Deterministic Sparse FFT Algorithms

Gerlind Plonka

Institute for Numerical and Applied Mathematics
University of Göttingen

based on joint work with
Katrin Wannenwetsch (Göttingen),
Annie Cuyt, and Wen-Shin Lee (Antwerpen)

Glasgow

June 2017
Deterministic Sparse FFT Algorithms

Outline

• Introduction
• Sparse FFT: A first trial
• General multi-scale algorithm
• Adaptive approach for stable reconstruction
• Vandermonde matrices on the unit circle
• Numerical results
• Summary
Let $\mathbf{x} = (x_j)_{j=0}^{N-1} \in \mathbb{C}^N$ be given.

Let $\omega_N := e^{-2\pi i/N}$ and

$$\hat{\mathbf{x}} := \mathbf{F}_N \mathbf{x} \quad \text{with} \quad \mathbf{F}_N := (\omega_N^{jk})_{j,k=0}^{N-1}.$$

Assume $\hat{\mathbf{x}} = \mathbf{F}_N \mathbf{x}$ is $M$-sparse, i.e., $\|\hat{\mathbf{x}}\|_0 := M$. Sparsity $M \leq N$ is unknown.

**Problem**

Find a stable deterministic algorithm to compute $\hat{\mathbf{x}}$ with a small number of arithmetical operations (sublinear sparse FFT).
Recent approaches

**Basis pursuit denoise.** Minimize $\|\hat{x}\|_1$ s.t. $\|A_L \hat{x} - x_L\|_2 \leq \sigma$

Chen, Donoho, Saunders (98); Donoho, Tanner (05); Candès, Donoho, Tao (06); Tropp (04,06); van den Berg, Friedlander (08,11); ...

**Random Fourier measurements $A_L = F_{N,L}$**

Candes, Tao (06); Rudelson, Vershynin (08); Rauhut (07); Foucart, Rauhut (13); ...

**Deterministic Fourier CS-matrices**

DeVore (07); Haupt, Applebaum, Nowak (10); Xu, Xu (13); ...

**Deterministic and randomized sparse FFT**

Iwen, Spencer (08); Akavia (08); Iwen (10,13), Hassanieh et al. (12); Gilbert et al. (14); Plonka, Wannenwetsch (16,17), Bittens (16), ...

**Prony approaches, Super-Resolution**

Roy, Kailath (89); Pereyra, Scherer (10); Heider, Kunis, Potts, Veit (13); Peter, Plonka (13); Candès, Fernandez-Granda (14); Potts, Volkmer, Tasche (16),...
Sparse FFT: A first trial

How to recover $\hat{x}$ if it contains only one nonzero component?

Let $e_j$, $j = 0, \ldots, N - 1$, be the unit vectors in $\mathbb{C}^N$.

$$\hat{x} = \hat{x}_{k_0} e_{k_0} = \begin{pmatrix} 0 \\ \vdots \\ \hat{x}_{k_0} \\ \vdots \\ 0 \end{pmatrix} \Rightarrow x = \hat{x}_{k_0} \mathbf{F}_N^{-1} e_{k_0} = \frac{\hat{x}_{k_0}}{N} \begin{pmatrix} \omega_0^0 \\ \omega_N^{-k_0} \\ \vdots \\ \omega_N^{-(N-2)k_0} \\ \omega_N^{-(N-1)k_0} \end{pmatrix}$$

We find

$$x_0 = \frac{1}{N} \hat{x}_{k_0}, \quad x_1 = \frac{1}{N} \hat{x}_{k_0} \omega_N^{-k_0}$$

Thus, two components of $x$ are sufficient to recover $\hat{x}$:

$$\hat{x}_{k_0} = N x_0, \quad \omega_N^{-k_0} = \frac{x_1}{x_0}.$$ 

Observe that for noisy data the determination of $k_0$ is not stable.
Stabilization of the approach

Let \( x \in \mathbb{C}^N \), \( N = 2^J \), and let \( \hat{x} \in \mathbb{C}^N \) be \( M \)-sparse, \( M \leq N \).

Consider the periodized vectors

\[
\hat{x}^{(j)} := (\hat{x}_k^{(j)})_{k=0}^{2^j-1} := \left( \sum_{\ell=0}^{2^j-j-1} \hat{x}_{k+2^j \ell} \right)_{k=0}^{2^j-1}.
\]

Then

\[
\hat{x}^{(J)} = \hat{x}, \quad \hat{x}^{(J-1)} = \left( \hat{x}_k + \hat{x}_{k+N/2} \right)_{k=0}^{N/2-1}, \ldots,
\]

\[
\hat{x}^{(1)} = \begin{pmatrix} \sum_{\ell=0}^{N/2-1} \hat{x}_{2\ell} \\ \sum_{\ell=0}^{N/2-1} \hat{x}_{2\ell+1} \end{pmatrix}^T, \quad \hat{x}^{(0)} = \sum_{\ell=0}^{N-1} \hat{x}_\ell.
\]

Further, let

\[
x^{(j)} := F_{2^j}^{-1} \hat{x}^{(j)} = 2^{J-j} \left( x_{2^j} \right)_{k=0}^{2^j-1}.
\]
Stabilized evaluation with $j + 1$ samples

Example. \( \hat{x} = \hat{x}^{(3)} = (0, 0, 0, 0, 0, 0, 1, 0)^T \) with \( k_0 = k_0^{(3)} = 6 \),
\[ \hat{x}^{(2)} = (0, 0, 1, 0)^T \] with \( k_0^{(2)} = 2 \),
\[ \hat{x}^{(1)} = (1, 0)^T \] with \( k_0^{(1)} = 0 \),
\[ \hat{x}^{(0)} = (1) \] with \( k_0^{(0)} = 0 \).

Idea. Compute \( k_0^{(j)} \) iteratively, starting with \( k_0^{(0)} = 0 \).

We observe
\[
k_0^{(j)} = \begin{cases} k_0^{(j+1)} & 0 \leq k_0^{(j+1)} \leq 2j - 1, \\ k_0^{(j+1)} - 2j & 2j \leq k_0^{(j+1)} \leq 2j+1 - 1. \end{cases}
\]

If \( \omega_{2j+1} := \frac{x_1^{(j+1)}}{x_0^{(j+1)}} = \frac{x_{2J-j-1}^{(j+1)}}{x_0^{(j+1)}} = \omega_{2j+1}^{-k_0^{(j)}} \) then \( k_0^{(j+1)} = k_0^{(j)} \).

If \( \omega_{2j+1} := \frac{x_1^{(j+1)}}{x_0^{(j+1)}} = \frac{x_{2J-j-1}^{(j+1)}}{x_0^{(j+1)}} = -\omega_{2j+1}^{-k_0^{(j)}} \) then \( k_0^{(j+1)} = k_0^{(j)} + 2j \).
General case: Recovery of $M$-sparse vectors

Let $x \in \mathbb{C}^N$, $N = 2^J$, and let $\hat{x} \in \mathbb{C}^N$ be $M$-sparse, $M \leq N$.

**Assumption:** There is no cancellation by periodizations of $\hat{x}$.

If $\hat{x}_k \neq 0$ is significant, then $\hat{x}_{k \mod 2^j} \neq 0$ is significant.

**Example**

$$\hat{x} = (\hat{x}_k)_{k=0}^{N-1}$$ with $\text{Re} \, \hat{x}_k \geq 0$, $\text{Im} \, \hat{x}_k \geq 0$, $k = 0, \ldots, N - 1$.

For example

$$\hat{x} = \hat{x}^{(3)} = (0, 0, 3, 0, 1, 0, -3, 0)^T$$
$$\hat{x}^{(2)} = (1, 0, 0, 0)^T$$

is not allowed!
General case: Recovery of $M$-sparse vectors

Idea

Iterative reconstruction of the periodized vectors $\hat{x}^{(j)}$ for $j = 0, 1, \ldots, J$.

Observations

1. $\hat{x}^{(j)}$ is $M_j$ - sparse: $M_0 \leq M_1 \leq \ldots \leq M_J = M$.

2. $\hat{x}^{(j+1)}_k + \hat{x}^{(j+1)}_{k+2^j} = \hat{x}^{(j)}_k$, \hspace{1cm} $k = 0, \ldots, 2^j - 1$. 
Example

\[
\hat{x}^{(5)}
\]

\[
\hat{x}^{(4)}
\]

\[
\hat{x}^{(3)}
\]

\[
\hat{x}^{(2)}
\]
Idea of the algorithm

1. Choose the sample $x_0$ and compute $\hat{x}^{(0)} = \sum_{k=0}^{N-1} \hat{x}_k = N x_0$.
   If $x_0 = 0$ then $\hat{x} = 0$ (no cancellation), $M = 0$, done.

2. If $x_0 > \epsilon$ then compute

   \[
   \hat{x}^{(1)} = \begin{pmatrix}
   \hat{x}_0^{(1)} \\
   \hat{x}_1^{(1)}
   \end{pmatrix} = \frac{N}{2} F_2 \left( \begin{array}{c} x_0 \\ x_{N/2} \end{array} \right).
   \]

   Then

   \[
   \hat{x}_0^{(1)} + \hat{x}_1^{(1)} = \hat{x}_0^{(0)},
   \]

   \[
   \hat{x}_0^{(1)} - \hat{x}_1^{(1)} = \frac{N}{2} x_{N/2}.
   \]

   If $\hat{x}_0^{(1)} = 0$ all even components of $\hat{x}$ vanish.

   If $\hat{x}_1^{(1)} = 0$ all odd components of $\hat{x}$ vanish.
General step

Let \( M_j \leq M \) be the number of significant entries of \( \hat{x}^{(j)} \).
Indices of non-zero components:

\[
0 \leq n_0 < n_1 < \ldots < n_{M_j-1} \leq 2^j - 1.
\]

We have

\[
\hat{x}^{(j+1)}_k + \hat{x}^{(j+1)}_{k+2^j} = \hat{x}^{(j)}_k, \quad k = 0, \ldots, 2^j - 1.
\]

Hence, only the components \( \hat{x}^{(j+1)}_{n_\ell} \) and \( \hat{x}^{(j+1)}_{n_\ell+2^j} \) are candidates for non-zero entries in \( \hat{x}^{(j+1)} \).

Hence \( M_{j+1} \leq 2M_j \) and only \( M_j \) “suitable” further conditions are needed to recover \( \hat{x}^{(j+1)} \).
Theorem (P., Wannenwetsch, Cuyt, Lee (2017))

Let $\hat{x}(j)$ be the periodized vectors with $\hat{x}^{(J)} = x$ satisfying the non-cancellation property. If $\hat{x}(j) \in \mathbb{C}^{2^j}$ is $M_j$-sparse with support indices

$$0 \leq n_0 < n_1 < \ldots < n_{M_j - 1} \leq 2^j - 1,$$

then $\hat{x}(j+1)$ can be uniquely recovered from $\hat{x}(j) \in \mathbb{C}^{2^j}$ and $M_j$ components of $x = \mathbf{F}_N^{-1}\hat{x}$, where the indices $k_0, \ldots, k_{M_j - 1}$ are taken from the set \{2$^{J-j-1}$(2$k$ + 1), $k = 0, \ldots, 2^j - 1$\} such that

$$\left(\omega_{2^j}^{-k_p n_r}\right)_{p, r=0}^{M_j-1} = \left(\exp\left(\frac{2\pi i k_p n_r}{2^j}\right)\right)_{p, r=0}^{M_j-1} \in \mathbb{C}^{M_j \times M_j}$$

is invertible and has small condition number. Then $\hat{x}(j+1)$ can be obtained from $\hat{x}(j)$ by solving a linear system of size $M_j$.

We need less than $M(2 + \log \frac{N}{M})$ signal values to recover $x$! We need $\mathcal{O}(M \log M \log N)$ arithmetical operations to compute $\hat{x}$ using inverse NFFT!
Remaining problem

For given $N = 2^j$, $M \leq N$ and given indices

$$0 \leq n_0 < n_1 < \ldots < n_{M-1} \leq N - 1,$$

how to choose a new set of indices

$$0 \leq k_0 < k_1 < \ldots < k_{M-1} \leq N - 1$$

such that

$$\left( \omega_N^{-k_p n_r} \right)^{M-1}_{p,r=0}$$

is optimally well conditioned?
Remaining problem

For given $N = 2^j$, $M \leq N$ and given indices

$$0 \leq n_0 < n_1 < \ldots < n_{M-1} \leq N - 1,$$

how to choose a new set of indices

$$0 \leq k_0 < k_1 < \ldots < k_{M-1} \leq N - 1$$

such that

$$\left( \omega_N^{-k_p n_r} \right)^{M-1}_{p,r=0}$$

is optimally well conditioned?

We strongly simplify the problem

Let $k_p := \sigma p \mod N$. How to choose $\sigma \in \{1, \ldots, 2^j\}$ such that

$$\left( \omega_N^{-k_p n_r} \right)^{M-1}_{p,r=0} = \left( \omega_N^{-\sigma n_r p} \right)^{M-1}_{p,r=0} = V_M$$

is optimally well conditioned?
Vandermonde matrices on the unit circle

We know

1. The Vandermonde matrix $V_M = (\omega_N^{\sigma n_r p})_{p,r=0}^{M-1}$ is invertible iff $\sigma n_r \text{ mod } N$ are pairwise distinct. 

Hence invertibility of $V_M$ already follows for $\sigma = 1$.

2. The condition number of $V_M$ strongly depends on the distribution of the values $\omega_N^{\sigma n_r}$, $r = 0, \ldots, M - 1$ on the unit circle.

3. $\text{cond } V_M = 1$ iff $\omega_N^{\sigma n_r}$ are equidistantly distributed on the unit circle (see e.g. Berman, Feuer (07)).

\[N = 32, \text{ left: } \sigma = 1, \text{ cond } V_5 = 8841, \text{ right: } \sigma = 6, \text{ cond } V_5 = 1.415\]
Conditions on $\sigma$

**Theorem (Moitra (2015))**

Let $0 \leq n_0 < n_1 < \ldots < n_{M-1} < N$ be a given set of indices. For a given $\sigma \in \{1, \ldots, N\}$ let

$$d_\sigma := \min_{0 \leq k < \ell \leq M-1} (\pm \sigma(n_\ell - n_k)) \mod N$$

be the smallest (periodic) distance between two indices $\sigma n_\ell$ and $\sigma n_k$, and assume that $d_\sigma > 0$. Then the condition number $\kappa_2(V_{M',M}(\sigma))$ of the Vandermonde matrix $V_{M',M}(\sigma) := (\omega_N^{\sigma n_\ell n_k})_{\ell=0, k=0}^{M'-1, M-1}$ satisfies

$$\kappa_2(V_{M',M}(\sigma))^2 \leq \frac{M' + N/d_\sigma}{M' - N/d_\sigma},$$

provided that $M' > \frac{N}{d_\sigma}$.

**Proof:** based on Hilbert’s inequality, see e.g. Moitra (2015).
Method to choose the optimal $\sigma$

Idea

Choose $\sigma$ such that for $N = 2^j$ the distance

$$d_\sigma := \min_{0 \leq k < \ell \leq M-1} (\pm \sigma |n_\ell - n_k|) \mod N$$

is maximal.

**Brute force method** $O(M2^j)$ operations at level $j = 0, \ldots, J - 1$.

**Open problem**

Is there a smart method to find the optimal $\sigma$ with $O(M^2)$ operations at each level?

**Up to now**

We have only heuristic algorithms to find an (almost) optimal $\sigma$ with $O(M^2)$ operations.
Worst case distance

To get $V_M = (\omega_N^{\sigma n_k p})_{p,k=0}^{M-1}$ with small condition number we want

$$d := \max_{\sigma} d_\sigma \approx \frac{N}{M}$$

where

$$d_\sigma := \min_{0 \leq k < \ell \leq M-1} (\pm \sigma |n_\ell - n_k|) \mod N.$$  

What is the worst case that can happen for $d$ and optimized $\sigma$?
Worst case distance

To get $V_M = (\omega_N^\sigma n_{kp})_{p,k=0}^{M-1}$ with small condition number we want

$$d := \max_\sigma d_\sigma \approx \frac{N}{M}$$

where

$$d_\sigma := \min_{0 \leq k < \ell \leq M-1} (\pm \sigma |n_\ell - n_k|) \mod N.$$ 

What is the worst case that can happen for $d$ and optimized $\sigma$?

**Theorem (P., Wannenwetsch (2017))**

For arbitrarily distributed $0 \leq n_0 < n_1 < \ldots < n_{M-1} \leq N - 1$ and optimally chosen $\sigma$ maximizing

$$d_\sigma = \min_{0 \leq k < \ell \leq M-1} (\pm \sigma |n_\ell - n_k|) \mod N$$

we have

$$d = \max_\sigma d_\sigma \geq \frac{N}{M^2}.$$
Worst case example

Let $N = 16$, $M = 4$, found indices $(n_0, n_1, n_2, n_3) = (0, 1, 3, 8)$.

$\sigma = 1$: $\Rightarrow d_1 = 1$

$\sigma = 3$: $\Rightarrow d_3 = 1$

$\sigma = 5$: $\Rightarrow d_5 = 1$

$\sigma = 7$: $\Rightarrow d_7 = 1$

Therefore $d = \frac{N}{M^2} = 1$. 
Numerical example

Let $N = 128$, $M = 4$

Number of different choices of ordered positions: $\binom{128}{4} = 10\,668\,000$

Cases for which $d \geq 16$: $10\,641\,376$ ($99.75\%$)
Cases for which $8 \leq d < 16$: $26\,624$ ($0.25\%$)
Cases for which $d < 8$: $0$

The worst case $d \approx \frac{N}{M^2}$ is rare!

To avoid bad condition numbers in these cases, we have two options:

a) We use further measurements to improve the condition number of the Vandermonde matrix.

b) We consider another strategy for extracting a suitable partial Fourier matrix (e.g. a second parameter $\sigma_2$ leading to a generalized Vandermonde-type matrix).
Numerical example: Adaptivity helps!

\( N = 16,384 \ (J = 14), \ M = 17 \) (adaptive versus nonadaptive)
active indices: 6, 7, 8, 9, 10, 11, 12, 13, 56, 57, 58, 79, 80, 81, 345, 1234, 1235

<table>
<thead>
<tr>
<th>( j )</th>
<th>( M )</th>
<th>( \sigma )</th>
<th>( \text{cond } V_M )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>8</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>13</td>
<td>3</td>
<td>11.64</td>
</tr>
<tr>
<td>6</td>
<td>16</td>
<td>3</td>
<td>51.17</td>
</tr>
<tr>
<td>7</td>
<td>17</td>
<td>11</td>
<td>97.37</td>
</tr>
<tr>
<td>8</td>
<td>17</td>
<td>22</td>
<td>97.37</td>
</tr>
<tr>
<td>9</td>
<td>17</td>
<td>44</td>
<td>97.37</td>
</tr>
<tr>
<td>10</td>
<td>17</td>
<td>88</td>
<td>97.37</td>
</tr>
<tr>
<td>11</td>
<td>17</td>
<td>285</td>
<td>14.41</td>
</tr>
<tr>
<td>12</td>
<td>17</td>
<td>570</td>
<td>14.41</td>
</tr>
<tr>
<td>13</td>
<td>17</td>
<td>203</td>
<td>7.98</td>
</tr>
<tr>
<td>14</td>
<td>17</td>
<td>406</td>
<td>7.98</td>
</tr>
</tbody>
</table>

used signal values: 181

adaptive choice of \( \sigma \)

<table>
<thead>
<tr>
<th>( j )</th>
<th>( M )</th>
<th>( \sigma )</th>
<th>( \text{cond } V_M )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>8</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>13</td>
<td>1</td>
<td>11.64</td>
</tr>
<tr>
<td>6</td>
<td>16</td>
<td>1</td>
<td>1.4425e + 05</td>
</tr>
<tr>
<td>7</td>
<td>17</td>
<td>1</td>
<td>8.8402e + 09</td>
</tr>
<tr>
<td>8</td>
<td>17</td>
<td>1</td>
<td>2.7140e + 07</td>
</tr>
<tr>
<td>9</td>
<td>17</td>
<td>1</td>
<td>4.5243e + 12</td>
</tr>
<tr>
<td>10</td>
<td>25</td>
<td>1</td>
<td>6.3748e + 15</td>
</tr>
<tr>
<td>11</td>
<td>39</td>
<td>1</td>
<td>1.2212e + 17</td>
</tr>
<tr>
<td>12</td>
<td>60</td>
<td>1</td>
<td>3.4276e + 16</td>
</tr>
<tr>
<td>13</td>
<td>114</td>
<td>1</td>
<td>2.1692e + 17</td>
</tr>
<tr>
<td>14</td>
<td>193</td>
<td>1</td>
<td>3.4942e + 17</td>
</tr>
</tbody>
</table>

used signal values: 351

nonadaptive choice of \( \sigma \)

Table 1: ???
Figure 1: Runtime comparison (in seconds) of the FFT (blue line) and our algorithm with $M = 5$ (red line), $M = 10$ (black dotted line), $M = 20$ (cyan dash-dots line) and $M = 30$ (green dashed line) for length $N = 2^j$ with $j = 12, \ldots, 22$. 
Summary

1. We propose a new multi-scale algorithm for sparse vector reconstruction.

2. The sparsity $M \leq N$ does not need to be known a priori.

3. We need less than $\min(M(2 + \log \frac{N}{M}), N)$ signal values for reconstruction.

4. We need less than $\min(O(M^2 \log N), O(N \log N))$ arithmetical operations for reconstruction (sparse FFT!).

5. At each iteration step only a linear system of size at most $M \times M$ needs to be solved.

6. Adaptivity is used to improve the numerical stability of the procedure.

7. Even a simple strategy optimizing only one parameter usually gives good numerical results.
References

- Gerlind Plonka, Katrin Wannenwetsch. 
  A deterministic sparse FFT algorithm for vectors with small support. 

- Gerlind Plonka and Katrin Wannenwetsch. 
  A sparse Fast Fourier algorithm for real nonnegative vectors. 

- Gerlind Plonka, Katrin Wannenwetsch, Annie Cuyt, Wen-Shin Lee. 
  Deterministic sparse FFT for M-sparse vectors. 

\thankyou