Uncertainty principles and sparse approximation

In this lecture, we will consider the special case where the dictionary Φ is composed of a pair of orthobases. We will see that our ability to find a sparse approximation to a signal (when one exists) hinges on the existence of an *uncertainty principle* between the two bases. Essentially, this uncertainty principle ensures that it would take many elements from one of the bases to represent the space spanned by a small number of elements from the other basis — signals that are concentrated in one of the bases must be spread out in the other. We will see how this type of relationship ensures that sparse combinations of the dictionary elements can only be written one way with a small number of terms.

<u>Spikes & Sinusoids</u>

Let's consider the special case where our dictionary is the union of two orthobases: the "spike" basis (identity) and the Fourier basis (sinusoids).

$$\Phi = [\, I \quad F \,] \quad \leftarrow n \times 2n \text{ matrix}$$

$F \rightarrow$ DFT matrix

$$F_{m\ell} = \frac{1}{\sqrt{n}} e^{j 2\pi (m-1)(\ell-1)/n} \qquad m,\ell = 1,\ldots,n$$

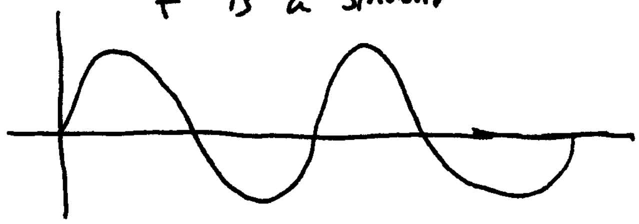
Note that these two bases are <u>completely</u> different.

- It takes n spikes to build up a single sinusoid.
- It takes n sinusoids to build up a single spike.

$\Rightarrow$ there is only one "good" (sparse) way to represent a 1-sparse signal

Ex.

f is a sinusoid



We can decompose f as

$$f = \Phi \alpha = [I \;\; F] \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 0 \\ \hline 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

single '1' at the corresp. frequency

Of course, there are many ways to decompose f, but this is the only one that uses fewer than n terms.

Out of the many decompositions of f, there is only one which is sparse.

We can generalize this statement using an uncertainty principle between the identity and Fourier bases.

Thm: Uncertainty Principle (Donoho & Stark, 1989)

Let $f \in \mathbb{C}^n$ be a discrete signal, and let $\hat{f} \in \mathbb{C}^n$ be its discrete Fourier transform.

Then

$$|\mathrm{supp}\, f| \cdot |\mathrm{supp}\, \hat{f}| \geq n$$

$$\Rightarrow \quad \underbrace{|\mathrm{supp}\, f|}_{\text{\# of nonzero terms in } f} + \underbrace{|\mathrm{supp}\, \hat{f}|}_{\text{\# of nonzero terms in } \hat{f}} \geq 2\sqrt{n}$$

pf:

Proof of Donoho-Stark Uncertainty Principle

Let T be a subset of the time domain
Let Ω be a subset of the frequency domain

Let $\Phi_{T\Omega} = [\underbrace{I_T}_{\substack{\text{columns}\\ \text{of identity}\\ \text{corresp. to } T}} \quad \underbrace{F_\Omega}_{\substack{\text{columns of}\\ \text{Fourier matrix}\\ \text{corresp. to } \Omega}}]$

If f is supported on T in the time domain
and $\hat{f} = F^* f$ is supported on Ω in the freq domain,

then
$$\Phi \begin{bmatrix} f \\ -\hat{f} \end{bmatrix} = \Phi_{T\Omega} \begin{bmatrix} f_T \\ -\hat{f}_\Omega \end{bmatrix} = 0$$

$f_T \in \mathbb{C}^{|T|}$ = restriction of f to T (throw away all zeros)

Similarly for $\hat{f}_\Omega \in \mathbb{C}^{|\Omega|}$

Thus, if all the eigenvalues of the $(|T|+|\Omega|)\times(|T|+|\Omega|)$ matrix $\Phi^*_{T\Omega}\Phi_{T\Omega}$ are > 0, then it is <u>impossible</u> to find an f supported on T s.t. $\hat{f}$ is supported on Ω.

13a

$$\Phi_{T\Omega}^* \Phi_{T\Omega} = \begin{bmatrix} I_T^* \\ F_\Omega^* \end{bmatrix} \begin{bmatrix} I_T & F_\Omega \end{bmatrix} = \begin{bmatrix} I_T^* I_T & I_T^* F_\Omega \\ F_\Omega^* I_T & F_\Omega^* F_\Omega \end{bmatrix}$$

$$= I + \underbrace{\begin{bmatrix} 0 & M \\ M^* & 0 \end{bmatrix}}_{= G}$$

$$= I + G$$

So all the eigenvalues of $\Phi_{T\Omega}^* \Phi_{T\Omega}$ are > 0 if $\underbrace{\|G\|}_{\substack{\text{maximum} \\ \text{singular value} \\ \text{(operator norm)}}} < 1$.

Note $\|G\|^2 = \|G^* G\| = \underbrace{\lambda_{max}}_{\text{maximum eigenvalue}}(G^* G)$

with $G^* G = \begin{bmatrix} MM^* & 0 \\ 0 & M^* M \end{bmatrix}$

$\Rightarrow \lambda_{max}(G^* G) = \lambda_{max}(MM^*)$ (which $= \lambda_{max}(M^* M)$)

So we need to derive conditions on T, Ω such that

$$\lambda_{max}(M^* M) < 1$$

13b

It is trivially true that

$$\lambda_{max}(M^*M) \le \text{Trace}(M^*M) = \underline{\sum_{w\in\Omega} (M^*M)_{w,w}}$$

↳ sum of values on diagonal

For each $w \in \Omega$

$$(M^*M)_{w,w} = \frac{1}{n}\sum_{t\in T} e^{-j\frac{2\pi w t}{n}} e^{j\frac{2\pi w t}{n}}$$

$$= \frac{|T|}{n}$$

$$\Rightarrow \text{Trace}(M^*M) = \frac{|\Omega|\cdot|T|}{n}$$

Thus, if

$$|\Omega|\cdot|T| < n$$

then $\text{Trace}(M^*M) < 1 \Rightarrow \|G\| < 1 \Rightarrow \lambda_{min}(\Phi_{T\Omega}^*\Phi_{T\Omega}) > 0$

$\Rightarrow$ it is impossible to find a signal concentrated on T in the time domain and Ω in the freq. domain

$\Rightarrow$ for any $f \in \mathbb{C}^n$

$$|\text{supp}\, f|\cdot|\text{supp}\, \hat{f}| \ge n$$

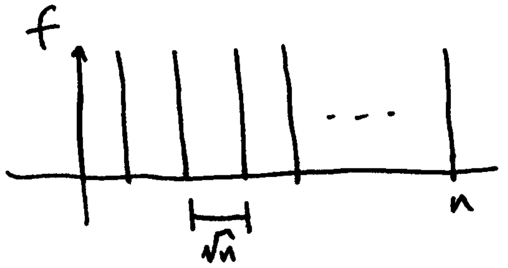
∎

13c

This UP is <u>sharp</u> in that there exists a signal for which it is met with equality

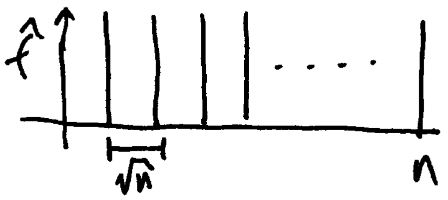
<u>Ex:</u> the Dirac comb

Say n is a square number. Construct f by placing $\sqrt{n}$ spikes spaced $\sqrt{n}$ apart



$|\mathrm{supp}\, f| = \sqrt{n}$

Then the DFT of f is the same thing:



$|\mathrm{supp}\, \hat{f}| = \sqrt{n}$

$$\Rightarrow \quad |\mathrm{supp}\, f| + |\mathrm{supp}\, \hat{f}| = 2\sqrt{n}$$

The UP has immediate consequences for the <u>uniqueness</u> of sparse decompositions.

Say we have an f which can be written as a sum of spikes on a set T + a sum of sinusoids with frequencies in Ω

$$f(t) = \sum_{\tau\in T} \alpha^1_\tau \delta(t-\tau) + \sum_{\omega\in\Omega} \alpha^2_\omega \frac{1}{\sqrt{n}} e^{j2\pi\omega t/n}$$

$$f = \Phi\alpha = [I \quad F]\begin{bmatrix}\alpha^1 \\ \hline \alpha^2\end{bmatrix}$$ where α^1 is nonzero on T
α^2 is nonzero on Ω.

Suppose that α has fewer than $\sqrt{n}$ non-zero terms:

$$|T| + |\Omega| < \sqrt{n}$$

Is there <u>another</u> way to write f with a comparable number of spikes and sinusoids?

<u>No.</u>

Why? Suppose there were another way, that is suppose $\exists\ \beta \neq \alpha$ such that

$$\Phi\beta = \Phi\alpha = f$$

and

$$|\text{supp}\ \beta^1| + |\text{supp}\ \beta^2| \leq |T| + |\Omega| < \sqrt{n}$$

Set $h = \alpha - \beta$, note that $h \in \text{Null}(\Phi)$

$$\Phi h = \Phi\alpha - \Phi\beta = f - f = 0$$

15

Also note that

$$|\text{supp } h| \le |T| + |\Omega| + |\text{supp } \beta^1| + |\text{supp } \beta^2|$$
$$< 2\sqrt{n}$$

But what do vectors in $\text{Null}(\Phi)$ look like?

$$\Phi h = 0 \Rightarrow [I \ \ F] \begin{bmatrix} h^1 \\ h^2 \end{bmatrix} = 0$$

$$\Rightarrow h^1 + F h^2 = 0 \quad \Rightarrow h^2 = -F^* h^1$$

$\Rightarrow$ they have the form

$$\left[\begin{array}{c} h^1 \\ \hline -\hat{h}^1 \end{array}\right]$$

By the UP, <u>any</u> such vector must have support at least $2\sqrt{n}$, which contradicts the above.

$\Rightarrow \alpha$ is the <u>only</u> decomposition which uses fewer than $\sqrt{n}$ terms.

As a direct result, if we observe f and there exists an expansion of f using fewer than $\sqrt{n}$ terms, solving

$$\min \|\beta\|_{\ell_0} \quad \text{s.t.} \quad \Phi\beta = f$$

will find it (and the solution is unique).

ℓ_1 & sparsest decomposition

It is also possible to develop an uncertainty principle for concentration in the ℓ_1-norm (see Donoho & Huo, 2001). Using this, it can be shown that if f has a decomposition α with

$$|\text{supp}\,\alpha| = |T| + |\Omega| \lesssim .9\sqrt{n}$$

the solution to

$$\min \|\beta\|_{\ell_1} \quad \text{s.t.} \quad \Phi\beta = f$$

(see work of Donoho-Huo, Elad-Bruckstein, Nielsen and Gribonval)

will be unique and equal to α

Summary

For an observed signal $f = \Phi\alpha$, where α has $\lesssim \sqrt{n}$ terms, we've learned

① α is the only way to decompose f w/ fewer than $\sqrt{n}$ terms

② α could be recovered with

$$\min \|\beta\|_{\ell_0} \quad \text{s.t.} \quad \Phi\beta = f$$

if we could solve this problem

③ Luckily, the convex problem

$$\min \|\beta\|_{\ell_1} \quad \text{s.t.} \quad \Phi\beta = f$$

will also recover α.

②,③ $\Rightarrow$ for sparse α, we can replace a discrete combinatorial optimization program with a convex one.

Pairs of Bases

The main results can be generalized past "spikes + Fourier".

Given a dictionary which is composed of two orthobases

$$\Phi = [\Psi_1 \quad \Psi_2]$$

define

$$\mu = \sqrt{n} \cdot \max_{\substack{\psi_1 \in \Psi_1 \\ \psi_2 \in \Psi_2}} |\langle \psi_1, \psi_2 \rangle|$$

Note that

$$1 \leq \mu \leq \sqrt{n}$$

Just as with spikes + Fourier, we can prove a UP:

Thm: (Elad & Bruckstein, '02)

Let $f \in \mathbb{C}^n$ be given, and set

$$\beta_1 = \Psi_1^* f \qquad (\Psi_1\text{-transform})$$

$$\beta_2 = \Psi_2^* f \qquad (\Psi_2\text{-transform})$$

Then

$$|\text{supp } \beta_1| \cdot |\text{supp } \beta_2| \geq n/\mu^2$$

$$\Rightarrow \quad |\text{supp } \beta_1| + |\text{supp } \beta_2| \geq 2\sqrt{n}/\mu$$

This leads directly to the uniqueness result:

<u>Thm:</u> Suppose $f = \Phi\alpha$, where

$$|\text{supp}\,\alpha| < \sqrt{n}/\mu$$

Then α is the <u>only</u> decomposition of f that uses fewer than $\sqrt{n}/\mu$ terms, and is the unique solution to

$$\min \|\beta\|_{\ell_0} \quad \text{s.t.} \quad \Phi\beta = f$$

An ℓ_1 result exists as well:

<u>Thm</u> (E+B '02):

If $f = \Phi\alpha$ with

$$|\text{supp}\,\alpha| < \frac{1}{2}\left(1 + \sqrt{n}/\mu\right) \qquad \left(\sim \sqrt{n}/\mu\right)$$

then α is the unique solution to the convex problem

$$\min \|\beta\|_{\ell_1} \quad \text{s.t.} \quad \Phi\beta = f$$

$\Rightarrow$ "untangling" the contributions from the two different bases requires η to be small

i.e. the bases must be "unsimilar"

Extensions:

We have seen that the UP for spikes & sinusoids is tight, since there exists a signal (Dirac comb) that meets the bound w/ equality.

But, it turns out we can derive much stronger uncertainty principles which hold for most sparse decompositions.

Thm: (Candès + R, '06)
Select a set T in the time domain and a set Ω in the frequency domain uniformly at random with

$$|T| + |\Omega| \leq \text{Const.} \frac{n}{\sqrt{\log n}} \qquad (\text{const} \sim 1/8)$$

Then with high probability, it is impossible to find a signal $f \in \mathbb{C}^n$ supported on T such that $\hat{f}$ is supported on Ω.

(In fact, it is impossible that even half of the energy of $\hat{f}$ can be supported on Ω.)

This theorem says that a UP holds for "generic" (i.e. minus special cases like the Dirac comb) signals up to sparsity $\sim n/\sqrt{\log n}$ (compare to $\sim\sqrt{n}$).

There are also associated ℓ_0 & ℓ_1 recovery results:

<u>Thm:</u> Let $\alpha = \begin{bmatrix}\alpha^1 \\ \alpha^2\end{bmatrix}$ be a sparse decomposition of $f = \Phi\alpha$, $\Phi = [I \;\; F]$, with α^1 supported on T and α^2 supported on Ω, where T & Ω are chosen as above. Then for the "vast majority" of such α, α will be the unique solution to

$$\min \|\beta\|_{\ell_0} \quad \text{s.t.} \quad \Phi\beta = f$$

where $f = \Phi\alpha$.

The ℓ_1 result requires a little bit more sparsity

<u>Thm:</u> Choose T, Ω uniformly at random with

$$|T| + |\Omega| \le \text{Const.} \frac{n}{\log n}$$

Then for the vast majority of α supported on $T \cup \Omega$, α is the unique soln. to

$$\min \|\beta\|_{\ell_1} \quad \text{s.t.} \quad \Phi\beta = f$$

where $f = \Phi\alpha$.

Extensions to general pairs of bases hold as well — replace n with n/μ^2 everywhere above.

Leaving out "special cases", ℓ_1 can effectively recover decompositions with sparsity on the order of $\sim n/\log n$ (not too far from n).

Bounds like these can also be found for general dictionaries Φ (not necessarily unions of $\perp$-bases) — the key parameter in this case is

$$\mu = \max_{\substack{\gamma,\alpha = 1,\dots,n \\ \gamma \neq \alpha}} |\langle \varphi_\gamma, \varphi_\alpha \rangle|$$

columns of Φ

(see work by Elad & Donoho '03, & Tropp '06).