University of Wrocław: Algorithms for Big Data (Fall'19) 02/12/2019

# Lecture 8: Sparse FFT

Lecturer: *Przemysław Uznański* Scribe: -

## 1 Sparse FFT - no noise

Assume wlog $n$ is power of two.

**Plan of solution:** Define $p_{d,\ell}(x) = \sum_{i:i \bmod 2^\ell = d} \hat{a}_i x^i$, and for short $p(x) = p_{0,0}(x) = \sum_i \hat{a}_i x^i$. and for a polynomial $f(x)$ we write for any polynomial $\|f\|_2^2 = \sum_i f_i^2$.

1. Define $S_\ell = \{i : \|p_{i,\ell}\|^2 > 0\}$
2. Given $S_\ell$, compute $S_{\ell+1}$: for each $d \in S_\ell$, test for $e \in \{d, d+2^\ell\}$ whether $\|p_{e,\ell+1}\|^2 > 0$ and if so, add $e$ to $S_{\ell+1}$.
3. $p_{i,\log n}(1) = \hat{a}_i$

The idea is to start at level 0 and proceed. The size of each $S_\ell$ is bounded by $k$, thus the total number of steps (2) done is $\mathcal{O}(k \log n)$.

**A few identities:**

By inverse Fourier transform

$$p(\omega^t)/\sqrt{n} = \frac{1}{\sqrt{n}} \sum_i \hat{a}_i \omega^{it} = a_{-t}$$

By Parseval's theorem

$$\|p\|_2^2 = \sum_i \hat{a}_i^2 = \sum_i a_i^2 = \frac{1}{n} \sum_i |p(\omega^i)|^2$$

so

$$\|p_{d,\ell}\|_2^2 = \frac{1}{n} \sum_{i=0}^{n-1} |p_{d,\ell}(\omega^i)|^2$$

Additionally

$$\begin{aligned}\frac{1}{2^\ell}\sum_{i=0}^{2^\ell-1} p(\omega^{t+i\frac{n}{2^\ell}})\omega^{-di\frac{n}{2^\ell}} &= \frac{1}{2^\ell}\sum_{i=0}^{2^\ell-1}\sum_{j=0}^{n-1} \hat{a}_j(\omega^{t+i\frac{n}{2^\ell}})^j\omega^{-di\frac{n}{2^\ell}} \\ &= \sum_{j=0}^{n-1}\hat{a}_j\omega^{tj}\frac{1}{2^\ell}\sum_{i=0}^{2^\ell-1}\omega^{i(j-d)\frac{n}{2^\ell}} \\ &= \sum_{j=0}^{n-1}\hat{a}_j\omega^{tj}[j-d=0 \bmod 2^\ell] \\ &= p_{d,\ell}(\omega^t)\end{aligned}$$

**Estimating sum via sampling:** Lets say we have $a_1, \ldots, a_n$ such that $|a_i| \le H$. We can use $\mathcal{O}(H^2/\varepsilon^2 \cdot \log n)$ samples to obtain $\pm\varepsilon$ estimate of $\frac{1}{n}\sum_i a_i$, or equivalently $\pm n\varepsilon$ estimate of $\sum_i a_i$, with proof via Hoeffding bound.

**First solution:** Denote $L = \max|\hat{a}_i|$ and $H = \max|p(\omega^t)|$. Observe that $p_{d,\ell}(\omega^t) \le H$. Since $p(\omega^t) = \sum_j \hat{a}_j\omega^{-tj}$, we have $H \le kL$.

1. Estimate $A_{d,\ell,t} \approx p_{d,\ell}(\omega^t)$ using $\mathcal{O}(H^4 \log n)$ samples of $a_{-t-i\frac{n}{2^\ell}}\omega^{-di\frac{n}{2^\ell}}$, up to error $\pm\frac{1}{16H}$.
2. $|A_{d,\ell,t}|^2 - |p_{d,\ell}(\omega^t)|^2 \le (|A_{d,\ell,t}| - |p_{d,\ell}(\omega^t)|)(|A_{d,\ell,t}| + |p_{d,\ell}(\omega^t)|) \le \frac{1}{16H}\cdot 2H \le \frac{1}{4}$
3. $\|p_{d,\ell}\|_2^2 = \frac{1}{n}\sum_{t=0}^{n-1}|p_{d,\ell}(\omega^t)|^2 = \frac{1}{n}\sum_{t=0}^{n-1}(|A_{d,\ell,t}|^2 \pm \frac{1}{4}) = (\sum_{t=0}^{n-1}|A_{d,\ell,t}|^2) \pm 1/4$
4. We have $|A_{d,\ell,t}|^2 = \mathcal{O}(H^2)$.
5. We sample $\mathcal{O}(H^4\log n)$ of $|A_{d,\ell,t}|^2$ to estimate $\|p_{d,\ell}\|_2^2$ up to $\pm\frac{1}{2}$.
6. In total $\mathcal{O}(H^8\log^2 n) = \mathcal{O}(k^8L^8\log^2 n)$ samples to compute such estimate.

Applying to our tree-traversal, we get $\mathcal{O}(k^9L^8\log^3 n)$ complexity. Once we have the indices of non-zero coefficients, we extract the exact values:

$$\hat{a}_d = p_{d,\log n}(\omega^0) = \frac{1}{n}\sum_{i=0}^{n-1} a_i\omega^{di}$$

via sampling.

**Second solution:** Let $T$ be a support of $\hat{a}_i$, that is $i \in T$ iff $\hat{a}_i \ne 0$. Using $\mathcal{O}(k^8\log^2 n)$ samples we estimate $\|p_{d,\ell}\|_2^2$ up to $\pm\frac{L^2}{16}$.

Using $\mathcal{O}(k^9\log^3 n)$ samples we find $T' \subseteq T$ such that if $|\hat{a}_i| \ge L/4$ then $i \in T'$, and $\hat{a}'_i = \hat{a}_i \pm L/4$, where $\hat{a}'_i$ is defined only over $T'$.

Recurse on $(\hat{a}'_i - \hat{a}_i)$. Specifically, this sequence is also $k$-sparse, and max-value is $L/2$.

This makes it so the total complexity is $\mathcal{O}(k^9\log^3 n\cdot\log_2 L)$.

# 2 Another algorithm, noisy case

## 2.1 $k = 1$

There is some heavy $\hat{a}_u$ such that $\sum_{u' \neq u} |\hat{a}_{u'}|^2 \leq \varepsilon |\hat{a}_u|^2$, for some small constant $\varepsilon$.

Idea: extract $u$ bit-by-bit.

**No noise:** If $u = 2v + b_0$ for $b_0 \in \{0, 1\}$, then $a_{n/2} = \frac{1}{\sqrt{n}} \hat{a}_u \omega^{un/2} = \frac{1}{\sqrt{n}} \hat{a}_u \omega^{vn + b_0 n/2} = \frac{1}{\sqrt{n}} \hat{a}_u \omega^{b_0 n/2} = \frac{1}{\sqrt{n}} \hat{a}_u (-1)^{b_0}$. Then, $a_0 = \frac{1}{\sqrt{n}} \hat{a}_u$. Thus we can use following test:

$$b_0 = 0 \qquad \text{iff} \qquad |a_0 - a_{n/2}| \leq |a_0 + a_{n/2}|$$

How to test for older bits? Assume wlog that $b_0 = 0$, since if $b_0 = 1$, we can always consider signal $a'$ defined as $a'_j = a_j \cdot \omega^j$, where $\hat{a}'_j = \hat{a}_{j-1}$. So $u = 4v' + 2b_1$, where $b_1 \in \{0, 1\}$. We then observe that $a_{n/4} = \frac{1}{\sqrt{n}} \hat{a}_u (-1)^{b_1}$, so the test is then

$$b_1 = 0 \qquad \text{iff} \qquad |a_0 - a_{n/4}| \leq |a_0 + a_{n/4}|$$

So we can proceed with all the bits in this manner.

**Noisy case:** Consider test for bit 0. If the noise is concentrated around $a_0$ and $a_{n/2}$, then such test fails. But we know that on average the noise is small. Thus we replace the test with a randomized one: pick $0 \leq r < n$ at random, and test:

$$b_0 = 0 \qquad \text{iff} \qquad |a_r - a_{r+n/2}| \leq |a_r + a_{r+n/2}|$$

(we can do many tests and pick majority vote) and in general

$$b_{i-1} = 0 \qquad \text{iff} \qquad |a_r - a_{r+n/2^i}| \leq |a_r + a_{r+n/2^i}|$$

of course assuming $b_0 = b_1 = .. = b_{i-2} = 0$, and changing the signal accordingly.

Why does it work?

Let $\hat{a}'$ be the output. We show that with ppb at least 3/4 there is $\|\hat{a} - \hat{a}'\|_2 \leq \varepsilon \|\hat{a} - \hat{a}^{(1)}\|_2$, where $\hat{a}^{(1)}$ is the top coefficient, so $\hat{a} - \hat{a}^{(1)}$ is the noise.

We rewrite

$$a_j = \frac{1}{\sqrt{n}} \hat{a}_u \omega^{uj} + \frac{1}{\sqrt{n}} \sum_{u' \neq u} \hat{a}_{u'} \omega^{u'j} = \frac{1}{\sqrt{n}} \hat{a}_u \omega^{uj} + \mu_j$$

so

$$a = F^{-1} \hat{a}^{(1)} + \mu$$

Looking at the error

$$\sum_{j=0}^{n-1} |\mu_j|^2 = \|\mu\|_2^2 = \|F^{-1}(\hat{a} - \hat{a}^{(1)})\|_2^2 = \|\hat{a} - \hat{a}^{(1)}\|_2^2 = \sum_{u' \neq u} |\hat{a}_{u'}|^2$$

Algorithm at single step compares $|a_k - a_\ell|$ vs $|a_k + a_\ell|$. We have

$$a_k - a_\ell = \frac{1}{\sqrt{n}} \hat{a}_u (\omega^{uk} - \omega^{u\ell}) + (\mu_k - \mu_\ell)$$

$$a_k + a_\ell = \frac{1}{\sqrt{n}} \hat{a}_u (\omega^{uk} + \omega^{u\ell}) + (\mu_k + \mu_\ell)$$

We know that $|\omega^{uk} \pm \omega^{u\ell}| \in \{0, 2\}$ so for the comparison to be done correctly, it is enough that $|\mu_k - \mu_\ell| + |\mu_k + \mu_\ell| \leq 2\frac{1}{\sqrt{n}}\hat{a}_u$, so $|\mu_k| + |\mu_\ell| \leq \frac{1}{\sqrt{n}}\hat{a}_u$.

We now have, since each index is picked with a random shift,

$$\mathbb{E}[|\mu_k|^2] = \frac{1}{n}\sum_{j=0}^{n-1} |\mu_j|^2 = \frac{1}{n}\sum_{u' \neq u} |\hat{a}_{u'}|^2 \leq \frac{1}{n}\varepsilon|\hat{a}_u|^2$$

so

$$\Pr[|\mu_k| \leq \frac{1}{2\sqrt{n}}|\hat{a}_u|] \leq \frac{\frac{1}{n}\varepsilon|\hat{a}_u|^2}{\frac{1}{4n}|\hat{a}_u|^2} = \frac{\varepsilon}{4}$$

So picking $\varepsilon = 1/2$ gives us by union bound $1/4$ ppb of success.

Now the trick is to amplify the ppb by repeating each test $\mathcal{O}(\log\log n)$ times and do the majority vote. This amplifies the ppb to $1/(4\log n)$, so by union bound the whole procedure is ok with $3/4$ ppb.

**Not covered:** how to extract the value of $\hat{a}_u$.