A FAST AND ROBUST PARADIGM FOR FOURIER COMPRESSED SENSING BASED ON CODED SAMPLING

Frank Ong[1], *Reinhard Heckel*[2], *Kannan Ramchandran*[1]

[1]Department of Electrical Engineering and Computer Sciences, University of California, Berkeley
[2]Department of Electrical and Computer Engineering, Rice University

ABSTRACT

First-order gradient methods are commonly used for compressed sensing reconstruction. However, for Fourier sampling systems, they require computing a large number of fast Fourier transforms (FFTs), which can be expensive in real-time applications. In this paper, instead of random sub-sampling, we use a sampling scheme inspired by coding theory from a recent sparse-FFT work of Pawar and Ramchandran [1]. In particular, we show that Iterative Soft Thresholding Algorithm (ISTA) applied on the Least Absolute Shrinkage and Selection Operator (LASSO) with the coded sampling provides an $O(\log n)$ per-iteration speedup over the standard iteration cost, where n is the signal length. Since the coded sampling operation deviates from the common randomized compressed sensing sampling, it is a priori unclear whether LASSO can recover sparse signals. We provide recovery guarantees for LASSO using the coded sampling guaranteed for an arbitrary signal-to-noise ratio. For a k-sparse signal and under a uniformly random sparsity model, we show that LASSO recovers the underlying signal from $O(k \log^4 n)$ measurements through the coded sensing system, with a reconstruction error that is proportional to the sparsity level and noise energy. Moreover, we demonstrate numerically computational speedups for using this scheme as well as lower MRI acquisition times.

Index Terms— Compressed sensing, Coded sampling, LASSO, FFAST, MRI

1. INTRODUCTION

In a variety of imaging applications, including magnetic resonance imaging (MRI), optical imaging, and astronomical imaging, images can be sparsely represented in a transform domain, and observed through the Fourier domain. Recent results in compressed sensing [2, 3] enable us to exploit this sparse structure to acquire and reconstruct signals from far fewer Fourier measurements than required by the Shannon-Nyquist theorem. In particular, the Least Absolute Shrinkage and Selection Operator (LASSO) [4] is commonly used to reconstruct sparse signals from randomly undersampled Fourier measurements. Algorithms for solving LASSO typically iteratively alternate between the spatial domain representation and the Fourier domain representation of the signals, and as a consequence perform a large number of fast Fourier transforms (FFTs). As a result, compressed sensing is challenging to apply in devices and acquisition systems demanding inexpensive, low-power or real-time signal analysis.

In an effort to make those fast Fourier transform steps more efficient, a number of works [5, 6, 7, 8, 9, 10, 11, 1] have proposed to compute a sparse discrete Fourier transform (DFT) with low sampling and computational complexity. In particular, an algorithm named FFAST (Fast Fourier Aliasing-based Sparse Transform) [1] was proposed to use structured uniform undersampling inspired by coding theory that enables us to exploit this structure for reconstruction speed. Albeit very fast, the sub-linear time FFAST framework is restricted to relatively high signal-to-noise ratio settings [12]. In this work, we remove this restriction, targeting any signal-to-noise ratio while reducing the reconstruction time to be linear in the ambient dimension.

Specifically, we propose to collect measurements via FFAST sampling, and reconstruct the signal by solving the LASSO-optimization problem with the Iterative Soft Thresholding Algorithm (ISTA) [13], and its accelerated version, the Fast Iterative Soft Thresholding Algorithm (FISTA) [14]. We achieve a $O(\log n)$ speedup per-iteration over conventional random Fourier sampling, where n is the signal length. We provide experiments that demonstrate a near 2x-speedup in compressed sensing MRI applications [15].

Since the FFAST sampling operation deviates from the common randomized compressed sensing sampling, it is a priori not clear whether LASSO can recover sparse signals from FFAST measurements. We provide recovery guarantees for LASSO using FFAST sampling. We show that for a k-sparse signal, under a uniformly random sparsity model, LASSO recovers the underlying signal from $O(k \log^4 n)$ measurements through the FFAST sensing system, with reconstruction error proportional to the sparsity level, and the noise energy.

Combining the results in this work, and the results in [1], the FFAST sensing system enables the following dual reconstruction scheme, illustrated in Figure 1: When the signal-to-noise is high, then the peeling decoder in [1] can be used to deliver the reconstruction result quickly with computational complexity $O(k \log^4 n)$. On the other hand, when the signal-to-noise is low, then LASSO can be used to deliver a more reliable reconstruction with per-iteration computational complexity of $O(n)$.

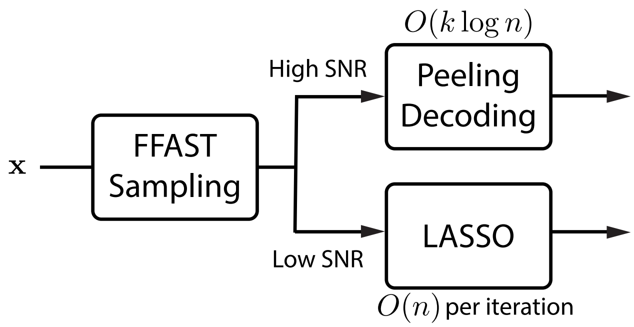

Fig. 1. Two reconstruction backends can be used for the FFAST sensing system. The peeling decoder proposed in [12] is fast, but fragile in low SNR. In this work, we show that FFAST sensing system also speeds up LASSO reconstruction while providing recovery guarantees for all SNR.

2. FFAST SENSING SYSTEM

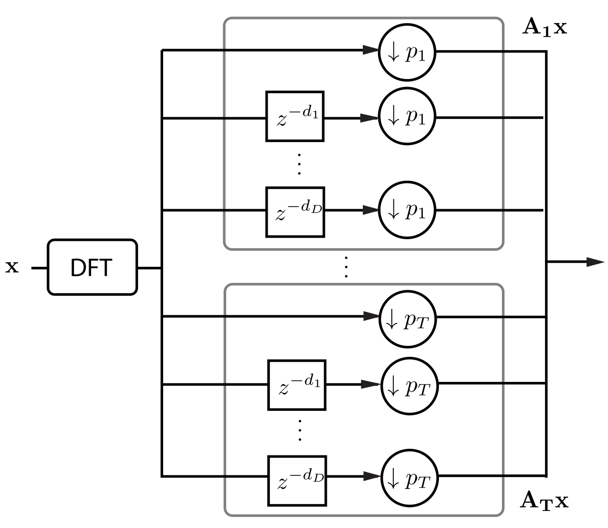

Fig. 2. Illustration of the FFAST sensing system, which consists of a DFT followed by a sequence of delays and uniform subsampling. FFAST sampling consists of T stages. Each t stage uniformly subsamples the DFT of the sparse signal with factor p_t, and random delays $\{d_1, \ldots, d_D\}$.

The FFAST sensing system was originally introduced and designed in Pawar and Ramchandran [1] using ideas from coding theory. Here we describe a perspective of the FFAST sensing system suitable to this paper, and refer the reader to their work for a different perspective.

Our goal is to design a Fourier sampling system for compressed sensing that induces simple structures in the signal domain. In the context of iterative methods, this enables us to efficiently go back and forth between the Fourier and signal domain. One way to achieve this is to use only stages of uniform sampling and delay, because both operations in Fourier domain induce computational efficient operations in the signal domain: uniform sampling in the Fourier domain induces aliasing in the signal domain, and delay in the Fourier domain induces linear phase shift in the signal domain. If we are able to use these two properties appropriately, we can bypass FFTs while still iteratively go between Fourier measurements and reconstructed signal.

Under the constraint of using only uniform sampling and delay as components, how can we ensure that the sensing system is suitable for compressed sensing? There are only two parameters we can design: the sampling factors, and the delay factors. For sampling factors, we note that there are certain subsampling combinations that are not suitable for compressed sensing. For example, a sampling scheme that subsamples by 2 and 4 violates the Nyquist theorem. To avoid these situations, we require the subsampling factors to be co-prime. That is, we consider undersampling factors $p_1, \ldots, p_T$, such that $\gcd(p_i, p_j) = 1$ if $i \neq j$. One consequence of using co-prime subsampling factors is that the aliasing induced in the signal domain does not overlap between different stages. As for the delay factors, we appeal to conventional compressed sensing principle and use a random sequence of delay factors, denoted as $\{d_1, \ldots, d_D\}$. Putting these design choices together, we obtain the FFAST sensing system illustrated in Figure 2.

Concretely, we consider a s-sparse signal $\mathbf{x}$ of length n, that is divisible by the co-prime under-sampling factors $p_1, \ldots, p_T$. Then let $W_n = e^{-\imath 2\pi/n}$, $\mathbf{A}_t \in \mathbb{C}^{nD/p_t \times n}$ be uniformly subsampled discrete Fourier matrices with subsampling factor p_t, and delays $\{d_1, \ldots, d_D\}$. Specifically for $i = 0, \ldots, n/p_t - 1$ and $j = 0, \ldots, n-1$,

$$\begin{bmatrix} [\mathbf{A}_t]_{(ip_t)j} \\ [\mathbf{A}_t]_{(ip_t+1)j} \\ \vdots \\ [\mathbf{A}_t]_{(ip_t+D-1)j} \end{bmatrix} = \frac{1}{\sqrt{n\kappa D}} \begin{bmatrix} W_n^{(ip_t+d_1)j} \\ W_n^{(ip_t+d_2)j} \\ \vdots \\ W_n^{(ip_t+d_D)j} \end{bmatrix}$$

where $\kappa := \sum_{t=1}^{T} 1/p_t$ is a normalization factor ensuring that the columns of the sensing matrix $\mathbf{A}$ defined below have unit norm. Then the signal model for the measured data is

$$\mathbf{y} = \mathbf{A}\mathbf{x}, \quad \mathbf{A} = \begin{bmatrix} \mathbf{A}_1 \\ \vdots \\ \mathbf{A}_T \end{bmatrix}.$$

We define $m = n\kappa D$ as the number of measurements, so that the matrix $\mathbf{A}$ is of size $m \times n$.

3. MAIN RESULTS

We consider the LASSO estimator to reconstruct our signal from the measurement $\mathbf{y}$:

$$\hat{\mathbf{x}} \in \arg\min_{\tilde{\mathbf{x}}} \frac{1}{2}\|\mathbf{A}\tilde{\mathbf{x}} - \mathbf{y}\|_2^2 + \lambda\|\tilde{\mathbf{x}}\|_1.$$

Here, $\mathbf{A}$ is the measurement matrix pertaining to the FFAST sensing system.

Amongst the fastest schemes to numerically solve this convex optimization problem are iterative first-order proximal gradient methods such as ISTA and FISTA. Specifically, ISTA performs the following update step:

$$\mathbf{x} \leftarrow \mathcal{T}_{\alpha\lambda}(\mathbf{x} - \alpha\mathbf{A}^H(\mathbf{A}\mathbf{x} - \mathbf{y})),$$

where $\mathcal{T}_\lambda$ is the soft thresholding operator, an element-wise operation defined as $[\mathcal{T}_\lambda(\mathbf{x})]_i = (|x_i| - \lambda)\mathrm{sgn}(x_i)$

Proposition 1. *Let $\mathbf{A}$ be the measurement matrix from the FFAST sensing system in Section 2. Then first-order gradient methods ISTA and FISTA have a one-time cost of computing $\mathbf{A}^H\mathbf{y}$ with complexity $O(n\log n)$ and per-iteration computational complexity of $O(n)$.*

In contrast, ISTA and FISTA with conventional Fourier compressed sensing system have the same one-time cost and per-iteration computational complexity of $O(n\log n)$ since they have to perform a FFT in each iteration. Hence, for a large number of iterations, the FFAST sensing system offers a $\log(n)$ speedup. Proposition 1 follows from normal matrix $\mathbf{A}^H\mathbf{A}$ being sparse, which results in low iteration complexity (see Section 4 for details).

Since the FFAST sensing system no longer follows the random undersampling scheme suggested by compressed sensing theory, it is natural to ask whether LASSO is still able to recover the underlying sparse signal given FFAST measurements. The following theorem answers this question in the affirmative.

Theorem 1. *Let $\mathcal{S}$ be a support set of cardinality s chosen uniformly at random from $[n] := \{1,\dots,n\}$, let $\mathbf{x} \in \mathbb{C}^n$ be supported on $\mathcal{S}$, and assume that the signs of the non-zero values of $\mathbf{x}$ are chosen uniformly at random. Consider $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{e}$, with $\mathbf{A}$ as described in Section 2 with delays $D \geq \log^2(n)$, and $\mathbf{e}$ is additive noise obeying $\|\mathbf{e}\|_2 \leq \eta$. Suppose that the number of measurements m satisfies,*

$$m \geq c_1 s \log^4(n). \tag{1}$$

Then, for an appropriately chosen regularization parameter λ, with probability at least $1 - 2n^{-4}$, we have that

$$\|\hat{\mathbf{x}} - \mathbf{x}\|_2 \leq c_2\eta\sqrt{s}.$$

Here, c_1 and c_2 are numerical constants.

For the noiseless case, in which $\eta = 0$, Theorem 1 guarantees that $\mathbf{x}$ is the unique solution to ℓ_1-minimization. In the noisy case, it guarantees stable recovery, even if the noise is chosen adversarially. An outline of Theorem 1 is provided in Section 5.

4. PROOF OF PROPOSITION 1

Considering the ISTA update step, we note that the term $\mathbf{A}^H\mathbf{y}$ can be computed one time and reused for each iteration. In particular, $\mathbf{A}^H\mathbf{y}$ can be computed using the FFT in $O(n\log n)$ time.

The $O(n)$ per-iteration complexity comes from the fact that the normal matrix $\mathbf{A}^H\mathbf{A}$ is sparse. To see this, we first note that the normal matrix for each stage t is given by:

$$[\mathbf{A}_t^H\mathbf{A}_t]_{ij} = \begin{cases} \frac{1}{\kappa p_t D}\sum_{k=1}^{D} W_n^{(i-j)d_k} & \text{if } \mathrm{mod}(i-j, \frac{n}{p_t}) = 0, \\ 0 & \text{else} \end{cases}.$$

The number of non-zero elements of each row of $\mathbf{A}_t^H\mathbf{A}_t$ is given by $|\{i : \mod(i, n/p_t) = 0, i = 0,\dots,n-1\}| = p_t$.

Since $\{p_1,\dots,p_T\}$ are co-prime, the off diagonals of the matrices $\{\mathbf{A}_t^H\mathbf{A}_t\}_{t=1}^T$ are disjoint. Hence, the overall normal matrix, which is given by, $\mathbf{A}^H\mathbf{A} = \sum_{t=1}^T \mathbf{A}_t^H\mathbf{A}_t$, has $n\sum_{t=1}^T p_t$ non-zero elements.

Under the assumption that the undersampling factors do not grow with n, then a matrix vector product with $\mathbf{A}^H\mathbf{A}$ costs $O(n)$ operations. Since the soft-thresholding operation also costs $O(n)$, the overall per-iteration complexity is $O(n)$.

5. PROOF OUTLINE OF THEOREM 1

In this section, we outline the proof for the recovery guarantees stated in Theorem 1. For ease of analysis, throughout this section, we consider the constrained form of the LASSO estimator:

$$\hat{\mathbf{x}} \in \arg\min_{\mathbf{x}} \|\mathbf{x}\|_1 \quad \text{subject to} \quad \|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2 \leq \eta.$$

Note that for λ chosen appropriately as a function of η, the constrained LASSO formulation and the original one are equivalent.

The proof of Theorem 1 relies on the following proposition.

Proposition 2. *Let* $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{e}$*, where* $\mathbf{e}$ *is additive noise obeying* $\|\mathbf{e}\|_2 \leq \eta$*, and* $\mathbf{x}$ *is* s*-sparse and supported on* $\mathcal{S}$*, and has random signs. Suppose that*

$$\left\|\mathbf{A}_{\mathcal{S}}^H \mathbf{A}_{\mathcal{S}} - \mathbf{I}\right\| \leq \frac{1}{2}, \tag{2}$$

and that, for some $\alpha \geq 0$*,*

$$\left\|\mathbf{A}_{\mathcal{S}}^H \mathbf{a}_\ell\right\|_2 \leq \alpha/2, \quad \text{for all } \ell \notin \mathcal{S}. \tag{3}$$

Here, $\mathbf{A}_{\mathcal{S}}$ *is the sub-matrix of* $\mathbf{A}$ *with columns in* S*, and* $\mathbf{a}_\ell$ *is the* ℓ*-th column of* $\mathbf{A}$*. Then, with probability at least* $1 - 2Ne^{-\alpha^{-2}/2}$*, we have that*

$$\|\hat{\mathbf{x}}_{lasso} - \mathbf{x}\|_2 \leq c_1 \eta \sqrt{s}.$$

Proof. Proposition 2 follows from a standard recovery condition in the theory of compressive sensing. Specifically, it follows from Theorem [16, Thm. 4.33] by setting $\mathbf{h} = (\mathbf{A}_{\mathcal{S}}^\dagger)^H \mathrm{sign}(\mathbf{x}_{\mathcal{S}})$. □

Theorem 1 now follows from proposition 2 by establishing that, provided that $\mathcal{S}$ is a support set chosen uniformly at random from $[n]$, and condition (1) holds true, then

$$\mathrm{P}\left[\left\|\mathbf{A}_{\mathcal{S}}^H \mathbf{A}_{\mathcal{S}} - \mathbf{I}\right\| \geq 1/2\right] \leq n^{-4} \tag{4}$$

and, for any $\ell \notin \mathcal{S}$,

$$\mathrm{P}\left[\left\|\mathbf{A}_{\mathcal{S}}^H \mathbf{a}_\ell\right\|_2 \geq 2\alpha\right] \leq n^{-5}. \tag{5}$$

By the union bound, this implies that conditions (2) and (3) hold with $\alpha = 1/\sqrt{10\log(n)}$ with probability at least $1 - 2n^{-5}$, as desired. The remainder of the proof is devoted to proving inequalities (4) and (5).

5.1. Proof sketch of inequality (4)

We use Geshgorin's disk theorem to obtain a lower bound on the smallest and largest eigenvalue of $\mathbf{A}_{\mathcal{S}}^H \mathbf{A}_{\mathcal{S}}$, which in turn yields an upper bound on $\left\|\mathbf{A}_{\mathcal{S}}^H \mathbf{A}_{\mathcal{S}} - \mathbf{I}\right\|$. In order to apply Geshgorin's disk theorem, consider any row of the $s \times s$ matrix $\mathbf{A}_{\mathcal{S}}^H \mathbf{A}_{\mathcal{S}}$, and note that the off-diagonal elements of that row consist of $s-1$ random entries of the off-diagonals of a row of $\mathbf{A}^H \mathbf{A}$. The sum of its off-diagonal elements therefore concentrates around its average using Bernstein's inequality. The off diagonal average is small when the number of delays is on the order of $\log^2 n$, which can be shown using Hoeffding's inequality.

5.2. Proof sketch of inequality (5)

We next upper-bound the term $\left\|\mathbf{A}_{\mathcal{S}}^H \mathbf{a}_\ell\right\|_2$. Fix $\ell \notin \mathcal{S}$. Recall that $\mathcal{S}$ is chosen uniformly at random. Thus, $\mathbf{A}_{\mathcal{S}}^H \mathbf{a}_\ell$ is the sum of $s = |\mathcal{S}|$ many entries chosen uniformly at random from the squared off-diagonal elements of any row of $\mathbf{A}^H \mathbf{A}$ (recall that $\mathbf{A}^H \mathbf{A}$ is a circulant matrix, and thus the *set* of off-diagonal elements in each row is equivalent). In order to upper-bound the probability of $\left\|\mathbf{A}_{\mathcal{S}}^H \mathbf{a}_\ell\right\|_2$ exceeding a certain value, we again apply Bernstein's inequality.

6. EXPERIMENTS

In this section, we provide empirical experiments on a real MRI dataset, to demonstrate the feasibility of using FFAST sampling with LASSO reconstruction. Since images are sparse in the wavelet domain, one difference in LASSO between the experiment here and previous sections is that we impose the ℓ_1 norm on the wavelet transform. In addition, the multi-channel acquisition MRI model is also incorporated in the measurement matrix.

Figure 3 compares the LASSO-FISTA reconstruction using the FFAST sampling and Poisson-disk sampling, a commonly used MRI sampling pattern, on a 8-channel 2D axial brain scan. The overall undersampling factor is 5.61. FISTA was run for 100 iterations. The run time for FFAST sampling and Poisson-disk sampling are 19 s, and 37 s respectively. The image with FFAST sampling achieved a PSNR of 31.60 dB, and with Poisson-disk sampling achieved a PSNR of 32.46 dB. While the PSNR with the FFAST sampling is slightly lower than that with Poisson-disk sampling, visually the images look very similar.

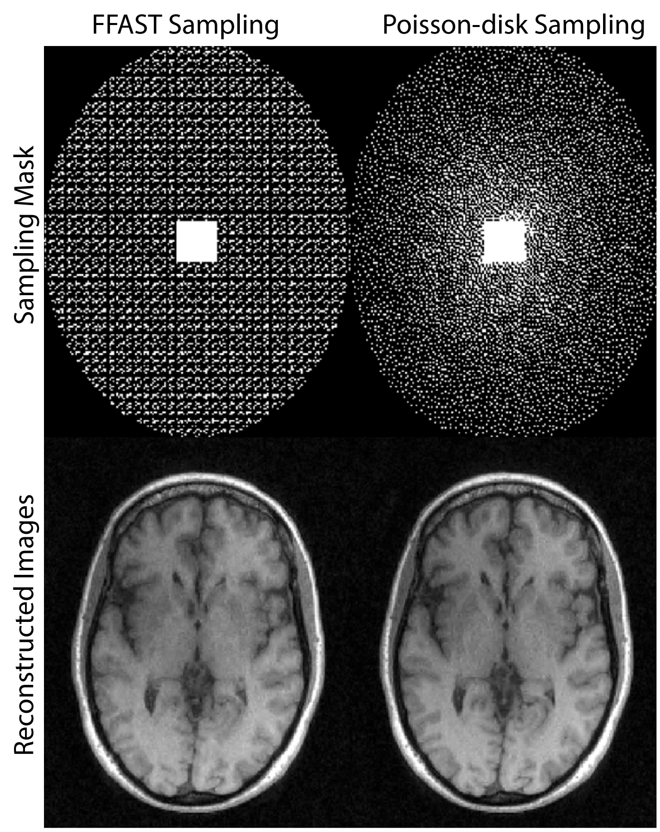

Fig. 3. Comparison LASSO reconstruction between FFAST and Poisson-disk sampling, a commonly used MRI sampling pattern, on a brain image. The overall acceleration factor is 5.61. The run time for FFAST sampling and Poisson-disk sampling are 19 s, and 37 s respectively.

7. REFERENCES

[1] Sameer Pawar and Kannan Ramchandran, "FFAST: an algorithm for computing an exactly k -Sparse DFT in $O(k \log k)$ Time," *IEEE Transaction on Information Theory*, vol. 64, no. 1, pp. 429–450, 2018.

[2] Emmanuel J Candès, Justin Romberg, and Terence Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Transactions on Information Theory*, vol. 52, no. 2, pp. 489–509, 2006.

[3] David L Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[4] Robert Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 267–288, 1996.

[5] Anna C Gilbert, Sudipto Guha, Piotr Indyk, S Muthukrishnan, and Martin Strauss, "Near-optimal sparse fourier representations via sampling," in *Proceedings of the thiry-fourth annual ACM symposium on Theory of computing*. ACM, 2002, pp. 152–161.

[6] Anna C Gilbert, S Muthukrishnan, and Martin Strauss, "Improved time bounds for near-optimal sparse fourier representations," in *Wavelets XI*. International Society for Optics and Photonics, 2005, vol. 5914, p. 59141A.

[7] Anna C Gilbert, Martin J Strauss, and Joel A Tropp, "A Tutorial on Fast Fourier Sampling," *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 57–66, Mar. 2008.

[8] Anna C Gilbert, Piotr Indyk, Mark Iwen, and L Schmidt, "Recent Developments in the Sparse Fourier Transform: A compressed Fourier transform for big data," *IEEE Signal Processing Magazine*, vol. 31, no. 5, pp. 91–100, Sept. 2014.

[9] Haitham Hassanieh, Piotr Indyk, Dina Katabi, and Eric Price, "Nearly optimal sparse fourier transform," in *the 44th symposium*, New York, New York, USA, May 2012, pp. 563–578, ACM.

[10] H Hassanieh, P Indyk, D Katabi, and E Price, "Simple and practical algorithm for sparse Fourier transform," *Proceedings of the twenty-third annual ACM-SIAM symposium on Discrete Algorithms*, 2012.

[11] P Indyk, M Kapralov, and E Price, "(Nearly) sample-optimal sparse Fourier transform," in *Proceedings of the Twenty-Fifth Annual ACM-SIAM Symposium on Discrete Algorithms*. 2014, pp. 480–499, SIAM.

[12] Sameer Pawar and Kannan Ramchandran, "R-FFAST: a robust Sub-Linear time algorithm for computing a sparse DFT," *IEEE Transaction on Information Theory*, vol. 64, no. 1, pp. 451–466, 2018.

[13] Ingrid Daubechies, Michel Defrise, and Christine De Mol, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," *Communications on Pure and Applied Mathematics: A Journal Issued by the Courant Institute of Mathematical Sciences*, vol. 57, no. 11, pp. 1413–1457, 2004.

[14] Amir Beck and Marc Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM journal on imaging sciences*, vol. 2, no. 1, pp. 183–202, 2009.

[15] Michael Lustig, David Donoho, and John M Pauly, "Sparse MRI: The application of compressed sensing for rapid MR imaging," *Magnetic Resonance in Medicine*, vol. 58, no. 6, pp. 1182–1195, 2007.

[16] Simon Foucart and Rauhut, Holger, *A Mathematical Introduction to Compressive Sensing*, Springer Berlin Heidelberg, 2013.