Group-Theoretic Methods for Parallel Computation of Convolution

Olga V. Klimova
Institute of Engineering Science
Ural Branch of the Russian Academy of Science
91 Pervomayskaya str., Ekaterinburg, 620219, Russia
Fax: (+7) 3432 –745-330
e-mail: ovs@ imach.uran.ru

1. Introduction

Convolution decomposition allowed creation of fast computation algorithms within the scope of sequential processing [1,2]. The proposed methods of convolution decomposition were aimed at the resolution of specific problems standing in the way of creating such algorithms. The methods consist in the determination of correspondence between the one-dimensional and two-(multi-) dimensional convolutions on the basis of the chinese remainder theorem or by increasing the initial convolution length. Thus fast algorithms using number theoretic transforms (NTT) and efficient algorithms for calculating short convolutions were created [3,4]. However, in one case, the decomposition methods made the algorithm structure redundant, whereas in other case, they imposed restrictions on the decomposition parameters, which need to be mutually prime numbers.

Parallel processing requires structural flexibility of algorithms, therefore the decomposition methods primordially characterized by redundancy and restrictions imposed on the parameters are not effective. However, parallel convolution algorithms are also to be created through decomposition. Moreover, the methods for parallel convolution decomposition must both solve the problems of developing fast algorithms and make them structurally nonredundant and parametrically adjustable to any degree of parallelism. Such methods were created on the basis of the group-theoretic approach to convolution decomposition [5]. Based on this approach, algorithms using NTT were considered in [6]. The group-theoretic approach proposed is complex in character, that is, it is orientated towards the decomposition of a number of basic functions of digital signal processing – convolution, correlation, discrete Fourier transform (DFT). The basic decomposition transforms of DFT and convolution within the scope of the group-theoretic approach were presented in [7]. The problem of DFT group-theoretic decomposition and DFT parallel computation were studied in detail in [8, 9]. This integrated approach has allowed one to develop a number of methods for parallel convolution computation based on the group-theoretic decomposition of DFT and convolution. These methods offer various fast parallel convolution algorithms. A set of such methods demonstrates a unified decomposition approach to the development of these algorithms. The objective of this paper is to develop a collection of methods for the parallel computation of convolution by generalizing and extending the results of the group-theoretic decomposition of DFT and convolution.

2. The method of group-theoretic decomposition of DFT

The method is not connected with the decomposition of the convolution function

$$C(t)=\sum_{q=0}^{h-1} x(t-q)y(q), \tag{2.1}$$

as it is primordially based on the transforms (DFT, NTT) displaying the cyclic convolution property [4,10,11]. Then the initial model of calculation $C(t)$ is

$$S_c(\alpha)=S_x(\alpha)\cdot S_y(\alpha), \tag{2.2}$$

where

$S_c(\alpha),\ S_x(\alpha),\ S_y(\alpha)$ are the transforms of the sequences $C(t),\ x(t),\ y(t)$ in the form

$$S_a(\alpha)=\sum_{t=o}^{N-1} a(t)\cdot W_N^{t\alpha}\ . \tag{2.3}$$

Equation (2.3) defines DFT when $W_N=e^{-2\pi i/N}$. The main body of computation is connected with the transition to the frequency domain and backwards. The proposed method of parallel computation of convolution is based on the realization of this transition over the fast parallel algorithms produced by the group-theoretic decomposition of DFT. There are two ways of realizing this decomposition, which correspond to two methods of DFT decomposition. The decomposition results in the decomposition forms of DFT representation. These DFT forms allow one to create fast parallel algorithms flexibly adjusted by parameters to different lengths and dimensions of signals and to different degrees and modes of parallelizm. The characteristics of the decomposition methods and the corresponding algorithms were described in [8, 9]. One of the important features of the group-theoretic decomposition is its ability to offer the unity of the structure of the algorithms created by different decomposition methods as well as the unity of algorithm structure for direct and inverse DFT computation.

The latter is important for the realization of the present method, when one and the same algorithm (with an accuracy of phase factors) is both for direct and inverse DFT. To complete the presentation of the method, we demonstrate some group-theoretic decomposition forms of DFT. The structural unity of the algorithms corresponding to them allows one to present only one of these forms. First we interpret some of the notions and symbols using in the decomposition form of DFT. The cyclic convolution (2.1) and DFT (2.3) are defined on the group $H=Z_N=\{M_N,+,-\}$ of order N. The group is interpreted as the interval

$M_N = [0, N-1]$ with the operation of modulo N addition + (or subtraction -) given on it. Suppose $N = h_1 \cdot h_2 \cdot \ldots \cdot h_k$. Then any numbers t, $\alpha \in [0, N-1]$ can be represented unambiguously as

$$t = t_k + t_{k-1} \cdot h_k + \ldots + t_1 \cdot h_2 \cdot \ldots \cdot h_{k-1} \cdot h_k , \tag{2.4}$$

$$\alpha = \alpha_1 + \alpha_2 \cdot h_1 + \ldots + \alpha_k \cdot h_{k-1} \cdot \ldots \cdot h_1 . \tag{2.5}$$

Let $N = \tilde{N}_0$, $\tilde{N}_{i-1} = h_{k-(i-1)} \cdot \tilde{N}_i \quad (i = 1,\ldots,k)$ and

$\tilde{\gamma}_{i-1} = \alpha_1 + \alpha_2 \cdot h_1 + \alpha_3 \cdot h_1 \cdot h_2 + \ldots + \alpha_{k-(i-1)} \cdot h_1 \cdot \ldots \cdot h_{k-i}$.

The group-theoretic decomposition of DFT enables one to obtain the following form of DFT representation:

$$\begin{aligned} S(\alpha) = {} & \sum_{t_2=0}^{h_k-1} W_{h_k}^{t_k \alpha_k} \cdot W_{\tilde{N}_0}^{t_k \tilde{\gamma}_1} \cdot \ldots \cdot \sum_{t_{k-(i-1)}=0}^{h_{k-(i-1)}-1} W_{h_{k-(i-1)}}^{t_{k-(i-1)} \alpha_{k-(i-1)}} \cdot W_{\tilde{N}_{i-1}}^{t_{k-(i-1)} \tilde{\gamma}_i} \cdot \ldots \\ & \ldots \cdot \sum_{t_2=0}^{h_2-1} W_{h_2}^{t_2 \alpha_2} \cdot W_{\tilde{N}_2}^{t_2 \alpha_1} \cdot \sum_{t_1=0}^{h_1-1} x(t_k,\ldots,t_{k-(i-2)},t_1) \cdot W_{h_1}^{t_1 \alpha_1} . \end{aligned} \tag{2.6}$$

The use of equation (2.6) in the time ↔ frequency transition has allowed us to construct fast parallel algorithms for convolution computation. Similar algorithms have been developed on the basis of NTT.

3. The method of group-theoretic decomposition of signal

The methods of parallel computation of convolution presented below are based on group-theoretic decomposition of convolution (2.1). The basic decomposition transforms of convolution resulting in new forms of convolution representation were described in [5-7]. These decomposition forms define the methods of parallel computation of convolution. Before presenting these decomposition forms, we interpret their principal elements used in the decomposition. We represent any number $q \in [0, N-1]$ in the form of equation (2.4) and define the subtraction (shift) operation $\ominus$ on M_N as

$$t \ominus q = \sum_{i=1}^{k} r_i h_{i+1} \cdot \ldots \cdot h_k$$

where

$$(t_i - q_i) = r_i (\mathrm{mod}\, h_i).$$

The addition $t \oplus q$ is defined similarly. The «$\oplus$» and «$\ominus$» operations define some group $H = \{M_N, \oplus, \ominus\}$ on the set M_N identifiable with the direct product of the cyclic groups Z_{h_i}

$$Z_{\vec{N}} = Z_{h_1} \times Z_{h_2} \times \ldots \times Z_{h_k} ,$$

where

$$Z_{h_i} = \{M_N, +, -\},$$

that is, it can be considered that $Z_{\vec{N}} = \{M_N, \oplus, \ominus\}$ owing to their isomorphism. On the interval M_N we study the cyclic convolution

$$C_{\vec{N}}(t) = \sum_{q \in H} x(t \ominus q) \cdot y(q) \tag{3.1}$$

given on the group $Z_{\vec{N}}$. Now the following symbols can be introduced

$$j_1 = t_k + t_{k-1} \cdot h_k + ... + t_2 \cdot h_3 \cdot ... \cdot h_{k-1} \cdot h_k,$$
$$N = N_0 = h_1 \cdot N_1, \qquad N_1 = h_2 \cdot ... \cdot h_k.$$

Let us next use the functions

$$x^*(t) = \begin{cases} x(t), & t = t_1 \cdot N_1 \\ 0 & t \neq t_1 \cdot N_1 \end{cases} \text{ and} \tag{3.2}$$

$$x_{j_1}{}^*(t) = \begin{cases} x(t + j_1), & t = t_1 \cdot N_1 \\ 0, & t \neq t_1 \cdot N_1 \end{cases}, \quad j_1 = 0,...N_1 - 1. \tag{3.3}$$

The functions (3.2) and (3.3) possess the following properties:

- equality of shifts

$$x^*(t - q) = x^*(t \ominus q); \tag{3.4}$$

- composition property

$$x(t) = \sum_{j_1=0}^{N_1-1} x^*_{j_1}(t - j_1). \tag{3.5}$$

The group-theoretic decomposition of convolution (2.1) based on the decomposition (3.5) of the signal $x(t)$ allows one to obtain the following form of convolution representation:

$$C(t) = \sum_{j_1=0}^{N_1-1} \sum_{q \in H} x^*_{j_1}(t \ominus q) y(q - j_1) = \sum_{j_1=0}^{N_1-1} C_{j_1}(t). \tag{3.6}$$

Equation (3.6) gives a formalized description of the method of parallel computation of convolution under study. The main characteristic feature of the method is the presence of N_1 independent computation processes for the functions $C_{j_1}(t)$. The functions $C_{j_1}(t)$ are equivalent on the groups with different structures, but the same order N:

$$Z_N, \quad Z_{h_1} \times Z_{N_1}, \quad Z_{h_1} \times Z_{\vec{N}_1},$$

due to their independence from the group shift operation. This independence stems from the property (3.4) of the functions (3.2), (3.3). Thus the decomposition (3.5) of the signal $x(t)$ has enabled the relations between the convolutions (2.1) and (3.1) to be established . Owing to the structure of the signals defining the functions $C_{j_1}(t)$, fast parallel computational algorithms for $C(t)$ based on different fast orthogonal transforms given on the group $Z_{h_1} \times Z_{\vec{N}_1}$ have been constructed within the scope of the method.

4. The method of group-theoretic decomposition of convolutions

Equation (3.6) can be considered as the basic decompositional form of the convolution (2.1) based on the functions $C_{j_1}(t)$. They can in turn be reduced to the following parallel form of representation:

$$C_{j_1}(t) = \sum_{q_1=0}^{h_1-1} \overset{*}{x}_{j_1}((t_1 - q_1)N_1) \cdot y(q_1 N_1 + p_1 - j_1), \tag{4.1}$$

where

$$t = t_1 N_1 + p_1, \quad q = q_1 N_1 + m_1, \quad t_1, q_1 \in [0, h_1 - 1], \quad p_1, m_1 \in [0, N_1 - 1].$$

In the derivation of eqn. (4.1), the property (3.4) of the functions $\overset{*}{x}_{j_1}(t)$ was used, which results in the equality between p_1 and m_1. Thus the values of any function $C_{j_1}(t)$ are formed by parallel computation of N_1 convolutions $C_{j_1 p_1}(t_1 N_1) = C_{j_1}(t_1 N_1 + p_1)$ of length h_1 of the signals $\overset{*}{x}_{j_1}(t_1 N_1)$ and $y(q_1 N_1 + p_1 - j_1) = y_{-j_1 p_1}(q_1 N_1)$. The method of parallel computation of convolution (2.1) is based on the group-theoretic decomposition (4.1) of the convolutions $C_{j_1}(t)$. The corresponding form of representing convolution (2.1) based on eqn. (3.6) is

$$C(t) = \sum_{j_1=0}^{N_1-1} \sum_{q_1=0}^{h_1-1} \overset{*}{x}_{j_1}((t_1 - q_1)N_1) \cdot y(q_1 N_1 + p_1 - j_1). \tag{4.2}$$

The fast parallel algorithms of convolution using efficient ways of computation of short convolutions of length h_1 have been developed on the basis of eqn. (4.2).

5. The group-theoretic method of reducing one-dimensional convolution to pseudo-two-dimensional one

Equation (4.2), hereinafter referred to as group-theoretic (GT-form), is the basis of transforms giving rise to the present method. Its form is almost similar to two-dimensional convolution:

$$\widehat{C}(t) = \widehat{C}(t_1 N_1 + p_1) = \sum_{j_1=0}^{N_1-1} \sum_{q_1=0}^{h_1-1} x\ ((t_1 - q_1)N_1 + j_1) \cdot y(q_1 N_1 + (p_1 \ominus j_1)). \tag{5.1}$$

Eqn. (4.2) and (5.1) differ in the character of group shift operations:

$$(p_1 - j_1) \text{ и } (p_1 \ominus j_1).$$

In equation (4.2) this shift is performed on the group Z_N of the order N, though in equation (5.1) it is performed on the group Z_{N_1} of the order N_1. By analyzing the function $y(q_1 N_1 + p_1 - j_1)$ from equation (4.2), we can notice that

- for $j_1 \le p_1$, $y(q_1 N_1 + p_1 - j_1) = y(q_1 N_1 + (p_1 \ominus j_1))$;
- for $j_1 > p_1$, $y(q_1 N_1 + p_1 - j_1) = y((q_1 - 1) \cdot N_1 + (p_1 \ominus j_1))$.

Then the GT - form (4.2) gives rise to a form different from the two-dimensional one (5.1) by one position shifts $(q_1 - 1)$ of $(N_1 - 1)$ sequences $y_{-j_1 p_1}(q_1 N_1)$ given on the group Z_{h_1}. The above-mentioned sequences are located at the points $j_1 > p_1$ in the computation of convolution along the coordinate p_1. This form of one-dimensional convolution (2.1) is referred to as pseudo-two-dimensional form. The corresponding method of parallel computation of convolution allows creation of fast parallel algorithms realizing the computation of two-dimensional convolution and the correction of the values of its samples. The recurrent application of this method makes it possible to use the Walsh transform [12] for the computation of convolution (2.1).

6. Conclusion

The methods of parallel computation of convolution illustrate great potentialities of the group-theoretic decomposition approach in the creation of effective parallel algorithms possessing the following properties:

- decomposition nonredundancy;
- absence of restrictions imposed on the decomposition parameters;
- parametric adjustability to any degree of parallelism;
- pertinence to the class of fast algorithms.

The unified approach to the development of a variety of such algorithms provides the unity of their common structure, multivariantness and succession of corresponding computation solutions. The possibility of creating such algorithms allows one to establish optimal relations between the algorithm and the architecture within parallel processing.

References

1. McClellan J.H., Rader C.M.: Number Theory in Digital Signal Processing, Prentice-Hall, Englewood Cliffs,N.J.,1979.
2. Nussbaumer H.J.: Fast Fourier Transform and Convolution Algorithms, Springer-Verlag, Berlin Heidelberg, 1982.
3. Agarwal R.C., Burrus C.S.: Fast One-Dimensional Convolution by Multidimensional Techniques. IEEE Trans. on Acoustics, Speech, and Signal Processing. ASSP-22 (1974) 1-10.
4. Agarwal R.C., Cooley J.W.: New Algorithms for Digital Convolution. . IEEE Trans. on Acoustics, Speech, and Signal Processing. ASSP-25 (1977) 392-409.
5. Klimova O.V.: Developing and Studying Architectural Methods for Parallel Multiprocessor System Designed to Compute Convolutions. Ph.D. diss., Inst. of Eng. Sci., Russian Ac. of Sci. (Ural Branch), Sverdlovsk (Ekaterinburg), 1988.

6. Klimova O.V.: Parallel Architecture of the Arbitrary-Length Convolution Processor with the Use of Rader Number Transforms. Izv. AN Tekhn. Kibernet. (Russia). 2 (1994) 183-191.
7. Klimova O.V.:Decomposition on a Group and Parallel Convolution and Fast Fourier Transform Algorithms. In Malyshkin V. ed., Pact-97, 4th Int'l Conference on Parallel Computing Technologies. Proceedings, pages 358-363. Springer-Verlag, Berlin, LNCS 1277.
8. Klimova O.V.: The Separating Decomposition of Discrete Fourier Transform and Vectorization of its Calculation. In Malyshkin V. ed., Pact-95, Third Int'l Conference on Parallel Computing Technologies. Proceedings, pages 241-245. Springer-Verlag, Berlin, LNCS 964.
9. Klimova O.V.: Group Theoretical Decomposition and Fast Parallel Algorithms for the Discrete Fourier Transform. Journal of Computer and Systems Sciences International. Vol. 36, No. 5 (1997) 802-806.
10. Rabiner L.R., Gold B.: Theory and Application of Digital Signal Processing, Prentice-Hall, Englewood Cliffs,N.J.,1975.
11. Pollard J.M.: The Fast Fourier Fransform in a Finite Field. Mathematics of Computation. 25 (1971) 365-374.
12. Elliot D.F., Rao K.R.: Fast transform: algorithms, analyses, applications, Academic Press, N.Y., 1982.