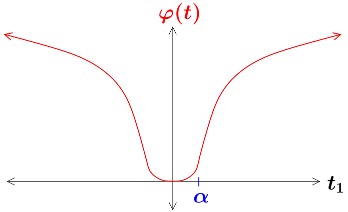


$\epsilon = 0.001$

Semiconvex regularization

Now we generalize an approach of J. Yang, W. Yin, Y. Zhang, and Y. Wang. Consider a mollified ℓ^p objective on $\mathbb{R}^2$:

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



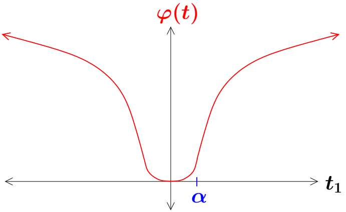
The parameters are chosen to make $\varphi \in C^1$.

Semiconvex regularization

Now we generalize an approach of J. Yang, W. Yin, Y. Zhang, and Y. Wang. Consider a mollified ℓ^p objective on $\mathbb{R}^2$:

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

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



Now we seek ψ 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 ℓ^p minimization problem, and replace

$$\min_u \sum_{i=1}^N \varphi((Du)_i) + (\mu/2)\|\Phi u - b\|_2^2$$

with the split version

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

which we solve by alternate minimization.

Easy iterations

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

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

This generalizes shrinkage (or soft thresholding) to ℓ^p.

Los Alamos NATIONAL LABORATORY EST. 1943

Operated by Los Alamos National Security, LLC for NNSA

Easy iterations

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

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

This generalizes shrinkage (or soft thresholding) to ℓ^p.

Holding w fixed, the u-problem is quadratic:

$$(\beta D^T D + \mu \Phi^T \Phi)u = \beta D^T w + \mu \Phi^T b.$$

Easy iterations

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

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

This generalizes shrinkage (or soft thresholding) to ℓ^p.

Holding w fixed, the u-problem is quadratic:

$$(\beta D^T D + \mu \Phi^T \Phi) u = \beta D^T w + \mu \Phi^T b.$$

If Φ is a Fourier sampling operator, we can solve this in the Fourier domain. This is very fast!

Enforcing equality

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

Typically one enforces $w = Du$ and $\Phi u = b$ by iteratively growing β and μ larger (continuation).

Enforcing equality

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

Typically one enforces $w = Du$ and $\Phi u = b$ by iteratively growing β and μ larger (continuation).

We get better results from Bregman iteration (generalizing T. Goldstein, S. Osher):

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

and update $B_w^{n+1} = B_w^n + w - Du$ (inner loop), $B_u^{m+1} = B_u^m + b - \Phi u$ (outer loop, if desired).

A surprise

- An early coding mistake led to the use of $p = -1/2$. This is what made the 9-line phantom reconstruction possible.

A surprise

- An early coding mistake led to the use of $p = -1/2$. This is what made the 9-line phantom reconstruction possible.
- Results decay slowly as p is decreased below $-1/2$.

A surprise

- An early coding mistake led to the use of $p = -1/2$. This is what made the 9-line phantom reconstruction possible.
- Results decay slowly as p is decreased below $-1/2$.
- Negative p-values were used by Rao and Kreutz-Delgado in an iteratively-reweighted least squares (IRLS) algorithm. Their negative p results were worse than their positive p results, which were hampered by getting stuck in local minima.

A surprise

- An early coding mistake led to the use of $p = -1/2$. This is what made the 9-line phantom reconstruction possible.
- Results decay slowly as p is decreased below $-1/2$.
- Negative p-values were used by Rao and Kreutz-Delgado in an iteratively-reweighted least squares (IRLS) algorithm. Their negative p results were worse than their positive p results, which were hampered by getting stuck in local minima.
- A mollified IRLS approach (with Wotao Yin) apparently avoids local minima, but negative p results are not better than positive p results.

Summary

- Nonconvex compressive sensing allows sparse signals to be recovered with even fewer measurements than "traditional" compressive sensing.
- Decreasing p also improves robustness to noise, and speeds up convergence.
- Regularizing the objective appears to keep algorithms from converging to nonglobal minima.
- For Fourier-sampling measurements, such as MRI, a very fast algorithm is available.

`math.lanl.gov/~rick`