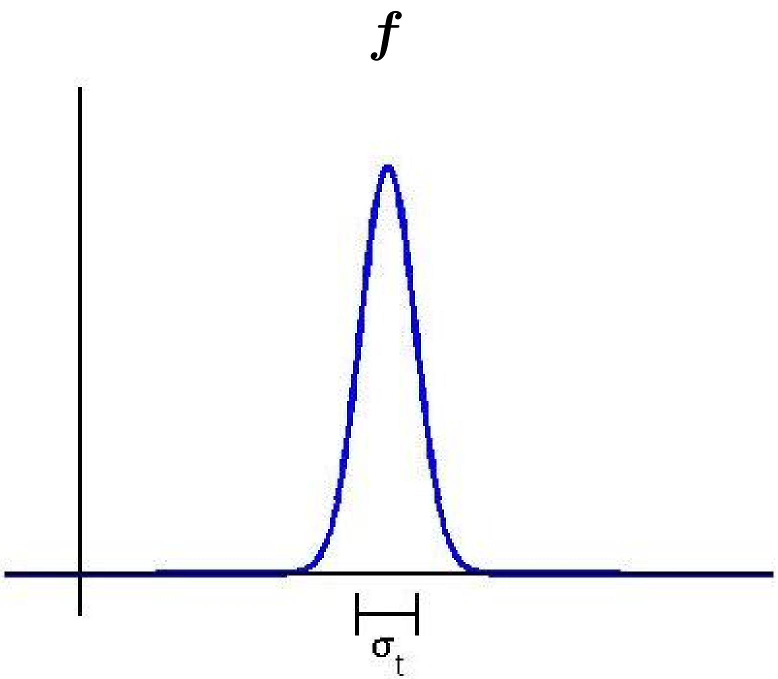
Robust Uncertainty Principles and Optimally Sparse Decompositions

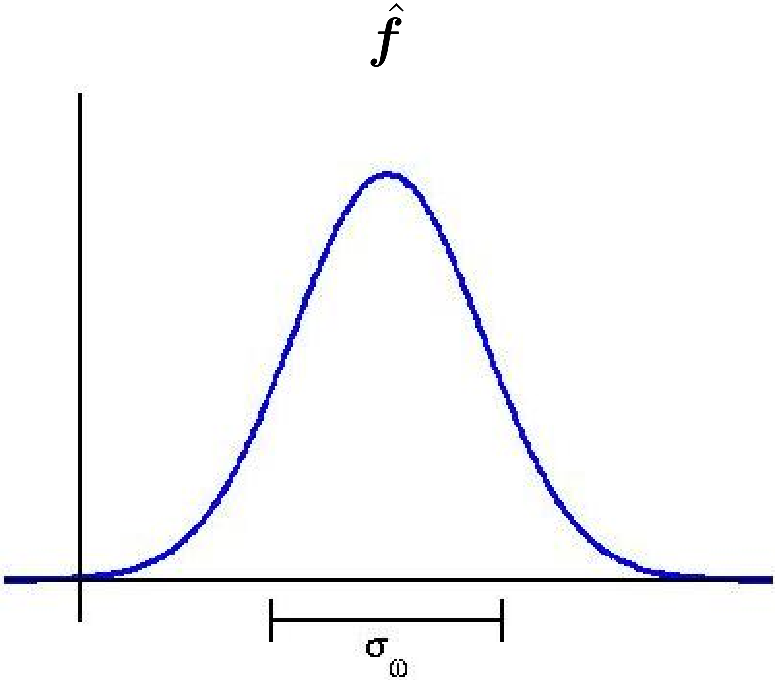
Justin Romberg, California Institute of Technology

Collaborators: *Emmanuel Candès* (Caltech), *Terence Tao* (UCLA)

Uncertainty Principles

- Heisenberg (1927)
  Uncertainty principle for continuous-time signals





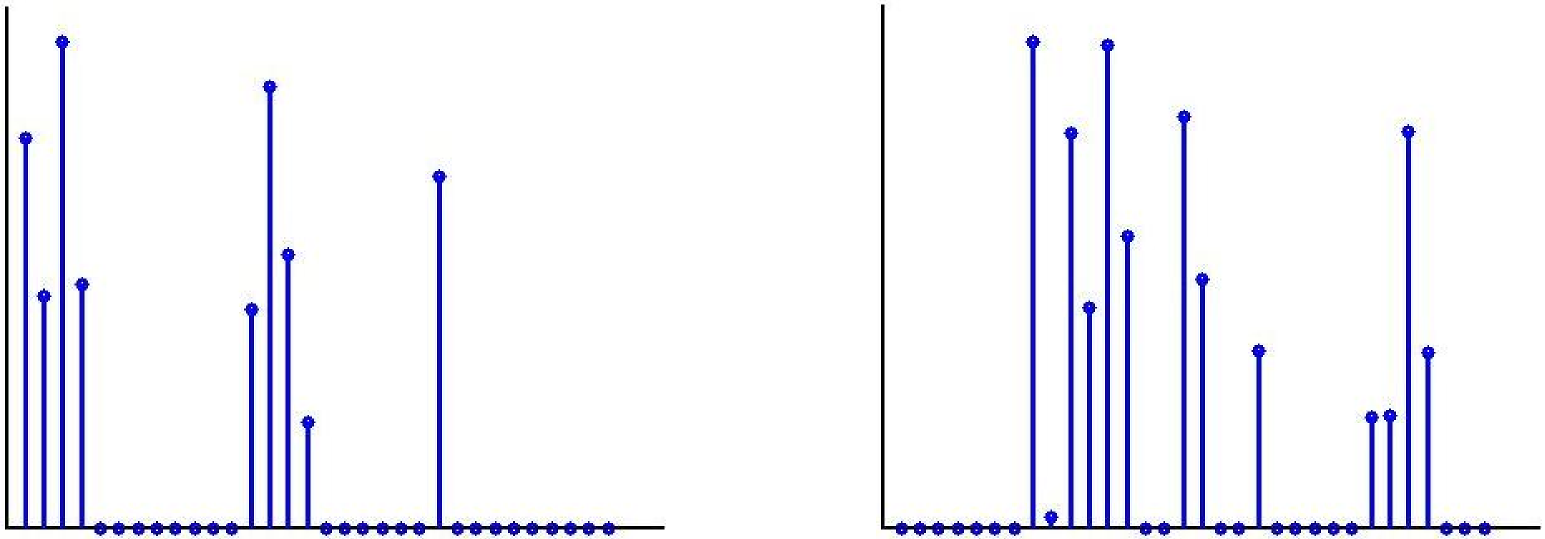
$$\boxed{\sigma_t \sigma_\omega \geq \frac{1}{4\pi}}$$

- Limits *joint resolution* in time and frequency

Uncertainty Principles

- Donoho and Stark (1989)
  Discrete uncertainty principle for $\mathbb{C}^{\boldsymbol{N}}$

$\boldsymbol{f}$ $\qquad\qquad$ $\hat{\boldsymbol{f}}$



$\operatorname{supp} \boldsymbol{f} = \boldsymbol{T}$ $\qquad\qquad$ $\operatorname{supp} \hat{\boldsymbol{f}} = \Omega$

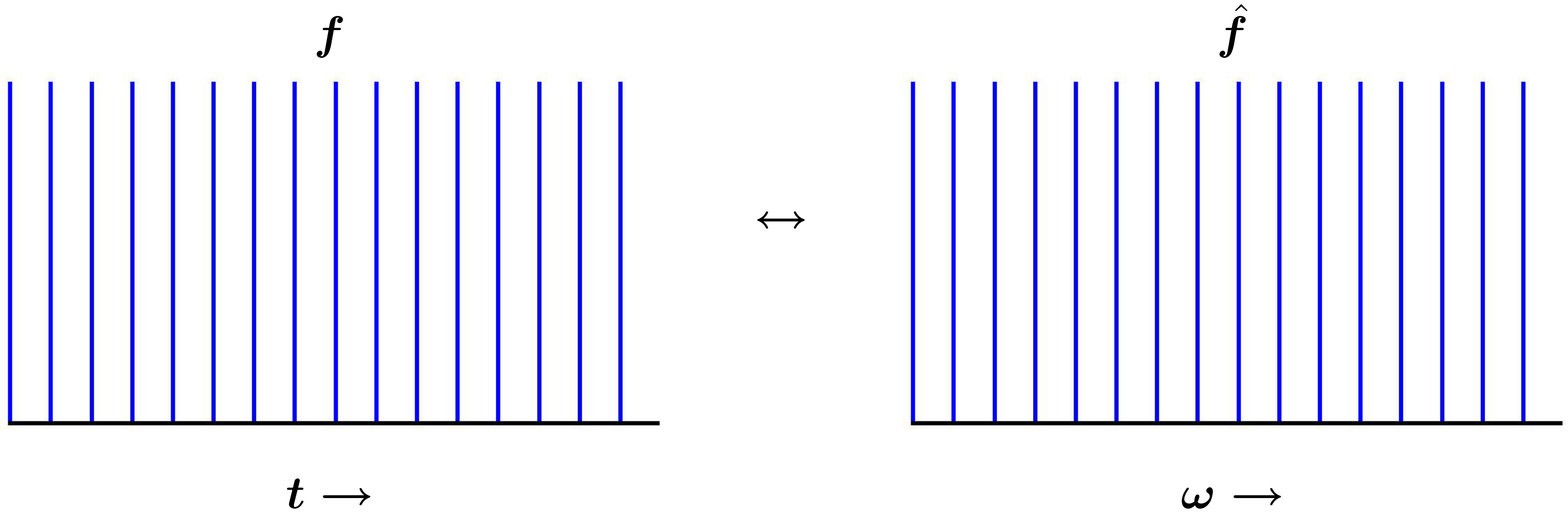
$$\boxed{|\boldsymbol{T}| + |\Omega| \geq 2\sqrt{\boldsymbol{N}}}$$

- Implications: recovery from partial information, unique sparse decompositions

- Generalization to pairs of bases $\boldsymbol{B}_1, \boldsymbol{B}_2$
  [Donoho,Huo,Elad,Bruckstein,Gribonval,Nielsen]

Dirac Comb

The discrete uncertainty principle is *exact*.



- $\sqrt{N}$ spikes spaced $\sqrt{N}$ apart
- Invariant under Fourier transform ($f = \hat{f}$)
- $|T| + |\Omega| = 2\sqrt{N}$

Sparse Representations

- Decompose f as a superposition of spikes and sinusoids

$$f(s) = \sum_{t \in T} \alpha_t \delta(t-s) + \sum_{\omega \in \Omega} \alpha_\omega e^{i\omega s}$$

$$T = \text{locations of spikes}, \quad \Omega = \text{frequencies of sinusoids}$$

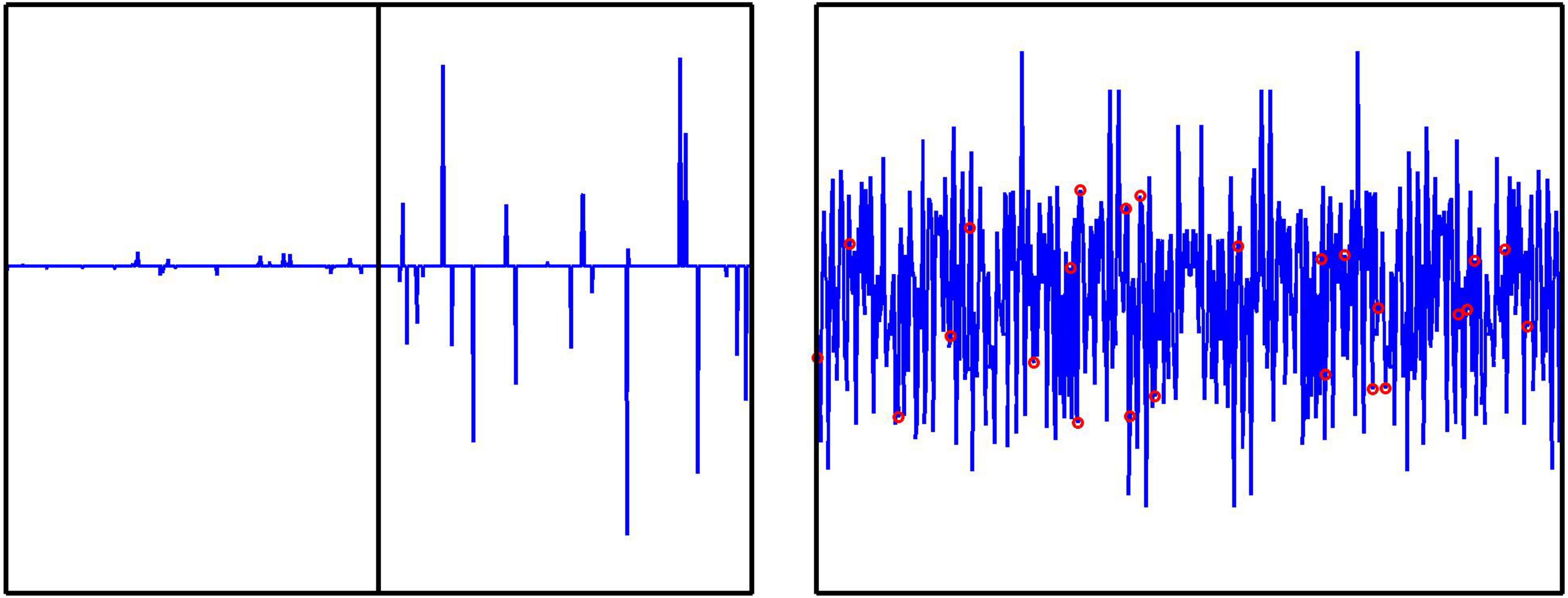
- Matrix form:

$$\Phi\alpha = f \quad \Leftrightarrow \quad \begin{pmatrix} I & F^* \end{pmatrix} \begin{pmatrix} \alpha_t \\ \alpha_\omega \end{pmatrix} = f$$

- Many solutions exist, we want the *sparsest*.

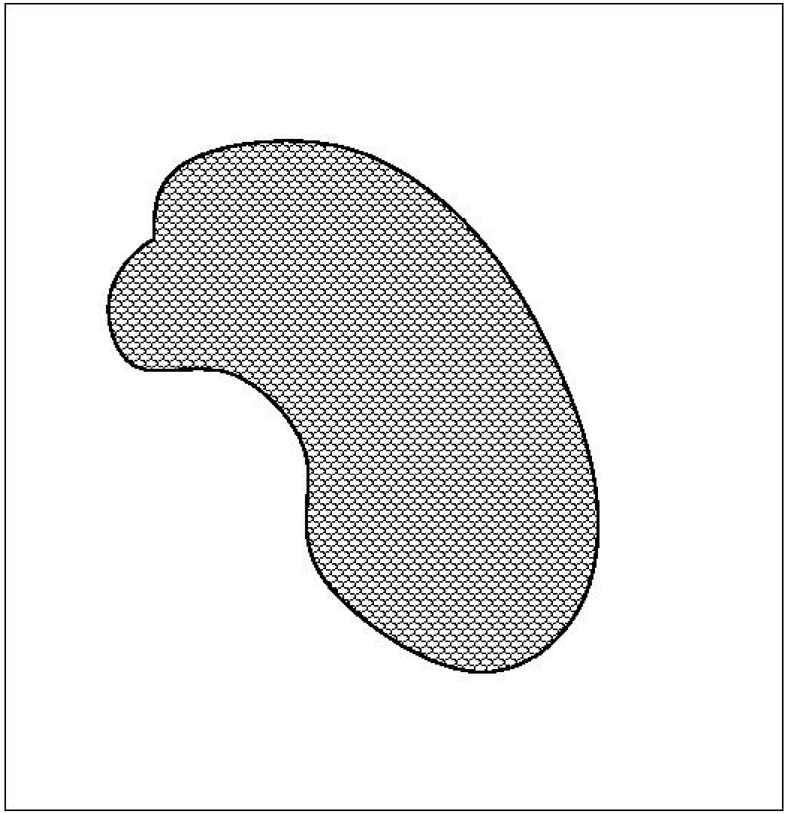
Example

coeffs $\boldsymbol{\alpha}$ signal $\boldsymbol{f} = \Phi\boldsymbol{\alpha}$



spikes freqs $\boldsymbol{t} \longrightarrow$

Image Processing and Sparse Representations



- Geometrical structure and texture are separate phenomena
- Each is sparse in a certain “basis”
- See also Geometric Separation of Donoho et. al, edge/texture separation of Starck et. al.

ℓ_0-minimization

- Dictionary Φ: columns are spikes and sinusoids

$$\Phi \boldsymbol{\alpha} = \boldsymbol{f} \quad \Leftrightarrow \quad \begin{pmatrix} \boldsymbol{I} & \boldsymbol{F}^* \end{pmatrix} \begin{pmatrix} \boldsymbol{\alpha}_t \\ \boldsymbol{\alpha}_\omega \end{pmatrix} = \boldsymbol{f}$$

- Observed signal $\boldsymbol{f}$

- Sparsest representation: solve (combinatorial problem)

$$\begin{array}{c} \min \|\boldsymbol{\alpha}\|_{\ell_0} \\ \text{s.t. } \Phi\boldsymbol{\alpha} = \boldsymbol{f} \end{array} \qquad (\ell_0\text{-min})$$

- $\|\boldsymbol{\alpha}\|_{\ell_0} = |\operatorname{supp} \boldsymbol{\alpha}|$ = number of non-zero terms

- When is ℓ_0-min unique?

Uncertainty Principles and Sparsity

[Donoho,Huo (2001)]

Fact: Say $\boldsymbol{\alpha}$ is a sparse decomposition of $\boldsymbol{f}$ $\qquad \|\boldsymbol{\alpha}\|_{\ell_0} = |\operatorname{supp}\boldsymbol{\alpha}| < \sqrt{\boldsymbol{N}}$.

$$\boldsymbol{f} = \Phi\boldsymbol{\alpha}$$

then $\boldsymbol{\alpha}$ is the *only* sparse decomposition.

Reason: Uncertainty Principle

$$|\operatorname{supp}\boldsymbol{f}| + |\operatorname{supp}\hat{\boldsymbol{f}}| \geq 2\sqrt{\boldsymbol{N}}$$

Say $\Phi\boldsymbol{\alpha}' = \boldsymbol{f}$ as well: $\boldsymbol{\alpha}' = \boldsymbol{\alpha} + \boldsymbol{\gamma}$ with $\Phi\boldsymbol{\gamma} = 0$

$$\begin{pmatrix} \boldsymbol{I} & \boldsymbol{F}^* \end{pmatrix} \begin{pmatrix} \gamma_1 \\ \gamma_2 \end{pmatrix} \rightarrow \quad \gamma_1 + \boldsymbol{F}^*\gamma_2 = 0 \Rightarrow \quad \boldsymbol{\gamma} = \begin{pmatrix} \gamma_1 \\ -\hat{\gamma}_1 \end{pmatrix} \Rightarrow \quad \|\boldsymbol{\gamma}\|_{\ell_0} \geq 2\sqrt{\boldsymbol{N}}$$

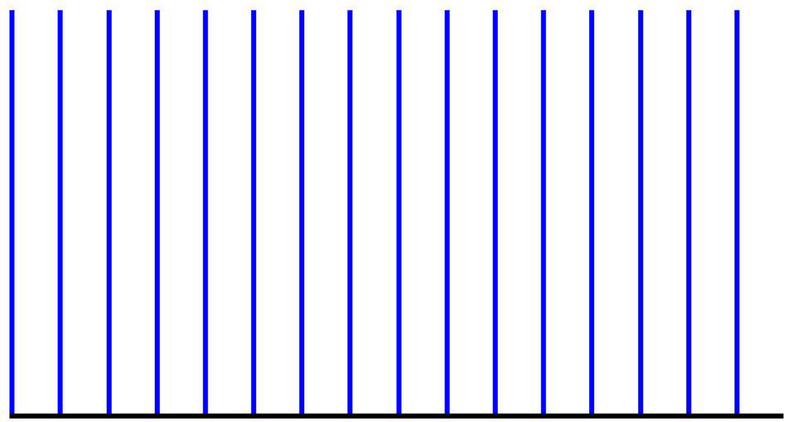
$$\Rightarrow \|\boldsymbol{\alpha}'\|_{\ell_0} = \|\boldsymbol{\alpha} + \boldsymbol{\gamma}\|_{\ell_0} > \sqrt{\boldsymbol{N}}$$

Consequence: $\boldsymbol{\alpha}$ is the *unique* ℓ_0 minimizer.

Dirac Comb

The “Dirac comb” has two representations of size $\sqrt{N}$.

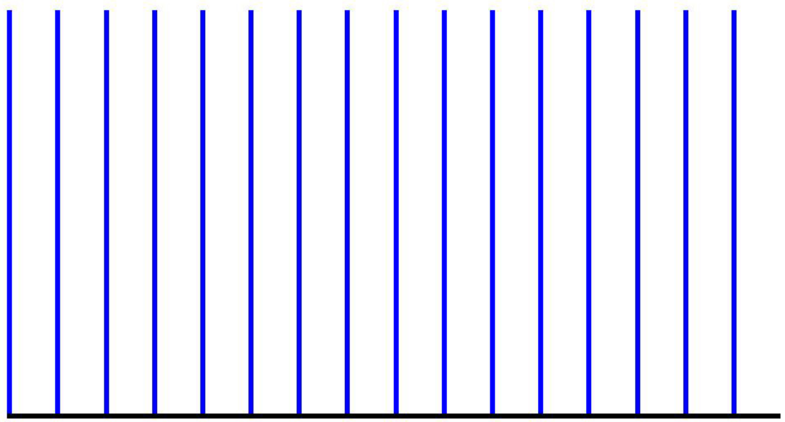
f



$t \longrightarrow$

$\longleftrightarrow$

$\hat{f}$



$\omega \longrightarrow$

sum of $\sqrt{N}$ spikes —or— sum of $\sqrt{N}$ sinusoids

Sharp Uncertainty Principle: N Prime

[Tao (2004)]

- For N prime, evenly spaced signals like the Dirac comb are impossible.
- The uncertainty principle is *significantly* more relaxed:

$$|\operatorname{supp} \boldsymbol{f}| + |\operatorname{supp} \hat{\boldsymbol{f}}| > \boldsymbol{N}$$

- Key: minors of the Fourier matrix have full rank

$$\boldsymbol{A} = \boldsymbol{R}_\Omega \boldsymbol{F} \boldsymbol{R}_{\boldsymbol{T}}^*$$

  $\boldsymbol{R}_{\boldsymbol{T}}, \boldsymbol{R}_\Omega$ are restriction operators.

- Compare to general UP: $\boldsymbol{N}$ vs. $2\sqrt{\boldsymbol{N}}$

ℓ_0 uniqueness: N Prime

Theorem: Let N be prime. Say $\boldsymbol{\alpha}^\sharp$ is supported on $\boldsymbol{T} \cup \Omega$ with

$$\Phi\boldsymbol{\alpha}^\sharp = \boldsymbol{f} \qquad |\boldsymbol{T}| + |\Omega| = \|\boldsymbol{\alpha}^\sharp\|_{\ell_0} < \boldsymbol{N}/2.$$

Then for all $\boldsymbol{\alpha}'$, $\quad \Phi\boldsymbol{\alpha}' = \boldsymbol{f}$, $\quad \boldsymbol{\alpha}' \neq \boldsymbol{\alpha}$

$$\|\boldsymbol{\alpha}'\|_{\ell_0} > \boldsymbol{N}/2$$

$\Rightarrow$ $\boldsymbol{\alpha}^\sharp$ is the *unique* ℓ_0 minimizer.

- Follows directly from relaxed UP
- Compare to general ℓ_0 uniqueness result: $\quad \boldsymbol{N}/2 \quad$ vs. $\quad \sqrt{\boldsymbol{N}}$

"Typical" Signals

- **Robust UP:** For an overwhelming percentage of $\boldsymbol{T}, \Omega$ with

$$|\boldsymbol{T}| + |\Omega| \sim \frac{\boldsymbol{N}}{\sqrt{\log \boldsymbol{N}}}$$

  it is *impossible* to find an $\boldsymbol{f}$ with $\quad \operatorname{supp} \boldsymbol{f} = \boldsymbol{T}, \quad \operatorname{supp} \hat{\boldsymbol{f}} = \Omega$.

- **ℓ_0-uniqueness:** Given "generic" $\boldsymbol{\alpha}$ on $\boldsymbol{T} \cup \Omega$, solving

$$\min \|\boldsymbol{\alpha}'\|_{\ell_0} \qquad \text{s.t.} \qquad \Phi\boldsymbol{\alpha}' = \boldsymbol{f}$$

  will recover $\boldsymbol{\alpha}$ *exactly*.

- **Tractability:** If $\boldsymbol{T}, \Omega$ are slightly smaller

$$|\boldsymbol{T}| + |\Omega| \sim \frac{\boldsymbol{N}}{\log \boldsymbol{N}}$$

  then there is a *tractable* algorithm to recover $\boldsymbol{\alpha}$ from $\boldsymbol{f}$.

- Similar results hold for pairs of bases $\boldsymbol{B}_1, \boldsymbol{B}_2$.

Probabilistic Framework

1. Generate $\boldsymbol{T}$, Ω with Bernoulli sequences

$$\boldsymbol{\chi}(\boldsymbol{t}), \boldsymbol{\chi}(\boldsymbol{\omega}) = \begin{cases} 1 & \text{w/ prob } \boldsymbol{\tau} \\ 0 & \text{w/ prob } 1 - \boldsymbol{\tau} \end{cases}$$

$$\begin{aligned} \boldsymbol{T} &= \{\boldsymbol{t} : \boldsymbol{\chi}(\boldsymbol{t}) = 1\} \\ \Omega &= \{\boldsymbol{\omega} : \boldsymbol{\chi}(\boldsymbol{\omega}) = 1\} \end{aligned}$$

   We will have $|\boldsymbol{T}|, |\Omega| \sim \boldsymbol{\tau N}$.

2. Generate $\boldsymbol{\alpha}$ on $\boldsymbol{T} \cup \Omega$ from a continuous distribution.

We derive bounds that hold with *very* high probability.

A Quantitative Robust Uncertainty Principle

Draw sets $\boldsymbol{T}, \Omega$ at random with $\boldsymbol{\tau} = \boldsymbol{C_M}/\sqrt{\log \boldsymbol{N}}$:

$$|\boldsymbol{T}|, |\Omega| \approx \boldsymbol{C_M} \frac{\boldsymbol{N}}{\sqrt{\log \boldsymbol{N}}}$$

Then with probability $1 - \boldsymbol{O}(\boldsymbol{N}^{-\boldsymbol{M}})$, for any $\boldsymbol{f}$ with

$$\operatorname{supp} \boldsymbol{f} = \boldsymbol{T}$$

most of the energy of $\hat{\boldsymbol{f}}$ lies outside of Ω:

$$\sum_{\boldsymbol{\omega} \in \Omega} |\hat{\boldsymbol{f}}(\boldsymbol{\omega})|^2 \ \leq \ \frac{1}{2} \|\hat{\boldsymbol{f}}\|^2.$$

- *No* $\boldsymbol{f} \in \mathbb{C}^{\boldsymbol{N}}$ exists with $\operatorname{supp} \boldsymbol{f} = \boldsymbol{T}, \quad \operatorname{supp} \hat{\boldsymbol{f}} = \Omega$.
- *For relatively large sets $\boldsymbol{T}$,Ω, it is impossible to concentrate $\boldsymbol{f}$ on $\boldsymbol{T}$ and $\hat{\boldsymbol{f}}$ on Ω.*

Key Estimate

- Again, $\boldsymbol{A} = \boldsymbol{R}_\Omega \boldsymbol{F} \boldsymbol{R}_{\boldsymbol{T}}^*$ plays a *key* role.

  Since $\operatorname{supp} \boldsymbol{f} = \boldsymbol{T}$, $\boldsymbol{R}_{\boldsymbol{T}}^* \boldsymbol{f}|_{\boldsymbol{T}} = \boldsymbol{f}$, where $\boldsymbol{f}|_{\boldsymbol{T}} = \boldsymbol{R}_{\boldsymbol{T}} \boldsymbol{f}$.

$$\begin{aligned} \sum_{\omega \in \Omega} |\hat{\boldsymbol{f}}(\boldsymbol{\omega})|^2 &= \langle \boldsymbol{R}_\Omega \boldsymbol{F}^* \boldsymbol{R}_{\boldsymbol{T}}^* \boldsymbol{f}|_{\boldsymbol{T}}, \boldsymbol{R}_\Omega \boldsymbol{F} \boldsymbol{R}_{\boldsymbol{T}}^* \boldsymbol{f}|_{\boldsymbol{T}} \rangle \\ &= \langle \boldsymbol{f}|_{\boldsymbol{T}}, \boldsymbol{A}^* \boldsymbol{A} \boldsymbol{f}|_{\boldsymbol{T}} \rangle \\ &= \|\boldsymbol{A}^* \boldsymbol{A}\| \cdot \|\boldsymbol{f}\|^2 \end{aligned}$$

- Need to show $\|\boldsymbol{A}^* \boldsymbol{A}\| \leq 1/2$ (bound the largest eigenvalue).

Generic ℓ_0 Uniqueness

Theorem: Let $\boldsymbol{T}, \Omega$ be sets for which the robust uncertainty principle holds,

$$|\boldsymbol{T}| + |\Omega| \approx \boldsymbol{C}_{\boldsymbol{M}} \frac{\boldsymbol{N}}{\sqrt{\log \boldsymbol{N}}}.$$

Choose $\boldsymbol{\alpha}$ on $\boldsymbol{T} \cup \Omega$ from a continuous distribution.
Then given $\boldsymbol{f} = \Phi \boldsymbol{\alpha}$, $\boldsymbol{\alpha}$ is the unique minimizer to

$$\min \|\boldsymbol{\alpha}'\|_{\ell_0} \qquad \text{s.t.} \qquad \Phi \boldsymbol{\alpha}' = \boldsymbol{f}$$

with probability 1.

- The uncertainty principle is fundamental, but the uniqueness is not immediate.

ℓ_1 minimization

- ℓ_0-min is highly non-convex
  $\longrightarrow$ *combinatorial* algorithm required to solve it in general.
- Instead, consider the convex program

$$\begin{aligned} \min \|\boldsymbol{\alpha}\|_{\ell_1} \\ \text{s.t. } \Phi\boldsymbol{\alpha} = \boldsymbol{s} \end{aligned} \qquad (\ell_1\text{-min})$$

- ℓ_1 norm also a sparsity measure
- ℓ_1-min *convex*
- When are the solutions to min-ℓ_0 and min-ℓ_1 the same?

Previous Work

- Donoho, Stark: uncertainty principles, sparse recovery via ℓ_1-min
- Donoho, Huo: ℓ_1-min for $|\operatorname{supp}\boldsymbol{\alpha}| \leq \frac{\sqrt{N}}{2}$
- Bruckstein, Elad: general basis pairs, ℓ_1-min for $|\operatorname{supp}\boldsymbol{\alpha}| \leq .9\sqrt{N}$
- Gribonval, Nielsen, Donoho, Elad: general dictionaries
- Tropp: greedy algorithms, robust recovery
- Gilbert, Strauss et al. (sparse Fourier approx, random algorithm)

Generic ℓ_1 recovery

Theorem: Draw sets T, Ω at random with $\tau = C_M / \log N$

$$|T| + |\Omega| \approx C_M \frac{N}{\log N}.$$

Choose α on $T \cup \Omega$ from a continuous distribution with uniform phase.
Then given $f = \Phi\alpha$, α is the unique minimizer to

$$\min \|\alpha'\|_{\ell_1} \qquad \text{s.t.} \qquad \Phi\alpha' = f$$

with probability $1 - O(N^{-M})$.

We can recover "sparse" decompositions with

$$|T| + |\Omega| \sim N / \log N$$

(compare to $\sim \sqrt{N}$)

Duality

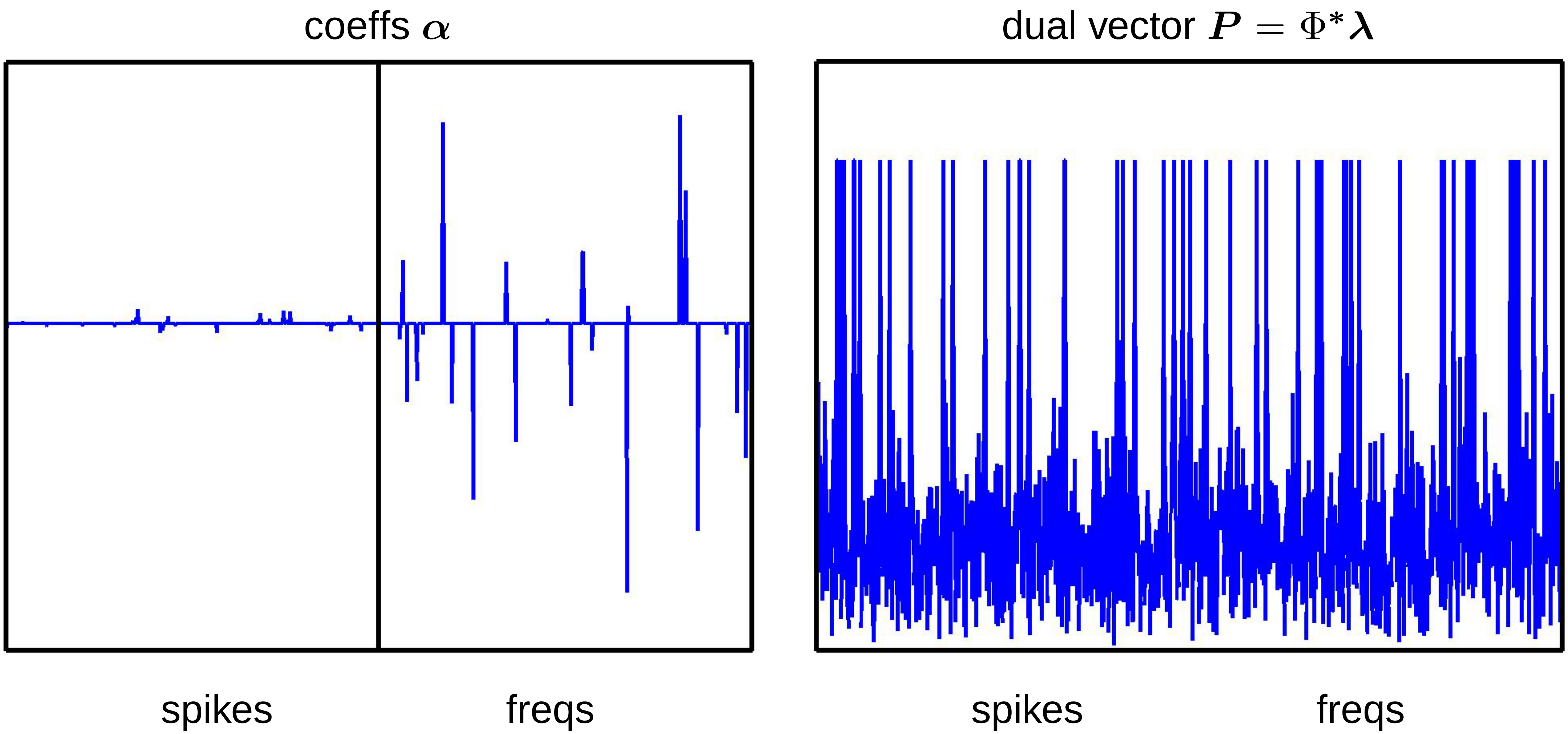
$$\min \|\boldsymbol{\alpha}\|_{\ell_1} \qquad (\ell_1\text{-min})$$
$$\text{s.t.} \quad \Phi\boldsymbol{\alpha} = \boldsymbol{f}$$

- $\boldsymbol{\alpha}^\sharp$ solves ℓ_1-min $\Leftrightarrow\ \exists\ \boldsymbol{\lambda}$ such that for $\boldsymbol{P}(\boldsymbol{k}) = (\Phi^*\boldsymbol{\lambda})(\boldsymbol{k})$

$$\boldsymbol{P}(\boldsymbol{k}) = \operatorname{sgn}(\boldsymbol{\alpha}^\sharp)(\boldsymbol{k}) \quad \text{on} \quad \boldsymbol{T} \cup \Omega \tag{1}$$
$$|\boldsymbol{P}(\boldsymbol{k})| < 1 \quad \text{on} \quad (\boldsymbol{T} \cup \Omega)^c. \tag{2}$$

- Sufficient: construct min-energy $\boldsymbol{\lambda}$ s.t. (1) is satisfied.
  Check:
  - if it exists
  - if (2) holds
- See also Fuchs (04)

Example


coeffs $\boldsymbol{\alpha}$
dual vector $\boldsymbol{P} = \Phi^* \boldsymbol{\lambda}$
spikes
freqs
spikes
freqs

Constructing the Dual Vector

$$\begin{array}{rcll} \boldsymbol{P}(\boldsymbol{k}) & \in & \mathrm{Range}(\Phi^*) & \\ \boldsymbol{P}(\boldsymbol{k}) & = & \mathrm{sgn}(\boldsymbol{\alpha}^\sharp)(\boldsymbol{k}) & \text{on} \quad \Gamma := \boldsymbol{T} \cup \Omega \\ |\boldsymbol{P}(\boldsymbol{k})| & < & 1 & \text{on} \quad (\boldsymbol{T} \cup \Omega)^{\boldsymbol{c}}. \end{array}$$

- Minimum energy $\boldsymbol{\lambda}$ with $\boldsymbol{R}_\Gamma \Phi^* \boldsymbol{\lambda} = \boldsymbol{R}_\Gamma \,\mathrm{sgn}(\boldsymbol{\alpha}^\sharp)$ ("least squares"):

$$\boldsymbol{\lambda} = \Phi \boldsymbol{R}_\Gamma^* (\boldsymbol{R}_\Gamma \Phi^* \Phi \boldsymbol{R}_\Gamma^*)^{-1} \boldsymbol{R}_\Gamma \,\mathrm{sgn}(\boldsymbol{\alpha}^\sharp)$$

- Need to invert

$$(\boldsymbol{R}_\Gamma \Phi^* \Phi \boldsymbol{R}_\Gamma^*)^{-1} = \begin{pmatrix} \boldsymbol{I} & \boldsymbol{A}^* \\ \boldsymbol{A} & \boldsymbol{I} \end{pmatrix}^{-1}, \qquad \|(\boldsymbol{R}_\Gamma \Phi^* \Phi \boldsymbol{R}_\Gamma^*)^{-1}\| = \frac{1}{1 - \sqrt{\|\boldsymbol{A}^* \boldsymbol{A}\|}}$$

with $\boldsymbol{A} = \boldsymbol{R}_\Omega \boldsymbol{F} \boldsymbol{R}_{\boldsymbol{T}}^*$.

- Eigenvalues of $\boldsymbol{A}^* \boldsymbol{A}$ again play a key role

$$A^*A = R_T F^* R_\Omega^* R_\Omega F R_T^* \qquad |T| \times |T|$$

Rewrite

$$A^*A = \frac{\Omega}{N} I - \frac{1}{N} H$$

where H is the *random matrix*

$$H(t,t') = \begin{cases} 0 & t = t' \\ -\sum_{\omega \in \Omega} e^{i\omega(t-t')} & t \neq t' \end{cases}$$

How big can we make $|T|$ and $|\Omega|$ and keep the eigenvalues under control?

$$\|H\| << |\Omega|$$

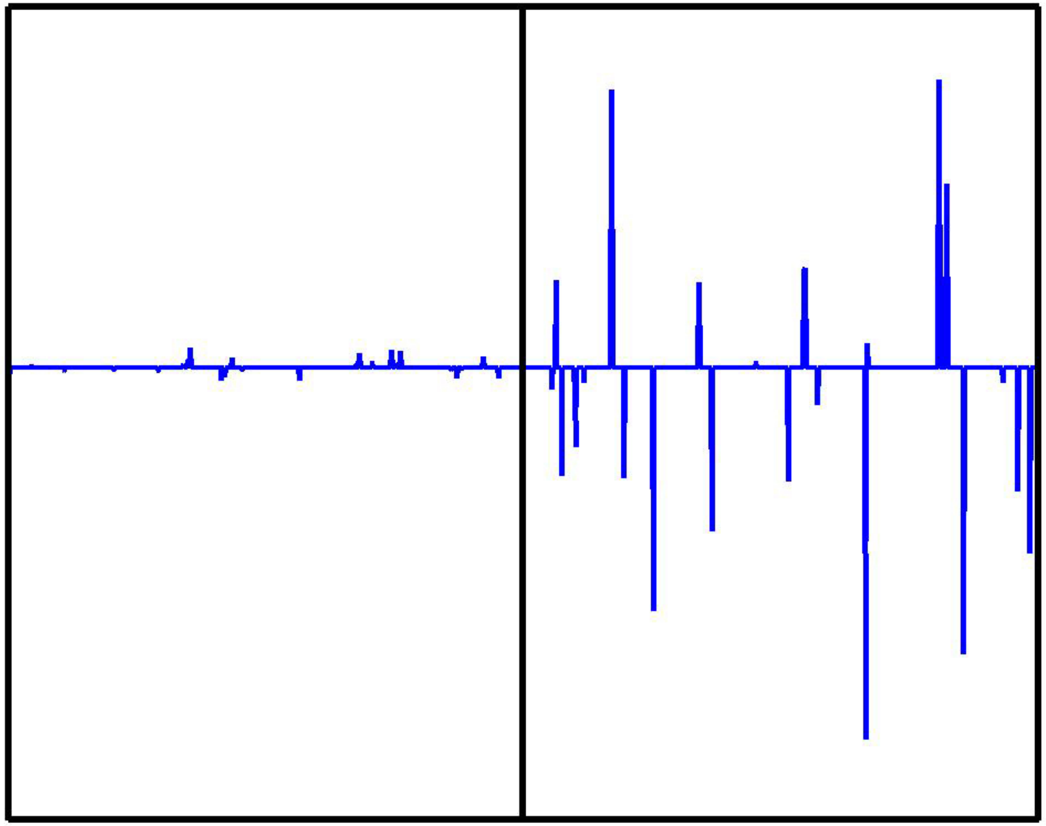
$|\boldsymbol{P}(\boldsymbol{k})| < 1$

$$\boldsymbol{P} = \Phi^*\Phi\boldsymbol{R}_\Gamma^*(\boldsymbol{R}_\Gamma\Phi^*\Phi\boldsymbol{R}_\Gamma^*)^{-1}\boldsymbol{R}_\Gamma \operatorname{sgn}(\boldsymbol{\alpha}^\sharp)$$

- $(\boldsymbol{R}_\Gamma\Phi^*\Phi\boldsymbol{R}_\Gamma^*)^{-1}$ is very well conditioned
- For $|\boldsymbol{T}|, |\Omega| \sim \boldsymbol{N}/\log \boldsymbol{N}$, the vector $(\boldsymbol{R}_\Gamma\Phi^*\Phi\boldsymbol{R}_\Gamma^*)^{-1}\boldsymbol{R}_\Gamma \operatorname{sgn}(\boldsymbol{\alpha}^\sharp)$ has small norm
- Use simple large deviation theory for $\Phi^*\Phi$ acting on a “small” vector

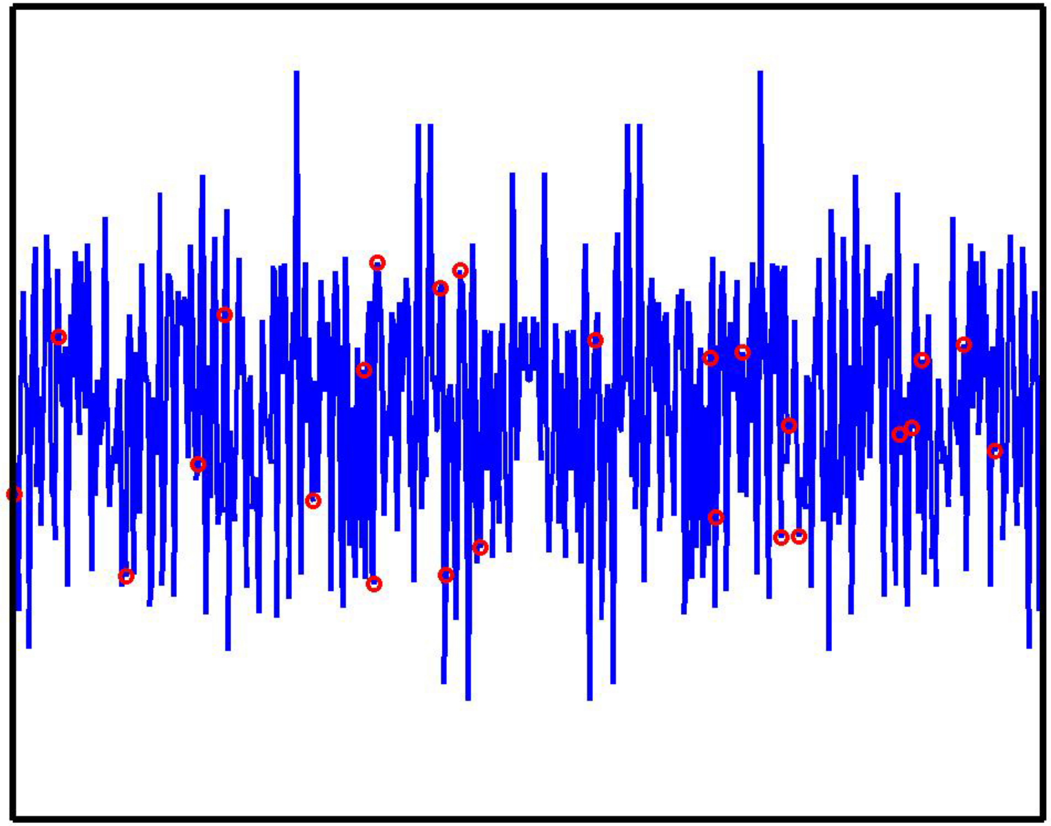
Example

coeffs α



spikes freqs

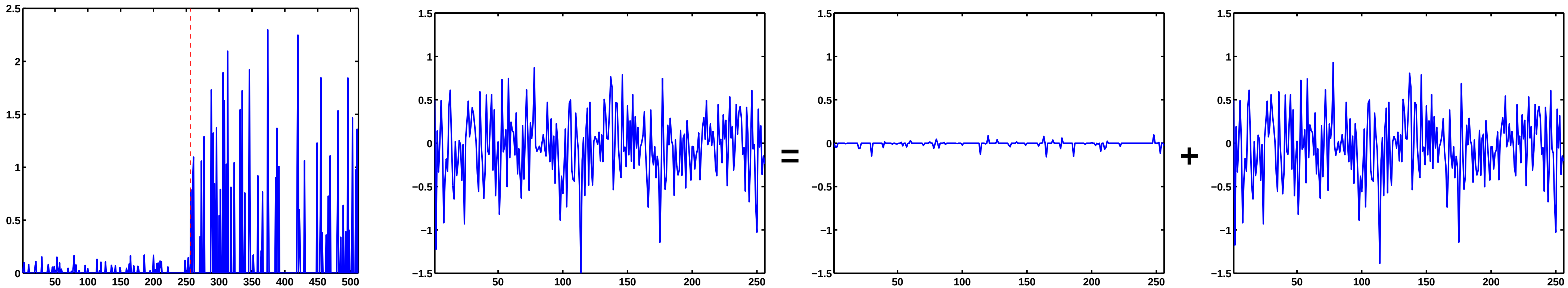
signal $f = \Phi\alpha$



$t \longrightarrow$

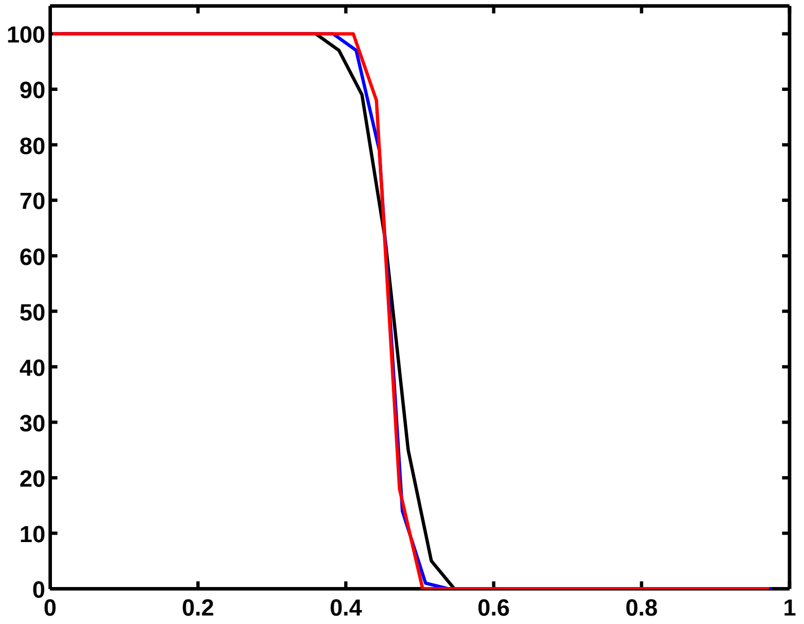
- In practice, we can recover "sparse" decompositions of size $\sim N/2$ (!!)

Example



Recovery Curves

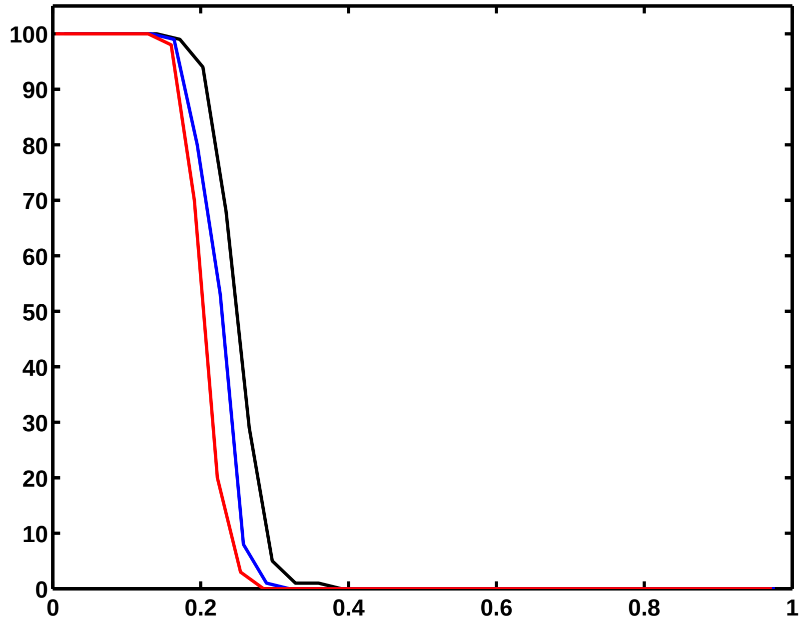
spikes/sinusoids



$|\Gamma|/N$

% success

sufficient condition



$|\Gamma|/N$

Pairs of Bases

Orthogonal bases $\boldsymbol{B}_1, \boldsymbol{B}_2$, $\Phi = \begin{pmatrix} \boldsymbol{B}_1^* & \boldsymbol{B}_2^* \end{pmatrix}$.

Define

$$\boldsymbol{\mu} = \sqrt{\boldsymbol{N}} \cdot \max \; |\langle \boldsymbol{b}_1, \boldsymbol{b}_2 \rangle| \qquad \boldsymbol{b}_1 \in \boldsymbol{B}_1, \; \boldsymbol{b}_2 \in \boldsymbol{B}_2.$$

With high probability:

- **RUP:** For $\boldsymbol{f} = \boldsymbol{B}_1 \boldsymbol{\alpha}_1, \boldsymbol{B}_2 \boldsymbol{\alpha}_2$, we have

$$|\operatorname{supp} \boldsymbol{\alpha}_1| + |\operatorname{supp} \boldsymbol{\alpha}_2| \sim \frac{\boldsymbol{N}}{\boldsymbol{\mu}^2 (\log \boldsymbol{N})^{\boldsymbol{p}}}$$

- ℓ_0 **uniqueness:** Given $\boldsymbol{\alpha} := \begin{pmatrix} \boldsymbol{\alpha}_1 & \boldsymbol{\alpha}_2 \end{pmatrix}$ supported on $\Gamma = \Gamma_1 \cup \Gamma_2$, set $\boldsymbol{f} = \Phi \boldsymbol{\alpha}$. For

$$|\Gamma| \sim \frac{\boldsymbol{N}}{\boldsymbol{\mu}^2 (\log \boldsymbol{N})^{\boldsymbol{p}}}$$

  $\boldsymbol{\alpha}$ will be the unique ℓ_0 minimizer.

- ℓ_1 **equivalence:** For Γ as above, $\boldsymbol{\alpha}$ will also be the unique minimizer of the convex ℓ_1 program.

Other Examples

- $\boldsymbol{B}_1 =$ spikes, $\boldsymbol{B}_2 =$ sinusoids
- $\boldsymbol{B}_1 =$ wavelets, $\boldsymbol{B}_2 =$ sinusoids
  at fine scales, wavelets look like spikes
- $\boldsymbol{B}_1 =$ spikes, $\boldsymbol{B}_2 =$ random basis
  performance is very close to spikes and sinusoids

Conclusions

- Sparse decompositions are *unique*
  - we can recover them from an observation of the signal
- Key: uncertainty principles
- The recovery is performed via convex optimization (min-ℓ_1)
- ”Typical” case bounds ($\sim N/\log N$) much more lenient than worst case ($\sim \sqrt{N}$)
- Derivation: show when the dual vector exists
- Bounds can be generalized to pairs of bases

Email: `jrom@acm.caltech.edu`