A Robust FFAST Scheme with Continuous Alphabet

Kabir Chandrasekher

Abstract

In this paper, we consider the sparse spectrum recovery problem in the presence of noise. Specifically, we consider a n length time domain signal x with k-sparse spectrum X and use a specific subsampling scheme to sample the time domain signal x. We extend the algorithms and methods of Pawar and Ramchandran [1] to the case where the DFT coefficients come from a continuous alphabet. To this end, the R-FFAST scheme induces a sparse-graph code, and recovery is done using the truncated peeling method introduced by Yin et. al [2] and the notion of bin peeling, which we introduce in this report. We maintain the decoding and sample complexity of the original work, using $D = \mathcal{O}(\log n \log^2 k)$ random delays for the procedure introduced in Pawar [3], noting that the same proofs extend easily to the case where frequency estimation is done using Kay's method instead of an exhaustive search.

1 Introduction

In this section, we review the R-FFAST method and outline the common notation. Consider a time domain signal x corrupted by noise such that we sample the noise corrupted version $y = x + w$, where the noise is assumed to be $\mathcal{CN}(0, I_{0,I_{n\times n}})$. We note that in the original work, it is assumed that the DFT coefficients are of the form $Ae^{j\phi}$, where $A = \sqrt{\rho}$ (ρ is a targeted SNR) and $\phi \in \{\frac{2\pi i}{M}\}_{i=0}^{M-1}$. We relax the latter assumption in this report, assuming ϕ is an arbitrary phase, rather than restricted to a discrete alphabet. The main result of the paper mimics that of Pawar and Ramchandran [1]:

Theorem 1.1. *Suppose X is the k-sparse DFT of an n-length signal x. The modified R-FFAST algorithm recovers a fraction $1-\delta$ of the non-zero DFT coefficients using $\mathcal{O}(k \log n \log^2 k)$ samples and time complexity $\mathcal{O}(n \log n \log^2 k)$ with probability at least $1 - \mathcal{O}(exp(-c_3(\delta)k^{c_2(\delta)})) + \mathcal{O}(\frac{1}{k})$.*

The proof is given in Section 3; note that the proofs given are sufficiently general to cover the case when frequency estimation is done using Kay's method rather than by exhaustive search, although for brevity we omit this proof. We now briefly outline the R-FFAST methodology. The idea is inspired by the original work of Pawar and Ramchandran [4] for computing a noiseless sparse DFT using sparse-graph codes. The main idea of this work is to decompose the signal length n into coprime factors $f_1, f_2, \dots, f_d$ such that $\prod_{i=1}^{d} f_i = n$ and subsample the original signal by each factor. Additionally, the signal is delayed and subsampled by each factor again. The decoder then takes short DFTs of each of these shorter signals and uses a ratio test to determine whether observations are zero-tons, singletons, or multi-tons. When the samples are corrupted by noise, the ratio test no longer suffices as it requires infinite precision. To cope with this, a number of random delays D are required. The subsampling scheme induces bin observations of the form $y_b = A_{i,j}X + w_b$, where $0 \leq i < d$ and $0 \leq j < D$ (for bin j in stage i). Here, $w_b \sim \mathcal{CN}(0, I_{D\times D})$ and we denote $A_{i,j}$ the bin measurement matrix. Additionally, we introduce the steering vector of ball ℓ, $a(\ell) = (e^{j2\pi\ell r_0/n}, e^{j2\pi\ell r_1/n}, \dots, e^{j2\pi\ell r_{D-1}/n})$. We now construct $A_{i,j}$ such that $A_{i,j}(\ell) = a(\ell)$ if $\ell \equiv j \mod f_i$ and 0 otherwise.

2 Algorithmic Methods

2.1 Truncated Peeling

In this section, we briefly review the peeling with truncation algorithm, as developed in Yin et. al [2]. We first describe the algorithm, including an example, and then detail the technical result, which is Lemma 2 in [2]. The idea of truncated peeling is very simple. Consider the R-FFAST sparse-graph construction, where there are k balls, $\mathcal{O}(k)$ bins, and each ball is thrown to d bins. It is possible in this construction that a particular bin contains $\mathcal{O}(k)$ contributions, in which case the error propagation is $\mathcal{O}(k)$ when peeling occurs. In the truncated peeling strategy, this propagation is limited by allowing each bin to peel up to a maximum L amount of contributions. Once L contributions are peeled, the bin is no longer used. An example of a decoder using the truncated peeling strategy is shown in Figure 1. The main technical lemma is the following:

Lemma 2.1. *Assume that we can always find the correct location indices of singleton balls. For any $\delta > 0$, when k is large enough, there exist proper parameters d and $M = \Theta(\log(\frac{1}{\delta})k)$, such that after $N(\delta)$ iterations of truncation peeling, the fraction of non-zero signal elements which are not detected is less than δ, with probability $\mathcal{O}(\exp\{-c_3(\delta)K^{c_4(\delta)}\})$. Here, $c_3(\delta), c_4(\delta) > 0$ are two quantities determined by δ.*

The proof of this Lemma is given in Appendix B of Yin et. al [2]. We omit it here, as this result remains unchanged in our situation.

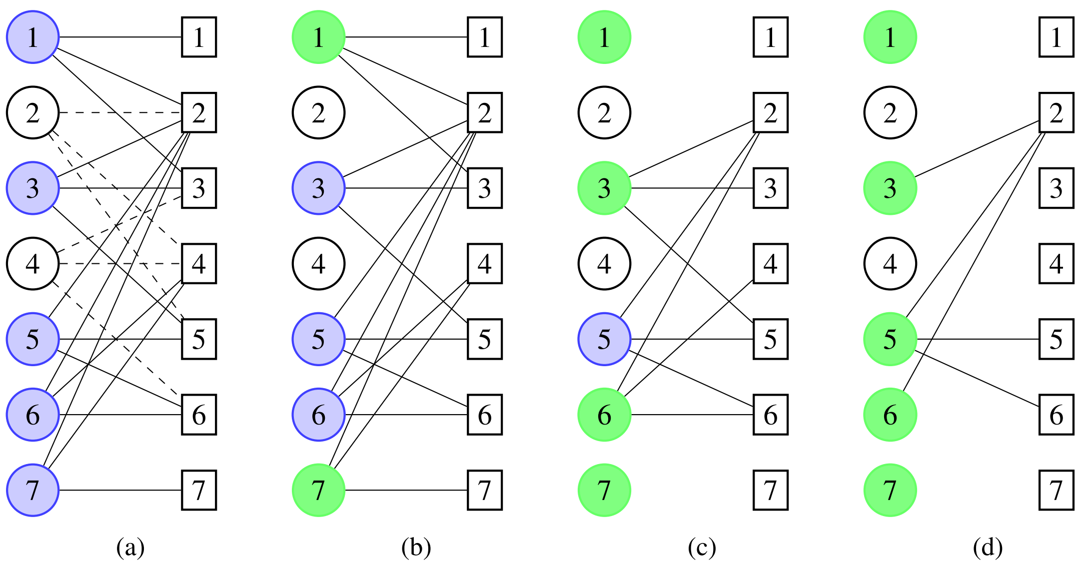

Figure 1: **Peeling with truncation.** In this example, the truncation level is set to be $L = 2$.

2.2 Bin Peeling

In this section, we introduce the notion of bin peeling and its consequences. The idea is extremely simple. Suppose that bin B is detected to be a singleton, passing the energy test. Then, estimate the value of the location of the corresponding ball and the value of the corresponding ball to be $(\hat{L}, \hat{V})$. When peeling, find the neighbors of the ball corresponding to location L, and peel the entire measurement of bin B instead of the estimated measurement $\hat{V}$. By doing so, instead of propagating the estimation error, the decoder propagates the ambient noise. A small example of this can be seen in Figure 2.

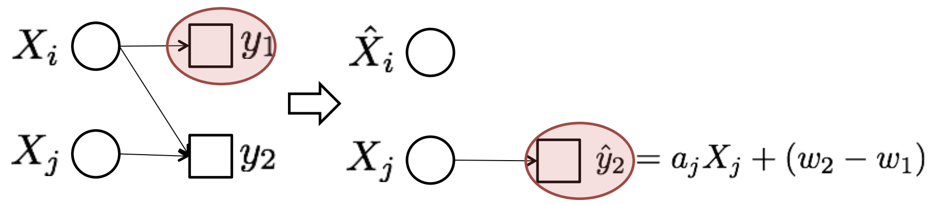

Figure 2: **Example of bin peeling.** In this example, the decoder determines that y_1 is a singleton, estimates the location and value pair $(\hat{L_{y_1}}, \hat{V_{y_1}})$, and peels off the measurement y_1 from bin y_2, leaving the estimate of ball X_i as $\hat{X}_i = \hat{L_{y_1}}$

3 Main Results

In this section, we show the main result of this report. We begin by outlining some tail bounds for χ^2 random variables that will be used repeatedly. Then, we show the reliability of bin estimation. We note that these proofs are extremely similar to those presented in Pawar and Ramchandran [1] and Pawar [3]. The proofs given were for noise distributed as $\mathcal{CN}(0, I_{D\times D})$. In the bin peeling, continuous alphabet case, we extend these proofs to the case where the noise is distributed as $\mathcal{CN}(0, \sigma^2 I_{D\times D})$.

3.1 Tail Bounds for χ^2 Random Variables

First, we state some lemmas (without proof) on concentration for χ^2 random variables:

Lemma 3.1. *Suppose X is a χ^2 random variable with k degrees of freedom and $\epsilon \in [0, 3]$, then:*

$$\mathrm{P}[X \geq k(1+\epsilon)] \leq exp\left(-\frac{k\epsilon^2}{18}\right) \tag{1}$$

Note that this follows from a bound on Lipschitz functions of Gaussian random variables and using the fact that the Euclidean norm is 1-Lipschitz. We additionally use the following lemma which is equation $(58b)$ in [5]:

Lemma 3.2. *Suppose X is noncentral χ^2 random variable with d degrees of freedom and noncentrality parameter $\nu \geq 0$. Then, for all $x > 0$, we have:*

$$\mathrm{P}[X \leq (d+\nu) - 2\sqrt{(d+2\nu)x}] \leq exp(-x) \tag{2}$$

3.2 Mutual Incoherence and Restricted Isometry Property

We now state (without proof) two useful lemmas from Pawar and Ramchandran [1] that we will make extensive use of:

Lemma 3.3. *The mutual incoherence $\mu_{max}(A_{i,j})$ of the bin-measurement matrix $A_{i,j}$, of the FFAST front-end with D delays, is upper bounded by:*

$$\mu_{max}(A_{i,j}) < 2\sqrt{\frac{\log(5n)}{D}}, \quad \forall i, j \tag{3}$$

with probability at least 0.2, where the delays $(r_0, r_1, \ldots, r_{D-1})$ are chosen uniformly at random from the set $\{0, 1, \ldots, n-1\}$.

We additionally have the following:

Lemma 3.4. *The bin-measurement matrix $A_{i,j}$, of the FFAST front-end with D delays, satisfies the following RIP condition for all X that have $\|X\|_0 \leq s$:*

$$D(1 - \mu_{max}(A_{i,j})(s-1))^{+} \|X\|^2 \leq \|A_{i,j}X\|^2 \leq D(1 + \mu_{max}(A_{i,j})(s-1)) \|X\|^2 \tag{4}$$

where the delays $(r_0, r_1, \ldots r_{D-1})$ are chosen uniformly at random from the set $\{0, 1, \ldots, n-1\}$.

3.3 Reliability Analysis

We must now show that bin processing is robust in the R-FFAST scheme. To this end, we define E_1, the error event that a bin processed by the R-FFAST scheme is decoded incorrectly. We additionally let E be the event that some bin is incorrectly decoded by the FFAST algorithm. We note that by the union bound:

$$\begin{aligned} \mathrm{P}(E) &< N(\delta) \cdot M \cdot \mathrm{P}(E_1) \\ &< \mathcal{O}\left(\frac{1}{k}\right) \end{aligned}$$

where $N(\delta)$ is the number of iterations (a constant), $M = \mathcal{O}(k)$ is the number of bins and $\mathrm{P}(E_1) < \mathcal{O}\big(\frac{1}{k^2}\big)$. This last fact remains to be shown and composes the rest of this section. We now establish the following lemma, which will be used many times in the following proofs:

Lemma 3.5. *For a complex vector $u \in \mathbb{C}^D$ and noise vector $w \sim \mathcal{CN}(0, \sigma^2 I_{D\times D})$, where σ^2 is finite, we have:*

$$\mathrm{P}(\|u+w\|^2 < (1+\gamma)D) < \mathit{exp}\left(-\frac{D\left[c_1 - \frac{\gamma - \|u\|^2}{\sigma^2 D}\right]^2}{2 + \frac{4\|u\|^2}{\sigma^2 D}}\right) \tag{5}$$

where $c_1 = \big(1 - \frac{1}{\sigma^2}\big)$.

Proof: Define $z = \frac{1}{\sigma}w$, and note that $z \sim \mathcal{CN}(0, I_{D\times D})$. Additionally, define $u^{'} = \frac{u}{\sigma}$. Thus, we see that:

$$\|u+w\|^2 = \sigma^2 \left\|u^{'} + z\right\|^2$$

And we have:

$$\begin{aligned} \mathrm{P}(\|u+w\|^2 < (1+\gamma)D) &= \mathrm{P}(\left\|u^{'} + z\right\| < \frac{(1+\gamma)}{\sigma^2}D) \\ &= \mathrm{P}(\left\|\sqrt{2}u^{'} + \sqrt{2}z\right\|^2 < \frac{2(1+\gamma)}{\sigma^2}D) \\ &< \exp\left(-\frac{D\big[c_1 - \frac{\gamma - \|u\|^2}{\sigma^2 D}\big]^2}{2 + \frac{4\|u\|^2}{\sigma^2 D}}\right) \end{aligned}$$

Where the last inequality follows directly from Lemma 3.2 $\square$

We now characterize the error events: let E_z be the event that a zero-ton is misclassified (and thus attempted to decode), E_{sz} be the event that a singleton is classified as a zero-ton, E_{ss} the event that a singleton is misclassfied as a different singleton, E_{sm} the event that a singleton is misclassified as a multi-ton bin, and E_m the event that a multi-ton bin is misclassified. With these definitions, we state and prove the following lemma:

Lemma 3.6. *The event E_1, denoting the event that a specific bin is processed incorrectly, satisfies the following bound:* $\mathrm{P}(E_1) < \mathcal{O}\big(\frac{1}{k^2}\big)$*, when the number of delays D used is* $\mathcal{O}(\log n \log^2 k)$.

Proof: We separately give bounds on the probabilities of each of the events $E_z, E_{sz}, E_{ss}, E_{sm}, E_m$ and use a union bound on each of the results. First, we show that $\mathrm{P}(E_z) < \mathcal{O}\big(\frac{1}{k^2}\big)$. We must show this for any iteration, with any amount of bins peeled, so we let the noise be $w \sim \mathcal{CN}(0, \sigma^2 I_{D\times D})$, where σ^2 is the finite variance. Now, define $z = \frac{1}{\sigma} w$ and note that $z \sim \mathcal{CN}(0, I_{D\times D}$. We are this interested in:

$$\begin{aligned}
\mathrm{P}(\|w\|^2 > (1+\gamma)D) &= \mathrm{P}(\|z\|^2 > \frac{(1+\gamma)}{\sigma^2} D) \\
&= \mathrm{P}(Z > \frac{2(1+\gamma)}{\sigma^2} D) \\
&\leq \exp\bigg(-\frac{D\big(\frac{2(1+\gamma)}{\sigma^2} - 1\big)}{18}\bigg)
\end{aligned}$$

Where we have let Z be a χ^2 random variable with $2D$ degrees of freedom, and in the last inequality, we have made use of Lemma 3.1. Note that the amount of balls peeled from a specific bin must be tracked so that the parameter γ may be tuned per bin. Additionally, one can see that if $D = \mathcal{O}(\log n \log^2 k)$, we have $\mathrm{P}(E_z) < \mathcal{O}\big(\frac{1}{k^2}\big)$. Next, we tackle E_{sz}. Letting y be the value of the specific singleton bin, we are interested in the probability $\mathrm{P}(\|y\|^2 < (1+\gamma)D)$. Thus, we have:

$$\begin{aligned}
\mathrm{P}(E_{sz}) &= \mathrm{P}(\|y\|^2 < (1+\gamma)D) \\
&= \mathrm{P}(\|X[\ell]a(\ell) + w\|^2 < (1+\gamma)D) \\
&< \exp\Bigg(-\frac{D\bigg[c_1 - \frac{\gamma - \|u\|^2}{\sigma^2 D}\bigg]^2}{2 + \frac{4\|u\|^2}{\sigma^2 D}}\Bigg)
\end{aligned}$$

where here we are using $c_1 = 1 - \frac{1}{\sigma^2}$, and have let $u = X[\ell]a(\ell)$. Now, we need to characterize the magnitude of this vector. We note that $\|a(\ell)\| = \sqrt{D}$, and let $c_2 = \|X[\ell]\|^2$. Thus, we have:

$$\mathrm{P}(E_{sz}) < \exp\Bigg(-\frac{D\big[c_1 - \frac{\gamma}{\sigma^2 D} + c_2\big]^2}{2 + \frac{4c_2}{\sigma^2}}\Bigg) \tag{6}$$

$$< \mathcal{O}\bigg(\frac{1}{k^2}\bigg) \tag{7}$$

where the last inequality follows since $D = \mathcal{O}(\log n \log^2 k)$. We now look at the event E_{ss}. Recall that this is the case where a singleton is mistaken for another singleton. Again letting y denote the measurement of the specific bin, i.e. $y = X[\ell]a(\ell) + w$, we have:

$$\begin{aligned}
\mathrm{P}(E_{ss}) &= \mathrm{P}(\big\|X[\ell]a(\ell) - X[\ell']a(\ell') + w\big\|^2 < (1+\gamma)D) \\
&= P(\|A_{i,j}v + w\|^2 < (1+\gamma)D)
\end{aligned}$$

Here, we have let $A_{i,j}$ contain $a(\ell)$ in column ℓ and $a(\ell')$ in column ℓ', and $v(\ell) = X[\ell], v(\ell') = X[\ell']$.

Making use of Lemmas 3.3, 3.4, we have:

$$\begin{aligned}\|A_{i,j}v\|^2 &\geq 2\|v\|^2 D(1-\mu_{\max}(A_{i,j}))_+ \\ &= 2c_2D(1-\mu_{\max}(A_{i,j}))_+ \\ &\geq 2c_2D(1-2\sqrt{\frac{\log(5n)}{D}})_+\end{aligned}$$

Using this and making use of Lemma 3.5, we have:

$$\mathrm{P}(E_{ss}) < \exp\left(-\frac{D\left[c_1 - \frac{\gamma}{\sigma^2 D} + \frac{2c_2}{\sigma^2}\left(1-2\sqrt{\frac{\log(5n)}{D}}\right)_+\right]^2}{2+\frac{8c_2}{\sigma^2}\left(1-2\sqrt{\frac{\log(5n)}{D}}\right)_+}\right)$$

Making note again that $D = \mathcal{O}(\log n \log^2 k)$, we have $\mathrm{P}(E_{ss}) < \mathcal{O}\left(\frac{1}{k^2}\right)$. Tackling the case where a singleton is misclassified as a multi-ton, we can see that:

$$\begin{aligned}\mathrm{P}(E_{sm}) &< \mathrm{P}(E_{sm}|\bar{E_{ss}}) + \mathrm{P}(E_{ss}) \\ &= \mathrm{P}(E_z) + \mathrm{P}(E_{ss}) \\ &< \mathcal{O}\left(\frac{1}{k^2}\right)\end{aligned}$$

where in the first inequality we have used the law of total probability and in the second inequality we have used our bounds for $\mathrm{P}(E_z)$ and $\mathrm{P}(E_{ss})$. Finally, we tackle the multiton case:

$$\begin{aligned}\mathrm{P}(E_m) &< \mathrm{P}(E_m|C<L) + \mathrm{P}(E_m|C \geq L) \\ &< \mathrm{P}(E_m|C<L)\end{aligned}$$

here we have C is the number of contributions to the specific bin and L is the maximum number of contributions under the truncated peeling strategy. Thus, we examine the term $\mathrm{P}(E_m|C<L)$. We see that the probability this bin is classified as a singleton is:

$$\begin{aligned}\mathrm{P}(E_m|C<L) &= \mathrm{P}(\|A_{i,j}v - X[\ell]a(\ell) + w\|^2 < (1+\gamma)D) \\ &< \exp\left(-\frac{D\left[c_1 - \frac{\gamma - \|u\|^2}{\sigma^2 D}\right]^2}{2+\frac{4\|u\|^2}{\sigma^2 D}}\right)\end{aligned}$$

where in the last inequality we make use of Lemma 3.5 and have let $u = A_{i,j}v - X[\ell]a(\ell)$. To finish, we examine this term:

$$\begin{aligned}\|A_{i,j}v - X[\ell]a(\ell)\|^2 &\geq Cc_2D(1-\mu_{\max}(A_{i,j})C)_+ \\ &> Cc_2D(1-2C\sqrt{\frac{\log(5n)}{D}})_+\end{aligned}$$

Thus, we have:

$$\mathrm{P}(E_m) < \mathrm{P}(E_m | C < L) \tag{8}$$

$$< \max_{2 \le C \le L} \exp\left(- \frac{D\left[c_1 - \frac{\gamma}{\sigma^2 D} + \frac{C c_2}{\sigma^2}\left(1 - 2\sqrt{\frac{\log(5n)}{D}}\right)_+ \right]^2}{2 + \frac{4 C c_2}{\sigma^2}\left(1 - 2\sqrt{\frac{\log(5n)}{D}}\right)_+} \right) \tag{9}$$

Again noting $D = \mathcal{O}(\log n \log^2 k)$, we have $\mathrm{P}(E_m) < \mathcal{O}\left(\frac{1}{k^2}\right)$. Thus, using a union bound over these error events, we see that $\mathrm{P}(E_1) < \mathcal{O}\left(\frac{1}{k^2}\right)$ and we are done $\square$.

We note that despite the length of the preceeding proof, much of the result comes out of the box from Pawar and Ramchandran [1], only needing to extend these results to the case where the noise variance is not unit, but rather some finite σ^2. The only remaining piece is to now show that under the truncated peeling strategy, the noise variance is indeed finite.

Lemma 3.7. *After $N(\delta)$ rounds of truncated peeling, with truncation level L, the maximum noise variance encountered is $\sigma^2 = L(L-1)^{N(\delta)-1}$.*

Proof (sketch): This can be seen by expanding the $N(\delta)$ neighborhood of a random bin. Noting that this neighborhood is tree-like with high probability, we consider the worst case where each right node has $\geq L$ contributions, and establish the maximum amount of peeling possible.

Now, we note that the proof of Theorem 1.1 follows immediately from Lemmas 2.1, 3.6, and 3.7.

4 Conclusions

In this report, we have shown that the R-FFAST algorithm, incorporating the elements of truncated peeling and bin peeling, is able to recover all but a fraction δ of the non-zero DFT coefficients, where the DFT coefficients come from a continuous alphabet using $D = \mathcal{O}(\log n \log^2 k)$ delays per bin. Because of the structure of the bin peeling, we were able to use many of the proofs and insights of Yin et. al [2] and Pawar and Ramchandran [1] almost as they are, only needing to adjust for the finite variance instead of unit variance case.

References

[1] S. Pawar and K. Ramchandran, "A robust sub-linear time r-ffast algorithm for computing a sparse dft," *arXiv preprint arXiv:1501.00320*, 2015.

[2] D. Yin, R. Pedarsani, X. Li, and K. Ramchandran, "Compressed sensing using sparse-graph codes for the continuous-alphabet setting," in *Communication, Control, and Computing (Allerton), 2016 54th Annual Allerton Conference on*. IEEE, 2016.

[3] S. A. Pawar, "Pulse: Peeling-based ultra-low complexity algorithms for sparse signal estimation," Ph.D. dissertation, University of California, Berkeley, 2013.

[4] S. Pawar and K. Ramchandran, "Computing a k-sparse n-length discrete fourier transform using at most 4k samples and O(klogk) complexity," in *IEEE Int. Symp. on Inf. Theory*. IEEE, 2013, pp. 464–468.

[5] M. Wainwright, "Information-theoretic bounds on sparsity recovery in the high-dimensional and noisy setting," in *2007 IEEE International Symposium on Information Theory*. IEEE, 2007, pp. 961–965.