Vítězslav Veselý

Algebraic theory of fast mixed-radix transforms. II. Computational complexity and applications

Archivum Mathematicum, Vol. 26 (1990), No. 1, 19--26

Persistent URL: `http://dml.cz/dmlcz/107365`

ARCHIVUM MATHEMATICUM (BRNO)
Vol. 26, No. 1 (1990), 19–26

ALGEBRAIC THEORY OF FAST MIXED-RADIX TRANSFORMS: II. COMPUTATIONAL COMPLEXITY AND APPLICATIONS

VÍTĚZSLAV VESELÝ

(Received June 26, 1986)

Abstract. This is a continuation of the part I [Arch. Math. (Brno), Vol. 25, No. 3(1989), 149–162] where new matrix operation (generalized Kronecker product) has been introduced for easy derivation of mixed-radix factorizations. In part II their computational complexity is studied and the effectivity of the new algebraic approach is demonstrated by deriving fast algorithm for a very general (recently introduced) concept of the parametric discrete Fourier transform.

Key words. Generalized Kronecker product of matrices, fast mixed-radix transform, fast Fourier transform, factorization of matrices, computational complexity.

MS Classification: 15 A 23, 15 A 04, 68 Q 25, 65 F 30, 65 T 05.

INTRODUCTION

Section 1 deals with computational complexity of *fast mixed-radix transforms* which have been introduced in [6]. In Section 2 we prove, among others, that the *parametric discrete Fourier transform* (DFT-P) recently introduced in [5] as an interesting generalization of the *discrete Fourier transform* is a *mixed-radix transform*, and illustrate how *parametric fast Fourier transforms* (FFT-P) may be derived using the new algebraic approach.

Throughout the paper we keep the notation of part I [6]. References relating to part I are prefixed by I.

1. COMPUTATIONAL COMPLEXITY OF FAST MIXED-RADIX TRANSFORMS

Let us have associated with a matrix $\mathbf{A} \in \mathscr{M}(N \times K)$ a suitable algorithm accomplishing the linear transform $\mathbf{y} = \mathbf{A}\mathbf{x}$ with at most $\alpha(\mathbf{A})$ scalar additions or subtractions and $\mu(\mathbf{A})$ scalar multiplications for any input vector $\mathbf{x}$. Then $\alpha(\mathbf{A})$ and $\mu(\mathbf{A})$ characterize additive and multiplicative complexity of the transform $\mathbf{A}$ if we do not distinguish between $\mathbf{A}$ and the corresponding algorithm.

Hereafter **A** and **B** stand for MRT matrices defined in I.2.7, and $o(\mathbf{A})$ and $o(\mathbf{B})$, $o \in \{\alpha, \mu\}$ for the computational complexity quantities of the corresponding FMRT from I.2.9. Thus we may write $o(\mathbf{A}) = \sum_{i=1}^{m} o(\mathbf{A}^{(i)})$ and $o(\mathbf{B}) = \sum_{i=1}^{m} o(\mathbf{B}^{(i)})$. As $o(\mathbf{C})$ is for any matrix **C** invariant with respect to its row and column permutations, all FMRTs derived from $\mathbf{A}^{(i)}$ and $\mathbf{B}^{(i)}$ by inserting factored identity matrix have the same complexity. Specifically $o(\mathbf{A}) = o(\mathbf{A}^{-}) = \sum_{i=1}^{m} o(\mathbf{A}^{-(i)})$ and $o(\mathbf{B}) = o(\mathbf{B}^{-}) = \sum_{i=1}^{m} o(\mathbf{B}^{-(i)})$. An inspection of I.2.10 shows that each $\mathbf{A}^{-(i)}(\mathbf{B}^{-(i)})$ may be decomposed into elementary transforms $\mathbf{A}_{i,k}(\mathbf{B}_{i,n})$. Consequently $o(\mathbf{A}^{-(i)}) = \sum_{k=0}^{K_{i+1,m}-1} N_{1,i-1} o(\mathbf{A}_{i,k})$ and $o(\mathbf{B}^{-(i)}) = \sum_{n=0}^{N_{i+1,m}-1} K_{1,i-1} o(\mathbf{B}_{i,n})$. Thus the resulting complexity of **A** and **B** may be expressed in terms of the complexity of elementary transforms as follows:

$$(1.1) \qquad o(\mathbf{A}) = \sum_{i=1}^{m} N_{1,i-1} \sum_{k=0}^{K_{i+1,m}-1} o(\mathbf{A}_{i,k}) \qquad \text{with } \mathbf{A}_{m,0} = \mathbf{A}_m$$

and

$$(1.2) \qquad o(\mathbf{B}) = \sum_{i=1}^{m} K_{1,i-1} \sum_{n=0}^{N_{i+1,m}-1} o(\mathbf{B}_{i,n}) \qquad \text{with } \mathbf{B}_{m,0} = \mathbf{B}_m .$$

As $\mathbf{A}_{i,k}$, $\mathbf{B}_{i,n} \in \mathcal{M}(N_i \times K_i)$, we have $\alpha(\mathbf{A}_{i,k})$, $\alpha(\mathbf{B}_{i,n}) \leqq N_i(K_i - 1)$ and $\mu(\mathbf{A}_{i,k})$, $\mu(\mathbf{B}_{i,n}) \leqq N_i K_i$ for each $k \in \mathbf{Z}_{K_{i+1,m}}$ and $n \in \mathbf{Z}_{N_{i+1,m}}$. Substituting this into (1.1) and (1.2), we arrive at the upper bounds

$$(1.3) \qquad o(\mathbf{A}) \leqq \sum_{i=1}^{m} N_{1,i-1} N_i (K_i - \delta_{o,\alpha}) K_{i+1,m} \leqq \sum_{i=1}^{m} N_{1,i} K_{i,m} = B(\mathcal{N}, \mathcal{K})$$

and

$$(1.4) \qquad o(\mathbf{B}) \leqq \sum_{i=1}^{m} K_{1,i-1} N_i (K_i - \delta_{o,\alpha}) N_{i+1,m} \leqq \sum_{i=1}^{m} K_{1,i} N_{i,m} = B(\mathcal{K}, \mathcal{N}).$$

Due to the symmetry of upper bounds $B(\mathcal{N}, \mathcal{K})$ and $B(\mathcal{K}, \mathcal{N})$ in (1.3) and (1.4) we shall deal only with (1.3) in further considerations. All stated later on will hold also for (1.4) when exchanging the roles of $\mathcal{N}$ and $\mathcal{K}$, i.e. of N_i and K_i for $i \in [1 : m]$.

We shall now investigate the following two aspects of the algorithm I.2.9 which are of practical interest:

1. *Is the term "fast" used for the algorithm I.2.9 justified? We are going to verify that $B(\mathcal{N}, \mathcal{K}) < NK$ except for unimportant special cases.*

1.1 Lemma. *Let $N_1, N_2 \in \mathbf{N}$ then it holds*

1° $N_1 + N_2 \leqq N_1 N_2$ iff $N_1 > 1$, $N_2 > 1$.

2° $N_1 + N_2 < N_1 N_2$ iff $N_1 > 1$, $N_2 > 2$ or $N_1 > 2$, $N_2 > 1$.

Proof. 1. Implication $\Rightarrow$: $N_1 = 1 \Rightarrow N_1 + N_2 = 1 + N_2 > N_2 = N_1 N_2$. The same is true for $N_2 = 1$ due to the symmetry. $N_1 = N_2 = 2 \Rightarrow N_1 + N_2 = N_1 N_2$.

2. Implication $\Leftarrow$: We can assume $2 \leqq N_1 \leqq N_2$ without loss of generality. Then $N_1 N_2 \geqq 2N_2 = N_2 + N_2 \geqq N_1 + N_2$. This inequality is sharp with $2 \leqq \leqq N_1 < N_2$ as well as with $2 < N_1 = N_2$. ∎

1.2 Theorem. *Let us consider the following conditions concerning $\mathcal{N}$ and $\mathcal{K}$:*

(i) $K_1 > 1, K_2 > 1, \ldots, K_{m-1} > 1, N_m > 1$.

(i′) $K_{m-1} > 2$ or $N_m > 2$ or $m > 2$ and there exists $i \in [2: m - 1]$ such that $N_i > 1$.

(ii) $K_1 > 1, N_2 > 1, \ldots, N_{m-1} > 1, N_m > 1$.

(ii′) $K_1 > 2$ or $N_2 > 2$ or $m > 2$ and there exists $i \in [2: m - 1]$ such that $K_i > 1$. Then the following holds:

(1) (i) or (ii) $\Rightarrow B(\mathcal{N}, \mathcal{K}) \leqq NK$. For $m = 2$ also the opposite is true, i.e. *(i) $\Leftrightarrow \Leftrightarrow$ (ii) $\Leftrightarrow B(\mathcal{N}, \mathcal{K}) \leqq NK$.*

(2) (i), (i′) or *(ii), (ii′) $\Rightarrow B(\mathcal{N}, \mathcal{K}) < NK$. For $m = 2$ also the opposite is true,* i.e. *(i), (i′) $\Leftrightarrow$ (ii), (ii′) $\Leftrightarrow B(\mathcal{N}, \mathcal{K}) < NK$.*

Proof. We proceed by induction on m.

1. $m = 2$: $B(\mathcal{N}, \mathcal{K}) - NK = N_1 K_1 K_2 + N_1 N_2 K_2 - N_1 N_2 K_1 K_2 = N_1 K_2 (K_1 + + N_2 - K_1 N_2)$ and the assertion is an immediate consequence of 1.1.

2. $m > 2$: First suppose that (*i*) is satisfied. Putting $\mathcal{N}' = (N'_1, \ldots, N'_{m-1}) = = (N_2, \ldots, N_m)$ and $\mathcal{K}' = (K_1, \ldots, K_{m-1})$, we get $B(\mathcal{N}, \mathcal{K}) - NK = \sum_{i=1}^{m} N_{1,i} K_{i,m} - - N_{1,m} K_{1,m} = N_1 K_m (\sum_{i=1}^{m} N_{2,i} K_{i,m-1} - N_{2,m} K_{1,m-1}) = N_1 K_m (\sum_{i=1}^{m-2} N'_{1,i-1} K_{i,m-1} + + (N'_{1,m-2} K_{m-1} + N'_{1,m-1}) - N'_{1,m-1} K_{1,m-1})$ where $N'_{1,i-1} K_{i,m-1} \leqq N'_{1,i} K_{i,m-1}$ for each $i \in [1: m - 2]$ and the inequality is sharp if $N_{i+1} = N'_i > 1$ for some $i \in [1: m - 2]$ (cf. (*i′*)). $N'_{1,m-2} K_{m-1} + N'_{1,m-1} = N'_{1,m-2}(K_{m-1} + N'_{m-1}) \leqq \leqq N'_{1,m-2} N'_{m-1} K_{m-1} = N'_{1,m-1} K_{m-1}$ by 1.1 and the inequality is sharp if $K_{m-1} > > 2$ or $N'_{m-1} = N_m > 2$ (cf. (*i′*)). Hence on the whole $B(\mathcal{N}, \mathcal{K}) - NK \leqq (<) \leqq \leqq (<) N_1 K_m (\sum_{i=1}^{m-1} N'_{1,i} K_{i,m-1} - N'_{1,m-1} K_{1,m-1}) = N_1 K_m (B(\mathcal{N}', \mathcal{K}') - - N'_{1,m-1} K_{1,m-1}) \leqq 0$. Here the former inequality is sharp if in addition to (*i*) also (*i′*) is satisfied, and the latter one holds by induction hypothesis because $\mathcal{N}'$ and $\mathcal{K}'$ satisfy (*i*). If (*ii*) holds (or in addition (*ii′*)) then (*ii*) and (*ii′*) is converted to (*i*) and (*i′*), respectively by exchanging the roles of N_i and K_{m+1-i} for $i \in [1: m]$, and the assertion is an immediate consequence of the evident equation $B(\mathcal{N}, \mathcal{K}) - - NK = B(\mathcal{K}s, \mathcal{N}s) - NK$. ∎

It remains to investigate the asymptotic behaviour of $B(\mathcal{N}, \mathcal{K})$ with N, K approaching infinity.

1.3 Theorem. *Let* $\{[m_i : M_i]\}_{i=1}^{\infty}$ *be an arbitrary but fixed sequence of intervals such that* $2 \leqq m_i \leqq M_i \leqq M$ *and* $1 < R \leqq m_i^2/M_i$ *is satisfied for each* $i \in \mathbf{N}$. *If* $\{\mathcal{N}^{(m)} = (N_1^{(m)}, \ldots, N_m^{(m)})\}_{m=2}^{\infty}$, $\{\mathcal{K}^{(m)} = (K_1^{(m)}, \ldots, K_m^{(m)})\}_{m=2}^{\infty}$, $\{N^{(m)} = N_{1,m}^{(m)}\}_{m=2}^{\infty}$ *and* $\{K^{(m)} = K_{1,m}^{(m)}\}_{m=2}^{\infty}$ *are sequences satisfying* $m_i \leqq N_i^{(m)}, K_i^{(m)} \leqq M_i$ *for each* $m \in \mathbf{N}$ *and* $i \in [1 : m]$ *then* $\dfrac{B(\mathcal{N}^{(m)}, \mathcal{K}^{(m)})}{N^{(m)}K^{(m)}} \leqq \dfrac{mM}{R^m}$, $\lim\limits_{m \to \infty} \dfrac{B(\mathcal{N}^{(m)}, \mathcal{K}^{(m)})}{N^{(m)}K^{(m)}} = 0$ *and the convergence is the faster the greater is* R.

Proof. As $N_i^{(m)}, K_i^{(m)} \geqq m_i \geqq 2$, we have $0 < \dfrac{B(\mathcal{N}^{(m)}, \mathcal{K}^{(m)})}{N^{(m)}K^{(m)}} < 1$ by 1.2 $(m > 2)$. From (1.3) we get further

$$\frac{B(\mathcal{N}^{(m)}, \mathcal{K}^{(m)})}{N^{(m)}K^{(m)}} \leqq \frac{\sum\limits_{i=1}^{m} M_i M_{1,m}}{\prod\limits_{i=1}^{m} m^2} \leqq \frac{mM}{(m_i^2/M_i)} \leqq$$

$$\leqq \frac{mM}{R^m} \to 0 \text{ with } m \to \infty \text{ because } R > 1. \blacksquare$$

We see by 1.3 that the algorithm I.2.9 is for large N and K the faster in comparison to direct computation of $\mathbf{Ax}$ the smaller are M_i, i.e. the smaller are the factors $N_i^{(m)}, K_i^{(m)}$ or the greater are m_i, i.e. the smaller is the range $M_i - m_i$ allowing the fluctuation of the factors $N_i^{(m)}, K_i^{(m)}$. Thus for a given $\{M_i\}_{i=1}^{\infty}$ best convergence rate is obtained with $N_{\min} \leqq m_i = N_i = K_i = M_i \leqq N_{\max}$, which gives, in view of (1.3), the bound of order $N \log N$ on operation counts, namely $o(\mathbf{A}) \leqq N_1 \ldots N_m(N_1 + \ldots + N_m - m\delta_{o,\alpha}) \leqq (N_{\max} - \delta_{o,\alpha})\, mN \leqq \leqq (N_{\max} - \delta_{o,\alpha})\, N \log_{N_{\min}}(N)$.

This bound is the best one with $M_i = N_i = 2$, $i \in [1 : m]$, which gives $(2 - \delta_{\bullet,\alpha})\, N \log_2 N$ for $N = 2^m$.

2. *What is the best (worst) ordering of factors in $\mathcal{N}$ and $\mathcal{K}$ minimizing (maximizing) the upper bound $B(\mathcal{N}, \mathcal{K})$?*

1.4 Lemma. *Let* $p_i \in \mathcal{P}([1 : m])$, $i \in [1 : m - 1]$, $m \geqq 2$ *stand for a transposition of* i *and* $i + 1$. *Then it holds*

1° $B(\mathcal{N}, \mathcal{K}) \leqq B(\mathcal{N}p_i, \mathcal{K})$ *iff* $N_i \leqq N_{i+1}$ *and*

2° $B(\mathcal{N}, \mathcal{K}) \geqq B(\mathcal{N}, \mathcal{K}p_i)$ *iff* $K_i \leqq K_{i+1}$,

where equality in 1° *or* 2° *is true iff* $N_i = N_{i+1}$ *or* $K_i = K_{i+1}$, *respectively.*

Proof.

1° $N_{1,j} = N_{p_i(1)} \cdots N_{p_i(j)}$ for each $j \in [1 : m]$, $j \neq i$ implies $B(\mathcal{N}, \mathcal{K}) - B(\mathcal{N}p_i, \mathcal{K}) = N_{1,i-1}N_iK_{i,m} - N_{1,i-1}N_{i+1}K_{i,m} = N_{1,i-1}(N_i - N_{i+1})\, K_{i,m} \leqq \leqq 0$ iff $N_i - N_{i+1} \leqq 0$.

2° follows analogically due to $K_{j,m} = K_{p_i(j)} K_{p_i(j+1)} \dots K_{p_i(m)}$, $j \in [1 : m]$, $j \neq i + 1$. ∎

1.5 Lemma. *Let $\mathscr{L} = (L_1, \dots, L_m)$, $m > 2$ be a* NS *such that $L_1 \leqq L_2 \leqq \dots \leqq L_m$ and $p \in \mathscr{P}([1 : m])$ an arbitrary permutation. Then there exists a sequence $\pi = \{p_n\}_{n=0}^{r}$, $p_n \in \mathscr{P}([1 : m])$ such that $p = p_0 p_1 \dots p_r$, $p_0 = 1$ and for $n > 0$ p_n is a transposition of i_n and $i_n + 1 (i_n \in [1 : m - 1])$ satisfying $L_{p_0 \dots p_{n-1}(i_n)} \leqq L_{p_0 \dots p_{n-1}(i_n+1)}$.*

Proof. For $p = 1$ we put $\pi = \{p_0\}$ and for $p \neq 1$ we proceed by induction on m:

1. $m = 2$: $p \neq 1 \Rightarrow p$ is a transposition $\Rightarrow \pi = \{p_0, p\}$ is the desired sequence because $L_{p_0(1)} = L_1 \leqq L_{p_0(2)} = L_2$.

2. $m > 2$: If $p(m) = m$ then $p = p_{1,m-1} \cup 1_{m,m}$ (see I.1.9) and we can put $\pi = \{p'_n \cup 1_{m,m}\}_{n=0}^{r}$ where $\pi' = \{p'_n\}_{n=0}^{r}$, $p'_n \in \mathscr{P}([1 : m - 1])$, $p_{1,m-1} = p'_0 \dots p'_r$ is a sequence having the desired properties with respect to $\mathscr{L}' = (L_1, \dots, L_{m-1})$. Such π' exists by induction hypothesis. Let $p(m) = i \neq m$. Clearly $p = q_1 q_2$ where q_1 is defined by $\mathscr{L} q_1 = (L_1, \dots, L_{i-1}, L_{i+1}, \dots, L_m, L_i)$ and $q_2 = q'_2 \cup 1_{m,m}$ by $\mathscr{L}' q'_2 = (L'_{q'_2(1)}, \dots, L'_{q'_2(m-1)}) = (L_{p(1)}, \dots, L_{p(m-1)})$ where $\mathscr{L}' = (L'_1, \dots, L'_{m-1}) = (L_1, \dots, L_{i-1}, L_{i+1}, \dots, L_m)$. We put $\pi = \{p_n\}_{n=0}^{m-i+r}$ where p_n is a transposition of $i + n - 1$ and $i + n$ for $n \in [1 : m - i]$ and $p_{m-i+k} = p'_k \cup 1_{m,m}$ for $k \in [1 : r]$ with sequence $\pi' = \{p'_k\}_{k=0}^{r}$, $p'_k \in \mathscr{P}([1 : m - 1])$, $q'_2 = p'_0 \dots p'_r$ having the desired properties with respect to $\mathscr{L}'$. Such π' exists again by induction hypothesis and $p = p_0 \dots p_{m-i+r}$ because $q_1 = p_0 \dots p_{m-i}$ and $q_2 = (p'_0 \cup 1_{m,m}) \dots (p'_r \cup 1_{m,m})$. It is easy to see that all p_n have the desired properties. ∎

1.6 Theorem. *Let $N_1 \leqq N_2 \leqq \dots \leqq N_m$ and $K_1 \geqq K_2 \geqq \dots \geqq K_m$. Then it holds $B(\mathscr{N}, \mathscr{K}) \leqq B(\mathscr{N}p, \mathscr{K}q) \leqq B(\mathscr{N}s, \mathscr{K}s)$ for each pair of permutations $p, q \in \mathscr{P}([1 : m])$.*

Proof. Let $\mathscr{L} = (L_1, \dots, L_m)$, $L_1 \leqq L_2 \leqq \dots \leqq L_m$ and $\mathscr{L}' = (L'_1, \dots, L'_m)$ be arbitrary. Then by 1.4 and 1.5 we have $B(\mathscr{L}, \mathscr{L}') \leqq B(\mathscr{L}p, \mathscr{L}')$ and $B(\mathscr{L}', \mathscr{L}) \geqq B(\mathscr{L}', \mathscr{L}p)$ for each permutation $p \in \mathscr{P}([1 : m])$. Hence $B(\mathscr{N}, \mathscr{K}) \leqq B(\mathscr{N}p, \mathscr{K}) = B(\mathscr{K}s, \mathscr{N}ps) \leqq B(\mathscr{K}s(sqs), \mathscr{N}ps) = B(\mathscr{K}qs, \mathscr{N}ps) = B(\mathscr{N}p, Kq)$ and $B(\mathscr{N}s, \mathscr{K}s) \geqq B(\mathscr{N}s, \mathscr{K}s(sq)) = B(\mathscr{N}s, \mathscr{K}q) = B(\mathscr{K}qs, \mathscr{N}) \geqq B(\mathscr{K}qs, \mathscr{N}ps) = B(\mathscr{N}p, \mathscr{K}q)$. ∎

2. PARAMETRIC DISCRETE FOURIER TRANSFORM

Discrete Fourier transform (DFT) is one of the most important linear transforms that are widely used in various applications. To make clear the benefits of the new algebraic approach, we shall give a simple derivation of the mixed-radix last Fourier transform (FFT) algorithm for a very general DFT concept, namely for that of the parametric discrete Fourier transform (DFT-P) which has been introduced in [5] recently. There DFT-P is defined by means of a square matrix over **C**, here we shall extend this notion to a rectangular matrix over any associative and commutative ring **R**.

2.1 Notation. If $\mathbf{A}_1 \in \mathscr{M}(N_1 \times K_1)$ and $\mathbf{A}_2 \in \mathscr{M}(N_2 \times K_2)$ then $\mathbf{A}_1 \cong \mathbf{A}_2$ means that $\mathbf{A}_1$ and $\mathbf{A}_2$ have equal elements in the first $N = \min(N_1, N_2)$ rows and $K = \min(K_1, K_2)$ columns.

2.2 Definition. *Parametric discrete Fourier transform.*

Let $W_N \in \mathbf{R}$ be N-th root of unity in $\mathbf{R}$ ($W_N^N 1$, $N \in = \mathbf{N}$, $N \geqq 2$) and Θ a rational parameter such that W_N^{Θ} exists in $\mathbf{R}$. *Parametric discrete Fourier transform* (with parameter Θ) is defined by a matrix $\mathbf{W}_{N,\Theta}$ or $\mathbf{W}_{\Theta,N} = \mathbf{W}_{N,\Theta}^T$ of size $N \times N$ where $W_{N,\Theta}(n, k) = W_N^{n(k+\Theta)}$ for each $n, k \in \mathbf{Z}_N$. If $\Theta = 0$ then we get the standard DFT and write simply $\mathbf{W}_N$ instead of $\mathbf{W}_{N,\Theta}$ or $\mathbf{W}_{\Theta,N}$. A matrix $\mathbf{A} \in \mathscr{M}(N' \times N'')$, $N', N'' \leqq N$ is said to define a *rectangular* DFT-P if $\mathbf{A} \cong \mathbf{W}_{N\,\Theta}$ or $\mathbf{A} \cong \mathbf{W}_{\Theta,N}$.

2.3 Theorem. *Let* $\mathscr{N} = (N_1, \ldots, N_m)$, $\mathscr{N}' = (N_1', \ldots, N_m')$, $\mathscr{N}'' = (N_1'', \ldots, N_m'')$, $N = N_{1,m}$, $N' = N_{1,m}'$ *and* $N'' = N_{1,m}''$ $(m \geqq 2)$ *where* $N_1', N_1'' \leqq N_1$ *and* $N_i = N_i' = N_i''$ *for* $i \in [2 : m]$. *Then it holds* $\mathbf{S}_{\mathscr{N}}^T \mathbf{W}_{N,\Theta} \cong \mathbf{A}_1 \otimes_R \ldots \otimes_R \mathbf{A}_m = \mathbf{A}$ *and* $\mathbf{W}_{\Theta,N} \bar{\mathbf{S}}_{\mathscr{N}} \cong \mathbf{B}_1 \otimes_L \ldots \otimes_L \mathbf{B}_m = \mathbf{B}$ *where* $\mathbf{B}_i = \mathbf{A}_i^T$ *and* $\mathbf{A}_i \in \mathscr{M}(N_i' \times N_{i,m}'')$, $i \in [1 : m]$ *are matrices having elements* $A_i(n_i, [k_i, \ldots, k_m]_{\mathscr{N}_{i,m}''}) = W_{N_{i,m}}^{n_i([k_i, \ldots, k_m]_{\mathscr{N}_{i,m}''} + \Theta)}$, $W_{N_{i,m}} = W_N^{N_{1,i-1}}$. *In particular* $\mathbf{A}_m = \mathbf{W}_{N_m,\Theta}$ *and* $\mathbf{B}_m = \mathbf{W}_{\Theta,N_m}$.

Proof. We proceed by induction on m.

1. $m = 2$: $W_N^{[n_2, n_1]_{\mathscr{N}'}([k_1, k_2]_{\mathscr{N}} + \Theta)} = W_N^{(n_2 N_1 + n_1)([k_1, k_2]_{\mathscr{N}} + \Theta)} = W_N^{n_1([k_1, k_2]_{\mathscr{N}} + \Theta} \cdot W_{N_2}^{n_2(k_1 N_2 + k_2 + \Theta)} = W_{N_{1,2}}^{n_1([k_1, k_2]_{\mathscr{N}''} + \Theta)} W_{N_2}^{n_2(k_2 + \Theta)} = A_1(n_1, [k_1, k_2]_{\mathscr{N}''}) A_2(n_2, k_2) = A([n_1, n_2]_{\mathscr{N}'}, [k_1, k_2]_{\mathscr{N}''})$.

2. $m > 2$: By induction hypothesis: $\mathbf{S}_{(N_1, N_{2,m})}^T \mathbf{W}_{N,\Theta} \cong \mathbf{A}_1 \otimes_R \tilde{\mathbf{A}}_2$ where $\mathbf{W}_{N_{2,m},\Theta} = \tilde{\mathbf{A}}_2 \in \mathscr{M}(N_{2,m} \times N_{2,m})$, $\mathbf{S}_{2,m}^T \tilde{\mathbf{A}}_2 = \mathbf{A}_2 \otimes_R \ldots \otimes_R \mathbf{A}_m$. Hence by I.2.6 $\mathbf{S}_{(N_1, N_{2,m})}^T \cdot \mathbf{W}_{N,\Theta} \cong \mathbf{A}_1 \otimes_R \mathbf{S}_{2,m}(\mathbf{A}_2 \otimes_R \ldots \otimes_R \mathbf{A}_m) = (\mathbf{I}_{N_1'} \otimes \mathbf{S}_{2,m})(\mathbf{A}_1 \otimes_R \ldots \otimes_R \mathbf{A}_m)$ and consequently $(\mathbf{I}_{N_1'} \otimes \mathbf{S}_{2,m}^T) \tilde{\mathbf{S}} \mathbf{W}_{N,\Theta} \cong \mathbf{A}_1 \otimes_R \ldots \otimes_R \mathbf{A}_m$, $\tilde{\mathbf{S}} \in \mathscr{M}(N_1' N_{2,m} \times N)$, $\tilde{\mathbf{S}} \cong \mathbf{S}_{(N_1, N_{2,m})}^T$. But $(\mathbf{I}_{N_1'} \otimes \mathbf{S}_{2,m}^T) \tilde{\mathbf{S}}$ are the first $N_1' N_{2,m}$ rows of $(\mathbf{I}_{N_1} \otimes \mathbf{S}_{2,m}^T) \cdot \mathbf{S}_{(N_1, N_{2,m})}^T = \mathbf{S}_{\mathscr{N}}^T$ in view of I.1.12 and due to the block diagonal form of $\mathbf{I}_{N_1} \otimes \mathbf{S}_{2,m}^T$. ∎

2.4 Corollary. *Fast parametric discrete Fourier transform (FFT-P).*
$\mathbf{A} = \mathbf{A}^{(m)} . \mathbf{A}^{(m-1)} \ldots \mathbf{A}^{(1)}$ *and* $\mathbf{B} = \mathbf{B}^{(1)} \mathbf{B}^{(2)} \ldots \mathbf{B}^{(m)}$ *where for* $i \in [1 : m]$ $\mathbf{B}^{(i)} = \mathbf{A}^{(i)T}$, $\mathbf{A}^{(i)} = \mathbf{D}^{(i)} \mathbf{W}^{(i)}$; $\mathbf{D}^{(i)} = \mathbf{I}_{N_{1,i-1}'} \otimes \tilde{\mathbf{D}}^{(i)}$, $\tilde{\mathbf{D}}^{(i)} = \operatorname{diag}(\tilde{D}_0^{(i)}, \ldots, \tilde{D}_{N_{i,m}'-1}^{(i)})$, $\tilde{D}_{[n_i, k]}^{(i)} = W_{N_{i,m}}^{n_i(k+\Theta)}$, $n_i \in \mathbf{Z}_{N_i'}$, $k \in \mathbf{Z}_{N_{i+1,m}}$; $\mathbf{W}^{(i)} = \mathbf{I}_{N_{1,i-1}'} \otimes \mathbf{W}_{N_i}' \otimes \mathbf{I}_{N_{i+1,m}}$, $\mathbf{W}_{N_i}' \in \mathscr{M}(N_i' \times N_i'')$, $\mathbf{W}_{N_i}' \cong \mathbf{W}_{N_i}$. *The elementary transforms attain the form:* $\mathbf{A}_{i,k} = \mathbf{D}_{i,k} \mathbf{W}_{N_i}'$, $\mathbf{D}_{i,k} = \operatorname{diag}(\tilde{D}_{[0,k]}^{(i)}, \tilde{D}_{[1,k]}^{(i)}, \ldots, \tilde{D}_{[N_i'-1,k]}^{(i)})$ *and* $\mathbf{B}_{i,k} = \mathbf{A}_{i,k}^T$ *for* $i \in [1 : m]$, $k \in \mathbf{Z}_{N_{i+1,m}}(\mathbf{A}_{m,0} = \mathbf{A}_m, \mathbf{B}_{m,0} = \mathbf{B}_m)$.

Proof. $A_i(n_i, [k_i, \ldots, k_m]) = W_{N_{l,m}}^{n_l(k_l N_{l+1,m}+[k_{l+1}, \ldots, k_m]+\Theta)} = W_{N_{l,m}}^{n_l([k_{l+1}, \ldots, k_m]+\Theta)}$. $\cdot\, W_{N_l}^{n_i k_l}$ for $i \in [1 : m-1]$ and $A_m(n_m, k_m) = W_{N_m}^{n_m(k_m+\Theta)} = W_{N_m}^{n_m\Theta} W_{N_m}^{n_m k_m} \Rightarrow \mathbf{A}_{i,k} = \mathbf{D}_{l,k}\mathbf{W}'_{N_l}$. Hence and by I.2.3 we get immediately $\mathbf{A}_l \otimes_R \mathbf{I}_{N_{l+1,m}} = \vec{\mathbf{A}}_i = \breve{\mathbf{D}}^{(i)}(\mathbf{W}'_{N_i} \otimes \mathbf{I}_{N_{l+1,m}})$, and finally by I.2.9 $\mathbf{A}^{(l)} = \mathbf{I}_{N'_{1,l-1}} \otimes (\mathbf{A}_i \otimes_R \mathbf{I}_{N_{l+1,m}}) = (\mathbf{I}_{N'_{1,l-1}} \otimes \tilde{\mathbf{D}}^{(l)})\,(\mathbf{I}_{N'_{1,l-1}} \otimes \mathbf{W}'_{N_l} \otimes \mathbf{I}_{N_{l+1,m}})$. ∎

2.5 *Applications of special cases of 2.3 and 2.4*

1. $\Theta = 0$, $\mathbf{R} = \mathbf{C}$: standard DFT $\mathbf{X} = \mathbf{W}_N \mathbf{x}$, $W_N = \exp(\pm i2\pi/N)$, $i = \sqrt{-1}$. We have $\mathbf{D}^{(m)} = \mathbf{I}_{N'}$ in this case.

a) $N'_1 = N''_1 = N_1$: $\mathbf{W}_N = \mathbf{S}_{\mathcal{N}}\mathbf{A}^{(m)}\mathbf{A}^{(m-1)} \ldots \mathbf{A}^{(1)}$ is the so—called *decimation in frequency* FFT (DIF FFT) algorithm, known also as *Sand—Tukey's* FFT. $\mathbf{W}_N = \mathbf{W}_N^T = \mathbf{B}^{(1)}\mathbf{B}^{(2)}, \ldots, \mathbf{B}^{(m)}\mathbf{S}_{\mathcal{N}}^T$ is the *decimation in time* FFT (DIT FFT), known also as *Cooley—Tukey's* FFT (cf. [1, 2, 3]).

b) $N'_1 < N_1$, $N''_1 = N_1$: DIF FFT *with decimation on output* or DIT FFT *with decimation on input*. Usually $N'_1 = 1$, which gives $\mathbf{A} = (\mathbf{I}_1 \otimes (\mathbf{A}_2 \otimes_R \ldots \otimes_R \mathbf{A}_m)) \cdot \tilde{\mathbf{D}}^{(1)}(\mathbf{W}'_{N_1} \otimes \mathbf{I}_{N_{2,m}})$ where $\mathbf{W}'_{N_1} = (1, 1, \ldots, 1) \in \mathcal{M}(1 \times N_1)$ and $\tilde{\mathbf{D}}^{(1)} = \mathbf{I}_{N_{2,m}}$ or equivalently $\mathbf{A} = (\mathbf{A}_2 \otimes_R \ldots \otimes_R \mathbf{A}_m)\,(\mathbf{I}_{N/N_1}, \ldots, \mathbf{I}_{N/N_1})$, which means that only one FMRT of order N/N_1 is to be accomplished. The decimated values $X([n_m, \ldots, n_2, 0]) = X(nN_1)$, $n \in \mathbf{Z}_{N_{2,m}}$ are obtained on output. $\mathbf{B} = \mathbf{A}^T = (\mathbf{I}_{N/N_1}, \ldots, \mathbf{I}_{N/N_1})^T (\mathbf{B}_2 \otimes_L \ldots \otimes_L \mathbf{B}_m) \Rightarrow \mathbf{X}$ is obtained repeating N_1-times the transform result of length N/N_1.

c) $N'_1 = N_1$, $N''_1 < N_1$: DIF FFT *with truncation on input* (the last $N - N'' = (N'_1 - N''_1)\,N_{2,m}$ elements of input vector $\mathbf{x}$ are zeros) or DIT FFT *with truncation on output* (cf. [2; p. 188]). Usually $N''_1 = 1$, which gives $\mathbf{A} = (\mathbf{I}_{N_1} \otimes (\mathbf{A}_2 \otimes_R \ldots \otimes_R \mathbf{A}_m))\,\tilde{\mathbf{D}}^{(1)}(\mathbf{W}'_{N_1} \otimes \mathbf{I}_{N_{2,m}})$ where $\mathbf{W}'_{N_1} = (1, 1, \ldots, 1)^T \in \mathcal{M}(N_1 \times 1)$ or equivalently $\mathbf{A} = (\mathbf{A}'_2\mathbf{D}'_0, \mathbf{A}'_2\mathbf{D}'_1, \ldots, \mathbf{A}'_2\mathbf{D}'_{N_1-1})^{BT}$ with $\mathbf{A}'_2 = \mathbf{A}_2 \otimes_R \ldots \otimes_R \mathbf{A}_m$ and $\mathbf{D}'_{n_1} = \operatorname{diag}(\tilde{D}^{(1)}_{[n_1,0]}, \ldots, \tilde{D}^{(1)}_{[n_1,N/N_1-1]})$, $n_1 \in \mathbf{Z}_{N_1}$. Clearly $\mathbf{D}'_0 = \mathbf{I}_{N/N_1}$ and $D^{(1)}_{[n',0]} = 1$. Thus $\mathbf{X}$ may be processed in parts performing the FMRT $\mathbf{A}'_2$ of order N/N_1 N_1-times, successively with data vectors $\mathbf{D}'_0\mathbf{x}', \mathbf{D}'_1\mathbf{x}', \ldots, \mathbf{D}'_{N_1-1}\mathbf{x}'$ where $\mathbf{x}'$ denotes the truncated input vector $\mathbf{x}$. $\mathbf{B} = \mathbf{A}^T = (\mathbf{D}'_0\mathbf{B}'_2, \mathbf{D}'_1\mathbf{B}'_2, \ldots, \mathbf{D}'_{N_1-1}\mathbf{B}'_2)$ where $\mathbf{B}'_2 \mp \mathbf{A}_2'^T = \mathbf{B}_2 \otimes_L \ldots \otimes_L \mathbf{B}_m \Rightarrow$ the truncated output $\mathbf{X}' = \sum_{n_1=0}^{N_1-1} \mathbf{D}'_{n_1}\mathbf{X}_{n_1}$ where $\mathbf{X}_{n_1}$ are outputs of FMRTs $\mathbf{B}'_2$ applied N_1-times on successive blocks of length N/N_1 of the input vector $\mathbf{S}_{\mathcal{N}}^T\mathbf{x}$. Thus $\mathbf{X}'$ may be processed in parts again.

d) $N'_1 < N_1$, $N''_1 < N_1$: DIF FFT *with decimation on output and truncation on input* or DIT FFT *with truncation on output and decimation on input*.

2. $\Theta > 0$, $\mathbf{R} = \mathbf{C}$, $N'_1 = N''_1 = N_1$: DFT-P introduced in [5]. Let $K \in \mathbf{N}$, then we have $W_{KN}^{[n,k]n'} = W_{KN}^{(nK+k)n'} = W_{KN}^{K(n+k/K)n'} \doteq W_N^{(n+\Theta_k)n'}$, $\Theta_k = k/K$ for each $k \in \mathbf{Z}_K$ and $n, n' \in \mathbf{Z}_N$. Hence $\mathbf{W}_{KN} \cong \mathbf{S}_{(K,N)}(\mathbf{W}_{\Theta_0,N}, \mathbf{W}_{\Theta_1,N}, \ldots, \mathbf{W}_{\bullet_{K-1},N})^{BT} \in \mathcal{M}(KN \times N)$ is DFT with truncation on input and similarly

$\mathbf{W}_{KN} \cong (\mathbf{W}_{N,\Theta_0}, \mathbf{W}_{N,\Theta_1}, \dots, \mathbf{W}_{N,\Theta_{K-1}})\, \mathbf{S}^T_{(K,N)} \in \mathcal{M}(N \times KN)$ is DFT with truncation on output. So we see that DFT with truncated input or output may be obtained in parts again, but this time performing K-times fast DFT-Ps of size $N \times N$, successively with parameters $\Theta_0, \Theta_1, \dots, \Theta_{K-1}$.

3. All what has been stated in 1. and 2. for $\mathbf{R} = \mathbf{C}$ is also true for *number theoretic transforms* ($\mathbf{R} = \mathbf{Z}_M$ or $\mathbf{R} =$ Galois field) and for *polynomial transforms* ($\mathbf{R} =$ residue ring of polynomials) – cf. [3, 4].

There arises a natural question: Which of the methods 1c) and 2 is computationally more effective if one wants to compute DFT of size $N_1 N_{2,m} \times N_{2,m}$ or of size $N_{2,m} \times N_1 N_{2,m}$? By 2.4 FMRTs for $\mathbf{A}'_2$ and $\mathbf{W}_{N_{2,m},\Theta_k}$ have the same structure except for the amount of multiplicative factors $\tilde{D}^{(i)}_{[n_i,0]} \neq 1$, $i \in [2 : m]$, $n_i \in \mathbf{Z}_{N_i}$. $\tilde{D}^{(i)}_{[n_i,0]}$, $n_i > 0$ is equal to unity for $\mathbf{A}'_2(\Theta = 0)$ but not equal to unity for $\mathbf{W}_{N_{2,m},\Theta_k}$, $k > 0\,(\Theta_k > 0)$. This means that the method 2 requires $\mu_2 = (N_1 - 1) \sum_{i=2}^{m} N_{2,i-1} \,.\, (N_i - 1)$ more such multiplications than method 1c). On the other hand 1c) requires in comparison to 2 $\mu_1 = (N_1 - 1)(N_{2,m} - 1)$ extra multiplications by $\mathbf{D}'_1, \dots, \mathbf{D}'_{N_1-1}$. But $\sum_{i=2}^{m} N_{2,i-1}(N_i - 1) = \sum_{i=2}^{m} N_{2,i} - \sum_{i=1}^{m-1} N_{2,i} = N_{2,m} - 1$ implies $\mu_1 = \mu_2$ and thus both methods have the same computational complexity.

REFERENCES

[1] E. O. Brigham, *The Fast Fourier Transform.* Prentice – Hall, Englewood Cliffs, New Jersey, 1974.

[2] Eh. E. Dagman; G. A. Kukharev, *Bystrye diskretnye ortogonal'nye preobrazovaniya* (Fast Discrete Orthogonal Transformations). Izdatel'stvo "Nauka", Sibirskoe otdelenie, Novosibirsk, 1983 (Russian).

[3] D. F. Elliott; K. R. Rao, *Fast Transforms, Algorithms, Analyses, Applications.* Academic Press, New York, London, 1982.

[4] H. J. Nussbaumer, *Fast Fourier Transform and Convolution Algorithms.* 2-nd ed., Springer-Verlag Berlin, Heidelberg, New York, 1982.

[5] V. A. Ponomarev; O. V. Ponomareva, *A Modification of Discrete Fourier Transform for Solution of Interpolation and Functional Convolution Problems.* Radiotekhn. i Elektron. 29 (1984), No. 8, 1561 – 1570 (Russian); translated as Radio Engrg. Electron. Phys. 29 (1984), No. 9, 79 – 88.

[6] V. Veselý, *Algebraic Theory of Fast Mixed-Radix Transforms: I. Generalized Kronecker Product of Matrices.* Arch. Math. (Brno), Vol. 25, No. 3 (1989), 149 – 162.

Vítězslav Veselý
Institute of Physical Metallurgy
Czechoslovak Academy of Sciences
616 62 Brno, Žižkova 22
Czechoslovakia