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

Problem

Let $\mathbf{x} = (x_j)_{j=0}^{N-1} \in \mathbb{C}^N$ be given.

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

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

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

Problem

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

Recent approaches

Basis pursuit denoise. Minimize $\|\widehat{\mathbf{x}}\|_1$ s.t. $\|\mathbf{A}_L\widehat{\mathbf{x}} - \mathbf{x}_L\|_2 \le \sigma$

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

Random Fourier measurements $\mathbf{A}_L = \mathbf{F}_{N,L}$

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 $\widehat{\mathbf{x}}$ if it contains **only one nonzero component**?

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

$$\widehat{\mathbf{x}} = \widehat{x}_{k_0}\,\mathbf{e}_{k_0} = \begin{pmatrix} 0 \\ \vdots \\ \widehat{x}_{k_0} \\ \vdots \\ 0 \end{pmatrix} \quad \Rightarrow \quad \mathbf{x} = \widehat{x}_{k_0}\,\mathbf{F}_N^{-1}\mathbf{e}_{k_0} = \frac{\widehat{x}_{k_0}}{N} \begin{pmatrix} \omega_N^0 \\ \omega_N^{-k_0} \\ \vdots \\ \omega_N^{-(N-2)k_0} \\ \omega_N^{-(N-1)k_0} \end{pmatrix}$$

We find

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

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

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

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

Stabilization of the approach

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

Consider the periodized vectors

$$\widehat{\mathbf{x}}^{(j)} := \left(\widehat{x}_k^{(j)}\right)_{k=0}^{2^j-1} := \left(\sum_{\ell=0}^{2^{J-j}-1} \widehat{x}_{k+2^j\ell}\right)_{k=0}^{2^j-1}.$$

Then

$$\widehat{\mathbf{x}}^{(J)} = \widehat{\mathbf{x}}, \qquad \widehat{\mathbf{x}}^{(J-1)} = \left(\widehat{x}_k + \widehat{x}_{k+N/2}\right)_{k=0}^{N/2-1}, \ldots,$$

$$\widehat{\mathbf{x}}^{(1)} = \left(\sum_{\ell=0}^{N/2-1} \widehat{x}_{2\ell}, \sum_{\ell=0}^{N/2-1} \widehat{x}_{2\ell+1}\right)^T, \qquad \widehat{\mathbf{x}}^{(0)} = \sum_{\ell=0}^{N-1} \widehat{x}_\ell.$$

Further, let

$$\mathbf{x}^{(j)} := \mathbf{F}_{2^j}^{-1} \widehat{\mathbf{x}}^{(j)} = 2^{J-j} \left(x_{2^{J-j}k}\right)_{k=0}^{2^j-1}.$$

Stabilized evaluation with $j+1$ samples

Example.
$$\widehat{\mathbf{x}} = \widehat{\mathbf{x}}^{(3)} = (0,0,0,0,0,0,1,0)^T \quad \text{with} \quad k_0 = k_0^{(3)} = 6,$$
$$\widehat{\mathbf{x}}^{(2)} = (0,0,1,0)^T \quad \text{with} \quad k_0^{(2)} = 2,$$
$$\widehat{\mathbf{x}}^{(1)} = (1,0)^T \quad \text{with} \quad k_0^{(1)} = 0,$$
$$\widehat{\mathbf{x}}^{(0)} = (1) \quad \text{with} \quad k_0^{(0)} = 0.$$

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

We observe

$$k_0^{(j)} = \begin{cases} k_0^{(j+1)} & 0 \le k_0^{(j+1)} \le 2^j - 1, \\ k_0^{(j+1)} - 2^j & 2^j \le k_0^{(j+1)} \le 2^{j+1} - 1. \end{cases}$$

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

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

General case: Recovery of M-sparse vectors

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

Assumption: There is no cancellation by periodizations of $\widehat{\mathbf{x}}$.

If $\widehat{x}_k \neq 0$ is significant, then $\widehat{x}^{(j)}_{k \bmod 2^j} \neq 0$ is significant.

Example

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

For example

$$\begin{aligned} \widehat{\mathbf{x}} = \quad \widehat{\mathbf{x}}^{(3)} &= (0, 0, 3, 0, 1, 0, -3, 0)^T \\ \widehat{\mathbf{x}}^{(2)} &= (1, 0, 0, 0)^T \end{aligned}$$

is **not allowed**!

General case: Recovery of M-sparse vectors

Idea

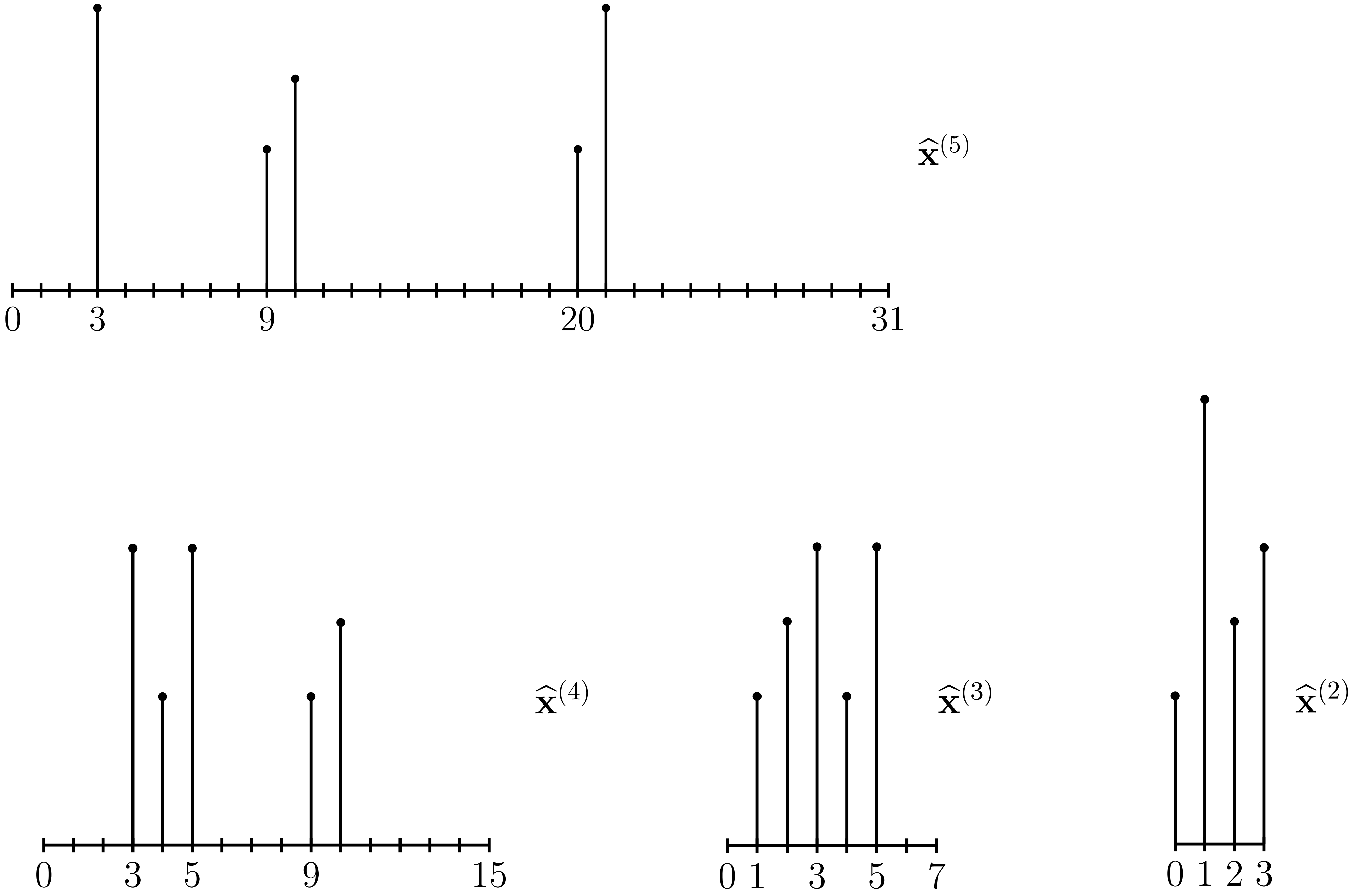
Iterative reconstruction of the periodized vectors $\widehat{\mathbf{x}}^{(j)}$ for $j = 0, 1, \ldots, J$.

Observations

1.
$$\widehat{\mathbf{x}}^{(j)} \text{ is } M_j - \text{sparse}: \; M_0 \le M_1 \le \ldots \le M_J = M.$$

2.
$$\widehat{x}_k^{(j+1)} + \widehat{x}_{k+2^j}^{(j+1)} = \widehat{x}_k^{(j)}, \qquad k = 0, \ldots, 2^j - 1.$$

Example



0
3
9
20
31
$\hat{\mathbf{x}}^{(5)}$
0
3
5
9
15
$\hat{\mathbf{x}}^{(4)}$
0
1
3
5
7
$\hat{\mathbf{x}}^{(3)}$
0
1
2
3
$\hat{\mathbf{x}}^{(2)}$

Idea of the algorithm

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

2. If $x_0 > \epsilon$ then compute

$$\widehat{\mathbf{x}}^{(1)} = \begin{pmatrix} \widehat{x}_0^{(1)} \\ \widehat{x}_1^{(1)} \end{pmatrix} = \frac{N}{2}\mathbf{F}_2 \begin{pmatrix} x_0 \\ x_{N/2} \end{pmatrix}.$$

   Then

$$\widehat{x}_0^{(1)} + \widehat{x}_1^{(1)} = \widehat{\mathbf{x}}_0^{(0)},$$
$$\widehat{x}_0^{(1)} - \widehat{x}_1^{(1)} = \tfrac{N}{2} x_{N/2}.$$

   If $\widehat{x}_0^{(1)} = 0$ all even components of $\widehat{\mathbf{x}}$ vanish.
   If $\widehat{x}_1^{(1)} = 0$ all odd components of $\widehat{\mathbf{x}}$ vanish.

General step

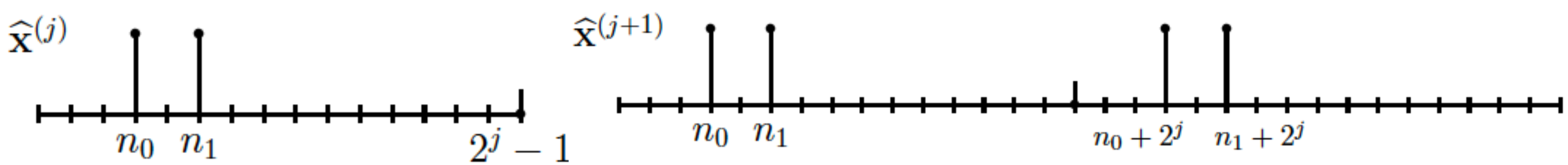
Let $M_j \leq M$ be the number of significant entries of $\widehat{\mathbf{x}}^{(j)}$.
Indices of non-zero components:

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

We have

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

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



Hence $M_{j+1} \leq 2M_j$ and only M_j "suitable" further conditions are needed to recover $\widehat{\mathbf{x}}^{(j+1)}$.

Theorem (P., Wannenwetsch, Cuyt, Lee (2017))

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

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

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

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

is invertible and has small condition number. Then $\widehat{\mathbf{x}}^{(j+1)}$ can be obtained from $\widehat{\mathbf{x}}^{(j)}$ 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 $\widehat{\mathbf{x}}$ 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

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

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

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

is optimally well conditioned?

We strongly simplify the problem

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

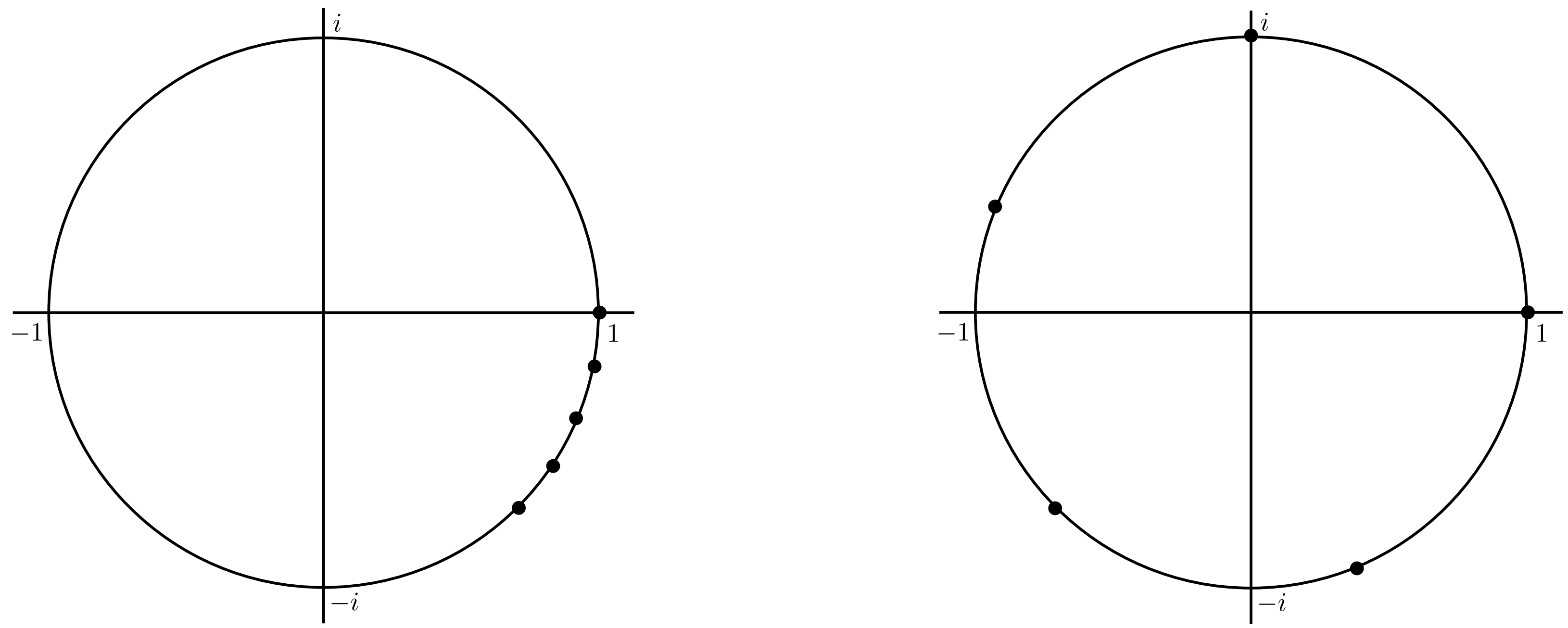
$$\left(\omega_N^{-k_p n_r}\right)_{p,r=0}^{M-1} = \left(\omega_N^{-\sigma n_r p}\right)_{p,r=0}^{M-1} = \mathbf{V}_M$$

is optimally well conditioned?

Vandermonde matrices on the unit circle

We know

1. The Vandermonde matrix $\mathbf{V}_M = \left(\omega_N^{\sigma n_r p}\right)_{p,r=0}^{M-1}$ is invertible iff $\sigma n_r \bmod N$ are pairwise distinct.
   Hence invertibility of $\mathbf{V}_M$ already follows for $\sigma = 1$.
2. The condition number of $\mathbf{V}_M$ strongly depends on the distribution of the values $\omega_N^{\sigma n_r}$, $r = 0, \ldots, M-1$ on the unit circle.
3. $\operatorname{cond} \mathbf{V}_M = 1$ iff $\omega_N^{\sigma n_r}$ are equidistantly distributed on the unit circle (see e.g. Berman, Feuer (07)).



$N = 32$, left: $\sigma = 1$, $\operatorname{cond} \mathbf{V}_5 = 8\,841$, right: $\sigma = 6$, $\operatorname{cond} \mathbf{V}_5 = 1.415$

Conditions on σ

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

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

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

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

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

Proof: based on Hilbert's inequality, see e.g. Moitra (2015).

Method to choose the optimal σ

Idea

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

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

is maximal.

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

Open problem

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

Up to now

We have only heuristic algorithms to find an (almost) optimal σ with $\mathcal{O}(M^2)$ operations.

Worst case distance

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

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

where

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

What is the worst case that can happen for d and optimized σ ?

Worst case distance

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

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

where

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

What is the worst case that can happen for d and optimized σ ?

Theorem (P., Wannenwetsch (2017)

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

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

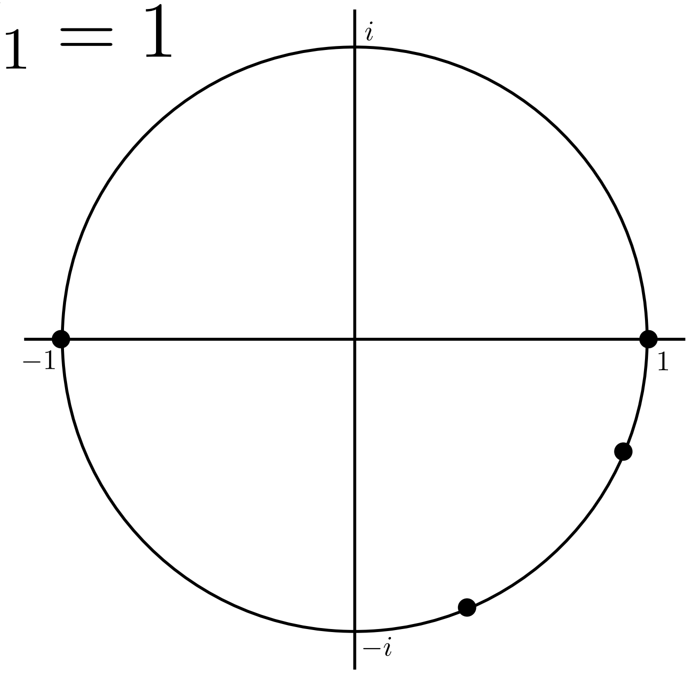
we have

$$d = \max_{\sigma} d_\sigma \ge \frac{N}{M^2}.$$

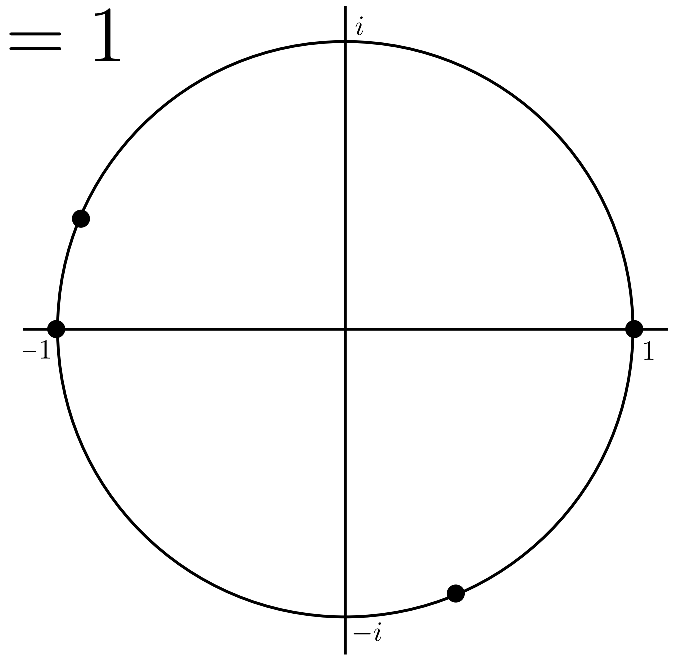
Worst case example

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

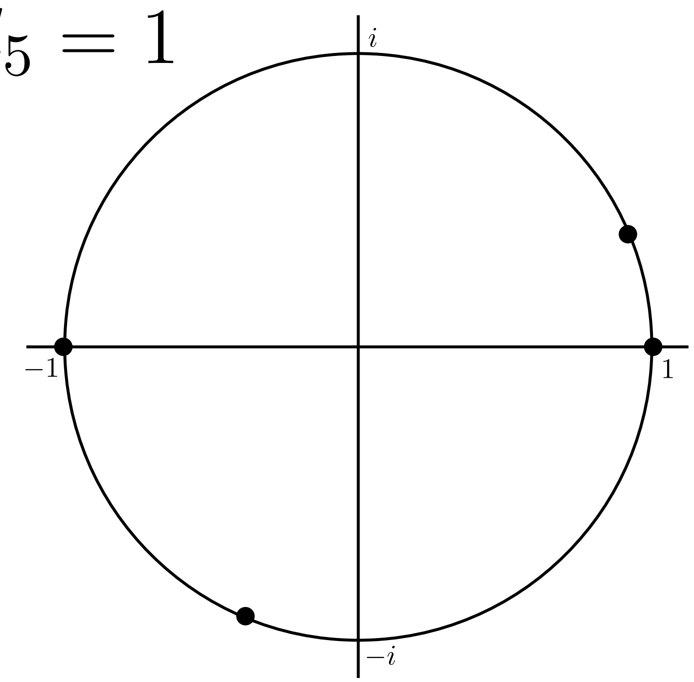
$\sigma = 1$: $\Rightarrow d_1 = 1$



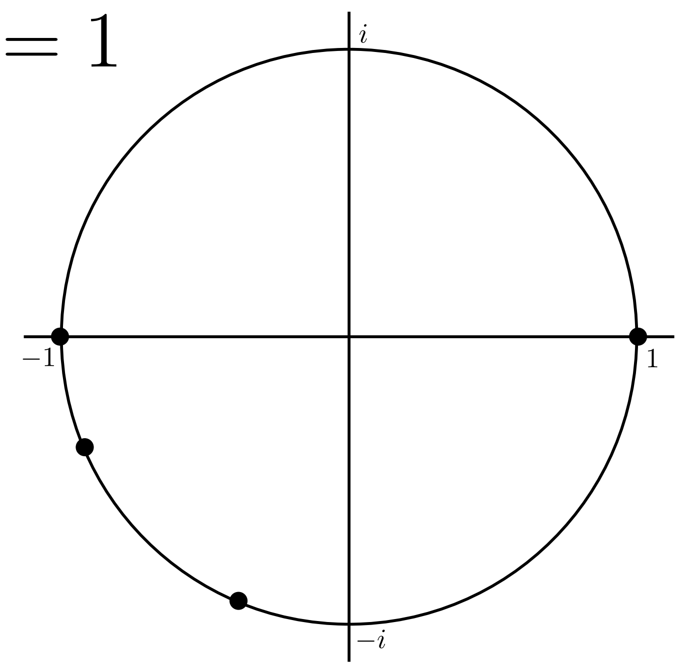
$\sigma = 3$: $\Rightarrow d_3 = 1$



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



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



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 σ_2 leading to a generalized Vandermonde-type matrix).

Numerical example: Adaptivity helps!

$N = 16\,384$ $(J = 14)$, $M = 17$ (adaptive versus nonadaptive)

active indices: $6, 7, 8, 9, 10, 11, 12, 13, 56, 57, 58, 79, 80, 81, 345, 1234, 1235$

j	M	σ	cond $\mathbf{V}_M$
1	1	1	1
2	2	1	1
3	4	1	1
4	8	1	1
5	13	3	11.64
6	16	3	51.17
7	17	11	97.37
8	17	22	97.37
9	17	44	97.37
10	17	88	97.37
11	17	285	14.41
12	17	570	14.41
13	17	203	7.98
14	17	406	7.98

used signal values: 181

adaptive choice of σ

j	M	σ	cond $\mathbf{V}_M$
1	1	1	1
2	2	1	1
3	4	1	1
4	8	1	1
5	13	1	11.64
6	16	1	$1.4425e+05$
7	17	1	$8.8402e+09$
8	17	1	$2.7140e+07$
9	17	1	$4.5243e+12$
10	25	1	$6.3748e+15$
11	39	1	$1.2212e+17$
12	60	1	$3.4276e+16$
13	114	1	$2.1692e+17$
14	193	1	$3.4942e+17$

used signal values: 351

nonadaptive choice of σ

Runtime experiments

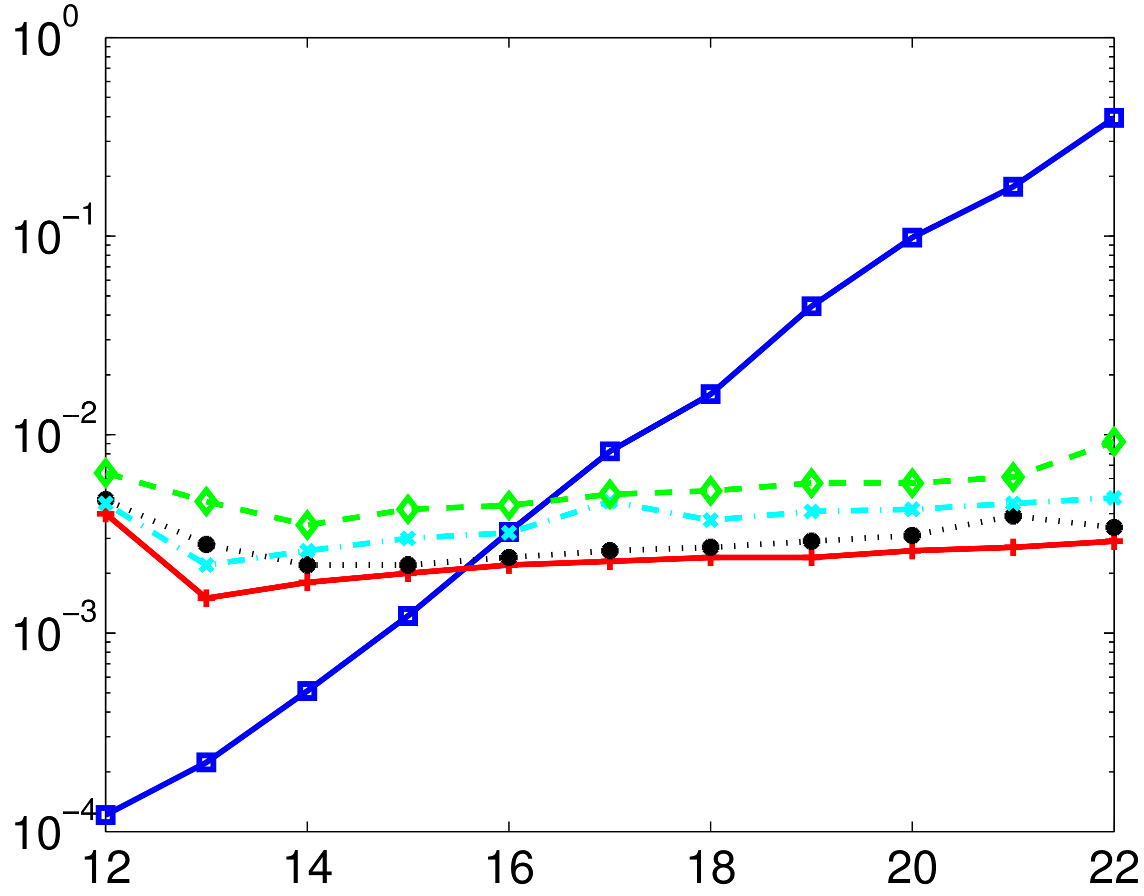


Figure 1: Runtime comparison (in seconds) of the FFT (blue line) and our algorithm with $M = 5$ (red line), $M = 10$ (black dotted line), $M = 20$ (cyan dash-dots line) and $M = 30$ (green dashed line) for length $N = 2^j$ with $j = 12, \ldots, 22$.

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 $\min(\mathcal{O}(M^2 \log N), \mathcal{O}(N \log N))$ 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.
  Numer. Algorithms **71**(4) (2016), 889-905.
- Gerlind Plonka and Katrin Wannenwetsch.
  A sparse Fast Fourier algorithm for real nonnegative vectors.
  J. Comput. Appl. Math. **321** (2017), 532-539.
- Gerlind Plonka, Katrin Wannenwetsch, Annie Cuyt, Wen-Shin Lee.
  Deterministic sparse FFT for M-sparse vectors.
  Numer. Algorithms, to appear.

\thankyou