Convex Demixing

Sharp Thresholds for Recovering Superimposed Signals

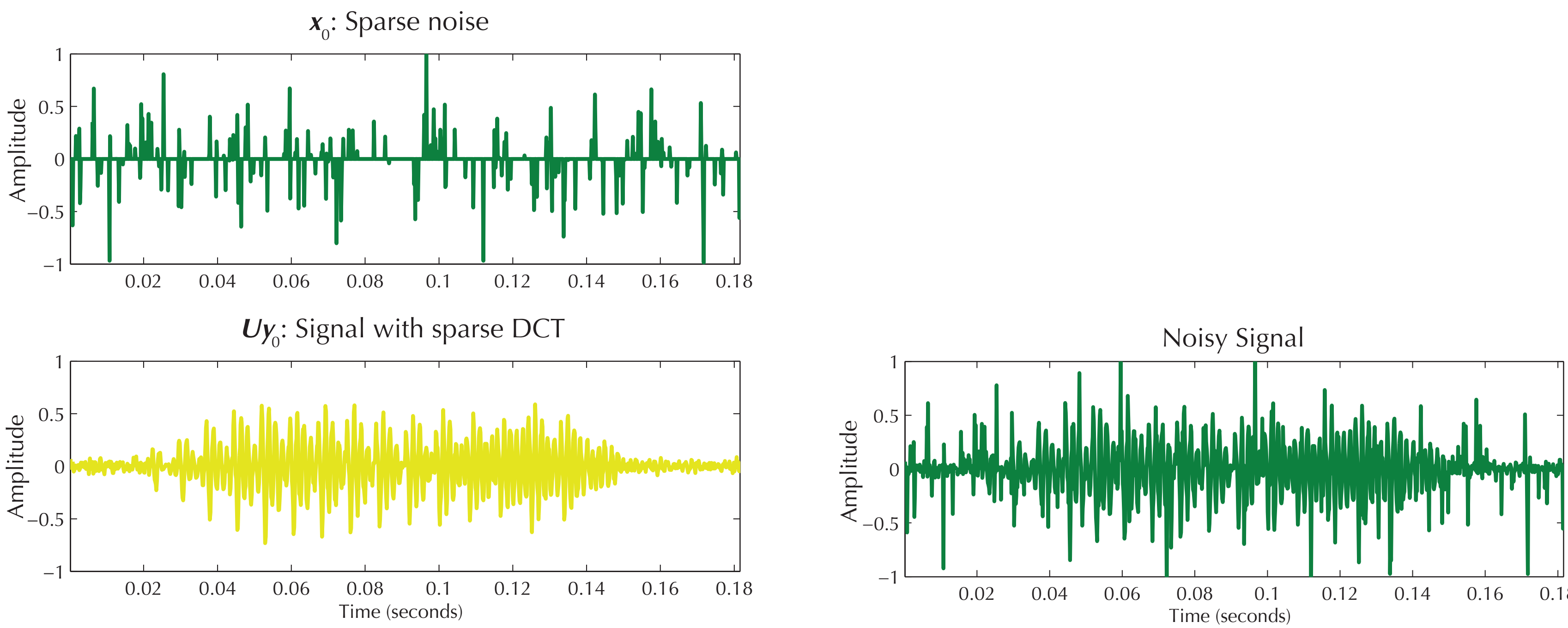
Joel A. Tropp
Michael B. McCoy

jtropp@cms.caltech.edu
mccoy@cms.caltech.edu

Computing + Mathematical Sciences
California Institute of Technology

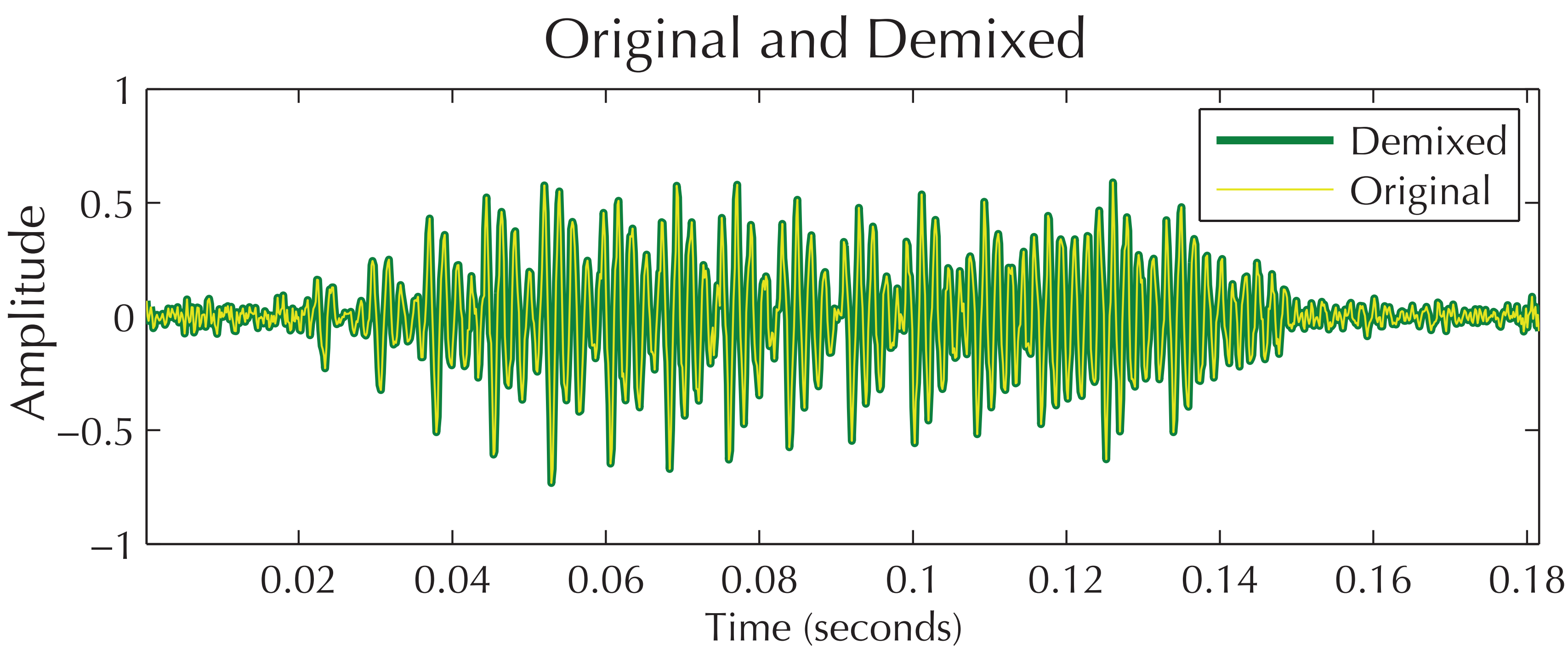
Research supported in part by ONR, AFOSR, DARPA, and the Sloan Foundation

Harmonic Signal + Impulsive Noise



Observation: $\boldsymbol{z}_0 = \boldsymbol{x}_0 + \boldsymbol{U}\boldsymbol{y}_0$ where $\boldsymbol{U}$ is the DCT

Convex Demixing Yields...



$$
\begin{array}{ll}
\text{minimize} & \|\boldsymbol{x}\|_{\ell_1} + \lambda \|\boldsymbol{y}\|_{\ell_1} \\
\text{subject to} & \boldsymbol{z}_0 = \boldsymbol{x} + \boldsymbol{U}\boldsymbol{y}
\end{array}
$$

Morphological Component Analysis

- Observe $\boldsymbol{z}_0 = \boldsymbol{x}_0 + \boldsymbol{U}\boldsymbol{y}_0$
- $\boldsymbol{U}$ is a known orthobasis; $\boldsymbol{x}_0$ and $\boldsymbol{y}_0$ are unknown vectors
- To identify this model, we can assume
  - **[Structure]** The vectors $\boldsymbol{x}_0$ and $\boldsymbol{y}_0$ are sparse
  - **[Incoherence]** Columns of $\boldsymbol{U}$ are weakly correlated with std basis
- Perform demixing using convex optimization:

$$\begin{array}{ll} \text{minimize} & \|\boldsymbol{x}\|_{\ell_1} + \lambda \|\boldsymbol{y}\|_{\ell_1} \\ \text{subject to} & \boldsymbol{z}_0 = \boldsymbol{x} + \boldsymbol{U}\boldsymbol{y} \end{array}$$

- **Application:** Astronomical image processing

[Refs] Starck et al. (2003), Starck et al. (2005)

Rank–Sparsity Decomposition

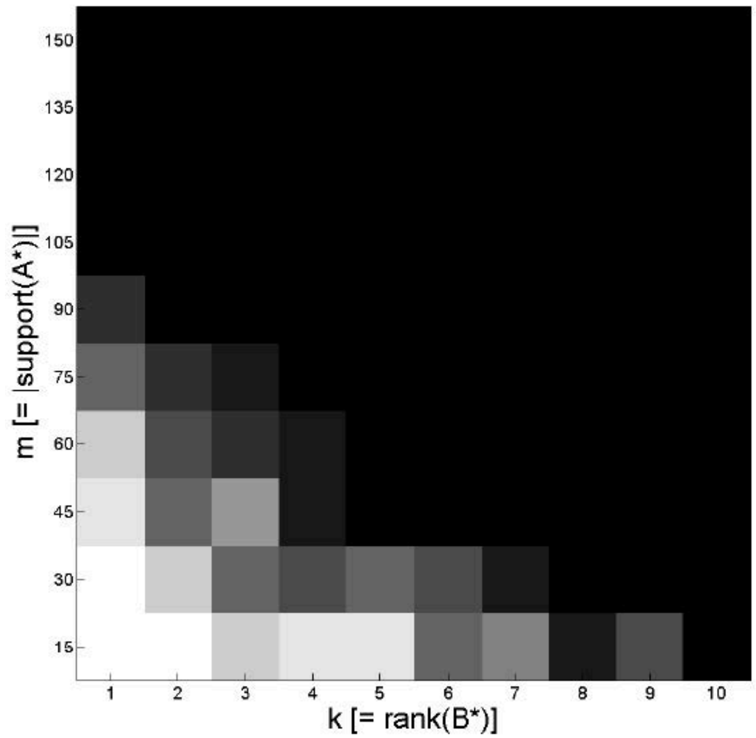
- Observe matrix $\boldsymbol{Z}_0 = \boldsymbol{X}_0 + \boldsymbol{Y}_0$
- To make this model identifiable, we can assume
  - **[Structure]** Matrix $\boldsymbol{X}_0$ is low rank and $\boldsymbol{Y}_0$ is sparse in std basis
  - **[Incoherence]** Singular vectors of $\boldsymbol{X}_0$ uncorrelated with std basis
- Perform demixing using convex optimization:

$$\begin{array}{rl} \text{minimize} & \|\boldsymbol{X}\|_{S_1} + \lambda \|\boldsymbol{Y}\|_{\ell_1} \\ \text{subject to} & \boldsymbol{Z}_0 = \boldsymbol{X} + \boldsymbol{Y} \end{array}$$

- **Application:** Identifying latent variables in graphical models

[Refs] Chandrasekaran et al. (2009), Candès et al. (2010)

A Phase Transition for Rank–Sparsity



$$\boldsymbol{Z}_0 = \boldsymbol{X}_0 + \boldsymbol{Y}_0 \quad \text{with} \quad k = \operatorname{rank}(\boldsymbol{X}_0) \quad \text{and} \quad m = \operatorname{nnz}(\boldsymbol{Y}_0)$$

[Source] Chandrasekaran et al. (2009)

Examples of Structural Penalties

Structure	Structural Penalty
Sparse vector	ℓ_1 norm
Sign vector	ℓ_∞ norm
Low-rank matrix	S_1 norm
Orthogonal mtx	S_∞ norm

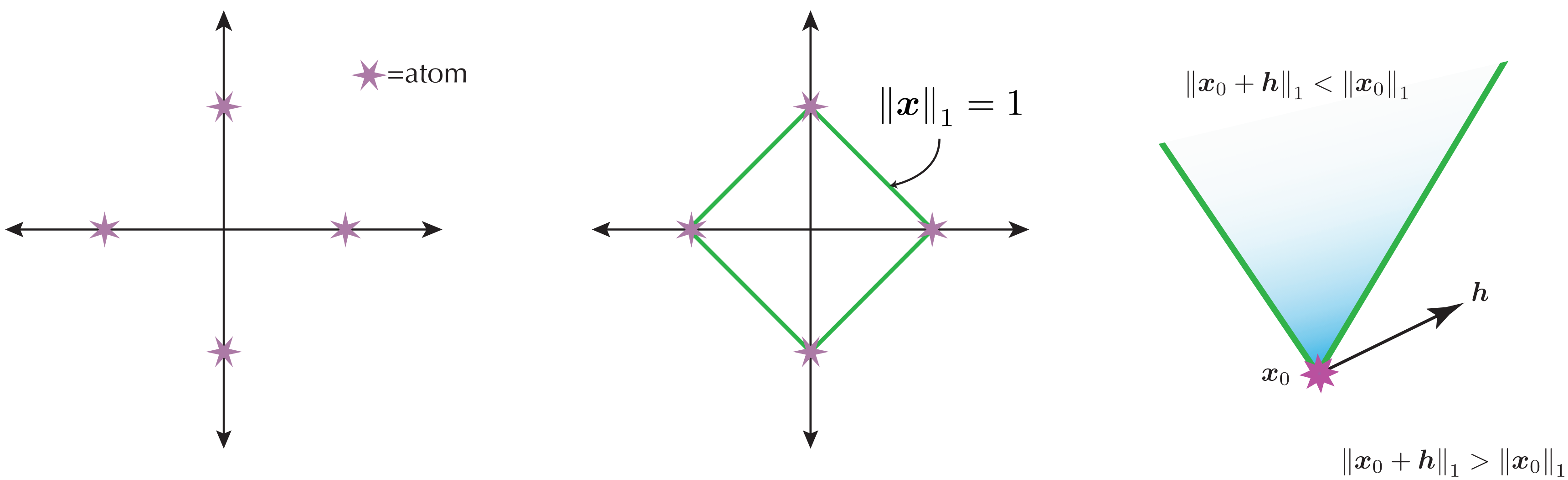
❧ Many, many others!

[Refs] DeVore & Temlyakov (1996), Temlyakov (2003), Chandrasekaran et al. (2010)

Small ℓ_1 Norm Reflects Sparsity

- A sparse vector $\boldsymbol{x}_0$ is a superposition of few standard basis vectors
- Consider the set $\mathscr{A} = \{\pm \mathbf{e}_k : k = 1, \dots, d\}$
- $\operatorname{conv} \mathscr{A}$ = smallest convex set containing $\mathscr{A}$
- The ℓ_1 norm has unit ball $\operatorname{conv} \mathscr{A}$, so

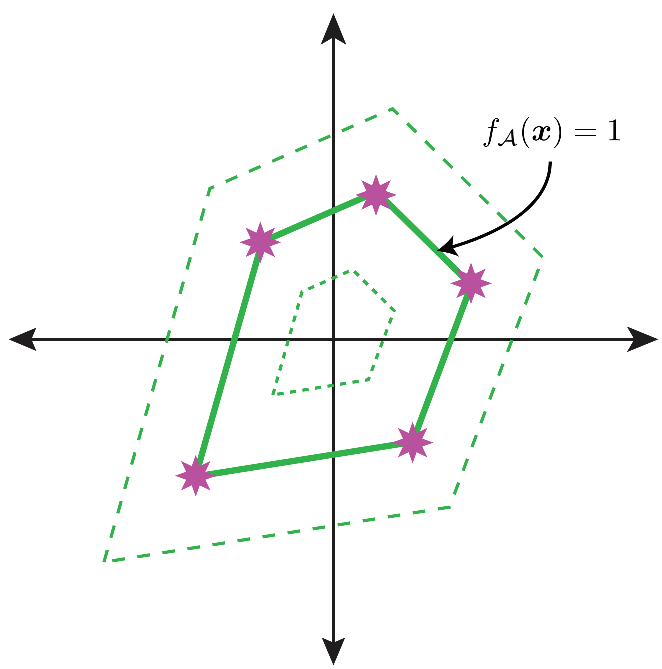
$$\|\boldsymbol{x}_0 + \boldsymbol{h}\|_1 > \|\boldsymbol{x}_0\|_1 \quad \text{for as many } \boldsymbol{h} \text{ as possible}$$

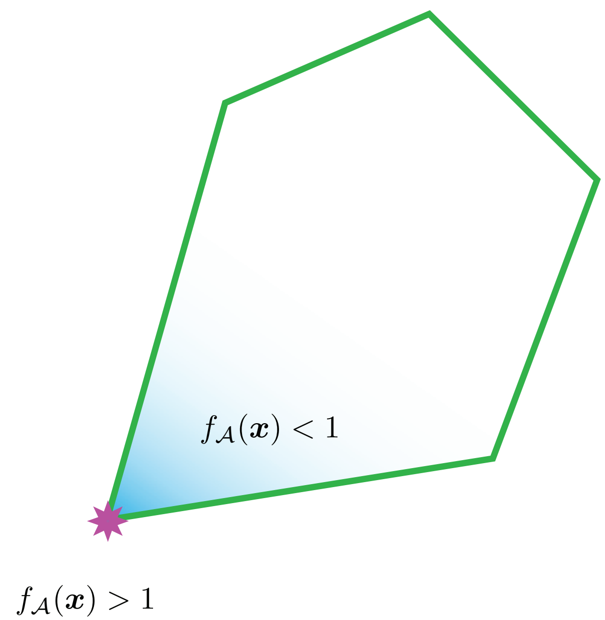


More General Structures

- Want to construct a convex function that reflects "structure"
- "Structure" = "a superposition of few atoms from a known set"
- Let $\mathscr{A}$ be a set of atoms. Define the *convex structural penalty*

$$f_{\mathscr{A}}(\boldsymbol{x}) := \inf\{t > 0 : \boldsymbol{x} \in t \cdot \mathrm{conv}(\mathscr{A})\}$$





Examples of Structural Penalties, Redux

Structure	Dictionary	Penalty
Sparse vector	$\mathscr{A} = \{\pm \mathbf{e}_i : i = 1, \dots, d\}$	ℓ_1 norm
Sign vector	$\mathscr{A} = \{\pm 1\}^d$	ℓ_∞ norm
Low-rank matrix	$\mathscr{A} = \{$unit norm, rank-one mtx$\}$	S_1 norm
Orthogonal mtx	$\mathscr{A} = \{$orthogonal mtx$\}$	S_∞ norm

- Every set of atoms gives a convex structural penalty!
- Not all convex penalties lead to tractable optimization problems...

[Refs] DeVore & Temlyakov (1996), Temlyakov (2003), Chandrasekaran et al. (2010)

An Abstract Demixing Problem

- Let $\boldsymbol{x}_0$ and $\boldsymbol{y}_0$ be structured vectors with convex penalties f and g
- Let $\boldsymbol{U}$ be a known orthogonal matrix
- Observe $\boldsymbol{z}_0 = \boldsymbol{x}_0 + \boldsymbol{U}\boldsymbol{y}_0$
- We pose the *convex demixing method*

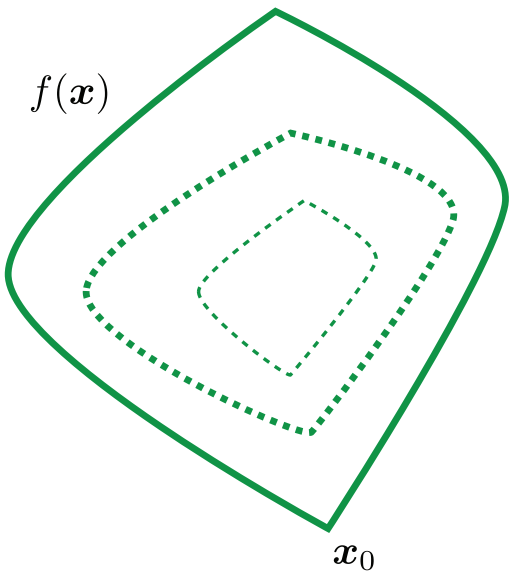
$$\begin{aligned} \text{minimize} \quad & f(\boldsymbol{x}) \\ \text{subject to} \quad & g(\boldsymbol{y}) \leq \alpha \quad \text{and} \quad \boldsymbol{z}_0 = \boldsymbol{x} + \boldsymbol{U}\boldsymbol{y} \end{aligned}$$

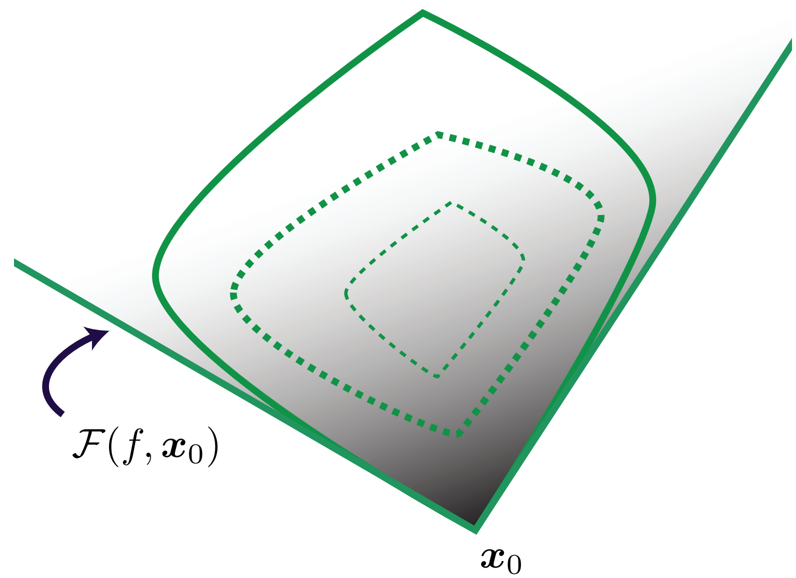
  where $\alpha = g(\boldsymbol{y}_0)$ is side information

- **Hope:** The pair $(\boldsymbol{x}_0, \boldsymbol{y}_0)$ is the unique solution

Geometry of Exact Recovery I

- The sublevel sets of a nonsmooth convex function are locally conic
- The *feasible cone* $\mathscr{F}(f, \boldsymbol{x}_0)$ is the convex cone generated by directions where f is locally nondecreasing at $\boldsymbol{x}_0$
- "Pointy" sublevel set = small feasible cone

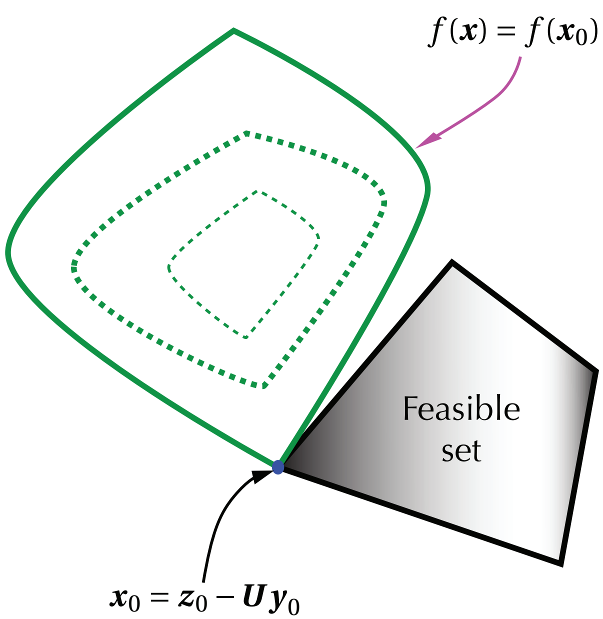


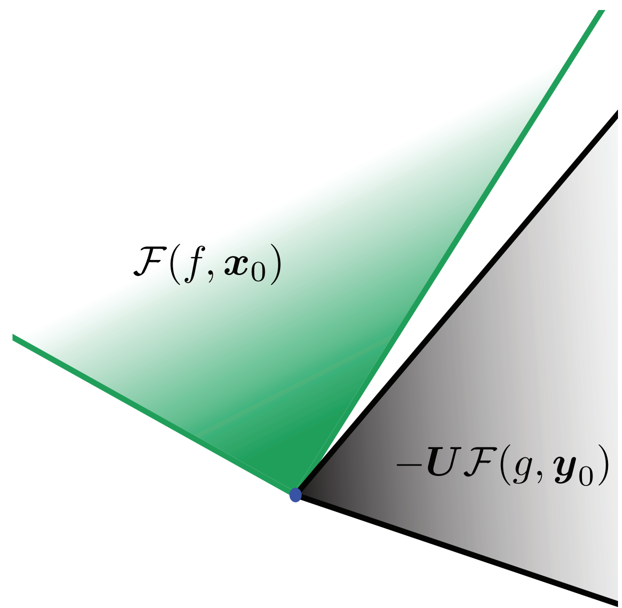


Geometry of Exact Recovery II

- The convex demixing method succeeds if and only if two feasible cones intersect trivially:

$$\mathscr{F}(f, \boldsymbol{x}_0) \cap (-\boldsymbol{U}\mathscr{F}(g, \boldsymbol{y}_0)) = \{\boldsymbol{0}\}$$





A Randomized Model for Incoherence

- We want to study the case where signal structures are oblique
- **Idea:** Use a random model for incoherence
- Let $\boldsymbol{Q}$ be a uniformly random orthogonal matrix, and we observe

$$\boldsymbol{z}_0 = \boldsymbol{x}_0 + \boldsymbol{Q}\boldsymbol{y}_0$$

- Convex demixing succeeds if and only if

$$\mathscr{F}(f, \boldsymbol{x}_0) \cap (-\boldsymbol{Q}\mathscr{F}(g, \boldsymbol{y}_0)) = \{\mathbf{0}\}$$

[Refs] Donoho & Stark (1989), Donoho & Huo (2001)

The Spherical Kinematic Formula

- Need to study when two randomly oriented cones strike
- There is an *exact* expression for this quantity!

Spherical Kinematic Formula

Let K and $\tilde{K}$ be closed convex cones, one of which is not a subspace

$$\mathbb{P}\left\{K \cap \boldsymbol{Q}\tilde{K} \neq \{\mathbf{0}\}\right\} = \sum_{j=0}^{d} (1 + (-1)^{j+1}) \sum_{i=j}^{d} v_i(K) \cdot v_{d-i+j}(\tilde{K})$$

where v_i is the ith *spherical intrinsic volume* for $i = 0, 1, \dots, d$

[Refs] Allendorfer & Weil (1943), Glasauer (1996)

Spherical Intrinsic Volumes

- The spherical intrinsic volumes v_i are measures of content for cones
- Let K be a polyhedral cone. Define

$$v_i(K) := \mathbb{P}\left\{ \mathbf{\Pi}_K(\boldsymbol{\omega}) \text{ lies inside an } i\text{-dimensional face of } K \right\}$$

  - $\mathbf{\Pi}_K$ is the Euclidean projection of a point onto the cone K
  - $\boldsymbol{\omega}$ is a standard Gaussian vector
- In non-polyhedral case, define v_i by approximating with polyhedral cones

[Refs] Glasauer (1996), Schneider & Weil (2008), Amelunxen (2011)

Example: The Nonnegative Orthant

$$\mathbb{R}^d_+ := \{\boldsymbol{x} \in \mathbb{R}^d : x_i \geq 0\}$$

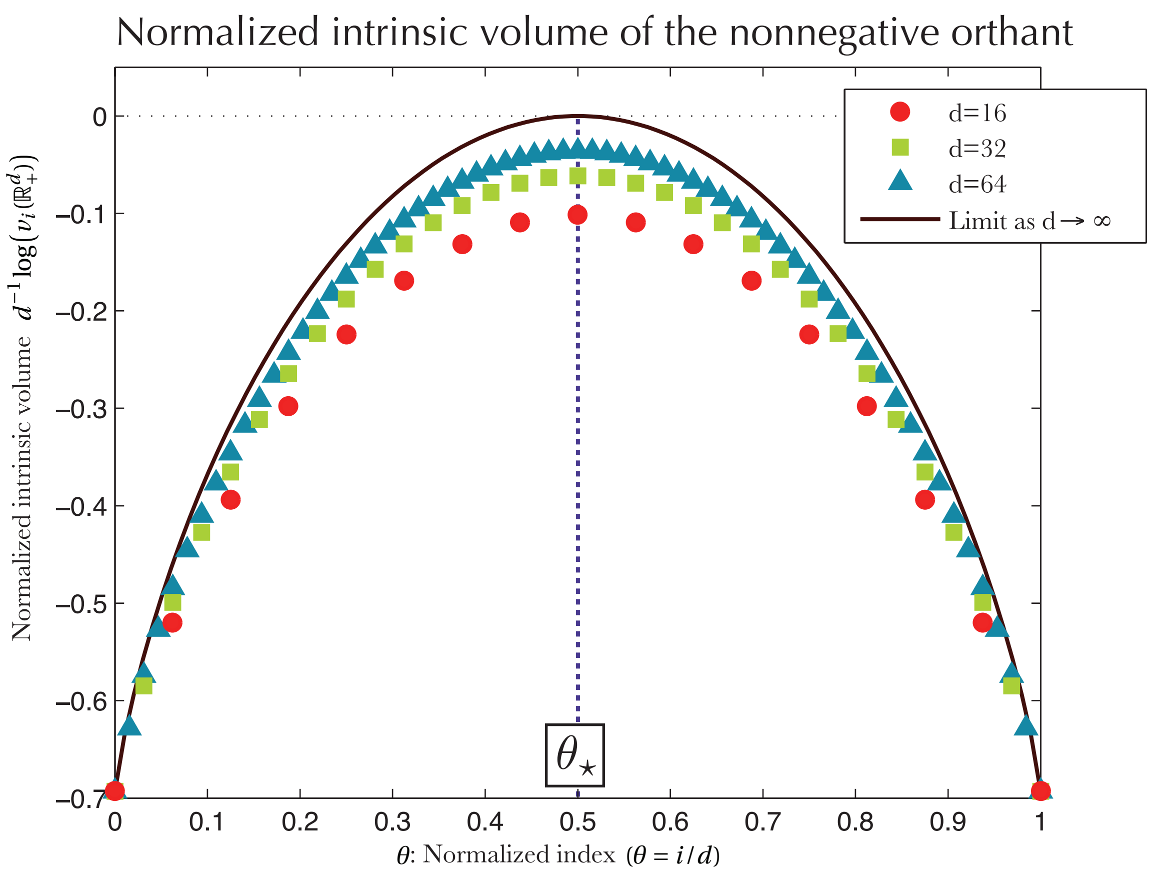
- Up to rotation, the orthant is the feasible cone of ℓ_∞ at a sign vector
- The projection of $\boldsymbol{x}$ onto the orthant satisfies $(\boldsymbol{\Pi}_+(\boldsymbol{x}))_i = \max\{x_i, 0\}$
- The i-dimensional faces contain vectors with exactly i positive entries

$$\begin{aligned} v_i(\mathbb{R}^d_+) &= \mathbb{P}\left\{\boldsymbol{\Pi}_+(\boldsymbol{\omega}) \text{ lies inside an } i\text{-dim. face of } \mathbb{R}^d_+\right\} \\ &= \mathbb{P}\left\{\boldsymbol{\omega} \text{ has exactly } i \text{ positive entries}\right\} = 2^{-d}\binom{d}{i} \end{aligned}$$

- For the normalized index $\theta = i/d$, the normalized intrinsic volume

$$d^{-1} \log v_i(\mathbb{R}^d_+) \to H(\theta) - \log(2) \quad \text{where} \quad H = \text{bit entropy}$$

Spherical Intrinsic Volumes are Peaked



- The value $\theta_\star$ is the *structural complexity* of the cone K
- For any $\theta > \theta_\star$, the intrinsic volume $v_{\theta d}(K) \leq \text{const} \cdot \mathrm{e}^{-c(\theta)\cdot d}$

Examples of Structural Complexity

Structure	Penalty	Parameter	Structural Complexity
Subspace	Indicator	$\sigma = \frac{\text{subspace dim}}{\text{ambient dim}}$	$\theta_\star = \sigma$
Sparse vector	ℓ_1 norm	$\tau = \frac{\text{sparsity}}{\text{dimension}}$	$\theta_\star = \varphi(\tau)$
Sign vector	ℓ_∞ norm	—	$\theta_\star = \frac{1}{2}$
Square, low-rank	S_1 norm	$\rho = \frac{\text{rank}}{\text{side length}}$	$\theta_\star \leq 6\rho - 5\rho^2$
Orthogonal mtx	S_∞ norm	—	$\theta_\star \leq \frac{3}{4}$

Structural complexity = "dimension" of cone / ambient dimension!

Demixing of Structured Incoherent Signals

Theorem 1. **Suppose that**

- $\boldsymbol{x}_0$ *and* $\boldsymbol{y}_0$ *are structured signals with convex penalties* f *and* g*;*
- $\boldsymbol{Q}$ *is a uniformly random orthobasis;*
- *We observe* $\boldsymbol{z}_0 = \boldsymbol{x}_0 + \boldsymbol{Q}\boldsymbol{y}_0$ *and side information* $\alpha = g(\boldsymbol{y}_0)$*;*
- *The structural complexities satisfy*

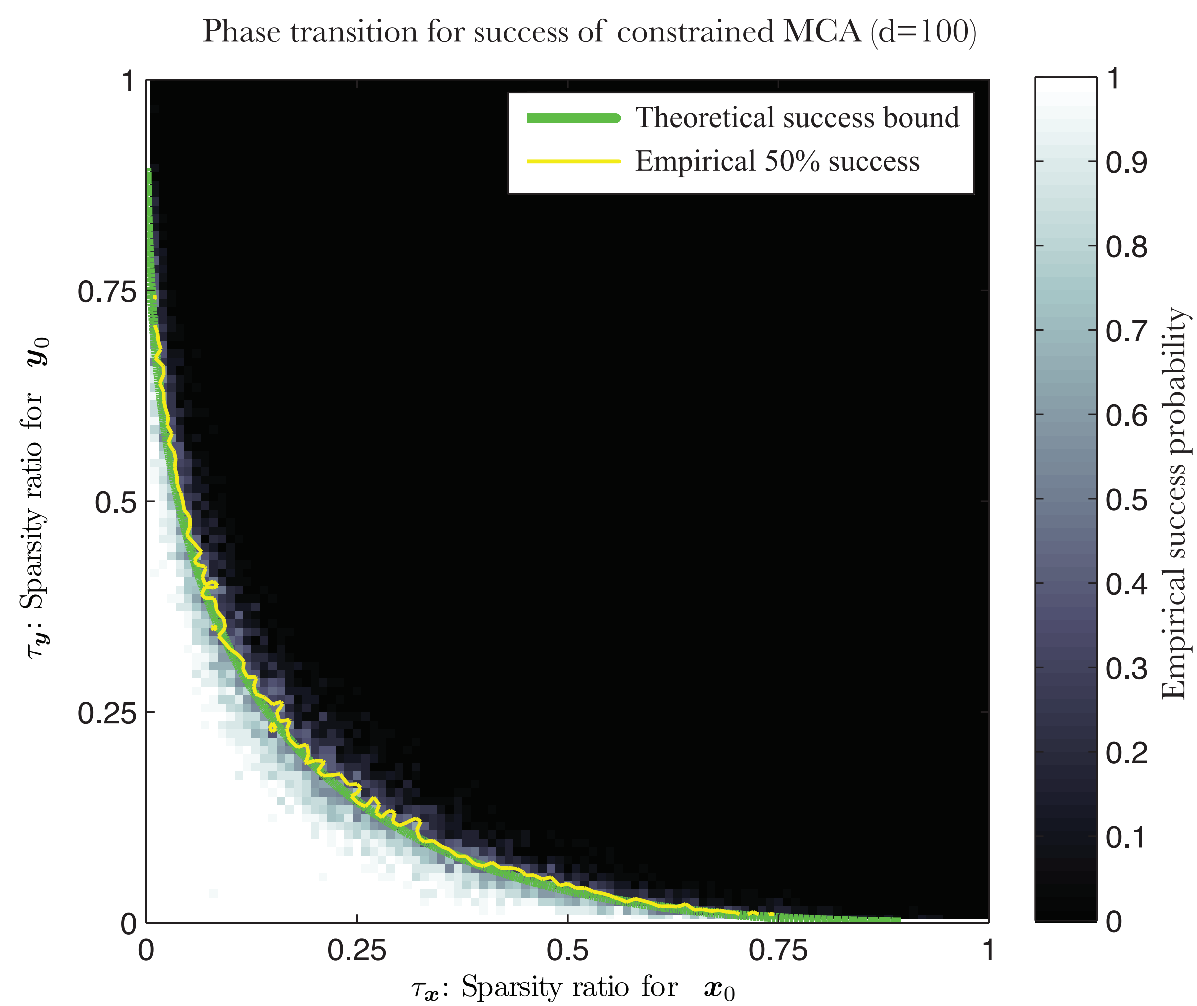
$$\theta_\star(f, \boldsymbol{x}_0) + \theta_\star(g, \boldsymbol{y}_0) < 1.$$

Then, *w.h.p. over* $\boldsymbol{Q}$, *the pair* $(\boldsymbol{x}_0, \boldsymbol{y}_0)$ *is the unique solution to*

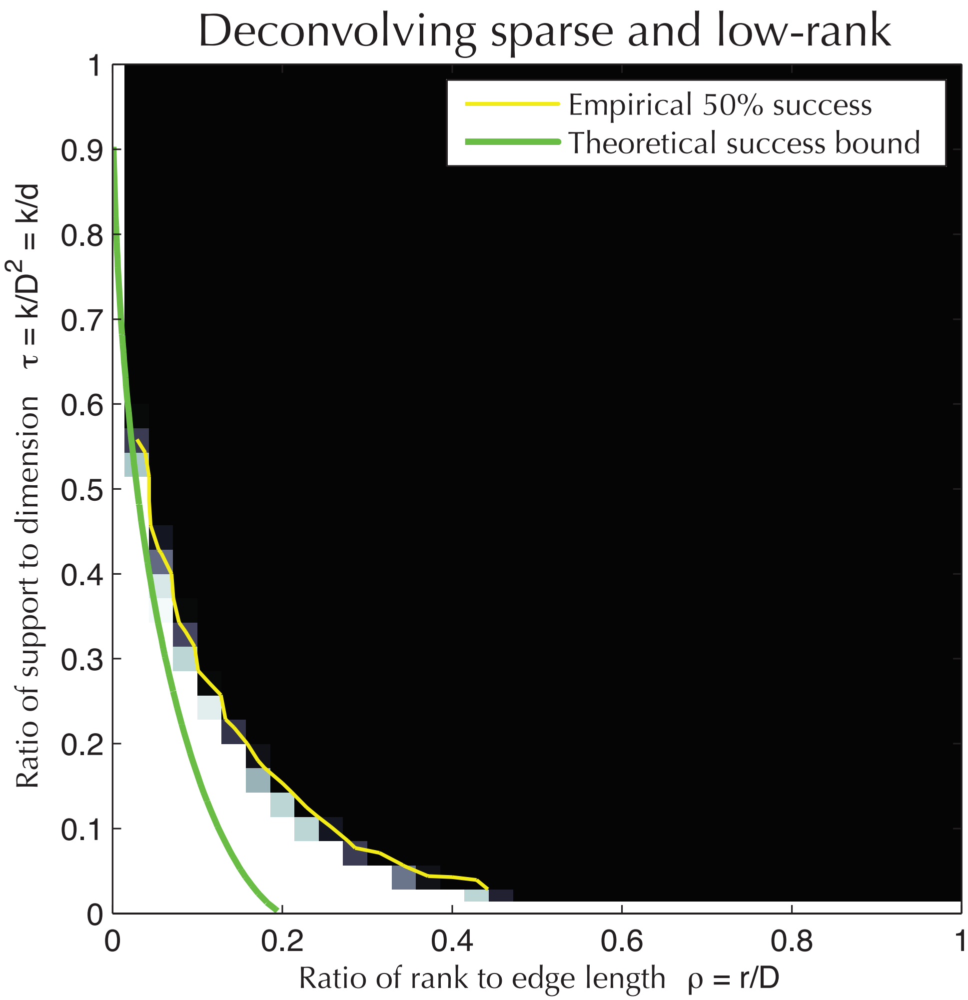
$$\begin{aligned} &\text{minimize} \quad f(\boldsymbol{x}) \\ &\text{subject to} \quad g(\boldsymbol{y}) \leq \alpha \quad \text{and} \quad \boldsymbol{z}_0 = \boldsymbol{x} + \boldsymbol{Q}\boldsymbol{y}. \end{aligned}$$

Conversely, *if* $\theta_\star(f, \boldsymbol{x}_0) + \theta_\star(g, \boldsymbol{y}_0) > 1$, *the optimization fails w.h.p.*

Example: Sparse + Sparse in Random Basis



Example: Low Rank + Sparse in Random Basis



Example: Spread-Spectrum Communications

- Want to transmit a binary message $\boldsymbol{m}_0 \in \{\pm 1\}^d$
- Modulate with a random matrix $\boldsymbol{Q}$, known at transmitter and receiver
- Receiver observes

$$\boldsymbol{z}_0 = \boldsymbol{Q}\boldsymbol{m}_0 + \boldsymbol{c}_0$$

  where $\boldsymbol{c}_0$ is a sparse corruption

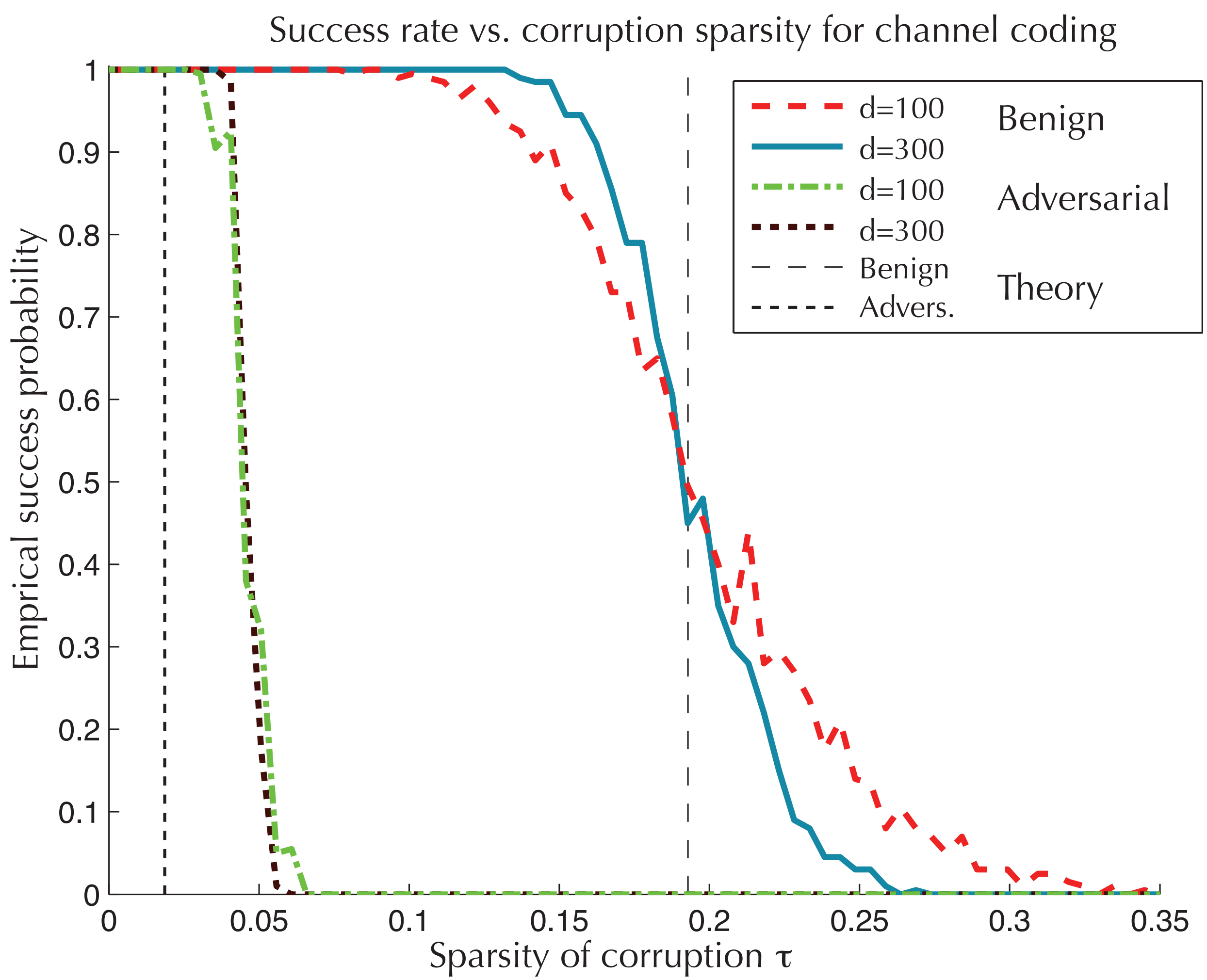
- Decode using convex demixing method

$$\begin{array}{ll} \text{minimize} & \|\boldsymbol{c}\|_{\ell_1} \\ \text{subject to} & \|\boldsymbol{m}\|_{\ell_\infty} \leq 1 \quad \text{and} \quad \boldsymbol{z}_0 = \boldsymbol{Q}\boldsymbol{m} + \boldsymbol{c} \end{array}$$

- When does the receiver correctly identify the true message?

[Refs] Wyner (1979), Donoho & Huo (2001)

Decoding Performance



To learn more...

E-mail: mccoy@cms.caltech.edu
jtropp@cms.caltech.edu

Web: http://users.cms.caltech.edu/~mccoy
http://users.cms.caltech.edu/~jtropp

Papers:

- MT, "Sharp recovery bounds for convex deconvolution, with applications." arXiv cs.IT 1205.1580
- More to come...