Fourier Sampling Numbers for Besov spaces

Jonathan W. Siegel

Texas A&M University

jwsiegel@tamu.edu

Workshop in Honor of Albert Cohen
July 3, 2025

Outline

General Approximation Theory Problem

- We consider the following general problem[1]:

$$\inf_{D,E} \sup_{f\in K} \|f - D(E(f))\|_X \tag{1}$$

- For various different norms X, model classes K, and restrictions on D and E we get a variety of problems:
  - D restricted to a certain type (e.g. piecewise polynomials, trigonometric polynomials, neural networks): approximation rates

[1] Albert Cohen, Ronald DeVore, Guergana Petrova, and Przemyslaw Wojtaszczyk. "Optimal stable nonlinear approximation". In: *Foundations of Computational Mathematics* 22.3 (2022), pp. 607–648.

General Approximation Theory Problem

- We consider the following general problem[1]:

$$\inf_{D,E} \sup_{f\in K} \|f - D(E(f))\|_X \tag{1}$$

- For various different norms X, model classes K, and restrictions on D and E we get a variety of problems:
  - D restricted to a certain type (e.g. piecewise polynomials, trigonometric polynomials, neural networks): approximation rates
  - D and E both linear: linear widths

[1] Albert Cohen, Ronald DeVore, Guergana Petrova, and Przemyslaw Wojtaszczyk. "Optimal stable nonlinear approximation". In: *Foundations of Computational Mathematics* 22.3 (2022), pp. 607–648.

General Approximation Theory Problem

- We consider the following general problem[1]:

$$\inf_{D,E} \sup_{f \in K} \|f - D(E(f))\|_X \tag{1}$$

- For various different norms X, model classes K, and restrictions on D and E we get a variety of problems:
  - D restricted to a certain type (e.g. piecewise polynomials, trigonometric polynomials, neural networks): approximation rates
  - D and E both linear: linear widths
  - D linear and E arbitrary: Kolmogorov widths

[1] Albert Cohen, Ronald DeVore, Guergana Petrova, and Przemyslaw Wojtaszczyk. "Optimal stable nonlinear approximation". In: *Foundations of Computational Mathematics* 22.3 (2022), pp. 607–648.

General Approximation Theory Problem

- We consider the following general problem[1]:

$$\inf_{D,E} \sup_{f \in K} \|f - D(E(f))\|_X \tag{1}$$

- For various different norms X, model classes K, and restrictions on D and E we get a variety of problems:
  - D restricted to a certain type (e.g. piecewise polynomials, trigonometric polynomials, neural networks): approximation rates
  - D and E both linear: linear widths
  - D linear and E arbitrary: Kolmogorov widths
  - D arbitrary and E linear: Gelfand widths

[1] Albert Cohen, Ronald DeVore, Guergana Petrova, and Przemyslaw Wojtaszczyk. "Optimal stable nonlinear approximation". In: *Foundations of Computational Mathematics* 22.3 (2022), pp. 607–648.

General Approximation Theory Problem

- We consider the following general problem[1]:

$$\inf_{D,E} \sup_{f\in K} \|f - D(E(f))\|_X \tag{1}$$

- For various different norms X, model classes K, and restrictions on D and E we get a variety of problems:
  - D restricted to a certain type (e.g. piecewise polynomials, trigonometric polynomials, neural networks): approximation rates
  - D and E both linear: linear widths
  - D linear and E arbitrary: Kolmogorov widths
  - D arbitrary and E linear: Gelfand widths
  - D arbitrary and E restricted linear: today

[1] Albert Cohen, Ronald DeVore, Guergana Petrova, and Przemyslaw Wojtaszczyk. "Optimal stable nonlinear approximation". In: *Foundations of Computational Mathematics* 22.3 (2022), pp. 607–648.

Model Class Assumptions

- We will consider the case $X = L_p(\Omega)$
  - $\Omega \subset \mathbb{R}^d$ or $\Omega = \mathbb{T}^d$

[2] Andrew R Barron. "Universal approximation bounds for superpositions of a sigmoidal function". In: *IEEE Transactions on Information theory* 39.3 (1993), pp. 930–945.

[3] Rahul Parhi and Robert D Nowak. "Banach space representer theorems for neural networks and ridge splines". In: *Journal of Machine Learning Research* 22.43 (2021), pp. 1–40.

Model Class Assumptions

- We will consider the case $X = L_p(\Omega)$
  - $\Omega \subset \mathbb{R}^d$ or $\Omega = \mathbb{T}^d$
- Classical smoothness assumptions:
  - f is in the unit ball of a Sobolev or Besov space, i.e.,

$$\|f\|_{B^s_r(L_q(\Omega))} \leq 1 \text{ or } \|f\|_{W^s(L_q(\Omega))} \leq 1 \tag{2}$$

  - Need $1/q - 1/p < s/d$ for compactness

[2] Andrew R Barron. "Universal approximation bounds for superpositions of a sigmoidal function". In: *IEEE Transactions on Information theory* 39.3 (1993), pp. 930–945.

[3] Rahul Parhi and Robert D Nowak. "Banach space representer theorems for neural networks and ridge splines". In: *Journal of Machine Learning Research* 22.43 (2021), pp. 1–40.

Model Class Assumptions

- We will consider the case $X = L_p(\Omega)$
  - $\Omega \subset \mathbb{R}^d$ or $\Omega = \mathbb{T}^d$
- Classical smoothness assumptions:
  - f is in the unit ball of a Sobolev or Besov space, i.e.,

$$\|f\|_{B^s_r(L_q(\Omega))} \leq 1 \text{ or } \|f\|_{W^s(L_q(\Omega))} \leq 1 \tag{2}$$

    - Need $1/q - 1/p < s/d$ for compactness
- Shallow neural network model classes:
  - Barron's class[2], Radon BV[3], etc.

[2] Andrew R Barron. "Universal approximation bounds for superpositions of a sigmoidal function". In: *IEEE Transactions on Information theory* 39.3 (1993), pp. 930–945.

[3] Rahul Parhi and Robert D Nowak. "Banach space representer theorems for neural networks and ridge splines". In: *Journal of Machine Learning Research* 22.43 (2021), pp. 1–40.

Model Class Assumptions

- We will consider the case $X = L_p(\Omega)$
  - $\Omega \subset \mathbb{R}^d$ or $\Omega = \mathbb{T}^d$
- Classical smoothness assumptions:
  - f is in the unit ball of a Sobolev or Besov space, i.e.,

$$\|f\|_{B^s_r(L_q(\Omega))} \leq 1 \text{ or } \|f\|_{W^s(L_q(\Omega))} \leq 1 \tag{2}$$

    - Need $1/q - 1/p < s/d$ for compactness
- Shallow neural network model classes:
  - Barron's class[2], Radon BV[3], etc.
- More general non-convex model classes
  - Ex: $K = \{1_C : \; C \subset \Omega \text{ is convex}\}$.

[2] Andrew R Barron. "Universal approximation bounds for superpositions of a sigmoidal function". In: *IEEE Transactions on Information theory* 39.3 (1993), pp. 930–945.

[3] Rahul Parhi and Robert D Nowak. "Banach space representer theorems for neural networks and ridge splines". In: *Journal of Machine Learning Research* 22.43 (2021), pp. 1–40.

Besov Spaces[4]

- The k-th order finite difference of a function f is defined by

$$\Delta_k^h f(x) = \begin{cases} \sum_{j=0}^k (-1)^j \binom{k}{j} f(x+jh) & x, x+h, ..., x+jh \in \Omega \\ 0 & \text{otherwise} \end{cases} \quad (3)$$

[4] Ronald A DeVore and Robert C Sharpley. "Besov spaces on domains in $\mathbb{R}^d$". In: *Transactions of the American Mathematical Society* 335.2 (1993), pp. 843–864.

Besov Spaces[4]

- The k-th order finite difference of a function f is defined by

$$\Delta_k^h f(x) = \begin{cases} \sum_{j=0}^k (-1)^j \binom{k}{j} f(x+jh) & x, x+h, ..., x+jh \in \Omega \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

- The k-th order L_q modulus of continuity is defined by

$$\omega_k(f,t)_q := \sup_{0<|h|\leq t} \left(\int_\Omega |\Delta_k^h f(x)|^q dx\right)^{1/q} \tag{4}$$

[4] Ronald A DeVore and Robert C Sharpley. "Besov spaces on domains in $\mathbb{R}^d$". In: *Transactions of the American Mathematical Society* 335.2 (1993), pp. 843–864.

Besov Spaces[4]

- The k-th order finite difference of a function f is defined by

$$\Delta_k^h f(x) = \begin{cases} \sum_{j=0}^{k} (-1)^j \binom{k}{j} f(x+jh) & x, x+h, ..., x+jh \in \Omega \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

- The k-th order L_q modulus of continuity is defined by

$$\omega_k(f,t)_q := \sup_{0<|h|\leq t} \left(\int_\Omega |\Delta_k^h f(x)|^q dx\right)^{1/q} \tag{4}$$

- Given parameters $1 \leq q, r \leq \infty$ and $s > 0$ we define the Besov norm of a function $f \in L_q(\Omega)$ via

$$\|f\|_{B_r^s(L_q(\Omega))} := \|f\|_{L_q(\Omega)} + \left(\int_0^\infty \left(\frac{\omega_k(f,t)_q}{t^s}\right)^r \frac{dt}{t}\right)^{1/r} \tag{5}$$

[4] Ronald A DeVore and Robert C Sharpley. "Besov spaces on domains in $\mathbb{R}^d$". In: *Transactions of the American Mathematical Society* 335.2 (1993), pp. 843–864.

Restrictions on Measurements

- Point samples, i.e., $\lambda_i(f) = f(x_i)$ for some points $x_i \in \Omega$
  - Sampling numbers

Restrictions on Measurements

- Point samples, i.e., $\lambda_i(f) = f(x_i)$ for some points $x_i \in \Omega$
  - Sampling numbers
- General linear measurements, i.e., $\lambda_i \in X^*$
  - Gelfand widths

Restrictions on Measurements

- Point samples, i.e., $\lambda_i(f) = f(x_i)$ for some points $x_i \in \Omega$
  - Sampling numbers
- General linear measurements, i.e., $\lambda_i \in X^*$
  - Gelfand widths
- Fourier measurements, i.e.,

$$\lambda_i(f) = \hat{f}(\xi_i) = \int_\Omega e^{2\pi \mathrm{i} \xi_i \cdot x} f(x) dx. \tag{6}$$

  - Models measurements made by MRI

Restrictions on Measurements

- Point samples, i.e., $\lambda_i(f) = f(x_i)$ for some points $x_i \in \Omega$
  - Sampling numbers
- General linear measurements, i.e., $\lambda_i \in X^*$
  - Gelfand widths
- Fourier measurements, i.e.,

$$\lambda_i(f) = \hat{f}(\xi_i) = \int_\Omega e^{2\pi \mathrm{i} \xi_i \cdot x} f(x) dx. \tag{6}$$

  - Models measurements made by MRI
- Radon measurements, i.e.,

$$\lambda_i(f) = \mathcal{R}(f)(\omega_i, b_i) = \int_{\Omega \cap \{\omega_i \cdot x = b_i\}} f(x) dx. \tag{7}$$

  - Models measurements made by CT

Sampling Numbers of Besov spaces

- We'll consider $K_q^s := \{f : \|f\|_{B^s_\infty(L_q(\Omega))} \leq 1\}$
  - Recall: Need $1/q - 1/p < s/d$ to ensure compactness in L_p

Sampling Numbers of Besov spaces

- We'll consider $K_q^s := \{f : \|f\|_{B^s_\infty(L_q(\Omega))} \leq 1\}$
  - Recall: Need $1/q - 1/p < s/d$ to ensure compactness in L_p
- Consider recovering f in L_p from each of the four types of measuments considered:
  - Point samples
  - General linear functionals
  - Fourier samples
  - Radon samples

Sampling Numbers of Besov spaces

- We'll consider $K_q^s := \{f : \|f\|_{B^s_\infty(L_q(\Omega))} \leq 1\}$
  - Recall: Need $1/q - 1/p < s/d$ to ensure compactness in L_p
- Consider recovering f in L_p from each of the four types of measuments considered:
  - Point samples
  - General linear functionals
  - Fourier samples
  - Radon samples
- In each case, we want the corresponding sampling numbers:

$$s_n(K_q^s)_{L_p} := \inf_{\substack{D,E \\ E \text{ restricted}}} \sup_{f \in K_q^s} \|f - D(E(f))\|_{L_p}. \tag{8}$$

  - Non-linear regime: $q < p$

General Recovery Algorithm

- Given a set of measurements $\Lambda = E(f)$, the radius of the smallest ball containing the set[5]

$$\{f \in K : \ E(f) = \Lambda\}, \tag{9}$$

  also called the Chebyshev ball, is the minimal reconstruction error
  - The center of (or any point in) the Chebyshev ball is a good estimate

[5] Charles A Micchelli, Th J Rivlin, and Shmuel Winograd. "The optimal recovery of smooth functions". In: *Numerische Mathematik* 26 (1976), pp. 191–200, Joseph Frederick Traub and Henryk Woźniakowski. "A general theory of optimal algorithms". In: (1980), Borislav Bojanov. "Optimal recovery of functions and integrals". In: *First European Congress of Mathematics Invited Lectures*. Springer. 1994, pp. 371–390, Peter Binev, Andrea Bonito, Ronald DeVore, and Guergana Petrova. "Optimal learning". In: *Calcolo* 61.1 (2024), p. 15.

General Recovery Algorithm

- Given a set of measurements $\Lambda = E(f)$, the radius of the smallest ball containing the set[5]

$$\{f \in K : \; E(f) = \Lambda\}, \tag{9}$$

also called the Chebyshev ball, is the minimal reconstruction error
  - The center of (or any point in) the Chebyshev ball is a good estimate
- In our case, we can easily find such an element by solving

$$\arg\min_{E(f)=\Lambda} \|f\|_{B^s_\infty(L_q(\Omega))}. \tag{10}$$

[5] Charles A Micchelli, Th J Rivlin, and Shmuel Winograd. "The optimal recovery of smooth functions". In: *Numerische Mathematik* 26 (1976), pp. 191–200, Joseph Frederick Traub and Henryk Woźniakowski. "A general theory of optimal algorithms". In: (1980), Borislav Bojanov. "Optimal recovery of functions and integrals". In: *First European Congress of Mathematics Invited Lectures*. Springer. 1994, pp. 371–390, Peter Binev, Andrea Bonito, Ronald DeVore, and Guergana Petrova. "Optimal learning". In: *Calcolo* 61.1 (2024), p. 15.

General Recovery Algorithm

- Given a set of measurements $\Lambda = E(f)$, the radius of the smallest ball containing the set[5]
$$\{f \in K : \ E(f) = \Lambda\}, \tag{9}$$
also called the Chebyshev ball, is the minimal reconstruction error
  - The center of (or any point in) the Chebyshev ball is a good estimate
- In our case, we can easily find such an element by solving
$$\arg \min_{E(f)=\Lambda} \|f\|_{B^s_\infty(L_q(\Omega))}. \tag{10}$$
- Further, the convexity and symmetry of K^s_q implies that
$$s_n(K^s_q)_{L_p} \eqsim \sup\{\|f\|_{L_p} : \ f \in K^s_q, \ E(f) = 0\} \tag{11}$$

[5] Charles A Micchelli, Th J Rivlin, and Shmuel Winograd. "The optimal recovery of smooth functions". In: *Numerische Mathematik* 26 (1976), pp. 191–200, Joseph Frederick Traub and Henryk Woźniakowski. "A general theory of optimal algorithms". In: (1980), Borislav Bojanov. "Optimal recovery of functions and integrals". In: *First European Congress of Mathematics Invited Lectures*. Springer. 1994, pp. 371–390, Peter Binev, Andrea Bonito, Ronald DeVore, and Guergana Petrova. "Optimal learning". In: *Calcolo* 61.1 (2024), p. 15.

Point Samples

- Consider recovering $f \in K_q^s$ from *point samples*
  - Need $s > d/q$ to ensure that point samples are well-defined
- Point sampling rates are[6]

$$s_n^P(K_q^s)_{L_p} \eqsim n^{-s/d+(1/q-1/p)_+} \tag{12}$$

  - Uniform grid of points is quasi-optimal
  - Rate deteriorates in the non-linear regime $q < p$

[6] Erich Novak and Hans Triebel. "Function spaces in Lipschitz domains and optimal rates of convergence for sampling". In: *Constructive approximation* 23 (2006), pp. 325–350, Jan Vybíral. "Sampling numbers and function spaces". In: *Journal of Complexity* 23.4-6 (2007), pp. 773–792, Andrea Bonito, Ronald DeVore, Guergana Petrova, and Jonathan W Siegel. "Convergence and error control of consistent PINNs for elliptic PDEs". In: *IMA Journal of Numerical Analysis* (2025), draf008.

Gelfand Widths

- Suppose we allow general linear functionals
- Optimal recovery is controlled by the Gelfand widths[7]:

$$s_n^G(K_q^s)_{L_p} \eqsim \begin{cases} n^{-s/d+(1/q-1/p)_+} & q \geq 2 \\ n^{-s/d+(1/2-1/p)_+} & 1 \leq q < 2. \end{cases} \tag{13}$$

  - When $1 \leq q, p \leq 2$ we get $O(n^{-s/d})$ even in the non-linear regime
  - In this regime we need a complicated random set of measurements

[7] George G Lorentz, Manfred von Golitschek, and Yuly Makovoz. *Constructive approximation: advanced problems*. Vol. 304. Citeseer, 1996.

Non-linear Approximation

- Why do we care about the regime $q < p$?

Non-linear Approximation

- Why do we care about the regime $q < p$?
- Suppose we consider $BV(\Omega) \subset B^1_\infty(L_1(\Omega))$ and

$$f(x) = \begin{cases} 1 & x \in C \\ 0 & x \notin C \end{cases} \tag{14}$$

  for some open set C (with nice boundary)
  - We have $f \in BV(\Omega)$
  - $BV(\Omega) \subset L_p$ if $p < \frac{d}{d-1}$

Non-linear Approximation

- Why do we care about the regime $q < p$?
- Suppose we consider $BV(\Omega) \subset B^1_\infty(L_1(\Omega))$ and

$$f(x) = \begin{cases} 1 & x \in C \\ 0 & x \notin C \end{cases} \tag{14}$$

  for some open set C (with nice boundary)
  - We have $f \in BV(\Omega)$
  - $BV(\Omega) \subset L_p$ if $p < \frac{d}{d-1}$
- Let us approximate f from:
  - Point samples, get error $O(n^{-1/d+(1-1/p)})$
  - General linear functionals, get error $O(n^{-1/d})$

Non-linear Approximation

- Notice that

$$|\{x : |f(x) - f_n(x)| \geq 1/2\}| \leq (2\|f - f_n\|_{L_p})^p \leq C \begin{cases} n^{-p/d+p-1} & \text{point samples} \\ n^{-p/d} & \text{general linear functionals} \end{cases} \tag{15}$$

Non-linear Approximation

- Notice that

$$
\begin{aligned}
|\{x : |f(x) - f_n(x)| \geq 1/2\}| &\leq (2\|f - f_n\|_{L_p})^p \\
&\leq C \begin{cases} n^{-p/d+p-1} & \text{point samples} \\ n^{-p/d} & \text{general linear functionals} \end{cases}
\end{aligned} \tag{15}
$$

  - With point samples we recover the boundary/edges up to accuracy $O(n^{-1/d})$ (with $p = 1$)
  - With general functionals we recover the boundary/edges up to accuracy $O(n^{-1/(d-1)})$ (with $p \to d/(d-1)$)

Non-linear Approximation

- Notice that

$$|\{x : |f(x) - f_n(x)| \geq 1/2\}| \leq (2\|f - f_n\|_{L_p})^p \leq C \begin{cases} n^{-p/d+p-1} & \text{point samples} \\ n^{-p/d} & \text{general linear functionals} \end{cases} \quad (15)$$

  - With point samples we recover the boundary/edges up to accuracy $O(n^{-1/d})$ (with $p = 1$)
  - With general functionals we recover the boundary/edges up to accuracy $O(n^{-1/(d-1)})$ (with $p \to d/(d-1)$)
- Non-linear approximation can recover edges to much higher accuracy!

Compressive Sensing[9]

- Recover a k-sparse vector $x \in \mathbb{C}^N$ from few measurements:

$$\hat{x} = \arg\min_{Ay=b} \|y\|_{\ell_1} \tag{16}$$

  - A is the measurement matrix, $b = Ax$ are the measurements

[8] Emmanuel J Candes and Terence Tao. "Decoding by linear programming". In: *IEEE transactions on information theory* 51.12 (2005), pp. 4203–4215.

[9] David L Donoho. "Compressed sensing". In: *IEEE Transactions on information theory* 52.4 (2006), pp. 1289–1306, Emmanuel J Candès, Justin Romberg, and Terence Tao. "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information". In: *IEEE Transactions on information theory* 52.2 (2006), pp. 489–509.

Compressive Sensing[9]

- Recover a k-sparse vector $x \in \mathbb{C}^N$ from few measurements:

$$\hat{x} = \arg\min_{Ay=b} \|y\|_{\ell_1} \tag{16}$$

  - A is the measurement matrix, $b = Ax$ are the measurements
- A satisfies the (s, δ) restricted isometry property (RIP)[8], i.e.,

$$(1-\delta)\|x\|_2 \leq \|Ax\|_2 \leq (1+\delta)\|x\|_2 \tag{17}$$

  for all s-sparse vectors

---

[8] Emmanuel J Candes and Terence Tao. "Decoding by linear programming". In: *IEEE transactions on information theory* 51.12 (2005), pp. 4203–4215.

[9] David L Donoho. "Compressed sensing". In: *IEEE Transactions on information theory* 52.4 (2006), pp. 1289–1306, Emmanuel J Candès, Justin Romberg, and Terence Tao. "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information". In: *IEEE Transactions on information theory* 52.2 (2006), pp. 489–509.

Null Space Property

- A sensing matrix satisfying the (s,δ)-RIP with $\delta \leq 1/4$ satisfies the following Null Space Property[10]:

$$\|x\|_2 \leq \frac{C}{\sqrt{s}}\|x\|_1 \text{ if } Ax = 0. \tag{18}$$

[10] Albert Cohen, Wolfgang Dahmen, and Ronald DeVore. "Compressed sensing and best k-term approximation". In: *Journal of the American mathematical society* 22.1 (2009), pp. 211–231, George G Lorentz, Manfred von Golitschek, and Yuly Makovoz. *Constructive approximation: advanced problems*. Vol. 304. Citeseer, 1996.

[11] Emmanuel J Candes and Terence Tao. "Decoding by linear programming". In: *IEEE transactions on information theory* 51.12 (2005), pp. 4203–4215, Richard Baraniuk, Mark Davenport, Ronald DeVore, and Michael Wakin. "A simple proof of the restricted isometry property for random matrices". In: *Constructive approximation* 28 (2008), pp. 253–263.

Null Space Property

- A sensing matrix satisfying the (s, δ)-RIP with $\delta \leq 1/4$ satisfies the following Null Space Property[10]:

$$\|x\|_2 \leq \frac{C}{\sqrt{s}}\|x\|_1 \text{ if } Ax = 0. \tag{18}$$

  - There exist matrices satisfying an (s, δ)-RIP with $O(s \log(N/s))$ rows[11]

[10] Albert Cohen, Wolfgang Dahmen, and Ronald DeVore. "Compressed sensing and best k-term approximation". In: *Journal of the American mathematical society* 22.1 (2009), pp. 211–231, George G Lorentz, Manfred von Golitschek, and Yuly Makovoz. *Constructive approximation: advanced problems*. Vol. 304. Citeseer, 1996.

[11] Emmanuel J Candes and Terence Tao. "Decoding by linear programming". In: *IEEE transactions on information theory* 51.12 (2005), pp. 4203–4215, Richard Baraniuk, Mark Davenport, Ronald DeVore, and Michael Wakin. "A simple proof of the restricted isometry property for random matrices". In: *Constructive approximation* 28 (2008), pp. 253–263.

Null Space Property

- A sensing matrix satisfying the (s, δ)-RIP with $\delta \leq 1/4$ satisfies the following Null Space Property[10]:

$$\|x\|_2 \leq \frac{C}{\sqrt{s}} \|x\|_1 \text{ if } Ax = 0. \tag{18}$$

  - There exist matrices satisfying an (s, δ)-RIP with $O(s \log(N/s))$ rows[11]
  - This gives sharp bounds on the Gelfand widths

[10] Albert Cohen, Wolfgang Dahmen, and Ronald DeVore. "Compressed sensing and best k-term approximation". In: *Journal of the American mathematical society* 22.1 (2009), pp. 211–231, George G Lorentz, Manfred von Golitschek, and Yuly Makovoz. *Constructive approximation: advanced problems*. Vol. 304. Citeseer, 1996.

[11] Emmanuel J Candes and Terence Tao. "Decoding by linear programming". In: *IEEE transactions on information theory* 51.12 (2005), pp. 4203–4215, Richard Baraniuk, Mark Davenport, Ronald DeVore, and Michael Wakin. "A simple proof of the restricted isometry property for random matrices". In: *Constructive approximation* 28 (2008), pp. 253–263.

Fourier CS Matrices

- Random Fourier matrices satisfy the RIP[12]
- Randomly sampled bounded orthogonal systems satisfy the Null Space Property[13]:
  - Let $\phi_1, ..., \phi_n$ be an orthonormal system in L_2 such that $\|\phi_i\|_{L_\infty} \leq C$.
  - Let $1 < k < n$ indices be chosen randomly (gives a set $|I_k| = k$). Then with probability at least $1/2$ we have

$$\left\|\sum_{i \notin I_k} a_i\phi_i\right\|_{L_2} \lesssim \mu(\log(\mu))^{5/2}\left\|\sum_{i \notin I_k} a_i\phi_i\right\|_{L_1} \quad (19)$$

where $\mu = \sqrt{\frac{n}{k}(\log k)}$, for all coefficients a_i

[12] Emmanuel J Candes and Terence Tao. "Near-optimal signal recovery from random projections: Universal encoding strategies?" In: *IEEE transactions on information theory* 52.12 (2006), pp. 5406–5425, Mark Rudelson and Roman Vershynin. "On sparse reconstruction from Fourier and Gaussian measurements". In: *Communications on Pure and Applied Mathematics: A Journal Issued by the Courant Institute of Mathematical Sciences* 61.8 (2008), pp. 1025–1045.

[13] Olivier Guédon, Shahar Mendelson, Alain Pajor, and Nicole Tomczak-Jaegermann. "Majorizing measures and proportional subsets of bounded orthonormal systems". In: (2008).

Continuous Compressed Sensing

- Traditional compressed sensing applies to sparse, discrete signals and discrete measurements

[14] Ben Adcock, Anders C Hansen, Clarice Poon, and Bogdan Roman. "Breaking the coherence barrier: A new theory for compressed sensing". In: *Forum of mathematics, sigma*. Vol. 5. Cambridge University Press. 2017, e4, Yaakov Tsaig and David L Donoho. "Extensions of compressed sensing". In: *Signal processing* 86.3 (2006), pp. 549–571.

Continuous Compressed Sensing

- Traditional compressed sensing applies to sparse, discrete signals and discrete measurements
- We're interested in continuous functions and continuous measurements
  - Some numerical analysis must be done

[14] Ben Adcock, Anders C Hansen, Clarice Poon, and Bogdan Roman. "Breaking the coherence barrier: A new theory for compressed sensing". In: *Forum of mathematics, sigma*. Vol. 5. Cambridge University Press. 2017, e4, Yaakov Tsaig and David L Donoho. "Extensions of compressed sensing". In: *Signal processing* 86.3 (2006), pp. 549–571.

Continuous Compressed Sensing

- Traditional compressed sensing applies to sparse, discrete signals and discrete measurements
- We're interested in continuous functions and continuous measurements
  - Some numerical analysis must be done
- Existing works[14] make much stronger assumptions on the target function than we need

[14] Ben Adcock, Anders C Hansen, Clarice Poon, and Bogdan Roman. "Breaking the coherence barrier: A new theory for compressed sensing". In: *Forum of mathematics, sigma*. Vol. 5. Cambridge University Press. 2017, e4, Yaakov Tsaig and David L Donoho. "Extensions of compressed sensing". In: *Signal processing* 86.3 (2006), pp. 549–571.

Multiscale Decomposition

- Recall, we want to find a set S of n frequencies such that

$$\max\{\|f\|_{L_p} : \ f \in K_q^s \text{ and } \hat{f}(k) = 0 \text{ for all } k \in S\} \tag{20}$$

  is minimized
- Let's consider just the case $q = 1$ and $1 \leq p \leq 2$

Multiscale Decomposition

- Recall, we want to find a set S of n frequencies such that

$$\max\{\|f\|_{L_p} : \; f \in K_q^s \text{ and } \hat{f}(k) = 0 \text{ for all } k \in S\} \tag{20}$$

  is minimized
- Let's consider just the case $q = 1$ and $1 \leq p \leq 2$
- Multiscale decomposition of f:

$$f = \sum_{i=0}^{\infty} f_i \tag{21}$$

  - Support of $\hat{f}_i$ contained in $S_i := \{k : \lfloor 2^{i-1} \rfloor \leq |k|_\infty \leq 2^{i+1}\}$
  - $\hat{f}(k) = 0$ implies $\hat{f}_i(k) = 0$
  - $\|f_i\|_{L_1} \leq C2^{-is}\|f\|_{B^s_\infty(L_1)} \leq C2^{-is}$

Main Bounds

- For each i, we now sample frequencies from S_i, either
    - All frequencies in S_i
    - Randomly sample $k_i > 1$ frequencies in S_i
    - None of the frequencies in S_i

Main Bounds

- For each i, we now sample frequencies from S_i, either
  - All frequencies in S_i
  - Randomly sample $k_i > 1$ frequencies in S_i
  - None of the frequencies in S_i
- The frequencies in S_i form a bounded orthogonal system, so that if f_i vanishes at these frequencies we get an error of
  - $\|f_i\|_{L_p} = 0$ if all frequencies are sampled

Main Bounds

- For each i, we now sample frequencies from S_i, either
  - All frequencies in S_i
  - Randomly sample $k_i > 1$ frequencies in S_i
  - None of the frequencies in S_i
- The frequencies in S_i form a bounded orthogonal system, so that if f_i vanishes at these frequencies we get an error of
  - $\|f_i\|_{L_p} = 0$ if all frequencies are sampled
  - $$\begin{aligned}\|f_i\|_{L_p} \leq \|f_i\|_{L_1}^{2/p-1}\|f_i\|_{L_2}^{2-2/p} &\leq C[\mu_i \log(\mu_i)^{5/2}]^{2-2/p}\|f_i\|_{L_1} \\ &\leq C\mu_i^{2(1-1/p)}\log(\mu_i)^{5(1-1/p)}2^{-is}\end{aligned} \tag{22}$$

    if k_i frequencies are sampled, where $\mu_i = \sqrt{\frac{2^{id}}{k_i}(\log k_i)}$

Main Bounds

- For each i, we now sample frequencies from S_i, either
  - All frequencies in S_i
  - Randomly sample $k_i > 1$ frequencies in S_i
  - None of the frequencies in S_i
- The frequencies in S_i form a bounded orthogonal system, so that if f_i vanishes at these frequencies we get an error of
  - $\|f_i\|_{L_p} = 0$ if all frequencies are sampled
  - $$\begin{aligned}\|f_i\|_{L_p} \leq \|f_i\|_{L_1}^{2/p-1}\|f_i\|_{L_2}^{2-2/p} &\leq C[\mu_i \log(\mu_i)^{5/2}]^{2-2/p}\|f_i\|_{L_1} \\ &\leq C\mu_i^{2(1-1/p)}\log(\mu_i)^{5(1-1/p)}2^{-is}\end{aligned} \tag{22}$$

    if k_i frequencies are sampled, where $\mu_i = \sqrt{\frac{2^{id}}{k_i}(\log k_i)}$
  - $\|f_i\|_{L_p} \leq C2^{id(1-1/p)}\|f_i\|_{L_1} \leq C2^{-i(s+d(1-1/p))}$ if no frequencies are taken

Optimal sampling strategy

- Based on the previous estimates, we optimize the sampling strategy as follows:
  - Choose all frequencies up to level i_0
  - Above i_0 select $k_i = 2^{i_0 d} 2^{-\alpha(i-i_0)}$ frequencies until $k_i < 2$
    - Here $0 < \alpha$ and $(d+\alpha)(1-1/p) < s$

Optimal sampling strategy

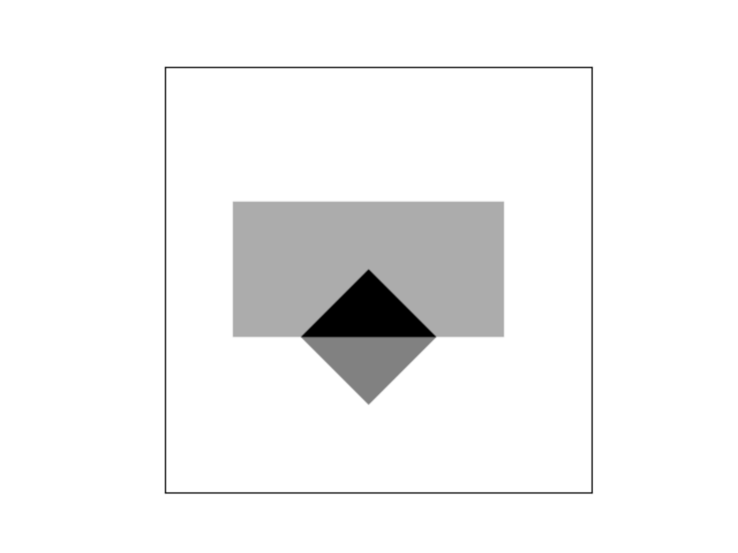
- Based on the previous estimates, we optimize the sampling strategy as follows:
  - Choose all frequencies up to level i_0
  - Above i_0 select $k_i = 2^{i_0 d} 2^{-\alpha(i-i_0)}$ frequencies until $k_i < 2$
    - Here $0 < \alpha$ and $(d+\alpha)(1-1/p) < s$
- Putting together the previous bounds, we get

$$\|f\|_{L_p} \leq \sum_{i=1}^{\infty} \|f_i\|_{L_p} \leq C 2^{-i_0 s} i_0^{(1-1/p)} \log(i_0)^{5(1-1/p)} \tag{23}$$

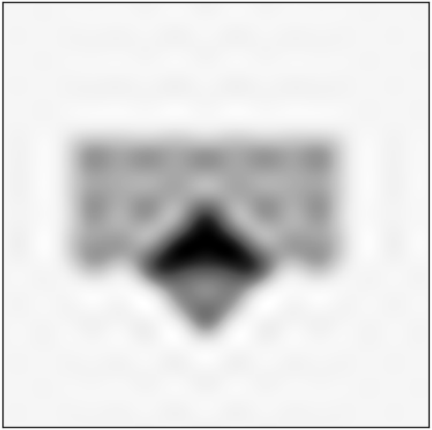
- Total number of Fourier measurements: $n \leq C 2^{i_0 d}$, so

$$s_n^G(K_1^s)_{L_p} \leq C n^{-s/d} \log(n)^{(1-1/p)} \log(\log(n))^{5(1-1/p)}. \tag{24}$$

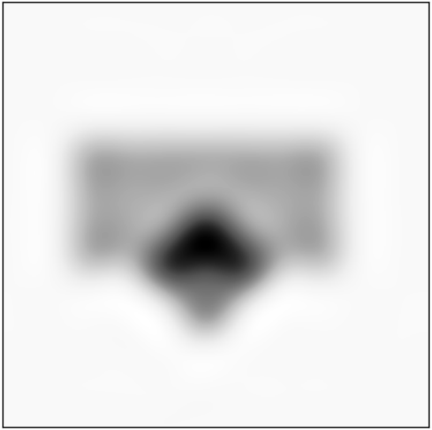
Ground Truth



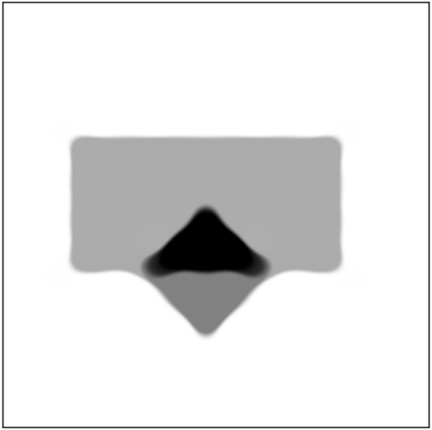
Fourier Sum (289 lowest frequencies)



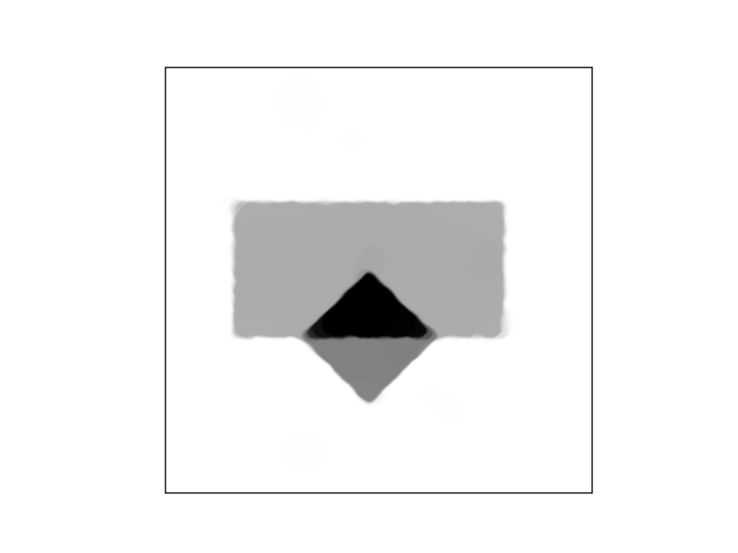
Smoothed Fourier Sum (289 lowest frequencies)



BV-norm Minimizer (289 lowest frequencies)



BV-norm Minimizer (289 hierarchically random)



Conclusion

- Non-linear compressive sampling is possible from Fourier measurements
- Open Problems:
  - What about Radon measurements?
  - What about noisy measurements?
  - What about other (even non-linear) measurements such as the magnitude of the Fourier coefficients, etc.

Happy Birthday Albert!